Technische Universität München
TUM Campus Straubing für Biotechnologie und Nachhaltigkeit

Deep Learning-based Weed Detection in Agricultural Images

Nikita Genze

Vollständiger Abdruck der von dem TUM Campus Straubing für Biotechnologie und Nachhaltigkeit der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Clemens Thielen

Prüfende der Dissertation:

1. Prof. Dr. Dominik Grimm
2. Prof. Dr. Kang Yu

Die Dissertation wurde am 21.03.2025 bei der Technischen Universität München eingereicht und durch den TUM Campus Straubing für Biotechnologie und Nachhaltigkeit am 29.05.2025 angenommen.

Deep Learning-based Weed Detection in Agricultural Images

by

Nikita Genze

Above the fields so vast and wide,
Our drones would drift, our hopes would ride.
Through golden maize and sorghum tall,
We traced the lines that mapped it all.

Each pixel marked with steady hand,
A map of growth, of soil, of land.
With careful strokes, we drew the seams,
Of nature's patchwork, stitched in dreams.

No rush, no haste, just artful care,
Each row and leaf, a tale laid bare.
Oh, days of masks and careful sight,
When fields were painted in soft light.

- ChatGPT

Abstract

Weed management is critical in modern agriculture, as weeds significantly reduce crop yields. They compete for essential resources and serve as hosts for pests and diseases. Traditional weed control relies on herbicides, but their excessive use has led to herbicide-resistant weed species, environmental pollution, and health concerns. Alternative farming systems, such as organic and conservation agriculture, provide sustainable solutions but often require intensive manual labor, which makes large-scale implementation difficult.

♦

Automation and Artificial Intelligence (AI) are increasingly integrated into modern agricultural practices addressing these challenges. Unmanned Aerial Vehicles (UAVs) enable high-resolution field scouting and offer a scalable solution to monitor weeds. However, technical challenges degrade their image quality, for example, due to motion blur and varying illumination. Deep learning (DL) has emerged as a powerful tool for analyzing agricultural imagery. Unlike traditional Computer Vision methods, DL techniques may adapt to environmental variations and improve the accuracy of weed detection tasks.

♦

This thesis advances UAV-based weed detection through three contributions: First, this study explores the impact of motion blur on weed segmentation and evaluates DL models. Here, it tests different DL architectures and feature extractors to determine their robustness under varying degrees of motion blur. The second contribution proposes a novel deblurring-assisted segmentation model to enhance weed detection accuracy in UAV images. This study evaluates a a shift in distribution that reduces DL models' performance when trained on sharp images and applied to motion-blurred UAV captures. The proposed model `DeBlurWeedSeg` utilizes a model specially designed for image deblurring. Afterward, a semantic segmentation model processes the images to classify their pixels as weeds, crops, or soil. This approach significantly improves weed detection in motion-blurred images compared to traditional models and offers a more scalable solution. The third contribution addresses the scarcity of large, high-quality datasets for training DL models. With the Moving Fields Weed Dataset (MFWD), this work presents a comprehensive dataset of 28 weed species captured under controlled greenhouse conditions. This dataset captures images with high spatial and temporal resolution to monitor weed species across different growth stages. Ultimately, it enables the training and evaluation of DL models for weed Classification, Object Detection, and Segmentation tasks.

♦

In conclusion, these contributions support the development of more accurate, scalable, and environmentally sustainable weed management systems to ultimately reduce dependence on herbicides and improve agricultural productivity.

Zusammenfassung

Entfernung von Beikraut ist in der Landwirtschaft von entscheidender Bedeutung, da Beikräuter Ernteerträge erheblich schmälern. Beikräuter konkurrieren um Ressourcen und dienen außerdem als Wirt für Schädlinge und Krankheiten. Traditionelle Beikrautbekämpfung basiert auf Herbizideinsatz, doch übermäßige Anwendung führte zu herbizidresistenten Beikrautarten, starker Umweltbelastung und gesundheitlichen Folgen. Alternative Anbausysteme, wie die ökologische und die konservierende Landwirtschaft, bieten nachhaltige Lösungen. Sie erfordern jedoch oft aufwändige manuelle Arbeit, was eine großflächige Umsetzung erschwert.

Automatisierung und künstliche Intelligenz (KI) werden zunehmend in der modernen Landwirtschaft eingesetzt. Unbemannte Luftfahrzeuge (UAVs) bieten eine skalierbare Lösung zur Beikrautkontrolle durch hochauflösende Fernerkundung an.

Allerdings beeinträchtigen technische Herausforderungen die Bildqualität, zum Beispiel aufgrund von Bewegungsunschärfe. Deep Learning (DL) hat sich als wirksames Tool für die Analyse von landwirtschaftlichen Bilddaten erwiesen. Im Gegensatz zu herkömmlichen Methoden können sich diese Techniken an Schwankungen in der Helligkeit anpassen und die Genauigkeit von Methoden der Beikrauterkennung verbessern.

In dieser Arbeit wird die UAV-basierte Beikrauterkennung durch drei Arbeiten weiterentwickelt: Der erste Beitrag untersucht die Auswirkungen von Bewegungsunschärfe auf Segmentierung von Beikraut und evaluiert die Robustheit von DL-Modellen. Diese Studie untersucht verschiedene DL-Architekturen und Feature-Extraktoren, um ihre Zuverlässigkeit bei Bewegungsunschärfe zu bestimmen. Der zweite Beitrag beschreibt ein neuartiges Modell zur Segmentierung von Bildern mit Bewegungsunschärfe, um die Genauigkeit der Beikrauterkennung in Drohnenaufnahmen zu verbessern. In dieser Studie wird eine Abweichung der Verteilung bewertet, die die Leistung der Modelle verringert, wenn sie auf scharfen Bildern trainiert und auf unscharfen Drohnenbildern angewendet werden. Das vorgeschlagene Modell `DeBlurWeedSeg` verwendet ein spezielles, zusätzliches Modell für die Entfernung von Unschärfe. Anschließend verarbeitet ein Modell zur Segmentierung die Bilder, um Beikraut, Nutzpflanze und Boden zu klassifizieren. Dieser Ansatz verbessert die Beikrauterkennung in bewegungsunscharfen Bildern im Vergleich zu herkömmlichen Modellen erheblich und bietet eine besser skalierbare Lösung. Der dritte Beitrag befasst sich mit dem Mangel an großen, hochwertigen Datensätzen für das Training. Mit dem Moving Fields Weed Dataset (MFWD) wird in dieser Studie ein umfassender Datensatz von 28 Beikrautarten vorgestellt, die unter kontrollierten Gewächshausbedingungen aufgenommen wurden. Dieser Datensatz erfasst Bilder mit hoher räumlicher und zeitlicher Auflösung, um Beikrautarten in verschiedenen Wachstumsstadien zu dokumentieren. Letztendlich ermöglicht er das Training und die Evaluierung von Modellen zur Bildklassifikation, Objekterkennung und Segmentierung. Zusammenfassend unterstützen diese Beiträge die Entwicklung genauerer, skalierbarer und ökologisch nachhaltiger Systeme zum Beikrautmanagement, um in Zukunft die Abhängigkeit von Herbiziden zu verringern und die landwirtschaftliche Produktivität zu steigern.

Acknowledgments

Therefore, I would like to express my deepest gratitude to my supervisor, Prof. Dr. Dominik Grimm. His guidance, support, and advice shaped my development as a researcher. His mentorship made this sometimes challenging journey not only possible but also truly rewarding. I am incredibly grateful for the warm and collaborative atmosphere at GrimmLab, where encouragement and teamwork are always present. A special thank you to Dr. Richa Bharti, one of the group's earliest members, for being an inspiring mentor and friend.

I would also like to thank Ashima Khanna, Josef Eiglsperger, Anna Fischer, and Jasmin Schneider. Your constant willingness to listen, share ideas, and offer encouragement means a lot to me. My deepest thanks go to Maura John and Sofia Martello for their invaluable help proofreading this thesis and all our discussions at work. Thank you, Jonathan Pirnay, especially for helping me with these pesky GPU errors that popped up from time to time. Thank you, Prof. Dr. Florian Haselbeck - I learned a lot from our discussions.

This work would not have been possible without the contributions of our project partners. A sincere thank you to Michael Grieb and Raymond Ajekwe from the Technology and Support Centre (TFZ) for conducting the drone missions, as well as to Jennifer Groth and Wouter Vahl from the Bavarian State Research Center for Agriculture for managing the Moving Fields facility. Our exchange has been incredibly valuable, especially in understanding the complexities of growing weeds systematically in a large-scale greenhouse. I learned that, in the world of weeds, they often grow when you least expect it - but I also discovered they can behave quite differently when you actually need them. I also want to express my gratitude to all those who contributed to annotating the datasets. Our painstaking work crafting pixel-perfect segmentation masks was as much an art as science.

Finally, I would like to thank Prof. Dr. Kang Yu for serving as the second reviewer of this thesis, and Prof. Dr. Clemens Thielen for being the chair of the examination committee. Special thanks to Marina Zapilko for her exceptional administrative support throughout this process.

To my dear friends and family, thank you for your unconditional support and the welcome distractions that always came when I needed them most. Lastly, to my love, Franziska: your unconditional support means the world to me. I can't find the right words to thank you for all the love and care you show me.

List of Publications

Main Publications

PUBLICATION A

Nikita Genze, Raymond Ajekwe, Zeynep Güreli, Florian Haselbeck, Michael Grieb & Dominik G. Grimm (2022). Deep learning-based early weed segmentation using motion blurred UAV images of sorghum fields. *Computers and Electronics in Agriculture*, 202, 107388

PUBLICATION B

Nikita Genze, Maximilian Wirth, Christian Schreiner, Raymond Ajekwe, Michael Grieb & Dominik G. Grimm (2023). Improved weed segmentation in UAV imagery of sorghum fields with a combined deblurring segmentation model. *Plant Methods*, 19(1), 87

PUBLICATION C

Nikita Genze, Wouter K. Vahl, Jennifer Groth, Maximilian Wirth, Michael Grieb & Dominik G. Grimm (2024). Manually annotated and curated Dataset of diverse Weed Species in Maize and Sorghum for Computer Vision. *Scientific Data*, 11(1), 109

Further Publications

FURTHER PUBLICATION A

Natalia Bercovich, **Nikita Genze**, Marco Todesco, Gregory L. Owens, Jean-Sébastien Légaré, Kaichi Huang, Loren H. Rieseberg & Dominik G. Grimm (2022). HeliantHOME, a public and centralized database of phenotypic sunflower data. *Scientific Data*, 9(1), 735

FURTHER PUBLICATION B

Larissa Barl, Betina Debastiani Benato, **Nikita Genze**, Dominik G. Grimm, Michael Gigl, Corinna Dawid, Chris-Carolin Schön & Viktoriya Avramova (2025). The combined effect of decreased stomatal density and aperture increases water use efficiency in maize. (To Appear in Scientific Reports)

Contents

1 Introduction

Agriculture aims to provide food, feed and fiber, and ensure food security [1]. However, it faces significant challenges as global food demand is rising due to population and consumption growth. Therefore, agricultural output needs to increase substantially [2, 3, 4, 5]. In the last centuries, agriculture continuously expanded by converting forests into farmland to boost food production [6, 7]. Nevertheless, this expansion impacts carbon emissions and contributes to climate change [8, 9, 10, 11, 12]. Moreover, it is limited by urbanization and infrastructure constrains, as nearly half of the habitable land is already used for agriculture [13]. Furthermore, fertile farmland is currently *decreasing*. Each year 1% of fertile farmland is lost due to various factors, including degradation and desertification [14]. Therefore, it is crucial to utilize fertile soil to its full potential by reducing yield loss [15]. Key techniques for achieving this include weed management using herbicides or through alternative farming systems.

1.1 Yield Loss

Yield loss is defined as the difference between possible yield of a healthy crop and actual yield of a stressed or diseased crop [16, 17]. On the one hand, crops might be either stressed by abiotic factors like missing irrigation, wrong temperatures or nutrient deficiency. On the other hand, the presence of harmful biotic factors like weeds, animals or pathogens might negatively affect crop growth. Research shows that weeds contribute significantly to yield loss [15]. Estimates indicate that they cause a loss of 30% - 60% of potential yield worldwide, as they compete for nutrients, water, space, and light [4, 3, 2, 18]. Moreover, weeds serve as refuges for pests such as insects, fungi or bacteria which may further harm crops. Finally, weed infestation interferes with harvesting, and reduces product quality by lowering protein levels in grain or decreasing fruit size [19].

♦

Weed management is crucial to maintain agricultural productivity and protect food security. Nonetheless, it is highly dependent on the farming system. Farmers may manage weeds using several methods: "preventative" methods focus on preventing weed establishment; "cultural" practices aim to maintain a clean field to reduce the weed seed bank; "mechanical" techniques include mulching, tilling, and cutting; "biological" approaches employ natural enemies like insects, grazing animals, or diseases; and "chemical" methods utilize herbicides [20]. In Conventional Agriculture, farmers primarily use herbicides or mechanical devices, because manual labor is expensive and limited [21]. Historically, improvements in cultivation and technology gave farmers the possibility to manage larger fields or cropping systems [22]. Nowadays, they may utilize additional technologies, like thermal or electrical weed control. Thermal strategies employ flame weeding, hot water, or steam application to damage plants through intense heat [23, 24]. Electrical weeding utilizes high-voltage to disrupt the cellular structure of weeds [25, 26].

Herbicides are chemicals used to kill or inhibit the growth of unwanted plants. Various types exist depending on the mode of action, selectivity, and time of application. Examples of modes of action include disrupting critical biological functions such as photosynthesis, enzyme activity, or cell division [27]. Additionally, these substances can be selective and target only specific weed species or non-selective, affecting all plant types they come into contact with. Farmers use them not only for early-season weed management but also shortly before harvest, for example, to reduce grain moisture in wheat [28] or control late-emerging weeds [29]. Over the last century, advancements in herbicides led to an increase of yield by more than 200% globally [30, 31, 32, 33]. Regardless, their excessive usage leads to amplified resistance in weeds, threatening to make most chemical substances non-functional by 2050 [34]. In the last century, genetically modified crops were developed to withstand herbicides, which led to their increased usage - especially glyphosate [35]. This significantly contributed to resistant weeds, and nowadays over 220 species are immune globally [36]. These substances do not only negatively impact the environment, such as groundwater pollution [37], soil contamination [38], loss of biodiversity [39], but also for humankind [40]. Their strong biological activity makes them toxic to humans [41] and animals [42, 43, 44, 45]. Here, research shows that some herbicides may cause cancer [46]. Due to these risks, governments strictly regulate them at the national and state levels, ban additional varieties from use, and limit their development [47]. For example, in 2019, the European Union announced the European Green Deal, a policy framework to reach climate neutrality by 2050 [48]. Their Farm-to-Fork strategy aims to reduce the impact of herbicides and fertilizers while increasing the use of alternative farming systems.

♦

Farming systems, such as Organic or Conservation Agriculture, gained more attention as concerns about the sustainability of Conventional Agriculture have risen [49]. Organic Agriculture has many advantages: it prohibits synthetic inputs like pesticides and mineral fertilizers [50]. Further, it enhances soil quality and plant diversity through crop rotations and biological processes [49]. Also, it boosts biodiversity and lessens environmental impact, making organic farming more sustainable. For example, soil in organic farming systems retains more water, which is more efficient during droughts and heavy rainfall [51, 52]. Nowadays, global demand for organic food is rising [53, 54]. However, organic farming commonly produces lower crop yields than conventional systems, necessitating larger land areas to sustain output [1, 50, 55]. Moreover, farming systems are complex to compare, as an increase in biodiversity might outweigh lost yield [56]. Therefore, more extensive case studies are needed, as many factors must be considered. Moreover, missing infrastructure, additional labor requirements, and economic barriers hinder market access and certification for organic practices [57]. Conservation Agriculture [58] aims to minimize soil disturbance and promote permanent soil cover and crop rotations to combat soil degradation [59]. Therefore, it enhances natural biological processes below and above the soil and makes nutrient and water intake more efficient. While organic farmers benefit from adopting these practices, no single farming method works for every field.

1.2 Automation in Agriculture

Automation addresses labor shortages since the Industrial Revolution, as engineers have developed machines to support agricultural production. Additionally, automation helps to make agriculture more sustainable as it may optimize synthetic inputs, such as fertilizers and herbicides, and conserve water. Historically, inputs were applied at averaged rates and scaled for the complete field using equipment that treats multiple crop rows at once [60]. Advancements in key technologies like Global Positioning System (GPS) enabled automated planters, thinners [61], and harvesters [22, 62]. Nowadays, farmers widely use them for pest control, fertilization, irrigation, sowing, and harvesting [3].

♦

More recently, Unmanned Ground Vehicles (UGVs) are being developed to perform various tasks in agricultural fields. They enhance plant breeding and Precision Farming through fine-scale sensing and actuation [63], which significantly reduces labor input. Additionally, they may be integrated into the workflow, from crop establishment over cultivation to harvest [64]. However, they face various challenges. The high variation in environmental conditions on the field makes autonomous driving difficult, as factors such as uneven terrain, changing weather conditions, and unpredictable obstacles require advanced sensing and adaptive control mechanisms. Also, automatic guidance relies on crop-row detection, which works well when crops contrast visually or spectrally with the soil. However, it becomes problematic when dealing with the complexity and diversity of intra-row weeds or different cropping systems. Due to these challenges, governments highly regulate UGVs while operating on the field [65]. Therefore, manufacturers must include safety features like slow velocity and the requirement of manually pre-defined drive paths.

♦

As weeds do not spread evenly in agricultural fields [22, 4], Site-Specific Weed Management (SSWM) may reduce their impact, as it targets weeds based on their precise location within a field [66, 67]. As an essential element in Precision Agriculture, it consists of three key components [68]: First, a `Weed Sensing System` is designed to identify and localize weeds and crops and measure their density and species composition parameters. Second, a `Weed Management Model` integrates knowledge about crop-weed competition, population dynamics, and the effectiveness of different control methods into a decision-making algorithm. This model aims to optimize weed treatment strategies based on the density and composition of weed species while also considering economic goals and environmental constraints. Lastly, a `Precision Weed Control Implement` directs weed removal based on the recommendations from the `Weed Management Model`. This process usually involves mechanical weed control using an UGV or chemical control through spot spraying systems attached to a tractor [69].

The goal is to optimize herbicide use, reduce costs, and minimize environmental impact. Several early studies removed economically significant weeds site-specifically. They concluded that higher net returns may be achieved [70, 71, 72, 73, 74]. However, the weed patches in these studies were identified and mapped manually by the farmer or crop consultant. Novel systems developed over time as hardware became more accessible. These range from determining the presence of a weed in an image to identifying specific weed species and finally characterizing and locating weed species in images [75]. Early developments in reflectance-based detection separated living plants from the soil [76]. However, changes in color during growth or ambient light demonstrated the need for more robust methods [77]. Hyper-spectral sensors, which cover hundreds of spectral bands, successfully address these challenges [78]. However, their high costs and intensive computing requirements pose significant barriers. Additionally, it is crucial to calibrate these sensors frequently [79]. In contrast, multi-spectral sensors collect only a few spectral bands [70], but their coarse spatial resolution limits early weed detection. Therefore, despite their vulnerability to illumination changes, visual sensors using the red-green-blue (RGB) color space remain popular in weed detection as they have a high spatial resolution and are inexpensive.

1.3 Recent Advancements in Weed Control

In the current century, advanced technology for automatic weed control has been shown under constrained conditions [80]. However, commercial-scale applicability to various crops, weed species, and growing conditions has yet to be demonstrated [34, 81]. While some commercial weed control robots exist, they struggle to accurately detect crops and weeds under varying or complex field conditions [82, 22]. Currently, the most critical technological bottleneck preventing the realization of fully automatic weed control is the lack of a robust `Weed Sensing System` that can detect and distinguish among closely related crop and weed species in the field environment [66, 83, 22, 84]. Therefore, weed detection with visual RGB sensors is highly dependent on controlled environments [75]. Thus, it is challenging to find practical algorithms to analyze images and identify weed pixels against the background in varying conditions [85, 86, 87].

♦

Remote Sensing collects and analyzes data about an object or environment from a distance using sensors on satellites, Unmanned Aerial Vehicles (UAVs), or aircraft [88]. It transforms modern agricultural practices, providing real-time data on crop health, soil conditions, and environmental factors. These technologies may capture RGB, multi-spectral, and hyper-spectral images at a wide range of scales to detect plant stress, monitor growth, and optimize irrigation [89, 90] by collecting high-resolution images across large agricultural fields [91, 92, 93, 94]. Compared to tractors or UGVs, UAVs avoid soil compaction and cover the same area in less time with high spatial and temporal precision [86, 3]. Nevertheless, their operation highly depends on the current weather, special training is necessary, and the battery pack's capacity limits their flight time [86]. Furthermore, images captured by UAVs vary in illumination, which poses a significant bottleneck to image processing.

Artificial Intelligence (AI) is enhancing agriculture through improvements in various domains [95], such as crop monitoring [96], insect monitoring [97], resource optimization, and autonomous machinery [98, 99]. Decision support systems and predictive analytics help farmers make data-driven choices [100], while robotics and Remote Sensing automate tasks such as planting, monitoring, and harvesting. This technological approach also strengthens climate resilience by managing water scarcity [101, 102] and monitoring environmental changes. In livestock management, AI contributes to healthier animals through scouting and tracking. Furthermore, AI optimizes post-harvest processes, forecasts market demand, and streamlines logistics within the supply chain [103, 104]. The continuous advancements of sophisticated sensors and algorithms are opening new grounds towards efficient agriculture while reducing the effects of human-driven climate change [95].

The combination of AI - especially Deep Learning (DL) - with Remote Sensing has a massive potential in weed detection. Farmers may scout their fields swiftly and manage weeds early in the growing season. Moreover, DL algorithms are more robust to environmental changes than classical approaches, as they extract features from underlying data.

1.4 Current Challenges

While this research field advances rapidly, it is still a novel discipline. Therefore, AI-based algorithms face several limitations that create not only technical but also societal challenges [105, 106, 107, 108]. In summary, we motivate this thesis based on the technical difficulties stemming from UAVs in natural environments and introduce solutions based on DL and discuss further societal challenges.

♦

The performance of weed detection algorithms varies significantly due to changes in natural agricultural environments. Here, climate, weather, or soil differences affect the resulting UAV captures. Climate affects the composition and growth patterns of weeds on the field. Moreover, some weed species might look differently in other geographic areas. They also change their appearance rapidly during growth. Therefore, the growth stage - a distinct phase of the plant's life cycle - is an essential variable for early weed detection. Usually, farmers remove weeds as soon as possible in the growing season to protect the sprouting crops. However, they cannot consistently achieve perfect results, leading to the presence of larger weeds during the season. Some weed plants might also survive harvest and seedbed preparation in the upcoming season. These factors lead to a wide variety of possible weed shapes.

The weather is another variable to consider. On the one hand, UAV images taken in bright sunlight may suffer from shadows, reflections, and contrast issues, while on the other hand, cloudy or low-light conditions can lead to under-exposure and loss of detail. Moreover, partly cloudy skies with frequent sun changes lead to inconsistent illumination in the images.

In addition, the soil plays a further critical role. Diverse weed species grow in various soil types, and the visual appearance of the soil varies due to different cultivation techniques and cropping practices. For instance, the presence of mixed vegetation in intercropping systems may significantly alter the soil's appearance.

As mentioned, Remote Sensing with UAVs is essential in advancing Precision Agriculture. However, their operation poses another critical challenge in weed detection tasks. First, the flight altitude has a direct effect on image resolution. Lower altitudes provide higher spatial resolution but cover smaller areas. Therefore, the total efficiency diminishes due to limited battery capacity. Higher altitudes cover a larger area but at a lower spatial resolution. Because of that, finer details that are necessary for early weed detection tasks might be overlooked. More importantly, motion-blur is an inherent issue with UAVs due to their movement, turbulences, wind pressure, and cameras with rolling shutters. While DL models may handle varying natural environments with additional data, distortions caused by blur remove crucial edge features, which reduce the overall accuracy.

Therefore, we address challenges with regard to motion-blur in two contributions. First, we investigate the robustness of DL models to varying levels of motion-blur in UAV images. Section 3.1 outlines the following publication in more detail:

Publication A

Nikita Genze, Raymond Ajekwe, Zeynep Güreli, Florian Haselbeck, Michael Grieb & Dominik G. Grimm (2022). Deep learning-based early weed segmentation using motion blurred UAV images of sorghum fields. *Computers and Electronics in Agriculture*, 202, 107388

In our second contribution, we *enhance* our model. Here, we develop a combined model that first sharpens the images and afterward distinguishes between crops, weeds, and soil. Section 3.2 details the results of the following publication:

Publication B

Nikita Genze, Maximilian Wirth, Christian Schreiner, Raymond Ajekwe, Michael Grieb & Dominik G. Grimm (2023). Improved weed segmentation in UAV imagery of sorghum fields with a combined deblurring segmentation model. *Plant Methods*, 19(1), 87

Further, we address a major bottleneck of DL models: data - more specifically - the lack of sufficient data. As it is commonly known, DL models require extensive labeled datasets to be effective [109, 110, 111, 112]. Significant breakthroughs in Computer Vision were only possible because of large datasets. The most prominent example is AlexNet, a Convolutional Neural Network (CNN) that substantially outperformed all other submissions in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012 [113]. However, labeled datasets are challenging to obtain in several domains. For example, the recording hardware in medical imaging is expensive [114] and should

be certified, which makes the image acquisition process difficult. Farmers may use inexpensive devices in agriculture, such as digital cameras, mobile phones, or consumer-grade UAVs. Here, the main bottleneck is obtaining the corresponding ground-truth. Meanwhile, most current weed detection studies involve small and tailored datasets for a specific crop and weed composition [115]. Here, the shape of the plants is limited, and DL models might be unable to detect them in different growth stages or other natural conditions. Related to this, agricultural datasets are not only limited in their size but also in their availability, which slows down research on this topic [116].
In summary, this contribution addresses two challenges. First, regarding data scarcity, we publish a large-scale image dataset containing over 94'000 high-definition images of 28 weed species commonly found in agricultural fields in Germany. We annotated and curated these images manually using FAIR [117] principles. Hence, our contribution benefits agricultural research in several ways: Thanks to a single data collection protocol and labeling standard, it enhances agricultural research by providing a consistent dataset with uniform features and labels. Additionally, the dataset captures diverse growing stages, as images were collected frequently - at least once daily - to capture subtle changes in fast-growing weeds. Using this data, we partly address the challenge of varying environments in agriculture by analyzing growth stages and multiple weed species. Therefore, we experimented with classifying 27 weed species with fine-grained distinction of their growth. Moreover, other researchers might use this dataset as a standard benchmark for developing and evaluating novel DL models, which is difficult with multiple small datasets and inconsistent labeling protocols. Further, multiple Computer Vision tasks in the weed detection domain could be evaluated, as we did not only publish bounding boxes but also segmentation masks (see Section 2.2 for more details). Finally, our dataset could give insights into the growth dynamics of different weed species as we tracked the growth of individual plants over time. Section 3.3 summarizes the following publication in more detail:

Publication C

Nikita Genze, Wouter K. Vahl, Jennifer Groth, Maximilian Wirth, Michael Grieb & Dominik G. Grimm (2024). Manually annotated and curated Dataset of diverse Weed Species in Maize and Sorghum for Computer Vision. *Scientific Data*, 11(1), 109

1.5 Further Contributions

This section briefly outlines two further contributions that co-evolved with the main contributions mentioned before. However, they are not directly related to the content of this thesis. First, we developed a public and centralized database for the common sunflower (*Helianthus annuus* var. *macrocarpus*). Here, the author managed and curated the phenotypic dataset and assisted with its implementation into the web service. These learnings helped manage large vision datasets, for example, our contribution outlined in Section 3.1.

Further Publication A

Natalia Bercovich, **Nikita Genze**, Marco Todesco, Gregory L. Owens, Jean-Sébastien Légaré, Kaichi Huang, Loren H. Rieseberg & Dominik G. Grimm (2022). HeliantHOME, a public and centralized database of phenotypic sunflower data. *Scientific Data*, 9(1), 735

Second, we developed a DL-based model to detect multiple stomata in maize microscopy images. Here, the author selected the final image dataset and trained several human oracles to annotate the ground-truth with high precision. Further, he trained several object detectors and evaluated their performance. Finally, the author edited the manuscript.

Further Publication B

Larissa Barl, Betina Debastiani Benato, **Nikita Genze**, Dominik G. Grimm, Michael Gigl, Corinna Dawid, Chris-Carolin Schön & Viktoriya Avramova (2025).The combined effect of decreased stomatal density and aperture increases water use efficiency in maize. (To Appear in Scientific Reports)

2 Materials & Methods

The upcoming chapter describes the materials and methods relevant to this thesis. After introducing our image acquisition process, we briefly overview various Computer Vision tasks. Next, we explain the Semantic Segmentation task in more detail, as it is a key component of this thesis. Further, we summarize the image de-blurring task. Finally, we outline critical elements involved in training DL models.

2.1 Image Acquisition

In our contributions, we collected images using two different methods. First, we conducted a large-scale experiment in an automated greenhouse to capture the growth of over 28 weed species with high spatial resolution. Second, we utilized a consumer-grade UAV to capture images of agricultural fields, as it is more efficient and represent a more realistic use case for future applications.

♦

First, we captured the initial growth dynamics of several weed species in a controlled environment. Therefore, we used the Moving Fields Facility in Freising, as it is equipped with a conveyor belt system, irrigation stations, and photo cabins, enabling automated plant movement and monitoring. Together with our partners from the Bavarian State Research Center for Agriculture (LfL), we selected weed species common to sorghum (*Sorghum bicolor*) and maize (*Zea mays*) fields in Germany based on their availability and compatibility with greenhouse conditions. We grew all plants as monocultures in micro-plots and automatically photographed them daily from sowing to harvest. Here, we adjusted the number of these micro-plots and the seed density per species based on differing germination rates of the seeds. Additionally, we assessed seedling emergence and removed any weeds if they overlapped or grew too large. In addition to weeds, we grew different varieties of maize and sorghum using the same protocol.

♦

Further, we prepared field experiments with our partners at the Technology and Support Centre in the Centre of Excellence for Renewable Resources (TFZ). Therefore, we used experimental field sites around Straubing. Here, the sorghum variety 'Farmsugro 180' was sown at 37.5 cm row spacing. Also, several weed species were present on these fields, consisting primarily of dicotyledons. Those include goosefoot (*Chenopodium album*), common gypsyweed (*Veronica officinalis*), wild chamomile (*Matricaria chamomilla*), and field thistle (*Cirsium arvense*).
We conducted automated flight missions with the camera pointing nadir (facing exactly downwards) with a consumer-grade UAV (DJI Mavic 2 Pro) and obtained high-resolution images with a Ground Sampling Distance (GSD) of 1 $\frac{mm}{px}$. Additionally, we used two different capturing settings, as shown in Figure 2.1a. With the first setting, 'Hover and Capture', the UAV stopped and stabilized before every image capture, which led to

sharp plant borders. However, the UAV did not stop during the second setting 'Capture at equal distance'; consequently, motion-blur was visible in the images. Figure 2.1b shows an example capture of the same area with both settings, while Figure 2.1c shows a comparison with different plant species.

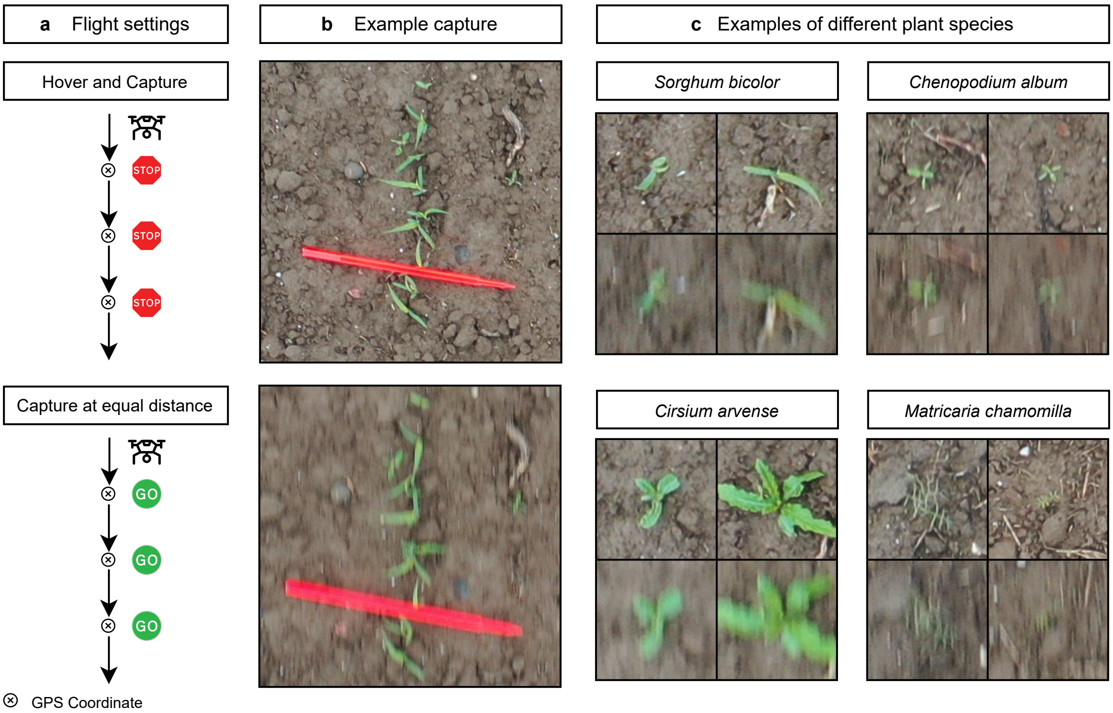

Figure 2.1: Different image capturing settings of an UAV. **a** 'Hover and Capture' leading to sharp plants due to stopped motion, while 'Capture at equal distance' does not stop the UAV leading to visible motion-blur. **b** Cropped example capture indicating the GSD with a red marking stick of 40 cm **c** Examples of different plants captured with both settings. This figure is adapted from the original Publication B [118].

2.2 Computer Vision Tasks

Computer Vision focuses on identifying specific objects, features, or activities within an image. Here, Supervised Learning - a subset of AI - is an approach where a model learns to map input features to output labels. This technique enables the development of robust models to address real-world problems across various tasks and is frequently used in weed detection [119, 120, 121, 78, 122, 123, 124]. Nevertheless, the complexity different tasks in Computer Vision can vary significantly, as illustrated in Figure 2.2.

Using an image as illustrated in Figure 2.2a, the least precise task is Image Classification (Figure 2.2b), which requires prior processing to utilize individual plant cut-outs and thus is not usable for weed detection tasks. Therefore, Object Detection (Figure 2.2c) is

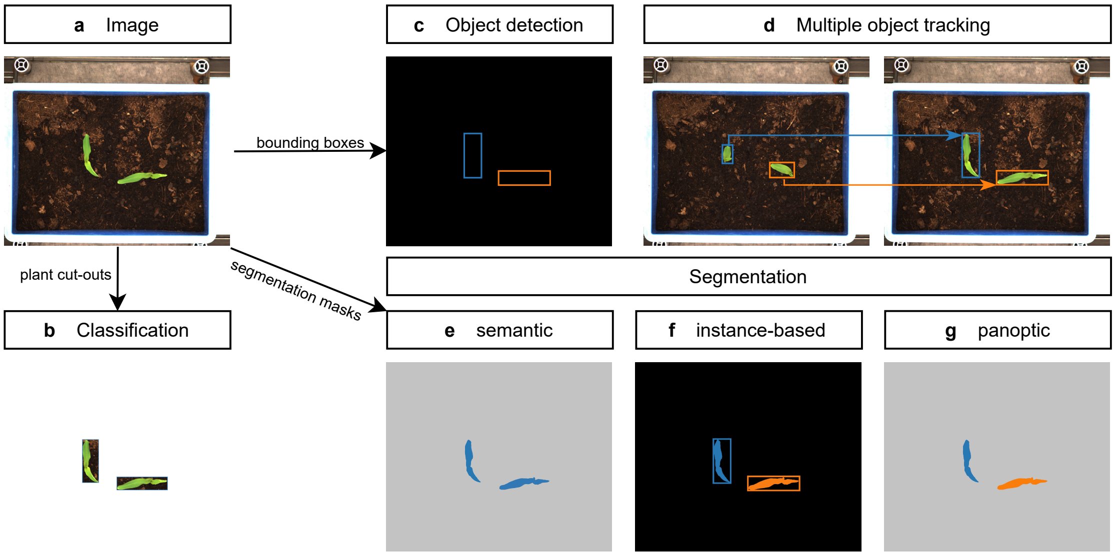

Figure 2.2: Comparison of different Computer Vision tasks used in plant phenotyping. Plants **a** can be classified **b** as individual cut-outs, detected by bounding boxes **c**, tracked through time **d**, or segmented containing pixel-wise information. This can be done either semantically **e**, by instance **f** or by the combination of both called panoptic segmentation **g**. This figure is adapted from the original Publication C [125] for better readability.

commonly employed, as it allows for the identification and localization of multiple plants simultaneously. Additionally, by analyzing a sequence of images, the growth dynamics of plants can be studied through Multiple Object Tracking (Figure 2.2d). However, specific weed management applications require a more detailed differentiation than what bounding boxes may provide. Here, segmentation may be suitable, as it divides images into more fine-grained sections. This task may be further divided into three sub-tasks, each differing in how they treat objects: Semantic Segmentation (Figure 2.2e) assigns each pixel of an image into specific categories. However, it does not differentiate between individual instances, which is necessary to count objects. To address this, Instance Segmentation (Figure 2.2f) can be employed. However, this task does not detect uncountable regions like sky or soil. Finally, Panoptic Segmentation (Figure 2.2g) combines Semantic and Instance Segmentation and can comprehensively represent the entire scene.

Annotation is the process of providing accurate ground-truth information, which varies significantly depending on the specific task. In Image Classification, a single class is assigned to each image. In contrast, object detection tasks require identifying and locating multiple objects by drawing bounding boxes around them and specifying their class. While bounding boxes offer a general outline, segmentation provides more detailed locations. Semantic segmentation involves labeling each pixel in an image with a specific class and creating a segmentation mask. Instance segmentation provides separate masks for each individual instance, allowing for even greater precision.

In our initial two contributions, we approach weed detection as a Semantic Segmentation task. Therefore, in the following section, we go into more detail about DL-based techniques for this purpose.

2.3 Semantic Segmentation Models

The evolution of DL has significantly advanced the field of Semantic Segmentation, with CNNs forming the backbone of state-of-the-art models. Since the seminal work of Long et al. with Fully Convolutional Networks (FCNs) [126], architectures such as `U-Net` [127] or `DeepLabV3+` [128] (see Figure 2.3) have set new benchmarks by utilizing hierarchical feature extraction, multiscale context aggregation, and encoder-decoder structures.

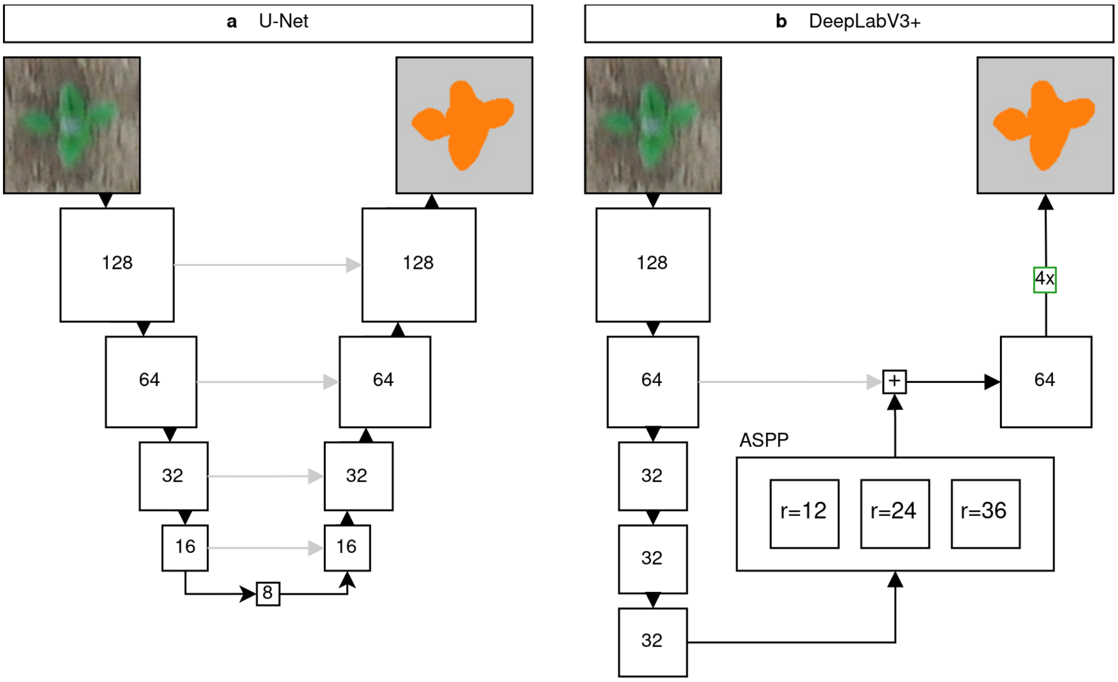

Figure 2.3: Comparison of two Semantic Segmentation models `U-Net` and `DeepLabV3+`. The spatial resolution is denoted inside the individual layers. Starting with an input image patch of 256×256 px^2, the spatial resolution of the feature maps is decreasing in the encoder (left side) and increasing in the decoder (right side). **a** `U-Net` with an encoder-decoder structure and skip connections (in gray). **b** `DeepLabV3+` with an additional Atrous Spatial Pyramid Pooling (ASPP) layer. Dilation rates in this layer are abbreviated with r.

Ronneberger et al. proposed `U-Net` with a focus on biomedical image segmentation. The model architecture is named for its characteristic `U` shape, which consists of two symmetrical paths: a contracting path (encoder) and an expansive path (decoder), as shown in Figure 2.3a. The model is designed to capture context and spatial information as it combines high-resolution features from the encoder with up-sampled feature maps

in the decoder. Each down-sampling step doubles the number of feature channels and allows the network to learn more complex features at deeper layers, as shown on the left side in Figure 2.3a. This path also reduces the spatial dimension of the feature maps due to max-pooling operations, resulting in a hierarchical representation of the input image.
At the bottom of the `U`, one bottleneck layer bridges the encoder and decoder. This layer captures the most abstract features. Next, the decoder recovers the spatial information and constructs the segmented output. It consists of several up-sampling blocks, which contain up-convolutions to double the spatial dimension. Skip connections concatenate high-resolution features from the encoder with the up-sampled output in the decoder and ensure that fine-grained details are preserved. Additionally, the decoder re-uses features from the encoder, making the model more efficient. The symmetrical architecture allows the model to efficiently learn global context and local details, which is essential for accurate segmentation. Additionally, its modular architecture enables easy integration with other DL models. Here, many variants of `U-Net` were proposed in different domains to enhance performance [129, 130, 131, 132, 133].
In contrast, `DeepLabV3+` is a more recent and more complex architecture [134], as shown in Figure 2.3b. As an advanced DL model for Semantic Segmentation, it improves upon `DeepLabV3` [135] by incorporating an encoder-decoder structure. Additionally, it uses an ASPP layer to capture multiscale contextual information using dilated convolutions with different rates. This type of convolution increases the receptive field while preserving high spatial resolution in the feature maps, which improves overall context understanding and accuracy in fine-grained segmentation.

♦

One possible variation replaces the original `U-Net` encoder with `Residual Networks (ResNets)` to enable deeper feature extraction without compromising training efficiency. The `ResNet` architecture [136] was introduced in 2015 by He et al. and has become one of the most influential CNN architectures in DL [137]. The key innovation of `ResNets` lies in residual learning, which helps mitigate the vanishing gradient problem. By enabling deeper networks, `ResNets` achieve state-of-the-art results in numerous Computer Vision tasks, including Image Classification, object detection, and segmentation, and have influenced many subsequent DL architectures.
The vanishing gradient problem is a well-known challenge in training deep Neural Networks, where gradients used for updating the model's weights become increasingly small as they propagate backward through the layers. These tiny updates can lead to slow convergence or cause the network to stop learning completely.
Therefore, `ResNets` add shortcut connections to pass over layers and allow gradients to flow directly through these shortcuts during back-propagation, similarly to skip connections of `U-Net`. This happens inside the key component of `ResNets`: a residual block, which consists of two or three convolutional layers, batch normalization, and Rectified Linear Unit (ReLU) activation. More importantly, a shortcut connection bypasses these layers and directly adds the block's input to the output. This design ensures that the network learns residual functions instead of trying to learn the entire transformation from input to output directly, which leads to overall larger gradients.

This network architecture consists of several stages, each containing multiple residual blocks. Similar to `U-Net`, as the network progresses through these stages, the number of filters in each convolutional layer increases while the spatial dimension decreases through strided convolutions. This hierarchical structure allows the network to capture complex patterns and hierarchical features in the data. The final layers consist of a global average pooling layer, which reduces each feature map to a single value, followed by a fully connected layer that produces the final output.

`ResNets` come in various depths, such as `ResNet-18`, `ResNet-34`, `ResNet-50`, `ResNet-101`, and `ResNet-152`, where the trailing number indicates the total amount of layers in the network. `ResNet-50` and deeper variants introduce an additional bottleneck design within the residual blocks to reduce the computational complexity.

Finally, `EfficientNet` [138] is another popular encoder, which employs a compound scaling method to balance network depth, width, and resolution. This approach results in an architecture that is both accurate and computationally efficient. However, when comparing `EfficientNets` with `ResNets`, the advantages of architectural changes may be less significant than improvements made through training or scaling strategies, such as using deeper `ResNet` variants [139].

2.4 Image De-blurring

This section goes into more detail about image quality, motion-blur and consequently recovering image quality through image de-blurring methods.

Key image quality attributes include sharpness, which indicates the level of detail in an image. Other important factors that affect image sharpness are lens design, sensor resolution, and environmental conditions. Additionally, image qualities such as contrast, brightness, color accuracy, and artifacts like noise or compression play a crucial role in the overall image perception. In summary, it is important to ensure that these factors are optimized for achieving high image quality, both technically and perceptually.

♦

Nevertheless, blur is a *common* artifact in photographing dynamic environments. Motion-blur occurs when the camera or the objects of interest move during capture, leading to smeared images, especially at high speeds or in low-light conditions. Typically, motion-blur comes jointly with out-of-focus blur, which arises from improper lens focus, and reduces image sharpness [140].

Motion-blur can be classified into being *uniform* or *non-uniform*, based on the type of movement during exposure. Their causes differ depending on motion patterns and scene conditions [141]. Uniform motion-blur typically results from linear and consistent movements. In contrast, non-uniform blur is caused by varying blur directions or intensities, for example, during camera rotation or out-of-plane movement. Correcting non-uniform blur is challenging because it requires spatially varying de-blurring techniques.

From the mathematical point of view, the blurring process can be simplified with the following equation [142], where a degraded image g results from convolving a sharp image f with a point-spread function (PSF) and adding noise.

$$g = f \circledast h + n$$

Blind and non-blind image de-blurring are two approaches to restore sharp images from blurred ones, differing mainly in their knowledge of the PSF. In non-blind image de-blurring, the PSF is known or pre-estimated, which simplifies the restoration process. Here, such algorithms may apply Wiener filter [143] or DL techniques [144, 145] to recover the sharp image. This technique generally produces more accurate results than blind de-blurring. However, its effectiveness depends on the accuracy of the provided PSF, and it is more suitable for controlled environments with known blur characteristics. In contrast, with blind image de-blurring, the PSF is unknown and is estimated along with the sharp image. This additional step makes the problem highly challenging and ill-posed, as multiple possible sharp images could correspond to the same blurred observation. In summary, blind de-blurring is widely applicable in real-world scenarios with unpredictable blur type and extent.

♦

Blind image de-blurring models are designed to reverse or mitigate the effects of blur. These models are usually trained on datasets of blurred and sharp image pairs to learn underlying patterns and characteristics of blur. Over the years, researchers introduced different DL-based approaches [146, 147, 148, 149]. Once trained, these models enhance sharpness and clarity in unknown degraded images. However, some challenges remain. One downside is the need for large, high-quality datasets with both - *naturally* blurred and sharp images - to train the models [150]. Additionally, de-blurring may not always fully recover image quality, especially when the blur is severe or occurs under complex environmental conditions. This complexity makes the solution space non-unique and sensitive to noise. For instance, large artifacts may be visible despite small errors in estimating the PSF.

♦

Recently, Chen et al. proposed a simple yet effective baseline for image de-blurring that surpasses state-of-the-art methods while significantly reducing computational complexity [151]. The authors introduce a `Nonlinear Activation Free Network (NAFNet)`, which removes conventional nonlinear activation functions and instead relies on element-wise multiplication only. By analyzing the complexity of existing models, the authors decompose their architectures into inter-block and intra-block complexity and gain performance without increasing the complexity of the system. Their baseline, derived from a `U-Net`-like structure with minimal modifications, outperforms previous state-of-the-art image de-blurring models. This model achieves comparable or superior results in their experiments with a small fraction (8.4%) of the computational cost.

2.5 Training Procedure

The following section outlines five key concepts for training Semantic Segmentation models: Dataset Preparation, Loss Function Design, Regularization, Training Configuration and Hyperparameter Optimization. Please refer to the individual contributions in the appendix for more details on how we applied these concepts.

Dataset Preparation In supervised weed detection, the dataset is typically divided into three subsets: the largest portion for training and the remaining two for validation and testing. A standard ratio is 70% for training, 20% for validation, and 10% for testing. However, the limited availability of labeled data can lead to discrepancies in the distributions of these subsets, which complicates the evaluation process. To address this issue, an alternative approach is to use k-fold cross-validation [152]. This method involves dividing the dataset into k subsets of approximately equal size. Here, each subset is used as a validation set once, while the other subsets are used for training. This setup yields k iterations, which reduces dependence on a single distribution. However, it is important to note that this approach will significantly increase the training time. Dataset stratification ensures that class distributions in training, validation, and test sets reflect the original dataset [153]. This strategy might mitigate issues with class imbalance and is useful in combination with small datasets where random splitting might exclude rare classes.

Loss Function Design The loss function is a crucial component in the training process. It guides the model and improves predictions by adjusting its parameters using optimization algorithms like gradient descent. As a scalar value, it quantifies the error by calculating the difference between predictions and the actual labels. The careful selection of a loss function is essential due to its dependency on the dataset, class distribution, and other task-specific requirements [154, 155]. For instance, pixel-wise cross-entropy is usually used in Image Classification tasks. However, this loss approaches the Semantic Segmentation task as a classification problem, which is unsuitable for high imbalance. In contrast, the Dice Loss is preferred for these tasks since it is region-based and can handle such imbalances [155].

Regularization A significant concern when training DL models is a phenomenon known as overfitting. It occurs when a model has a high capacity, and learns not only the underlying pattern of the data, but also the noise and irrelevant details. Consequently, its performance on the validation set generally decreases. Overfitting is more common when the dataset is small and noisy, which is typically the case in weed detection tasks. Researchers employed various regularization techniques to mitigate this issue. These techniques either reduce the dimensionality of the parameter space or diminish the effective size of each dimension [156]. Examples include L1 [157] and L2 [158] regularization, also called weight decay. These techniques add a penalty term to the loss function that discourages large weights. L1 regularization reduces *some* weights to zero, while L2 regularization generally makes all weights smaller. Another method is dropout

[159], which randomly turns off specific units in the hidden layers during training. This process prevents, that features become overly dependent on each other. Early Stopping [160] is another helpful strategy; it terminates the training process when the validation error is as low as possible. Finally, data augmentation is a standard method to artificially increase the size of the training dataset. Examples are random transformations, such as flipping, rotating, cropping, scaling, or adding noise.

Training Configuration CNNs rely on several key parameters to guide the training process and optimize performance. The learning rate controls the size of weight updates during backpropagation and influences the speed and stability of convergence. The batch size determines the number of training samples processed before updating the model's parameters and affects both memory usage and the stability of the gradient computation. The number of epochs defines how many times the entire dataset is passed through the network during training, which has an impact on the model convergence and potential overfitting.
A learning rate scheduler [161] offers several benefits by dynamically adjusting the learning rate during training. Controlling the learning rate ensures that the model converges efficiently. Starting with a larger step size helps to escape suboptimal local minima. Gradually reducing it, refines the model's parameters. This adaptive approach enhances convergence stability and reduces the risk of overshooting the optimal solution. A well-designed learning rate schedule helps mitigate oscillations during training, especially in high-dimensional loss landscapes, and promotes smoother convergence [162]. It also prevents the model from stagnating early by maintaining sufficiently large updates in the initial phases and adjusting the learning rate more fine-grained later in the training progresses. As a result, learning rate schedulers often lead to improved generalization and shorter training times.
DL models trained on large datasets capture hierarchical features that generalize across related domains. Early convolutional layers learn essential patterns like edges, textures, and shapes, providing a solid foundation for new tasks and minimizing the need for extensive labeled data and resources [163]. Therefore, we employ transfer learning [164] in our work. Here, we start our training process with pre-trained weights and fine-tuned them on our weed detection datasets.

Hyperparameter Optimization While training CNNs, the primary goal is to achieve optimal generalization performance, which means reducing error on examples of the validation or test set. Therefore, researchers frequently employ hyperparameter optimization. Unlike parameters (i.e., weights) of a model, hyperparameters determine *how* the model learns. They are typically not directly transferable across different architectures and datasets [165, 166, 167], meaning they must be re-optimized for each new task. There are several methods available to adjust hyperparameters: Grid Search is the simplest method and works well when a few hyperparameters have a similar impact on the validation score [168]. In contrast, Random Search is more effective when the magnitudes of influence among the hyperparameters are imbalanced, which is often the case as the number of parameters increases [169]. However, it's usually more efficient to reduce the hyperparameter space using prior knowledge and sound assumptions. Finally, another approach is to use Bayesian Optimization [170], which selects a new set of hyperparameters based on data from previous experiments conducted during the optimization process. However, this method requires an objective function to guide the sampling of new hyperparameters.

2.6 Performance Evaluation

A proper performance evaluation is essential for robust DL models in weed detection. The following sections provide details about how we evaluated the models developed in this research. Here, we focus on both: Semantic Segmentation and De-blurring.

2.6.1 Semantic Segmentation Metrics

Unlike traditional Image Classification tasks, which predict labels for entire images, Semantic Segmentation produces a pixel-level output. This granularity makes the evaluation process more complex. Therefore, it is crucial to choose the correct evaluation metrics. This section summarizes the key evaluation metrics Confusion Matrix, Per-pixel Metrics, Sørensen-Dice Coefficient and concludes with Visual Assessment.

The Confusion Matrix provides a comprehensive overview of a model's predictions compared to the ground-truth. Figure 2.4 shows an example for three classes.

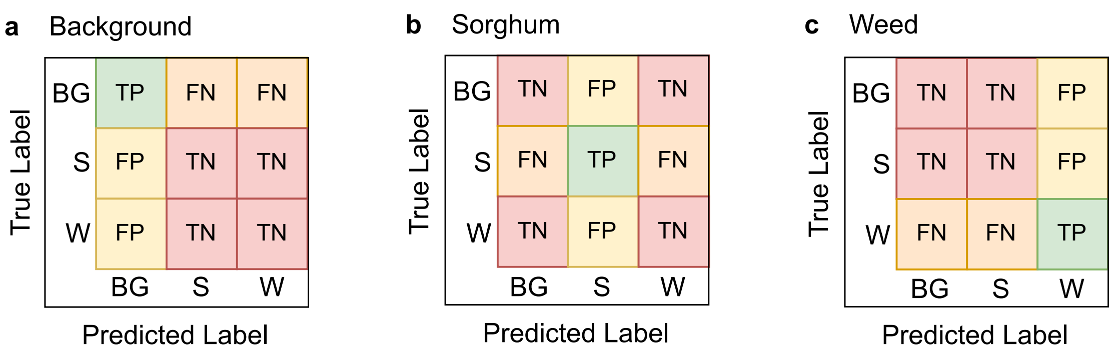

Figure 2.4: Confusion matrix showing positives and negatives for the three classes, **a** background, **b** sorghum and **c** weed.

The matrix consists of rows and columns representing the classes. In Semantic Segmentation, each cell in the matrix indicates the number of pixels belonging to the actual class (represented by the row) and that the model classifies as the predicted class (represented by the column). This setup allows for a detailed view of misclassifications for each specific class. True Positives (TP) are represented along the diagonal of the matrix. False Positives (FP) are found in the off-diagonal entries of the corresponding columns, while False Negatives (FN) are located in the off-diagonal entries of the corresponding rows.

Imbalanced class distributions lead to Confusion Matrices dominated by high counts of frequent classes (e.g., soil) and negligible entries for rare classes, making it difficult to assess the model's performance for underrepresented categories. Additionally, boundary pixels - where objects transition into different classes - often result in ambiguities and contribute to significant misclassification rates, skewing the Confusion Matrix. Choosing the right averaging method for performance metrics is essential, especially when dealing with multiple classes. Two common approaches are macro-average and weighted average. Macro-averaging first calculates the metric independently for each class and subsequently takes the average of these values, providing a balanced view. As weeds are occurring less frequently, macro-averaged metrics might be more suitable, as they give more importance to smaller classes, even if they contribute minimally to the overall dataset. Weighted average gives a realistic measure of the model's performance since each class is weighted proportionally based on its pixel count. It minimizes the influence of noisy or poorly segmented classes if they have fewer pixels, making it more robust for large-scale applications. However, it has a potential bias toward larger classes while masking poor performance in less frequent classes. Additionally, weighted average gives a less intuitive comparison across classes. However, high intuition about a metric is crucial in applications like agriculture, where smaller classes (e.g., weeds) are often as important as larger ones (e.g., crops). For semantic segmentation with high class balance, macro-average offers a more balanced evaluation by treating all classes equally.

Per-Pixel Metrics such as accuracy, precision, recall and the F1-score can be calculated from the Confusion Matrix to evaluate the model's overall performance as well as its performance on a per-class basis. Pixel Accuracy measures the ratio of correctly predicted pixels to the total number of pixels. While simple and intuitive, pixel accuracy can be misleading with class imbalance, as correctly classifying the background class can result in high accuracy despite poor performance in less frequent classes. Precision ($\frac{TP}{TP+FP}$) indicates how many of the predicted positive pixels were actually correct. In contrast, recall ($\frac{TP}{TP+FN}$) measures how many of the actual positive pixels were detected by the model. Both metrics provide insights into the types of errors, such as over-segmentation (many FP) or under-segmentation (many FN). Finally, the F1-score is the harmonic mean of precision and recall. It is mathematically equivalent to the Sørensen-Dice Coefficient (DC), but they are used in different contexts with different interpretations.

The Sørensen-Dice Coefficient is a widely used metric to evaluate the similarity between predicted and ground-truth segmentations [171]. Originating from the work of Lee Raymond Dice and Thorvald Sørensen in the mid-20th century, this coefficient quantifies the *overlap* of predicted and actual segmentations. Its sensitivity to class imbalance and small segmented regions makes it especially valuable in domains such as medical image analysis or Remote Sensing [172, 173]. It ranges from 0 to 1, and for the task of binary segmentation is defined as follows:

$$DC = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}$$

The Visual Assessment of predictions is a vital evaluation technique in Semantic Segmentation tasks with high class imbalance. While it is not strictly a metric, this additional step identifies errors that automated metrics might overlook. Here, experts analyze the model's predictions by manually comparing them directly to the ground-truth. Visual inspection reveals how well a model segments different minority classes and ensures that its performance remains reliable across all classes, regardless of their frequency. It is time-consuming and subjective; however, it highlights segmentation inconsistencies, confirms the accuracy of critical classes, and supports the refinement of the model for better real-world performance.

2.6.2 Image Quality Metrics

This section summarizes three main metrics for evaluating image quality in motion-blurred images: Peak Signal-to-Noise Ratio (PSNR), Structural SIMilarity Index (SSIM) and Learned Perceptual Image Patch Similarity (LPIPS).

The Peak Signal-to-Noise Ratio (PSNR) is widely used to assess the degree of distortion or degradation in an image. It measures the ratio between the maximum possible power of a signal in the original unblurred image and the power of the noise - blur for example. It is computed using the Mean Squared Error (MSE) between corresponding pixels of the original and the distorted image:

$$PSNR = 10 * log_{10} * (\frac{MAX^2}{MSE})$$

Here, MAX is the maximum pixel value, e.g., 255 for an 8-bit grayscale image. A higher PSNR value indicates less distortion and better quality, while a lower value reflects significant degradation. Nevertheless, this metric has limitations when applied to subjective visual perception; it assumes pixel-wise fidelity, which might not correlate with human perception of blur. For instance, an image with high-frequency details lost to blur may exhibit a similar MSE and PSNR to an image with global brightness distortion, yet the perceived quality differs significantly. Therefore, while PSNR is effective for objective evaluations and algorithm comparisons, complementary perceptual metrics like SSIM are often used alongside it for a holistic assessment of image blur.

The Structural SIMilarity Index (SSIM) is a perceptual metric that measures the similarity between two images. Designed to align with human visual perception, SSIM effectively assess image blur. Here, it highlights how features like edges and textures are altered, as blurring smooths sharp edges and reduces structural similarity. Its values range from -1 to 1, where 1 indicates identical images and values closer to 0 or negative indicate greater dissimilarity. This metric analyzes small local regions of the original and blurred images using three components: comparison of luminance (l), contrast (c), and structural information (s). Additional parameters α, β and γ control the relative importance of each component, as shown in the following formula:

$$SSIM(x,y) = [l(x,y)^{\alpha} \cdot c(x,y)^{\beta} \cdot s(x,y)^{\gamma}]$$

The Learned Perceptual Image Patch Similarity (LPIPS) is an advanced perceptual metric developed to evaluate the visual similarity between two images. Unlike traditional metrics like PSNR or SSIM, which rely on pixel-level or hand-crafted features, LPIPS measures similarity by comparing high-dimensional feature embeddings extracted from pre-trained CNNs. These embeddings make LPIPS particularly effective in capturing perceptual differences caused by image distortions, such as blur, as it aligns more closely with human visual perception.

To calculate LPIPS, a pre-trained network processes the original and distorted images and extracts feature maps at multiple layers. Afterward, the differences between these feature maps are weighted and aggregated, where a lower value indicates greater similarity and higher values suggest perceptual differences. This metric is especially effective when fine textures and details are lost due to blur and is used in multiple image-to-image tasks such as image restoration [174] and super-resolution [175].

3 Results

The results of this thesis are based on the three main contributions stated in Section 1.4. In this chapter, we provide bibliographic information, detail each author's contributions, and summarize their contents.

3.1 Weed Segmentation using Motion Blurred UAV Images

This section summarizes the following publication. The complete version can be found in Appendix A.

PUBLICATION A

Nikita Genze, Raymond Ajekwe, Zeynep Güreli, Florian Haselbeck, Michael Grieb & Dominik G. Grimm (2022). Deep learning-based early weed segmentation using motion blurred UAV images of sorghum fields. *Computers and Electronics in Agriculture*, 202, 107388

3.1.1 Author Contributions

Dominik G. Grimm, Nikita Genze and Michael Grieb conceived and designed the study. Raymond Ajekwe conducted the drone flights and the image acquisition. Nikita Genze and Zeynep Güreli labeled the dataset. Raymond Ajekwe and Michael Grieb guided the labeling process with domain expertise. Nikita Genze implemented the machine-learning pipeline and conducted all computational experiments with support from Florian Haselbeck. Nikita Genze analyzed the results with help from Dominik G. Grimm. Nikita Genze and Dominik G. Grimm wrote the manuscript with contributions from all authors. All authors read and approved the final manuscript.

3.1.2 Summary

As discussed in Section 2.4, one significant challenge in UAV-based weed detection is motion-blur, which can be caused by external factors like wind or internal factors like the UAV's flight velocity.
In this contribution, we trained a DL model to segment weeds and sorghum in motion-blurred images. Therefore, we gathered a dataset of images captured on sorghum fields. The level of motion-blur in the images varied due to the UAV's speed, mainly when it changed flight direction. The sorghum plants were at growth stage 17 according to the BBCH scale [176]. Additionally, we annotated images from two other missions to evaluate our model's performance on different growth stages (BBCH 15 and 19).
Further, we approached weed detection as a Semantic Ssegmentation task to accurately identify weeds and crop plants. We selected six different architectures designed explicitly

for this task. Additionally, we chose four feature extractors that utilize residual connections, each differing in the number of parameters. Here, we trained our models using four-fold cross-validation, along with a separate hold-out test-set, and we optimized 50 different hyperparameter configurations. Ultimately, the `U-Net` model architecture with `ResNet-34` was chosen, as it yielded the best performance.
We tested this model with three different and unseen sets of images. First, we selected a test-set from the same UAV mission of BBCH 17 to ensure that our model could generalize well on similar data. Here, a DC of over 99 % and a macro-averaged F1-score of approximately 90 % indicated strong segmentation performance. Next, we used an additional dataset featuring sorghum plants at a later growth stage (BBCH 19). In this dataset of 110 image patches, the illumination conditions were similar; however, the weed composition and the morphology of the sorghum plants differed. According to the confusion matrix, our model correctly predicted 77.6 % of all weed pixels. Additionally, it correctly classified 85.8 % of the sorghum pixels. This indicates an overall good performance.
Finally, we tested a third set of 110 image patches at an earlier growth stage of sorghum (BBCH 15), where also the illumination differed. Here, the model's performance decreased in predicting sorghum to 50 %, while the accuracy on weeds remained high. Further analysis of the misclassifications showed that mostly parts of a plant were falsely predicted as weeds.

3.2 Deblurring-Assisted Weed Segmentation Model

In this section the following publication is summarised. A copy of the full version can be found in Appendix B.

PUBLICATION B

Nikita Genze, Maximilian Wirth, Christian Schreiner, Raymond Ajekwe, Michael Grieb & Dominik G. Grimm (2023). Improved weed segmentation in UAV imagery of sorghum fields with a combined deblurring segmentation model. *Plant Methods*, 19(1), 87

3.2.1 Author Contributions

Dominik G. Grimm and Nikita Genze conceived and designed the study. Raymond Ajekwe conducted the drone flights and the image acquisition. Christian Schreiner and Maximilian Wirth labeled the dataset with supervision of Nikita Genze. Raymond Ajekwe and Michael Grieb guided the labeling process with domain expertise. Nikita Genze implemented the machine-learning pipeline and conducted all computational experiments. Nikita Genze and Dominik G. Grimm analyzed the results. Nikita Genze and Dominik G. Grimm wrote the manuscript with contributions from all authors. All authors read and approved the final manuscript.

3.2.2 Summary

In the second publication, we explored the generalization abilities of models trained under ideal conditions. Therefore, We trained `WeedSeg`, a baseline Semantic Segmentation model consisting of a `U-Net` architecture with `ResNet-34` as feature extractor, using images collected during an optimized flight mission with the 'Hover and Capture' setting. Here, we ensure sharp and well-focused UAV captures, which is beneficial for a quicker and more reliable annotation process. However, the flight mission takes significantly more time, as the UAV has to stop frequently. While `WeedSeg` performs well in predicting weeds in sharp test images, its accuracy significantly declines when presented with motion-blurred images. This decline may be attributed to the change in data distribution and the loss of clear edges, which serve as important features for prediction. Additionally, combining both sharp and motion-blurred image patches in the training data resulted in a model with reduced generalization abilities. To address this issue, we propose a combined de-blurring and Semantic Segmentation model called `DeBlurWeedSeg`. In this approach, we employ `NAFNet` [151], a computationally efficient model for image restoration, as described in Section 2.4. This model generates sharpened versions of the motion-blurred input images as a prior step. As a result, `DeBlurWeedSeg` demonstrated significantly improved performance in predicting both motion-blurred and sharp image patches.

3.3 Moving Fields Weed Dataset

This section summarizes the following publication. The complete version can be found in Appendix C.

PUBLICATION C

Nikita Genze, Wouter K. Vahl, Jennifer Groth, Maximilian Wirth, Michael Grieb & Dominik G. Grimm (2024). Manually annotated and curated Dataset of diverse Weed Species in Maize and Sorghum for Computer Vision. *Scientific Data*, 11(1), 109

3.3.1 Author Contributions

Dominik G. Grimm, Nikita Genze and Michael Grieb conceived and designed the study. Jennifer Groth and Wouter K. Vahl acquired the plants' seeds. Jennifer Groth and Wouter K. Vahl designed the experiments and were responsible for the image acquisition. Nikita Genze and Maximilian Wirth labeled the dataset. Michael Grieb guided the labeling process with domain expertise. Nikita Genze was responsible for quality control of the data. Nikita Genze and Dominik G. Grimm wrote the manuscript with contributions from all authors. All authors read and approved the final manuscript.

3.3.2 Summary

Another challenge with DL in the context of weed detection in agriculture is the lack of publicly available data, which is needed to test the generalization abilities of the models. These datasets must be diverse and of high quality to ensure robust model evaluation. Generating images of agricultural fields may be difficult; for instance, using UAVs is not always possible due to weather constrains and the small time-window for sensible weed removal. Additionally, the annotation process and quality control are time-consuming and demanding. Making this type of data publicly available accelerates training of new models and enhances reproducibility in agricultural research, as the results could be verified more quickly during the review process. We address this gap by publishing a comprehensive dataset that documents the growth of 28 different weed species. This dataset, named Moving Fields Weed Dataset (MFWD), contains over 94,000 high-quality images. Therefore, we utilized a high-throughput phenotyping facility, Moving Fields, at the Bavarian State Research Center for Agriculture in Freising to automatically capture all plant images. The plants were cultivated in boxes, also called micro-plots, with only one species planted per plot. During our experiment, we adjusted the number of boxes for each species, particularly focusing on those with low germination rates. The plants were photographed with a GSD of 0.17 $\frac{mm}{px}$. With this experimental setup, we also monitored the growth of over 5'000 individual plants with high spatial and temporal resolution. Here, the same micro-plots were captured at least daily. With such temporal

resolution, researcher may analyze the growth dynamics of different weed species and ultimately slow their growth in agricultural fields.
A significant part of this contribution is the rich manually annotated ground-truth suitable for diverse computer vision tasks (as described in Section 2.2). Therefore, we used the open-source software Computer Vision Annotation Tool (CVAT)[1], as it provided an easy-to-use interface for tracking plants over multiple images.
Finally, we conducted a baseline experiment for weed classification in sorghum. Therefore, we selected 27 different classes with fine-grained growth stages and trained different `EfficientNet` and `ResNet` models with a stratified subset of this dataset. Our final `EfficientNet_b0` model achieves a F1-score of approximately 90 % when classifying around 33'000 test images, indicating good performance.

[1] https://www.cvat.ai/

4 Discussion

Here, we discuss our findings with respect to the current challenges of UAVs and DL in weed detection, as outlined in Section 1.4. Therefore, we draw inspiration from the works outlined in Chapter 3 [118, 177, 125]. We structure this chapter in the following way: First, we interpret our results more broadly. Then, we discuss the limitations of our work. Afterward, we outline practical implications that stem from our contributions. Finally, we place our findings into a broader agricultural context.

4.1 Result Interpretation

In our first contribution, we trained DL models to segment weeds and sorghum from the soil. Therefore, we conducted experiments with a consumer-grade UAV, which is small, lightweight, and affordable. Meanwhile, due to governmental regulations, the training necessary to operate an UAV depends on its weight. Therefore, a larger UAV would be more costly and require additional training, more safety regulations, and a substantially longer setup before missions. With our UAV, we could swiftly plan and conduct all missions during the growing season. However, the resulting images were degraded by motion-blur, possibly stemming from a combination of the following reasons:
First, the camera quality of a lightweight UAV might be inferior due to cost- and weight-saving. Second, the gimbal might not stabilize fast or precisely enough, leading to slight camera movements, especially in windy situations. Finally, the automatic capturing settings of the UAV adjust the exposure and shutter speed dependent on the current illumination levels. Here, these settings, together with the high flight velocity and low altitude, might have had a negative effect on the shutter speed and resulted in visible motion-blur. It also explains why the captures were less blurred when the UAV was slowing down to turn at the corners of the flight plan, as illustrated in Supplementary Figure S4 of Publication A. Our trained model could distinguish weeds from crop plants despite such degraded images. Nevertheless, larger and advanced UAVs might mitigate the effect of motion-blur and capture higher-quality images. Additionally, UAV pilots might fly with a lower velocity or the 'Hover and Capture' flight mode.

♦

Further, a small amount of weeds from the previous growing season was present in our experimental fields, which could not be removed before sowing. They could be considered outliers due to their larger size, but our model still predicted their shape reasonably well. Instead, small weeds were more challenging due to several reasons. Not only were they difficult to annotate, as they occupied only a few pixels, and their green color was blurred with the surroundings. But also our model was not able to detect them with high precision, which is a general issue in several Computer Vision tasks [178, 179, 180] due to indistinguishable features and limited context information [181]. Additionally, small objects contribute to class imbalance [182, 183, 184]. More specifically, as crop and weed plants often contain only a few pixels in UAV images, traditional CNNs might fail to extract enough distinct features, which is amplified because of missing edge features due to blur. Further, the number of background pixels is considerably higher, reaching up to 98% of the image area. Here, on the one hand, background samples are easy to distinguish, as they contain many homogeneous pixels. On the other hand, foreground samples are scarce and complex shaped, leading to difficulties during training. Moreover, this bias might hinder optimization by focusing on easy examples.

♦

Our contributions address this issue by Under- and Oversampling, as they are well-tested techniques [185]. Undersampling removes patches without annotated objects from our training dataset to balance the class distribution (Publication A). Oversampling adds more training examples of the majority class in each batch to make the training process more efficient (Publication C). Nevertheless, it also increases the training time, as the information of the minority class is copied multiple times [186]. Additionally, we use various metrics to evaluate our results (see Section 2.6). However, we observe discrepancies between them and our qualitative visual assessment of the predictions. Most metrics are derived from the confusion matrix, which is based on pixels-space. Nevertheless, the high percentage of the background class distorts these metrics. Also, weighting (i.e., macro-averaging) the results to better reflect the inherent class imbalance is insufficient. Here, we utilize DC, which is specially designed for tasks with high class imbalance [171]. In our experiments, a high DC value better reflects the results and is sufficient to guide the training process, leading to sufficiently well detected weeds in general. However, by assessing the predictions visually, we found most misclassified pixels at the plants' borders, indicating possible alternatives for evaluation. Therefore, metrics focusing on the objects' boundaries might make the results more interpretable [187]. However, they might require more precise instance-based segmentation masks, which are costly. Moreover, different loss functions could be used in conjunction with these metrics, which might help train models more efficiently. For instance, the focal loss adds a modulating term to the cross-entropy loss, so the training process focuses on complex misclassified examples [188]. Also, in the agricultural context, a boundary loss [189] or surface loss [190] might be better suited than the dice loss [155] to measure the misalignment between predicted and actual plant boundaries more precisely. In fact, the selection of a fitting loss function depends on several factors like differentiability and optimization stability and needs to be considered carefully.

♦

Finally, we test our model - trained on BBCH 17 - on two different growth stages: BBCH 15 and BBCH 19. These additional UAV test-sets also differ in illumination due to different sun intensities during our flight missions. In both cases, our model demonstrates strong performance in predicting the weed class. Also, it predicts the sorghum class well in images of BBCH 19, which indicates that the model might adapt to larger growth stages of sorghum. However, the model predicts sorghum poorly in images of BBCH 15. This drop in performance might be due to morphological changes, such as fewer visible leaves compared to BBCH 17.

In our second publication, we investigate weed segmentation models trained under idealized conditions - without motion-blur - and their ability to segment weeds in scenarios with and without visible blur. In the first experiment, we have trained a baseline weed segmentation model, `WeedSeg`, only on sharp image patches. Consequently, its performance is significantly lower when presented with motion-blurred images. This performance drop might be due to the following reason: As CNNs extract features from the training dataset, they only learn these specifics and this distribution. However, their performance degrades drastically when trained models are presented with data that deviates from this distribution - whether due to environmental changes or domain shifts [191]. Changes in the level of motion-blur also change the data distribution because blur removes key features like plant edges, which are essential in training the model. One fundamental challenge with such Out-of-distribution (OOD) data is that CNNs tend to make overconfident predictions even when the input is vastly different from the training dataset. In these scenarios, DL models often assign high probabilities to incorrect classes [192, 193, 194], as they miss explicit mechanisms to recognize distributional shifts. Researchers proposed various strategies to mitigate this challenge [195, 196]. One common approach is to incorporate explicit OOD detection mechanisms, such as monitoring activation patterns in the final layers of a network [197] or using statistical techniques to measure confidence scores [198]. Another approach involves training models with adversarial examples or collecting diverse datasets to enhance robustness and improve generalization [199]. Some methods also employ generative models or contrastive learning [200] to better distinguish between in-distribution and OOD samples [201].

♦

For further experiments, we collected a dataset of matching sharp and motion-blurred image patches. However, after training `WeedSeg` on this large dataset, it performed similarly to the first experiment, which indicates that training with combined sharp and motion-blurred images is not a suitable strategy. We treated each image patch independently during training and randomly sampled the training-set, which might be improved with a more refined data sampling scheme, for instance, by using curriculum learning [202] or more sophisticated models [203].

In our contribution, we separate the task of weed segmentation in motion-blurred UAV images into two steps. Here, we first enhance the image quality with `NAFNet`, a CNN specifically designed and trained to remove blur. This model properly sharpens the images and additionally removes high-frequency noise. Consequently, `WeedSeg` - trained on sharp image patches only - segments weeds and crops from the soil. The final model, `DeBlurWeedSeg`, performs well on sharp as well as motion-blurred image patches, which is important in settings with unknown blur levels. While other approaches rely on additional mechanisms to detect motion-blur, the strength of our method lies in its efficiency and speed. It is also a major improvement to our first contribution, as collecting large datasets with different levels of motion-blur is not feasible in agriculture.

In the next paragraphs, we focus on results based on greenhouse images. As our final contribution, we generated imaging data using a large-scale greenhouse experiment. The resulting dataset enabled us to train DL models for the multi-class classification of 27 plant species. As part of fine-grained classification tasks [204], the main challenge here is distinguishing between subtle differences in highly similar categories.

One aspect is inter-class similarity, as weed detection tasks aim to distinguish species that share similar visual or structural features [205]. This similarity makes it challenging for models to separate different classes, leading to misclassifications, lower model confidence, and overall increased error rates in automated weed detection systems [206]. In this work, we address this challenge by capturing the growth of weed and crop species that are visually similar, especially in the early growth stages. Likewise, intra-class variance [207] refers to variations within a single class due to several factors: illumination, occlusion, growth stage, viewpoints, or environmental conditions. To provide high intra-class variability within our dataset, we tracked the growth of over 5'000 individual plants with high temporal resolution. Nevertheless, there are limitations to our work, which we discuss next.

4.2 Limitations

Our final contribution examined inter-class similarity and intra-class variance in fine-grained Image Classification. By effectively handling class imbalances through oversampling and weighted F1-score validation, our `EfficientNet_b0` model achieves an F1-score of approximately 90 % in classifying around 33'000 test images into 27 classes. Similarly, our best `ResNet-10` model achieves almost 89%, indicating competitive performance.

However, there are some limitations regarding this dataset. First, we could only capture this dataset in a greenhouse with artificial illumination, which contrasts with variable illumination on the field. Additionally, we manually removed highly overlapping weeds, leading to a simplified weed detection scenario. Moreover, some weed species could only sprout occasionally in our experiments, which results in a long-tail distribution of the dataset. Second, certain weed species did not germinate at all. Therefore, we could not include every relevant species in our dataset, as there are more than 250 harmful weeds in agriculture [208]. Finally, our dataset may not be directly usable with OOD data, such as with different crops or data acquisition protocols. Nevertheless, it represents a major step towards large datasets in weed detection. Here, it is essential to capture more data of a wide range of weed species, natural agricultural environments, cropping systems, and farming practices over an extended time period. Only then it might be possible to train a robust `Weed Sensing System` [209].

Here, UAVs could accelerate this research and provide a large-scale image dataset for weed detection. However, this dataset must include challenging scenarios due to inconsistent lighting, shadows, occlusion from overlapping leaves, and background clutter for various staple crops and weed species. As already mentioned in Section 1.4, multiple small and crop-specific datasets are available for weed detection tasks, which we compare in Publication C. Here, most datasets lack variability in species, samples, or growth stages [125]. Therefore, combining these datasets would be an essential first step to large-scale weed detection. However, these datasets differ remarkably due to inconsistencies in imaging conditions, for example, their resolution, GSD, or image modality. Likewise, our contributions employed two different methods to capture the datasets, as explained in Section 2.1. Here, possible solutions might leverage DL models with attention modules [210] or transformer-based architectures [211] to map these differences in feature space.

♦

Domain Adaptation is another technique to combine information from different datasets [212, 213], as it aims to bridge the gap between domains by enabling models to generalize across variations in data distribution. In essence, this technique lowers the amount of training samples from additional environments (target domains), and uses already existing datasets as source domains. Unsupervised Domain Adaptation [214, 215, 216] is used when no or little data in the target domain is available. Techniques such as adversarial learning [217], feature alignment [218, 219], and self-supervised learning [220] reduce discrepancies between source and target domains and ensure robust accuracy across diverse conditions. For instance, self-supervised learning utilizes image datasets without labels to learn more general representations. Here, models like SimCLR [200] or MoCo [221] use a contrastive learning framework and yield similar performance to ImageNet pre-training. However, more research is needed to answer questions about transferability in specialized domains [222] like medical histology [223] or Remote Sensing [224]. Therefore, our MFWD might be a possible starting point, as it relates to the weed detection task more closely than the ImageNet dataset.

♦

Another issue with different datasets are task-specific ground-truth labels. Therefore, we annotated MFWD with various computer vision tasks in mind. Additionally, our UAV datasets could be used for Object Detection, as researchers may automatically transform our segmentation masks into bounding boxes. Nevertheless, these datasets also have their limitations. First, they were derived from limited UAV missions focusing on crop plants at particular growth stages. Further, detecting weeds at a crop's growth stage of BBCH 17 (see Publication A) is impractical for SSWM, as the plants are already too large for effective weed management. Ideally, data should be collected while crop plants emerge (BBCH 12-14). However, during these early growth stages, weeds were tiny and in low numbers, making them invisible in our motion-blurred UAV captures. The reason for the small amount of weeds might be proper weed management practices on our experimental fields. Therefore, we utilized a higher growth stage in Publication A and designed additional experiments with a lower growth stage (BBCH 13) in Publication B. Another limitation of our UAV datasets is the limited variability of weed species. Here, dicot weeds - mainly white goosefoot (*Chenopodium album L.*) and two thistle species (*Onopordum acanthium L.* or *Cirium arvense L.*) - were dominant. However, the weed composition is dependent on multiple variables, such as region, temperature, and soil [225], and is currently changing due to climate change [226].

While we conducted our UAV missions on relatively small experimental fields, we were able to select a low flight altitude and capture tiny weeds with a precise GSD of 1 $\frac{mm}{px}$. Consequently, the ground coverage is low, which might not work in large farming systems. Here, greater flight altitudes are required, which demands better imaging sensors or trade-offs in weed detection capabilities. Arguably, it might be possible to neglect tiny weeds, as they might not significantly influence the weed pressure at field-scale, and current solutions for spot-spraying and mechanical weed control are less precise. However, these small weeds can grow quickly and may negatively affect the crops later in the growing season. Another factor is the GPS uncertainty, meaning that the position of the weeds could not be mapped accurately. Therefore, identification on a plant level is usually sufficient.

Nevertheless, due to greater flight altitudes in favor of efficiency, crops and weeds would fall into the category of small objects. As mentioned in the previous section, small objects are a general issue in several Computer Vision tasks. Here, for example, encoder-decoder-based Semantic Segmentation models extract high-level semantic features in the encoder by progressively down-sampling the input image. This kind of feature extraction leads to the removal of small object features in deeper layers. One solution includes the usage of atrous convolutions, for instance, in `DeepLabV3+`, as they expand the receptive field without losing spatial resolution. However, this model architecture had a smaller DC than `U-Net` in our experiments. This result might indicate that there are additional factors to consider when selecting model architectures for small weed detection. Therefore, researchers introduced novel architectures, which use shallow and parallel encoder structures to maintain high-resolution features [227, 228].

♦

In our experimental setup, we focused on smaller models that could be trained and deployed without server-grade equipment. Therefore, we used small image patches of 256×256 px^2 in our first contribution and 128×128 px^2 in our second one. Consequently, there might be multiple partly visible plants on the border of these patches, which might affect the model's performance, as partly cut objects are more difficult to classify. A possible solution might be overlapping patches made by a moving window over the complete UAV capture [229]. However, overlaps have a higher demand for computational resources and increase the duration of training and inference. Another solution might be using larger image patches, which would substantially increase the model's size.

♦

In our studies, we employed pre-training as a form of transfer learning. Here, the model starts training with pre-computed ones derived from the ImageNet benchmark dataset [230], and not with randomly initialized weights for the encoder. However, not all features are equally transferable from one task to another [231]. Depending on the domains or tasks, the knowledge transfer might be difficult and yield sub-par results compared to randomly initialized weights [232]. Nevertheless, pre-computed weights reduce the training time until convergence and transfer learning is a well-established paradigm in DL [233, 234, 109], especially in low-data regimes.

Finally, we focused our experiments on Semantic Segmentation in UAV captures. While our predictions display relative positions of weeds and crops in pixel space, we do not compute their GPS positions. Therefore, orthophotos [235] might be more suitable due to following reasons: they remove sensor tilt and lens distortion from the images and adjust them based on terrain relief. The resulting uniform scale across an area shows the true geographical locations of objects [236]. Moreover, they are widely used in Remote Sensing applications, like cartography, urban planning, or environmental monitoring, as they combine multiple areal images. However, they face several limitations due to high processing and storage requirements, resolution constraints, and geometric distortions [237, 238]. In our case, another challenge is the low GSD of 1 $\frac{mm}{px}$, as factors like terrain variations cause misalignments in the final results. Moreover, motion-blur significantly impacts the quality of an orthophoto due to lower geometric accuracy and feature distinction and hinders the calculation of an orthophoto. Here, a possible solution might be linear interpolation, as the area of one capture results at around 27.5 m^2 (6.4 m $\times$ 4.3 m). Therefore, interpolating the relative positions of weeds in an image might be a valuable approximation for broad weed management zones.

4.3 Practical Implications

Our research presents significant practical implications in weed sensing. First, our DL models segment weeds and sorghum in UAV images with high precision, even under challenging conditions, multiple growth stages, and illumination. Additionally, our models may accurately detect intra-row weeds and partly occluded plants. This capability is crucial for advancing autonomous weeding systems. Second, our introduction `DeBlurWeedSeg`, a novel model integrating de-blurring and segmentation, further enhances weed detection in dynamic environments where UAVs and UGVs operate under conditions that degrade image quality, like motion, wind, and uneven terrain. Additionally, since our model is trained only on sharp image patches, it eliminates the need for labor-intensive annotations of blurred images, making it more practical for real-world deployment. In future, combining UAVs and UGVs presents a highly efficient weed management strategy. UAVs could rapidly map weed locations and transmit this data to UGVs, which then might navigate on optimized driving paths and target only specific weeds. This method minimizes unnecessary driving on the field, which saves time and energy, and reduces soil disturbance. Third, we support the development of more robust DL models by providing high-quality and high-variability data with our MFWD. This research enables farmers and agronomists to adapt our dataset in UAV-based or robotic weed management systems, and precisely identify weed species at various growth stages. Therefore, they can easily filter the dataset for specific weed species and fine-tune DL models for their needs. Additionally, this dataset might act as a benchmark for consistent and reproducible results, because we published ground-truths for different computer vision tasks using a standardized protocol. Finally, our relatively small model architectures enable future fine-tuning on affordable consumer-grade hardware, which makes it accessible to everyone. Finally, farmers retain complete control over their data and reduce their reliance on costly computing equipment. Overall, these advancements might contribute to cost savings, reduced herbicide use, lower environmental impact, and increased crop yields.

4.4 Broader Agricultural Context

DL-based weed detection is transforming Precision Agriculture by enabling automated and high-accuracy weed detection. However, several additional barriers must be overcome before widespread adoption becomes feasible [239].
One major issue is determining the required level of precision for effective plant identification. While modern object detectors are accurate in weed detection, it remains unclear what degree of precision is necessary to balance cost, practicality, and environmental sustainability [240]. Additionally, treatment methods for managing weeds can vary significantly in spatial resolution [67]. The least precise option is treating broad management zones or weed patches, which may be tailored to the specifications of the `Precision Weed Control Implement`. In contrast, the most accurate approach involves treating individual weeds. Broad management zones are most effective for large, stable weed patches, while higher resolution is more suitable for smaller patches dispersed throughout the field. Therefore, weed maps generated by UAVs can be utilized effectively. However, to treat individual plants, ultra-high-resolution images are necessary, typically captured in real-time by UGVs. This time constraint is due to the rapid growth and noticeable differences in appearance over a short period [241]. To achieve efficient weed management, all three key components - `Weed Sensing System`, `Weed Management Model` and `Precision Weed Control Implement` - require seamless, near real-time conversion of raw data into actionable information and final removal [242]. For instance, the herbicide mixture may be adjusted in real-time based on the distribution of different weed species [243]. Here, robust DL models are essential for the real-time identification and localization of weeds. Many research studies have demonstrated that object detectors of the You Only Look Once (YOLO) family [244], may detect weeds nearly in real-time [245, 246, 247, 248, 249]. While our studies investigated Semantic Segmentation models for precise weed detection, they cannot process images in real-time. Here, up-sampling and per-pixel classification is computationally complex compared to Object Detection tasks, which may process a whole UAV capture and output a fixed set of bounding boxes. Therefore, further research is needed to design and integrate precise models into UGVs and enable real-time actions for weed control [69, 69].

♦

The successful adoption of DL in agriculture depends on a complex interplay of economic, technological, regulatory, and agro-ecological factors [250, 251, 239, 252]. One primary factor is cost, including initial investment, maintenance, and operational expenses [253, 254], which often make UGVs more feasible for large-scale farms. In contrast, smaller farms struggle with financial constraints, as they lack access to funding, subsidies, and technical support. In contrast, farm labor shortages and increasing costs drive interest in automated systems. However, these technologies require specialized knowledge for operation and dataset management. Subsequently, training programs for farmers are crucial [242, 255, 243]. Ultimately, from a technological standpoint, the adoption depends on reliable, accurate, and interoperable system, as they must seamlessly integrate with existing farm machinery and sensors [243]. Moreover, concerns about data ownership, privacy, and cybersecurity create barriers to trust [256, 257, 258]. Additionally, regulatory constraints, including UAV regulations, and missing data transfer protocols are further limitations. Also, legal issues with autonomous UGVs, liability, and negligence are not clearly defined, which creates uncertainty for farmers, technology developers, and consumers [259, 260, 261]. Further, the dependency on such technologies poses risks, as connectivity failures or hardware malfunctions could disrupt agricultural activities [262, 263], especially in rural areas with low infrastructure [239, 264, 95]. Finally, environmental and agro-ecological factors complicate the integration of DL systems. Here, the economic benefits of Precision Agriculture remain inconsistent as they vary based on farm size, technology type, and management system [250, 68]. To ensure adoption, new agronomic practices identifying suitable weed management for a given crop, soil, and climate combination are necessary [17]. Addressing these challenges through standardized protocols, regulatory frameworks, and interdisciplinary collaboration will help develop resource-efficient and innovative solutions for weed management that align with sustainability.

5 Conclusion

This thesis explores DL-based weed detection with a focus on motion-blurred UAV images, model generalization, and dataset creation. We present three key contributions: (1) investigating the robustness of DL models to varying motion-blur levels, (2) developing a deblurring-assisted segmentation model, and (3) introducing the Moving Fields Weed Dataset (MFWD) for improved model training and evaluation. In this chapter, we conclude with the main highlights of this thesis.

♦

One of the major obstacles in UAV-based weed detection is motion-blur, caused by factors such as wind, flight speed, low illumination, and sudden UAV direction changes. Motion-blur degrades image quality, reduces feature sharpness, and impacts the accuracy of DL models. Our study evaluates the robustness of segmentation models when trained on sharp but tested on motion-blurred images, revealing a significant drop in performance due to shifts in distributions. Here, traditional segmentation models struggle with blurry images, as they rely on precise edges and fine textures for distinguishing between crops, weeds, and soil.

♦

To address this issue, we introduce `DeBlurWeedSeg`, a deblurring-assisted weed segmentation model. Our approach integrates `NAFNet`, a computationally efficient de-blurring model, as a step before weed segmentation. Our results show that `DeBlurWeedSeg` significantly improves weed segmentation performance, especially when tested on motion-blurred images. Additionally, our method may be trained on sharp images only while still achieving high accuracy on sharp *and* motion-blurred UAV captures, which eliminates the need for collecting and annotating motion-blurred datasets.

♦

A major limitation with DL in agricultural research is the lack of publicly available, diverse, and high-quality datasets. To overcome this limitation, our study presents the Moving Fields Weed Dataset (MFWD), which contains over 94,000 images of 28 weed species. This dataset benefits research in multiple ways: First, it was collected using a high-throughput phenotyping facility, where plants were photographed under controlled conditions with consistent lighting and background. Second, it provides detailed ground-truth annotations, including bounding boxes and segmentation masks, making it a valuable resource for training and evaluating different weed detection models. Third, it contributes to agricultural research by covering multiple weed species with various growth stages. And finally, it might serve as a foundation for future research, as we ensured standardized labeling and data collection protocols.

The findings in this thesis highlight several areas for future research in UAVs and DL: Further datasets of many natural environments in agriculture are essential for robust weed detection. Additionally, Domain Adaptation might be a key component for better model transferability across different crops, locations, and environmental conditions. However, more research and large-scale datasets are necessary obtain a robust `Weed Sensing System` for SSWM. Moreover, implementing robust real-time weed detection systems is crucial for practical field deployment.

♦

In conclusion, advancements in UAV technology will further expand the capabilities of DL models with higher-quality and user-friendly sensors. These advancements will allow non-experts to use advanced hyper-spectral, multi-spectral, and thermal imaging for weed detection. In the long term, fully automated UAV pipelines for flight planning, execution, and real-time data analysis will simplify weed detection. Further, the fusion of stationary (e.g., soil sensors, weather stations) and mobile (e.g., UAVs, autonomous tractors, satellites) data sources will enable continuous model updates for high-value agricultural areas. Finally, as data collection at the farm-level becomes more detailed and reliable, DL will be crucial in optimizing resource use, reducing environmental impact, and improving overall farm efficiency.

A Publication A

The following pages show the full version of the published manuscript:

Publication A

Nikita Genze, Raymond Ajekwe, Zeynep Güreli, Florian Haselbeck, Michael Grieb & Dominik G. Grimm (2022). Deep learning-based early weed segmentation using motion blurred UAV images of sorghum fields. *Computers and Electronics in Agriculture*, 202, 107388

Computers and Electronics in Agriculture 202 (2022) 107388

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Deep learning-based early weed segmentation using motion blurred UAV images of sorghum fields

Nikita Genze [a,b,*], Raymond Ajekwe [d], Zeynep Güreli [a,b], Florian Haselbeck [a,b], Michael Grieb [d], Dominik G. Grimm [a,b,c,*]

[a] Technical University of Munich, TUM Campus Straubing for Biotechnology and Sustainability, Bioinformatics, Schulgasse 22, 94315 Straubing, Germany
[b] Weihenstephan-Triesdorf University of Applied Sciences, Bioinformatics, Petersgasse 18, 94315 Straubing, Germany
[c] Technical University of Munich, Department of Informatics, Boltzmannstr. 3, 85748 Garching, Germany
[d] Technology and Support Centre in the Centre of Excellence for Renewable Resources (TFZ), Schulgasse 18, 94315 Straubing, Germany

ARTICLE INFO

Keywords:
Deep learning
Weed detection
Weed segmentation
UAV
Precision agriculture

ABSTRACT

Weeds are undesired plants in agricultural fields that affect crop yield and quality by competing for nutrients, water, sunlight and space. For centuries, farmers have used several strategies and resources to remove weeds. The use of herbicide is still the most common control strategy. To reduce the amount of herbicide and impact caused by uniform spraying, site-specific weed management (SSWM) through variable rate herbicide application and mechanical weed control have long been recommended. To implement such precise strategies, accurate detection and classification of weeds in crop fields is a crucial first step. Due to the phenotypic similarity between some weeds and crops as well as changing weather conditions, it is challenging to design an automated system for general weed detection. For efficiency, unmanned aerial vehicles (UAV) are commonly used for image capturing. However, high wind pressure and different drone settings have a severe effect on the capturing quality, what potentially results in degraded images, e.g., due to motion blur. In this paper, we investigate the generalization capabilities of Deep Learning methods for early weed detection in sorghum fields under such challenging capturing conditions. For this purpose, we developed weed segmentation models using three different state-of-the-art Deep Learning architectures in combination with residual neural networks as feature extractors.

We further publish a manually annotated and expert-curated UAV imagery dataset for weed detection in sorghum fields under challenging conditions. Our results show that our trained models generalize well regarding the detection of weeds, even for degraded captures due to motion blur. An UNet-like architecture with a ResNet-34 feature extractor achieved an F1-score of over 89% on a hold-out test-set. Further analysis indicate that the trained model performed well in predicting the general plant shape, while most misclassifications appeared at borders of the plants. Beyond that, our approach can detect intra-row weeds without additional information as well as partly occluded plants in contrast to existing research.

All data, including the newly generated and annotated UAV imagery dataset, and code is publicly available on *GitHub*: https://github.com/grimmlab/UAVWeedSegmentation and Mendeley Data: https://doi.org/10.17632/4hh45vkp38.4.

Abbreviations: SSWM, Site Specific Weed Management; ML, Machine Learning; DL, Deep Learning; SVM, Support Vector Machine; RF, Random Forest; CNN, Convolutional Neural Network; FCN, Fully Convolutional Network; GSD, Ground Sampling Distance; OBIA, Object Based Image Analysis; NIR, Near Infra-Red; GIMP, GNU Image Manipulation Program; TL, Transfer Learning; CRF, Conditional Random Field; TPE, Tree Parzen Estimator; DS, Dice Score or Sorensen-Dice Coefficient; TP, True Positives; FP, False Positives; FN, False Negatives; ASPP, Atrous Spatial Pyramid Pooling; CLAHE, Contrast Limited Adaptive Histogram Equalization.

* Corresponding authors at: Technical University of Munich, TUM Campus Straubing for Biotechnology and Sustainability, Bioinformatics, Schulgasse 22, 94315 Straubing, Germany.

E-mail addresses: nikita.genze@tum.de (N. Genze), dominik.grimm@hswt.de (D.G. Grimm).

https://doi.org/10.1016/j.compag.2022.107388
Received 18 February 2022; Received in revised form 31 August 2022; Accepted 8 September 2022
Available online 18 September 2022

1. Introduction

Weeds are undesired plants in agricultural fields that affect crop yield and quality by competing for nutrients, water, sunlight and space (Patel and Kumbhar, 2016). For centuries, farmers have used several strategies and resources to remove weeds. Some of these strategies include mechanical (mowing or tilling) and chemical (herbicides) control. The use of herbicides remains the most common control strategy as it is less time consuming and labor-intensive (Aktar et al., 2009). However, recent studies show that this strategy can impact negatively on the biota and the surrounding environment over time (Holt, 2004; Who, 1990). These disadvantages lead to a social pressure requesting less usage (Ridgway et al., 1978). To reduce the amount of herbicide and impact caused by uniform spraying, site-specific weed management (SSWM) through variable rate herbicide application and mechanical weed control using an autonomous field robotic system have long been recommended (Fernández-Quintanilla et al., 2018). These approaches heavily depend on weed sensing and mapping technology. To achieve this, an accurate as well as efficient detection and classification of weeds in crop fields is a crucial first step (Liu and Bruch, 2020).

Owing to the phenotypic similarity between some weeds and crops as well as the dynamic weather conditions, designing an automated system for general weed detection in crop fields is challenging (Hasan et al., 2021). Furthermore, variability in the emergence time of different weeds at various crop growth stages makes it rather complicated to design a generalized model. Most studies in weed detection have leveraged plant vegetation indices (Hamuda et al., 2016; Meyer and Neto, 2008) and hand-engineered features with a focus on classical Machine Learning (ML) methods, such as Support Vector Machines (SVM) and Random Forests (RF) (Zheng et al., 2017). For example, Kazmi *et al.* employed conventional image processing based on vegetation indices for creeping thistle weed detection in sugar beet fields (Kazmi et al., 2015). Relatedly, Islam *et al.* reported weed detection in chili plants using a similar technique (Islam et al., 2021).

Deep Learning (DL), particularly Convolutional Neural Networks (CNNs), are capable of extracting relevant features from raw images and were used extensively in image classification in recent years (LeCun and Bengio, 1998). They generalize well across illuminations and obstructions and were applied for several agricultural tasks (Genze et al., 2020; Kussul et al., 2017; Grinblat et al., 2016; Chen et al., 2017). These properties make them well suited for weed classification using drone images from fields, which comprise several changing conditions, such as illumination. Furthermore, recent surveys show a superiority of DL-based approaches compared to classical ML (Liu and Bruch, 2020; Kamilaris and Prenafeta-Boldu, 2018; Kamilaris et al., 2017; Alom et al., 2019). Additionally, several authors have concluded that hand engineered features tend to be less precise and hard to generalize compared to DL-based approaches (Penatti et al., 2015; Huang et al., 2018).

One downside of CNNs is the amount of data needed for generalization (Sun et al., 2017). There are several publicly available datasets for weed detection, as summarized in a recent review (Lu and Young, 2020). However, most of these datasets were collected using high-resolution cameras mounted on field robots or pulled carts. DL models have been applied for the task of weed detection on different crops. For example, Ramirez *et al.* used a dataset called WeedMap (Sa et al., 2018), consisting of high-resolution orthomosaic images of a sugar beet field in Germany. The authors compared SegNet (Badrinarayanan et al., 2015) and a modified U-Net (Ronneberger et al., 2015) with a recent segmentation architecture called DeepLabv3 (Chen et al., 2017). DeepLabv3 led to a higher classification accuracy due to a greater spatial context (Ramirez et al., 2020). Huang *et al.* studied weeds in rice fields and showed that Fully Convolutional Networks (FCNs) (Long et al., 2015) outperform other methods in accuracy and efficiency for drone imagery with a ground sampling distance (short: GSD, distance between adjacent pixel centers) of 0.3 cm (Huang et al., 2018; Huang et al., 2018).

Regarding data generation, most weed detection studies rely on object-based image analysis (OBIA) (Lam et al., 2021; Hay and Castilla, 2006), unsupervised (Dian Bah et al., 2018; dos Santos Ferreira et al., 2017) or semi-supervised (Pérez-Ortiz et al., 2015; Lottes and Stachniss, 2017) approaches for semantic segmentation. The main reason for this is the large effort of a manual pixel-based labeling. Beyond that, additional information like the spatial distribution of the crop rows (Lottes and Stachniss, 2017; Onyango and Marchant, 2003; Milioto et al., 2018) or different sensors (e.g. near infrared (NIR) and multispectral) are widely used (Lottes et al., 2018). By using different light spectra, it is easier to separate vegetation from soil (Yeom et al., 2019), thus making the ground truth generation faster. One downside of current NIR and multispectral sensors is their low GSD compared to RGB sensors. Although NIR and multispectral sensors provide more information about the vegetation compared to RGB sensors, the GSD is one of the most critical parameters for accurate weed detection. Besides the difficulty of detecting small weeds captured by drones with low GSD, it is more time-consuming and error-prone to generate the ground truth.

For a real-world application, an efficient way for image capturing is needed, for which drones can be utilized (Mukherjee et al., 2019). On the one hand, a higher throughput and temporal resolution are the main advantages compared to ground robots. On the other hand, image capturing using drones is prone to motion blur, e.g. due to wind-related displacements. This is an additional challenge for a robust weed classification. Only a limited number of datasets have been generated using drone captures, due to the time-consuming generation and labeling of such datasets.

In this study, we investigate weed segmentation capabilities of DL-based approaches using real-world motion blurred drone captures in sorghum fields. For this purpose, we generated a manually curated and expert-labeled UAV dataset of early weeds and sorghum plants. Our dataset is fully and manually annotated with the highest precision possible without relying on OBIA or threshold-based segmentation techniques. This dataset enabled us to evaluate state-of-the-art DL architectures in combination with residual neural networks of different sizes as feature extractors for early weed and sorghum segmentation. We further, evaluated the generalization abilities of our trained models with respect to different growth stages of sorghum.

2. Material and methods

In the next section, we provide insights into the data acquisition and annotation. Afterwards, we outline the machine learning models, the hyperparameter optimization as well as all used evaluation metrics. Finally, we describe the experimental setup.

2.1. Data acquisition and annotation

Images collected by an UAV on an experimental sorghum field in Southern Germany provided data for this study. The sorghum variety “Farmsugro 180” was sown at 37.5 cm row spacing with a seeding density of 25 seeds per m^2. During image capture, several weed species were present on the field. These weeds comprised mostly of dicotyledons namely, Goosefoot (*Chenopodium album* L.), Field pennycress (*Thlaspi arvense*), Wild chamomile (*Matricaria chamomilla*), Common gypsyweed (*Veronica officinalis*) and Cotton thistle (*Onopordum acanthium*). Beyond that, a consumer-grade drone “DJI Mavic 2 Pro” fitted with a 20 MP Hasselblad camera (L1D-20c) that captures images with a resolution of 5472x3648 $pixels^2$ was used. Automated drone flight with camera pointing nadir and a capture overlap of ten percent was carried out at a flight altitude of five meters above ground level. At this altitude, the corresponding GSD was one millimeter, and therefore precise enough to recognize sorghum and weeds in early growth stages. A set of 60 high resolution images was captured in late spring when sorghum was at the growth stage 17 (according to the BBCH scale (Hess et al., 1997). Thereby, a time- and battery-efficient setting was selected with a nonstop capturing at 60 points of equal distance along the flight path, which leads to motion blur in the collected images.

In this work, we defined the weed detection task as distinguishing weeds from sorghum plants and soil pixel by pixel (weed segmentation). To enable supervised ML, we assigned one of these three classes (weed, sorghum, soil) to every pixel of each collected image. For the annotation process, we used the open-source GNU Image Manipulation Program (GIMP). Agronomy experts guided the labeling process to ensure a high-quality ground truth. Fig. 1 shows an example of a drone capture and its related pixel-wise ground truth.

About 6.300 patches from 19 images were annotated resulting in over 3.000 sorghum and 3.000 weed instances, as summarized in Table 1. We divided each image in non-overlapping patches with a resolution of 256x256 pixel for an efficient training process. Additionally, all non-square patches (image and labels) were padded with zeros.

Several factors influenced the degree of motion blur in the images, e. g. the current velocity of the drone or external factors such as wind. The blurriness of images varied between different captures, e.g. some captures were slightly sharper after the rotation of the drone at an edge of the flight plan, because the drone was flying slower and stabilized itself after the rotation. A detailed evaluation of the blurriness of our captures can be found in Supplementary Text 1. Another challenge in this real-world datasets are overlapping plants. This might lead to multiple plants being labeled and counted as one instance.

To evaluate the generalization capabilities of our model, we labeled two additional UAV missions, as shown in Table1. These images were captured during different UAV flights on another part of the field and show sorghum at the growth stages BBCH 15 and 19. More detailed information and statistics about the dataset can be found in the Supplementary Figure S1.

Due to the pixel-wise annotation of the datasets, it can be easily adapted to other tasks, such as patch-based classification or object detection using bounding boxes. These additional labels can be generated automatically using our fine-grained labeled segmentation masks.

2.2. Machine learning and experimental setup

The architecture of our weed segmentation model consists of a feature extractor and a semantic segmentation architecture.

For our weed segmentation model, we combined these six semantic segmentation architectures with four residual neural networks of different sizes as feature extractors (He et al., 2016; Veit et al., 2016). Residual networks combat the vanishing gradient problem by introducing identity skip connections and allowing data to flow from any layer directly to any subsequent layer. This enabled researchers to build deeper networks with hundreds of layers efficiently. In addition, they utilize convolutions with strides instead of pooling layers for downsampling. The feature extractors ResNet-18, 34, 50 and 101 were used in this study, which consist of 18, 34, 50 and 101 layers, respectively.

Table 1
Summary statistics of the datasets used in this study.

Dataset Name	Sorghum BBCH	Annotated Patches	Sorghum Instances	Weed Instances
sorghum_17	17	6270	3056	3060
sorghum_15	15	110	115	429
sorghum_19	19	110	99	981

With respect to the semantic segmentation architecture, we investigated six state-of-the-art DL architectures, namely UNet (Ronneberger et al., 2015), DeepLabv3+ (Chen et al., 2017) and four variants of FCN (Long et al., 2015), since these models have already been successfully used for other crops (Ramirez et al., 2020; Huang et al., 2018). The main aspects of the model architectures are illustrated in Fig. 2.

FCNs perform convolutions, pooling and upsampling and can produce a spatial segmentation map of the same size as the input image and scale across different image sizes, as indicated in Fig. 2a-d. Semantic information from deep and coarse layers can be combined with appearance information from shallow and fine layers in the network. The FCN-32s model does not combine information of different layers, but predicts one segmentation mask from the last layer of the network and then calculates the result using bilinear interpolation with a factor of 32 (Fig. 2a). In the FCN-16s model, the predictions of the last two layers are combined. Therefore, the predictions from the last layer must be upsampled by a factor of two first and then combined with the feature maps of the second-last layer. In addition, one-by-one convolutions are used to match the depth of both feature maps before combining them. Finally, the result is upsampled by a factor of 16 (Fig. 2b). In the FCN-8s model, the two last feature maps are concatenated first (as in the FCN-16s model). This result gets upsampled by a factor of two again to combine it with the third-last layer. Finally, this output can only be upsampled by eight to match the spatial resolution of the input image (Fig. 2c). The FCN-8s model is able to predict more fine-grained segmentation maps by combining predictions from three different layers. This helps in training a model that is able to predict finer details. Alternatively, dilated (atrous) convolutions (Fisher and Koltun, 2016) can be used to preserve the spatial resolution. In addition, they use a greater receptive field to cover more information in the layer. This led to the architecture FCN-8s+ dilation, as shown in Fig. 2d. This architecture is similar to FCN-32s, as no information from intermediate layers is used, but it only needs to interpolate the predictions bilinear with a factor of eight.

The UNet architecture was initially developed for biomedical tasks and has been adapted to different domains. Existing research shows that

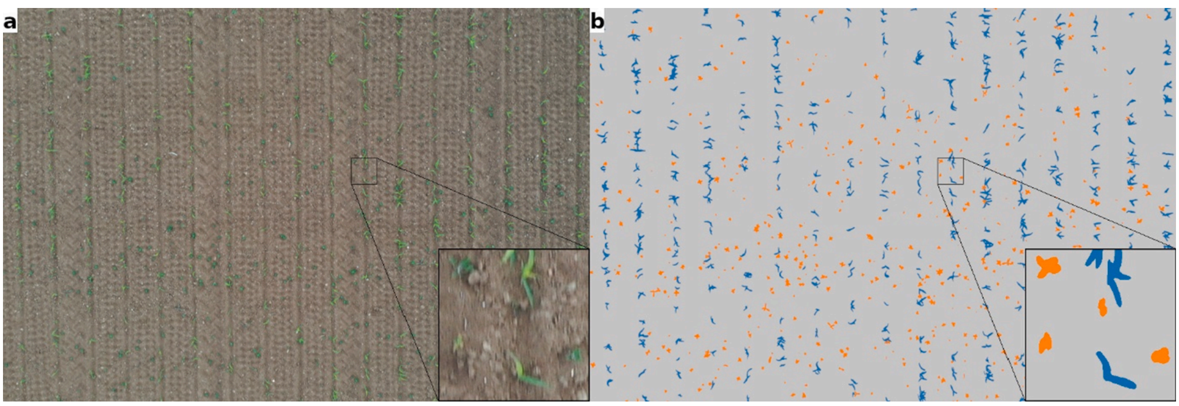

Fig. 1. Example drone capture and related ground truth for the dataset *sorghum_17*. a UAV capture from one of our sorghum fields **b** Pixel-wise ground truth with background in gray, sorghum plants in blue and weed plants in orange.

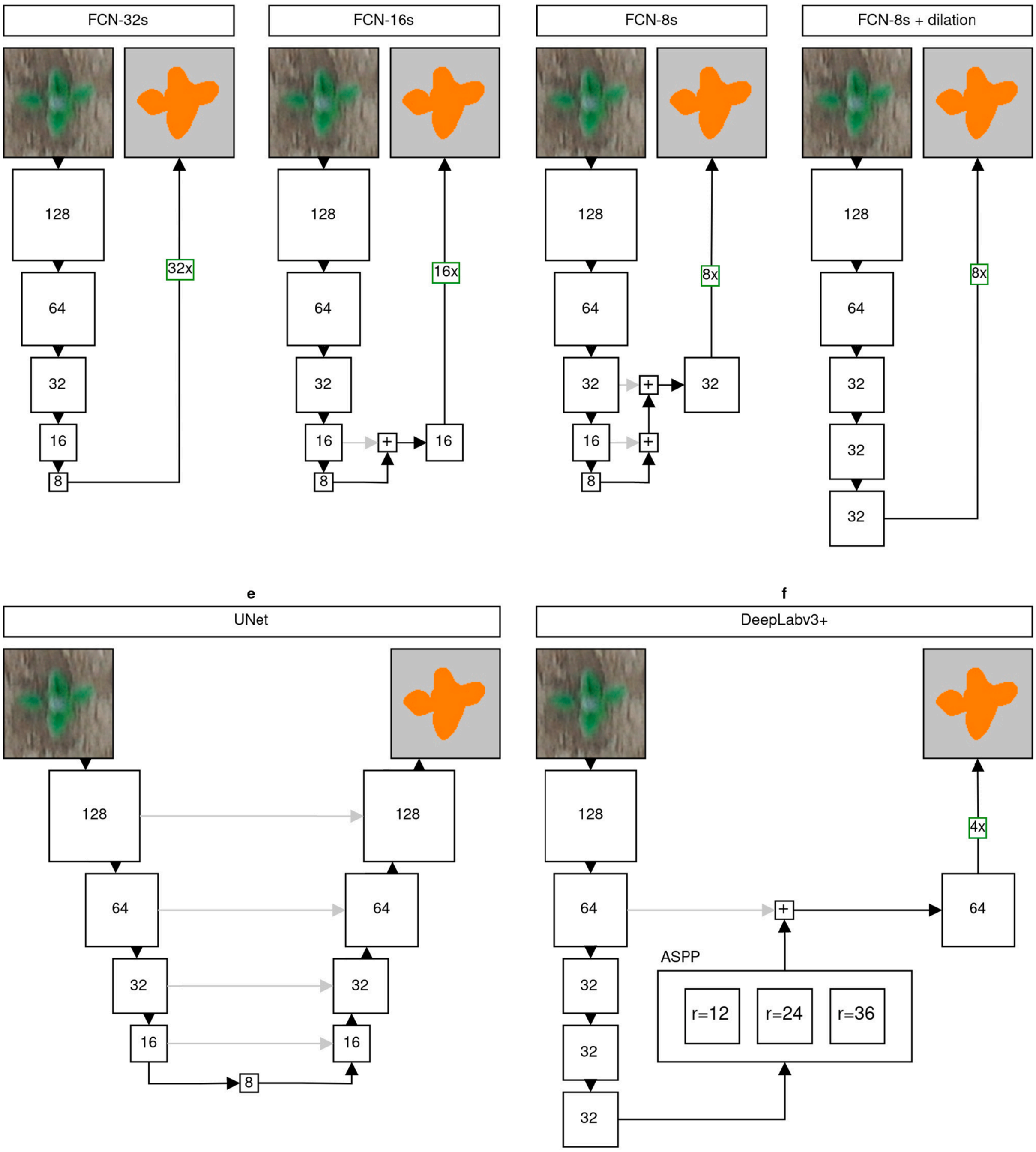

Fig. 2. Model diagrams of the different semantic segmentation architectures used in this study. Numbers indicate the spatial resolution of the feature maps. Skip connections are shown as gray arrows. The concatenation of feature maps is shown as plus. Upsampling operations before the concatenation is omitted for readability. Final bilinear interpolation is shown in a green box indicating the upsampling factor. **a-c** Modifications of the FCN architecture with different strides. **d** FCN architecture using dilated convolutions, preserving the spatial resolution in the last layers. **e** UNet architecture using skip connections. **f** DeepLabv3+ architecture using Atrous Spatial Pyramid Pooling (ASPP) with atrous rates of r = 12, 24 and 36.

this model achieved good segmentation results even with little training data by relying on a vigorous data augmentation pipeline. Skip connections between encoder and decoder are used to fuse information of all layers of the encoder with the corresponding layers in the decoder (Fig. 2e). In addition, the feature maps are upsampled more gradually in the decoder compared to FCN. Finally, the upsampling is done using learnable filters instead of a fixed bilinear interpolation.

DeepLabv3+ is a model architecture presented by Chen *et al.* that improves the coarse spatial resolution - caused by repeated pooling layers - by using dilated (atrous) convolutions. These special operations adjust the field-of-view of the filter, as indicated in Fig. 2f. The Atrous Spatial Pyramid Pooling (ASPP) module uses multiple parallel-dilated convolutions with different rates to consider objects at different scales (Chen et al., 2017). Similar to FCN and UNet, an encoder-decoder architecture is used to recover the lost spatial information and refine the predictions at the object boundaries.

Furthermore, early stopping (Prechelt et al., 2012) and a learning rate scheduler were implemented to lower the computational cost and avoid overfitting. The Adam (Kingma and Ba, 2014) optimizer was used with different initial learning rates. Following the principles of risk minimization (Vapnik, 1992), a differentiable version of the Sørensen-Dice Coefficient (DS) was used as a loss function to update the weights in the training process. The same loss was applied to validate the performance of the trained models. This metric is commonly used to measure the similarity of two samples in the evaluation of semantic segmentation tasks (Bertels et al., 2019) and has been shown to be robust to class imbalance (Sudre et al., 2017). It is defined as follows:

$$DS = \frac{2*TP}{2*TP + FP + FN} \quad (1)$$

where TP is the number of true positives, FP is the number of false positives and FN is the number of false negatives. The batch size was fixed to 100 for each combination of feature extractor and segmentation architecture.

Hyperparameters can greatly influence the predictions of a DL-based model. There are several strategies for optimizing these, e.g. Grid or Random Search (Bergstra and Bengio, 2012). Contrary to these strategies, Tree-structured Parzen Estimator (TPE) (Bergstra and Bengio, 2012) selects a new set of hyperparameters based on previous experiments conducted in an optimization study. Using an objective function, a new set of hyperparameters is sampled to achieve a high probability of a good score. This strategy requires the construction of a probabilistic model for the optimization process, but is efficient due to the automatic selection of promising settings. For this reason, we employed this strategy to optimize the hyperparameters summarized in Table 2.

All models were trained and evaluated based on patches of 256x256 pixel[2]. We excluded patches that contained only background or small plant instances (>99% only background pixels). This resulted in 2156 patches from 12 UAV captures. Patches that were zero-padded in an earlier step made up 143 patches (=6.63%). Padding does not affect the classification accuracy (Hashemi, 2019) itself, but is needed to use the whole dataset more efficiently. The data was split into four folds for cross validation and an additional hold-out test-set for estimating the final model performance, as shown in Fig. 3. The averaged performance on the validation sets was used to determine the best performing hyperparameter combinations for each weed segmentation model. A total of 50 different hyperparameter sets were evaluated. The setup leading to the smallest error averaged over all validation sets was trained again on the complete training and validation data. Finally, we evaluated the performance of the resulting model on the hold-out test-set.

We report the evaluation metrics precision, recall and the F1-score on a pixel basis:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

Table 2
Summary of the hyperparameters used and optimized in this study.

Hyperparameter	optimized	Range	Description
initial learning rate	yes	10^{-4} – 10^{-2}	learning rate at the start of model training. It is sampled from a log-uniform distribution.
learning rate decay	yes	0.9–0.1	rate of the learning rate decay used when the loss function plateaus. The step size was set to 0.1.
learning rate scheduler patience	no	5	reduce the learning rate, if the validation loss did not improve over this number of epochs
batch size	no	100	Number of patches that can be sent through the network in one forward pass
early stopping patience	no	10	stop the training loop, if the validation loss did not improve over this number of epochs

$$F1 - score = \frac{Precision*Recall}{Precision + Recall} \quad (4)$$

Further, as we observed a high-class imbalance, we calculate the macro-averaged metrics, as weight-averaged metrics favor the majority class (background). Weight-averaged metrics consider the proportion for each label in the dataset. As the weed class is occurring the least, but is the most important one, macro-averaged metrics are more suitable, as they reflect the arithmetic mean of all classes.

The final performance was calculated after combining all patches into individual images. This procedure is independent of the batch size and the size of the patches, thus ensures the comparison to different other weed detection approaches in future studies.

Beyond that, data augmentation (Shorten and Khoshgoftaar, 2019) is a common technique in image processing to enlarge the amount of training data. It is widely used, as the performance of a ML-based model depends on the amount of available data. We used several ways to augment data, which are summarized in Table 3. In addition to the standard translational augmentations, we used Contrast Limited Adaptive Histogram Equalization (CLAHE) (Zuiderveld, 1994) to enhance the local contrast of the image patches together with a channel-wise normalization.

We trained our weed segmentation models using Transfer Learning (TL) (He et al., 2016). We initialized the feature extractors using weights pre-trained on ImageNet (Deng et al., (2009–2009)). The decoder was initialized using He initialization as proposed by He et al. (He et al., (2015–2015)).

All code is implemented in Python 3.8 using the packages numpy (Mukherjee et al., 2019); pytorch (Hess et al., 1997); torchvision (Marcel and Rodriguez, 2010); albumentations (He et al., 2016); optuna (Veit et al., 2016); kornia (Fisher and Koltun, 2016) and scikit-learn (Prechelt et al., 2012). Optimizations were conducted using an Ubuntu 20.04 LTS machine with 104 CPU Cores, 756 GB memory and four GeForce RTX 3090 GPUs. Each model was trained on a single GPU. Models that did not fit on this GPU were trained on another machine using a NVIDIA A40 GPU with 48 GB VRAM.

3. Results

In this section, we first provide an overview of our results. Then, we analyze the predictions of our best performing weed segmentation model in more details. Finally, we discuss our results.

3.1. Results overview

First, we evaluated the architecture choice of the segmentation model in combination with four ResNet feature extractors employing a fourfold cross-validation. In Table 4, we show the results on the validation sets for these experiments using the Sørensen-Dice Coefficient (DS) summarized over all 50 trials, see Section 2.2. In summary, UNet with ResNet-34 as feature extractor performed best.

In general, the deepest feature extractor ResNet-101 performed worse for all segmentation models. The best performing feature extractor was dependent on the selected segmentation architecture. We observed minor differences in the objective value when comparing different feature extractors while using the same semantic segmentation model. In addition, there were minor differences in the objective values between different segmentation models UNet and DeepLabv3+. When using dilated convolutions in the feature extractor with the FCN-8s architecture (no concatenation of intermediate feature maps), the objective values had only minor differences compared to UNet and DeepLabv3+. This indicates a minor effect of the selection of the semantic segmentation model.

Both, FCN-16s and FCN-8s combined multiple output layers together and then upsampled the result. Here we observed a higher objective value in general, meaning worse predictions. In addition, the standard

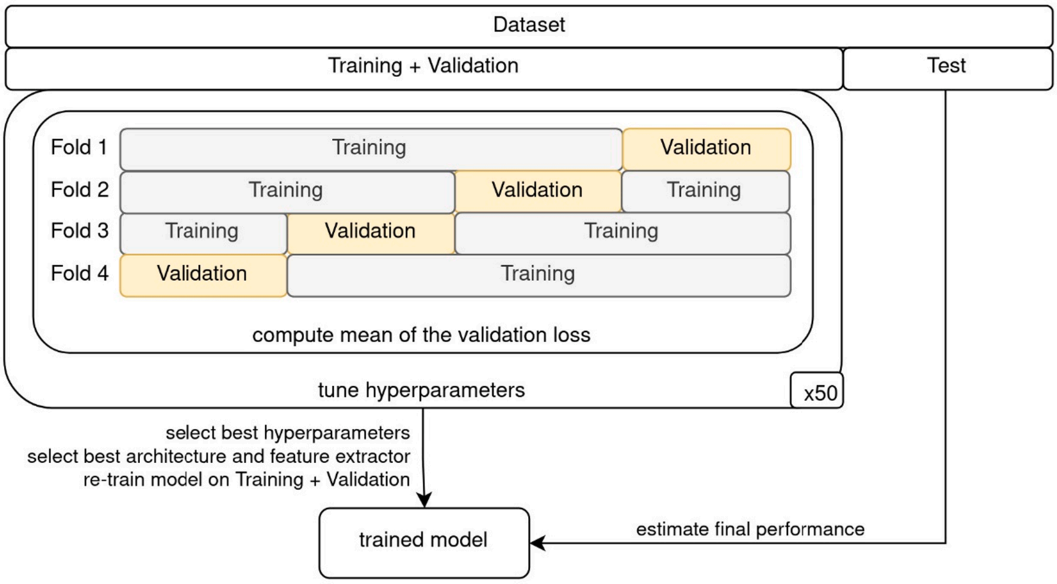

Fig. 3. Overview of the model selection and evaluation process conducted for each weed segmentation model.

Table 3
Data augmentation techniques used in this study.

Augmentation Transform	Description
Horizontal flip	Flips the image horizontally
Vertical flip	Flips the image vertically
RandomRotate90	Randomly rotates the image by a factor of 90 degrees
Transpose	Swaps rows and columns of the image
CLAHE	Applies CLAHE to the input image

Table 4
Best evaluation results on the validation sets for each weed segmentation model based on the Sørensen-Dice Coefficient for 50 trials. Standard deviation is provided in brackets. The best performing feature extractor for each segmentation architecture is denoted in bold.

	ResNet-18	ResNet-34	ResNet-50	ResNet-101
FCN-32s	0.0281 (0.0061)	0.0285 (0.0061)	**0.0275 (0.0056)**	0.0278 (0.0059)
FCN-16s	**0.0226 (0.16)**	0.0232 (0.16)	0.0255 (0.16)	0.0270 (0.16)
FCN-8s	**0.0134 (0.16)**	0.0136 (0.17)	0.0159 (0.17)	0.0160 (0.18)
FCN-8s + dilation	0.00999 (0.0095)	0.0101 (0.0044)	**0.00951 (0.0062)**	0.00963 (0.0052)
UNet	0.00923 (0.0041)	**0.00910 (0.0022)**	0.00913 (0.0033)	0.00926 (0.0035)
DeepLabv3+	0.0118 (0.0018)	0.00945 (0.0022)	**0.00939 (0.0022)**	N.A.*

* More than 48 GB of GPU memory were required.

deviation between the trials is approximately 28 times higher for FCN-16s and FCN-8s compared to FCN-32s. This holds true for all feature extractors examined in this study.

The best combination UNet with ResNet-34 was more robust during hyperparameter optimization, indicated by a small standard deviation of 0.0022.

3.2. *Best performing model*

Furthermore, the final performance was estimated on the above-described hold-out test-set using the best performing weed segmentation model (UNet with ResNet-34), which we re-trained on the complete training and validation data using the hyperparameter values that led to the lowest error. The final model has high macro-averaged evaluation values on the hold-out test-set, as shown in Table 5. Every metric is above 86% indicating sufficient generalization capabilities.

Fig. 4a shows the model's accuracy more precisely on a per-class basis by using a normalized confusion matrix. The soil was almost correctly predicted with high accuracy of 99.9%. The model classified 86.1% of all sorghum pixels correctly while misclassified 11.3% as background and 2.6% as weed. More importantly, 72.7% of all weed pixels were classified correctly, while predicting 23.2% as background and only 4.1% as sorghum. Additional per-class metrics can be found in Supplementary Table S5.

In addition, we tested the generalization abilities of our model (trained on sorghum at BBCH 17) using two different growth stages of sorghum (BBCH 15 and 19). Most importantly, our weed segmentation model trained on BBCH 17 is still able to detect weeds at least as well on images with lower or higher growth stages of sorghum (see Fig. 4b and 4c). Interestingly, for sorghum at BBCH 19, 77.6% of all weed pixels were classified correctly (i.e. an increase of 4.8% compared to results shown in Fig. 4a). However, the predictions on sorghum for different growth stages dropped to 50.0% for BBCH 15 (a decrease of 36.9%), mostly predicting weeds correctly. However, for BBCH 19 there was a decrease of only 1.1% to 85.8%, indicating consistent high performance on larger BBCH stages of sorghum.

Beyond that, we performed a qualitative analysis of the results on the hold-out test set and the additional patches from different growth stages. In general, the predictions of our weed segmentation model are accurate for the hold-out test set (as shown in Fig. 5a-f).

Most plants were segmented correctly, capturing the general shape of the plants. However, we observed most confusions at the borders of the plants. Unfortunately, these misclassifications of mainly the borders cannot be captured by the used evaluation metrics. Furthermore, most pixels belonging to "old weeds" (i.e. larger weeds that have been already present when sowing the crop) are predicted correctly, as shown in Fig. 5b, i and k. Only some parts of these instances were mis-classified as

Table 5
Macro-averaged results on the hold-out test-set in percent for the best performing weed segmentation model with UNet and ResNet-34.

DS	Precision	Recall	F1-score
99.69	93.01	86.25	89.37

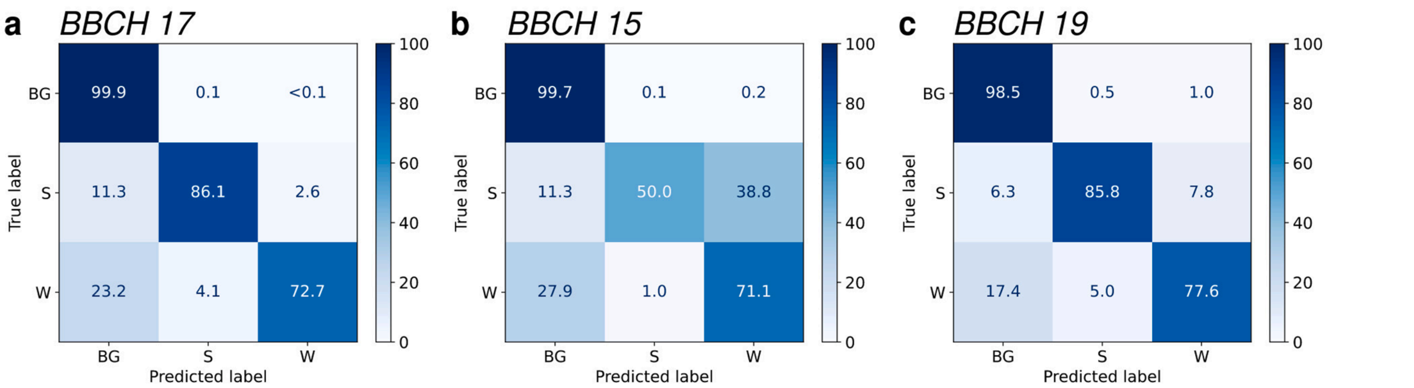

Fig. 4. Normalized confusion matrix by support size in percent shows the pixel-based classification results on the hold-out test-set. Background/soil is denoted by BG, sorghum by S and weed by W. **a** Sorghum at growth stage BBCH 17. **b** Sorghum at growth stage BBCH 15. **c** Sorghum at growth stage BBCH 19.

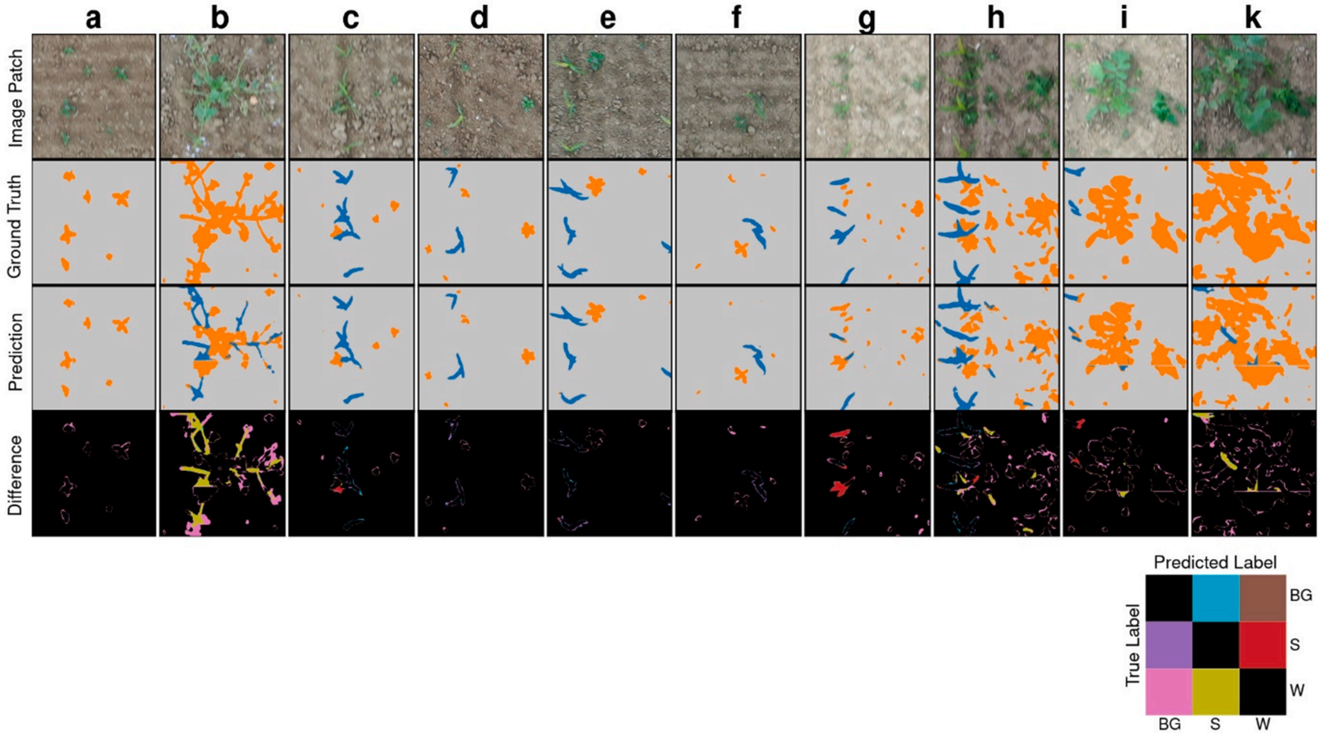

Fig. 5. Qualitative results on the hold-out test-set. Image patches of size 400x400 pixel^2 are cropped from each test image to show more details. Background (BG) is colored in gray, sorghum (S) in blue and weed (W) pixels in orange. The difference map shows only the misclassifications between ground truth and prediction.**a-c** Examples with a high degree of motion blur. **a** The general shape of weeds is predicted correctly. **b** Large weed plants that could not be removed before sowing the field. Most pixels are predicted correctly. There are small artifacts visible, which are due to the patching process. **c** Weeds intersecting with sorghum plants were predicted correctly. **d-f** Examples with a low degree of motion blur showing weeds and sorghum plants from different captures. The general shape is predicted correctly, where **d** is a patch from test_04, **e** from patch test_05 and **f** from patch test_06. **g-k** Comparison of two additional drone flights with different growth stages and different illumination. **g** Weed infestation when sorghum was at BBCH15, showing leaves of sorghum being predicted as weed. **h** The same excerpt as in c when sorghum was at BBCH 19. **i** Large weed plants that could not be removed before sowing the field is predicted mostly correct. Sorghum was at BBCH 15. **k** The same patch when sorghum was at BBCH 19.

sorghum, even though not many “old weed” instances were in the training set. Additionally, intra-row weeds were correctly detected, even if they are in close proximity or overlapping the crop (see Fig. 5c).

Plants in patches with a higher relative sharpness (for the sharpness calculations see Supplementary Text 1) were also predicted correctly, as shown in Supplementary Figure S5 and Fig. 5d-f. This indicates a high generalization ability between different degrees of motion blur.

For the additional growth stages, the main drawback of this model was the prediction of sorghum plants as weeds for BBCH 15 (Fig. 5g and 5i). This might be because of morphological changes due to growth (i.e. additional leaves were visible in BBCH 17 compared to BBCH 15). Here, in most of the examples only a part of the plant was misclassified, but occasionally there were also complete sorghum plants predicted as weed. The general weed distribution and the shape of large weeds was predicted correctly. Only some parts of weeds were predicted as sorghum. For BBCH19, the model was still able to predict weeds correctly, despite the fact, that they had grown larger than the weeds the model was trained on. In addition, “old weeds” that grew even larger were mostly predicted correctly, as shown in Fig. 5k. Some artifacts are visible especially for larger plants, which are due to the patching process, as one plant was partly cut in different patches resulting in small parts present on the patch border. These small parts were then not predicted correctly, as denoted in yellow in Fig. 5i and k.

Finally, the images of BBCH 15 and BBCH 19 were captured with different illumination settings than the model was trained on. Interestingly, weeds in these captures were still predicted with high precision. This indicates that our model might be applicable to captures of different image quality.

The complete dataset and code are publicly available on Mendeley-Data[1] and in our *GitHub* repository.[2]

3.3. Discussion

In this paper, we present a DL-based approach to weed detection and segmentation in sorghum fields under real-world conditions using motion blurred UAV images. Our dataset was generated with a high GSD of one millimeter by conducting drone missions from a low altitude (five meters) in sorghum fields. This might be contrary to several other approaches, which employed a higher flight altitude to ensure higher throughput. Yet, this high GSD allowed us to label and predict smaller weeds potentially resulting in a more accurate weed detection model. In addition, the ground truth was generated using a precise approach with a pixel-based semantic segmentation and expert-curated annotations for the complete dataset. There are several color- or threshold-based approaches, as discussed in a recent survey (Hamuda et al., 2016). The quality of those segmentations might vary considerably with the imaging and weather conditions, potentially resulting in a wrong ground truth, especially when using drone captures degraded by motion blur. As mentioned in this survey, changing weather and imaging conditions should not influence DL-based approaches, as one could use data augmentation to adjust the exposure and saturation.

Due to wind-related displacements, motion blur is a common challenge in drone imagery of agricultural landscapes. However, the effect of real motion blur on weed detection models is not studied widely. In this work, we show that DL models are able to accurately segment weeds from crop in blurry drone images independent on the degree of motion blur. The main bottleneck when developing new supervised DL models is the amount of data to label. We show that sorghum plants were segmented with a high precision in general when using around 3.000 labeled instances. One reason for this might be the size and the shape of the sorghum plants in comparison with weed instances. The crop plants were larger and had a smaller variance in shape. Additionally, overlapping sorghum plants were predicted sufficiently. Sharp edges were not visible due to motion blur and therefore the model could not detect any edges to predict the correct class. It had to focus on different features of the plants to discriminate between sorghum and weed. The model learned the background class well and was able to generalize on different backgrounds. Our dataset had few instances of “old weeds” (large weeds already present when sowing the crop) that were outliers from the normal weed’s distribution. Although the shape was not predicted correctly, still most of these weed pixels were correct. Our model did not segment small weeds, as this is also a general bottleneck in DL approaches and implies that conducting UAV missions with a higher GSD is favorable. However, the overall spatial weed distribution still matches the ground truth, which is the main objective in SSWM. More important, our model is not dependent on the detection of crop rows and is able to detect intra-row weeds, as shown in Fig. 5c and d. Consequently, crops that were not sown in precise rows could be detected using our model.

In this study, UNet with ResNet-34 performed best. This indicates that a large feature extractor (ResNet-101) is overfitting towards the training data more easily. There were slight differences in performance when comparing the architectures of the segmentation models. UNet was only slightly better than DeepLabv3+ and FCN-8s with dilated convolutions, indicating that the choice of the semantic segmentation model is not important when using UAV captures with a high GSD of 0.1 cm. The design choice of using dilated convolutions in the FCN-8s model achieved the highest relative improvement. The UNet architecture yielded the best results, indicating that a more gradual upsampling with skip connections is more favorable than atrous convolutions used in DeepLabv3+.

In our study, we observed a discrepancy between evaluation metrics and a qualitative assessment of the actual predictions. The confusion matrix treats every pixel equally, which might not be useful in weed segmentation approaches. By further analyzing the model’s predictions, we could conclude that most mis-classified pixels were on the border of plants. However, with a real-world SSWM application in mind, plant borders are not that important, as the general shape is sufficient to generate accurate weed maps. A correct classification of crop and weed pixels is more important, as this is the main challenge. Furthermore, rare occurring but large weeds and the high percentage of the background class distorted the results obtained from the confusion matrix, as misclassification of those weeds had a higher impact. In our dataset, 98% of all pixels were background, which had a significant impact on per-pixel metrics and the confusion matrix.

The main goal of weed detection for SSWM is to generate a weed density map, so weeds can be removed in downstream tasks. These tasks (i.e., weed removal by a field robot or herbicide application) are not pixel-precise, so arguably a coarser ground truth might be also sufficient while lowering the annotation cost and time. Our dataset could act as a base for further weed research using drones, as a coarser ground truth (i.e. for object detection or patch-based classification) could be generated automatically. One major difficulty in using a pixel-based segmentation approach is the ambiguity of the ground truth labels due to a subjectivity in the labeling process. This ambiguity is challenging for model training and evaluation especially when using real-world motion blurred imagery.

The application of our model on other BBCH stages showed good generalization capabilities towards the larger growth stage (BBCH 19). When presented with imagery of a smaller growth stage (BBCH 15), our model wrongly predicted around 39% of sorghum pixels as weed. Further analysis showed that a majority of the misclassifications were present on smaller sorghum plants where only two or three leaves were visible. This indicates that our model needs to be re-trained with UAV imagery of different growth stages in order to be more robust and applicable for SSWM.

Some errors in the predictions were also due to the patch generation process. They happen only, when there is a small part of a plant present on a patch border. The network was trained on mostly intact plants, so when presented with a partly cut plant, it might misclassify it. In addition, background pixels are mostly predicted in close proximity of the patch borders. These errors are not severe, as they only occur for larger plants that were separated in the patching process and might be solved by using a sliding window approach and generating patches with overlap.

While our model performed sufficiently well on our hold-out test-set, there are several factors that were not evaluated in this study. Our test-set was captured on a certain agricultural field. However, the weed flora might be different in other regions or fields. Thus, it is likely that there are several other weed species growing on different agricultural landscapes. In addition, the sampling time (time of the capture) of the different datasets is similar, as we conducted the drone missions at noon. Different sampling times (morning, noon or evening) and different weather conditions (sunny, cloudy) might have an effect on the illumination (white balance, color temperature) of the resulting captures, which might have a negative effect on the model’s performance. Nevertheless, our experiments on the datasets sorghum_15 and sorghum_19 with different illuminations indicate good generalization of our model. In this study, we briefly investigate the generalization capabilities of our model on two different growth stages. Our results

[1] https://doi.org/10.17632/4hh45vkp38.4.
[2] https://github.com/grimmlab/UAVWeedSegmentation.

suggest that it might be necessary to fine-tuning our model on different growth states which would improve the overall performance. A larger dataset with additional growth stages could be beneficial to yield a unified model for early weed detection across different stages. Additionally, drone settings, i.e. flight speed and altitude might play an important role on the quality of the captures and might influence the predictions as well. Therefore, we plan to generate a more diverse dataset to evaluate different model architectures on a wider scope, dealing with a more diverse weed flora, weather conditions and drone settings.

4. Conclusion

In this study we present a machine learning framework enabling early weed and sorghum detection and segmentation in real-world, degraded drone images. The model has high segmentation abilities and can detect intra-row weeds as wells as overlapping plants in captures with different degrees of motion blur. Further, a qualitative analysis of the actual predictions indicates that the general form of most plants can be detected accurately, with minor problems at borders of the plants. As discussed, plant borders are not that important in a downstream application for site specific weed management.

Additionally, we show that our method can be applied on different BBCH stages of sorghum. However, the predictions on smaller sorghum plants in an early growth stage dropped and might lead to an overestimation of weed instances. Nevertheless, our work shows promising results and give first insights that we plan to extend in future research.

With this study, we publish the first dataset for weed detection in sorghum, which can be used as a basis for future research. Additionally, we made the code of our proposed method publicly available to ensure the reproducibility of results and allow the comparison in future studies.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The generated and labeled dataset and the code for our proposed machine learning–based model can be found on Mendeley data https://doi.org/10.17632/4hh45vkp38.4 and GitHub: https://github.com/grimmlab/UAVWeedSegmentation.

Acknowledgments

This work was supported by the Weihenstephan-Triesdorf University of Applied Sciences and the Technical University of Munich, Campus Straubing for Biotechnology and Sustainability.

Funding

Funding for the research presented in this paper is provided by the Bavarian State Ministry for Food, Agriculture and Forests within the EWIS project (Funding ID: G2/N/19/13).

Authors' contributions

DGG, NG, MG conceived and designed the study. RA conducted the drone flights and the image acquisition. NG and ZG labeled the dataset. RA and MG guided the labeling process with domain expertise. NG implemented the machine-learning pipeline and conducted all computational experiments with support from FH. NG analyzed the results with help from DGG. NG and DGG wrote the manuscript with contributions from all authors. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.
Consent for publication.
Not applicable.
Competing interests.
The authors declare that they have no competing interests.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compag.2022.107388.

References

Aktar, M.W., Sengupta, D., Chowdhury, A., 2009. Impact of pesticides use in agriculture: their benefits and hazards. Interdiscip Toxicol 2, 1–12. https://doi.org/10.2478/v10102-009-0001-7.

Alom, M.Z., Taha, T.M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M.S., Hasan, M., Van Essen, B.C., Awwal, A.A.S., Asari, V.K., 2019. A state-of-the-art survey on deep learning theory and architectures. Electronics (Switzerland) 8 (3), 292.

Badrinarayanan, V., Kendall, A., Cipolla, R., 2015. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation.

Bergstra, J., Bengio, Y., 2012. Random Search for Hyper-Parameter Optimization. J. Mach. Learn. Res. 13, 281–305.

Bertels, J., Eelbode, T., Berman, M. et al., 2019. Optimizing the Dice Score and Jaccard Index for Medical Image Segmentation: Theory and Practice. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 11765 LNCS:92–100. https://doi.org/10.1007/978-3-030-32245-8_11.

Chen, L.-C., Papandreou, G., Schroff, F. et al., 2017. Rethinking Atrous Convolution for Semantic Image Segmentation. 2331–8422.

Chen, S.W., Shivakumar, S.S., Dcunha, S., Das, J., Okon, E., Qu, C., Taylor, C.J., Kumar, V., 2017. Counting Apples and Oranges With Deep Learning: A Data-Driven Approach. IEEE Rob. Autom. Lett. 2 (2), 781–788.

Deng, J., Dong, W., Socher, R., et al., 2009. ImageNet: A large-scale hierarchical image database. In: In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 248–255.

Dian Bah, M., Hafiane, A., Canals, R., 2018. Deep learning with unsupervised data labeling for weed detection in line crops in UAV images. Remote Sens. 10, 1–22. https://doi.org/10.3390/rs10111690.

dos Santos Ferreira, A., Matte Freitas, D., Gonçalves da Silva, G., Pistori, H., Theophilo Folhes, M., 2017. Weed detection in soybean crops using ConvNets. Comput. Electron. Agric. 143, 314–324.

Fernández-Quintanilla, C., Peña, J.M., Andújar, D., Dorado, J., Ribeiro, A., López-Granados, F., Smith, R., 2018. Is the current state of the art of weed monitoring suitable for site-specific weed management in arable crops? Weed Res. 58 (4), 259–272.

Fisher, Yu, Vladlen Koltun, 2016. Multi-Scale Context Aggregation by Dilated Convolutions. In: Yoshua Bengio, Yann LeCun (Eds.). 4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings.

Genze, N., Bharti, R., Grieb, M., Schultheiss, S.J., Grimm, D.G., 2020. Accurate machine learning-based germination detection, prediction and quality assessment of three grain crops. Plant Methods 16 (1). https://doi.org/10.1186/s13007-020-00699-x.

Grinblat, G.L., Uzal, L.C., Larese, M.G., Granitto, P.M., 2016. Deep learning for plant identification using vein morphological patterns. Comput. Electron. Agric. 127, 418–424.

Hamuda, E., Glavin, M., Jones, E., 2016. A survey of image processing techniques for plant extraction and segmentation in the field. Comput. Electron. Agric. 125, 184–199.

Hasan, A.S.M.M., Sohel, F., Diepeveen, D., Laga, H., Jones, M.G.K., 2021. A survey of deep learning techniques for weed detection from images. Comput. Electron. Agric. 184, 106067.

Hashemi, M., 2019. Enlarging smaller images before inputting into convolutional neural network: zero-padding vs. interpolation. J. Big Data 6. https://doi.org/10.1186/s40537-019-0263-7.

Hay, A.A.A., Castilla, G., 2006. Object-based image analysis: strengths, weaknesses, opportunities and threats (SWOT). Int. Arch. Photogramm., Remote Sens. Spat. Inform. Sci. 36, 4.

He, K., Zhang, X., Ren, S. et al., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2016-Decem:770–778. https://doi.org/10.1109/CVPR.2016.90.

He, K., Zhang, X., Ren, S., et al., 2015. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In: 2015 IEEE International Conference on Computer Vision (ICCV). IEEE, pp. 1026–1034.

Hess, M., Barralis, G., Bleiholder, H., Buhr, L., Eggers, T.H., Hack, H., Stauss, R., 1997. Use of the extended BBCH scale - General for the descriptions of the growth stages of mono- and dicotyledonous weed species. Weed Res. 37 (6), 433–441.

Holt, J.S., 2004. Principles of Weed Management in Agroecosystems and Wildlands Author (s): Jodie S. Holt Source: Weed Technology, 2004, Vol. 18, Invasive Weed

Symposium (2004), pp. 1559–1562 Published by: Cambridge University Press on behalf of the Weed Science. Weed Technology 18:1559–1562.
Huang, H., Deng, J., Lan, Y., Yang, A., Deng, X., Zhang, L., Gonzalez-Andujar, J.L., 2018. A fully convolutional network for weed mapping of unmanned aerial vehicle (UAV) imagery. PLoS ONE 13 (4), e0196302.
Huang, H., Lan, Y., Deng, J., Yang, A., Deng, X., Zhang, L., Wen, S., 2018. A semantic labeling approach for accurate weed mapping of high resolution UAV imagery. Sensors (Switzerland) 18 (7), 2113.
Islam, N., Rashid, M.M., Wibowo, S., Xu, C.-Y., Morshed, A., Wasimi, S.A., Moore, S., Rahman, S.M., 2021. Early weed detection using image processing and machine learning techniques in an australian chilli farm. Agriculture (Switzerland) 11 (5), 387.
Kamilaris, A., Prenafeta-Boldu, F.X., 2018. Deep learning in agriculture: A survey. Comput. Electron. Agric. 147, 70–90. https://doi.org/10.1016/j.compag.2018.02.016.
Kamilaris, A., Kartakoullis, A., Prenafeta-Boldú, F.X., 2017. A review on the practice of big data analysis in agriculture. Comput. Electron. Agric. 143, 23–37. https://doi.org/10.1016/j.compag.2017.09.037.
Kazmi, W., Garcia-Ruiz, F.J., Nielsen, J., Rasmussen, J., Jørgen Andersen, H., 2015. Detecting creeping thistle in sugar beet fields using vegetation indices. Comput. Electron. Agric. 112, 10–19.
Kingma DP, Ba J (2014) Adam: A Method for Stochastic Optimization.
Kussul, N., Lavreniuk, M., Skakun, S., Shelestov, A., 2017. Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. IEEE Geosci Remote Sens. Lett 14 (5), 778–782.
Lam, O.H.Y., Dogotari, M., Prüm, M., Vithlani, H.N., Roers, C., Melville, B., Zimmer, F., Becker, R., 2021. An open source workflow for weed mapping in native grassland using unmanned aerial vehicle: using Rumex obtusifolius as a case study. Eur. J. Remote Sens. 54 (sup1), 71–88.
LeCun, Y.A., Bengio, Y., 1998. Convolutional networks for images, speech, and time series.
Liu, B., Bruch, R., 2020. Weed Detection for Selective Spraying: a Review. Curr. Rob. Reports 1, 19–26. https://doi.org/10.1007/s43154-020-00001-w.
Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: CVPR 2015, pp. 3431–3440.
Lottes, P., Stachniss, C., 2017. Semi-supervised online visual crop and weed classification in precision farming exploiting plant arrangement. In: IEEE International Conference on Intelligent Robots and Systems, 2017-Septe. Institute of Electrical and Electronics Engineers Inc, pp. 5155–5161.
Lottes, P., Behley, J., Milioto, A., Stachniss, C., 2018. Fully convolutional networks with sequential information for robust crop and weed detection in precision farming. IEEE Rob. Autom. Lett. 3 (4), 2870–2877.
Lu, Y., Young, S., 2020. A survey of public datasets for computer vision tasks in precision agriculture. Comput. Electron. Agric. 178, 105760 https://doi.org/10.1016/j.compag.2020.105760.
Marcel, S., Rodriguez, Y., 2010. Torchvision the machine-vision package of torch. In: Del Bimbo, A., Chang, S.-F., Smeulders, A. (Eds.). Proceedings of the international conference on Multimedia - MM '10. ACM Press, New York, New York, USA, pp. 1485.
Meyer, G.E., Neto, J.C., 2008. Verification of color vegetation indices for automated crop imaging applications. Comput. Electron. Agric. 63, 282–293. https://doi.org/10.1016/j.compag.2008.03.009.
Milioto, A., Lottes, P., Stachniss, C., 2018. Real-Time Semantic Segmentation of Crop and Weed for Precision Agriculture Robots Leveraging Background Knowledge in CNNs. In: Proceedings - IEEE International Conference on Robotics and Automation. Institute of Electrical and Electronics Engineers Inc, pp. 2229–2235.
Mukherjee, A., Misra, S., Raghuwanshi, N.S., 2019. A survey of unmanned aerial sensing solutions in precision agriculture. J. Netw. Comput. Appl. 148, 102461 https://doi.org/10.1016/j.jnca.2019.102461.
Onyango, C.M., Marchant, J.A., 2003. Segmentation of row crop plants from weeds using colour and morphology. Comput. Electron. Agric. 39, 141–155. https://doi.org/10.1016/S0168-1699(03)00023-1.
Patel, D.D., Kumbhar, B.A., 2016. Weed and its management: A major threats to crop economy. Journal of. Pharmaceut. Sci. Biosci. Res. 6, 453–758.
Penatti, O.A.B., Nogueira, K., dos Santos, J.A., 2015. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 44–51.
Pérez-Ortiz, M., Peña, J.M., Gutiérrez, P.A., Torres-Sánchez, J., Hervás-Martínez, C., López-Granados, F., 2015. A semi-supervised system for weed mapping in sunflower crops using unmanned aerial vehicles and a crop row detection method. Appl. Soft Comput. 37, 533–544.
Prechelt, L., 2012. Early Stopping — But When? In: Montavon, G., Orr G.B., Müller, K.-R. (Eds.). Neural Networks: Tricks of the Trade, vol. 7700. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 53–67.
Ramirez, W., Achanccaray, P., Mendoza, L.F., et al., 2020. Deep convolutional neural networks for weed detection in agricultural crops using optical aerial images. Int. Arch. Photogramm., Remote Sens. Spat. Inform. Sci. - ISPRS Arch. 42, 551–555. https://doi.org/10.5194/isprs-archives-XLII-3-W12-2020-551-2020.
Ridgway, R.L., Tinney, J.C., MacGregor, J.T., Starler, N.J., 1978. Pesticide use in agriculture. Environ. Health Perspect. 27, 103–112.
Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W.M. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, vol 9351. Springer International Publishing, Cham, pp. 234–241.
Sa, I., Popović, M., Khanna, R., Chen, Z., Lottes, P., Liebisch, F., Nieto, J., Stachniss, C., Walter, A., Siegwart, R., 2018. WeedMap: A large-scale semantic weed mapping framework using aerial multispectral imaging and deep neural network for precision farming. Remote Sens. 10 (9), 1423.
Shorten, C., Khoshgoftaar, T.M., 2019. A survey on Image Data Augmentation for Deep Learning. J. Big Data 6. https://doi.org/10.1186/s40537-019-0197-0.
Sudre, C.H., Li, W., Vercauteren, T., et al., 2017. (2017) Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. Deep Learn. Med. Image Anal. Multimodal Learn. Clin. Decis. Support 2017, 240–248. https://doi.org/10.1007/978-3-319-67558-9_28.
Sun, C., Shrivastava, A., Singh, S. et al., 2017. Revisiting Unreasonable Effectiveness of Data in Deep Learning Era. In: ICCV 2017, pp. 843–852.
Vapnik, V., 1992. Principles of risk minimization for learning theory. Adv. Neural Inform. Process. Syst. 831–838.
Veit, A., Wilber, M., Belongie, S., 2016. Residual networks behave like ensembles of relatively shallow networks. Adv. Neural Inform. Process. Syst. 550–558.
Who, 1990. Public health impact of pesticides used in agriculture. World Health Organization, Geneva.
Yeom, J., Jung, J., Chang, A., Ashapure, A., Maeda, M., Maeda, A., Landivar, J., 2019. Comparison of Vegetation Indices Derived from UAV Data for Differentiation of Tillage Effects in Agriculture. Remote Sens. 11 (13), 1548.
Zheng, Y., Zhu, Q., Huang, M., Guo, Y.a., Qin, J., 2017. Maize and weed classification using color indices with support vector data description in outdoor fields. Comput. Electron. Agric. 141, 215–222.
Zuiderveld, K., 1994. Contrast Limited Adaptive Histogram Equalization. In: Graphics Gems. Elsevier, pp. 474–485.

Supplementary Information

Deep Learning-based Early Weed Segmentation using Motion Blurred UAV Images of Sorghum Fields

Nikita Genze[1,2,*], Raymond Ajekwe[4], Zeynep Güreli[1,2], Florian Haselbeck[1,2], Michael Grieb[4], Dominik G. Grimm[1,2,3,*]

[1] Technical University of Munich, TUM Campus Straubing for Biotechnology and Sustainability, Bioinformatics, Schulgasse 22, 94315 Straubing, Germany

[2] Weihenstephan-Triesdorf University of Applied Sciences, Bioinformatics, Petersgasse 18, 94315 Straubing, Germany

[3] Technical University of Munich, Department of Informatics, Boltzmannstr. 3, 85748 Garching, Germany

[4] Technology and Support Centre in the Centre of Excellence for Renewable Resources (TFZ), Schulgasse 20, 94315 Straubing, Germany

[*] Correspondence: nikita.genze@tum.de, dominik.grimm@hswt.de

Supplementary Information

Dataset Statistics

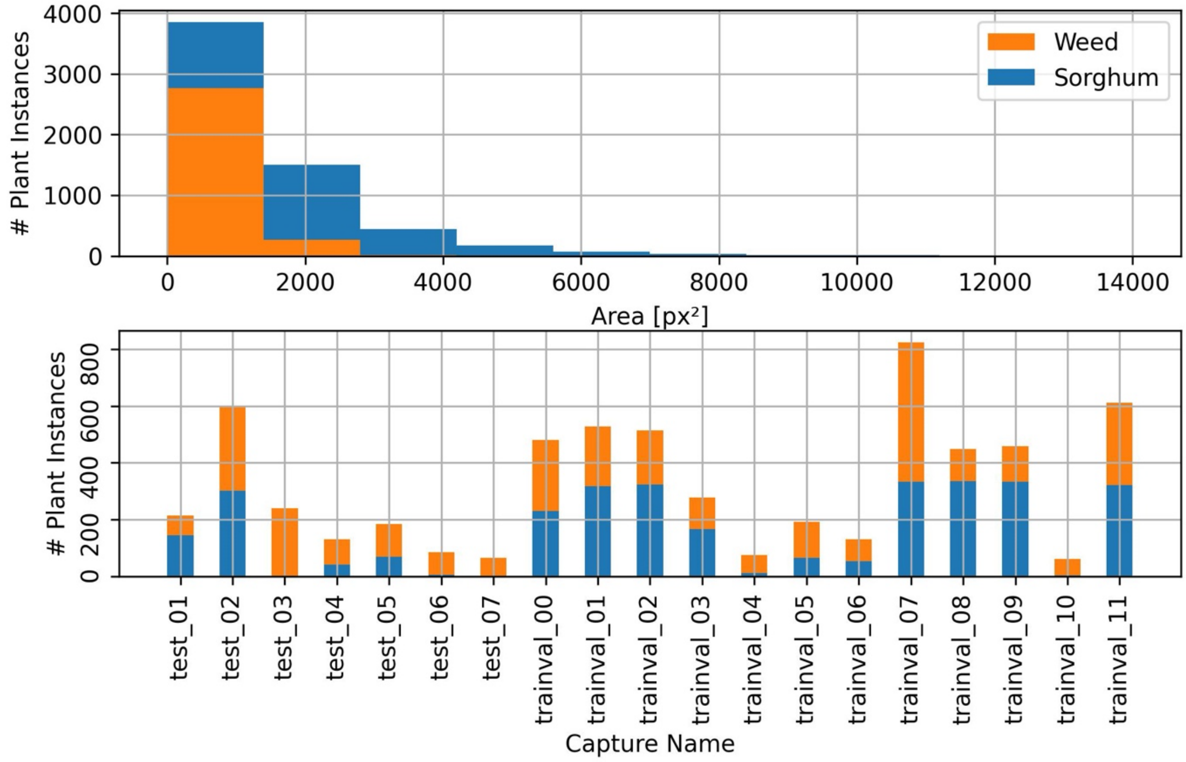

Figure S1: **Statistics of the sorghum_17 dataset**. Top: Number of plant instances by area. Bottom: Number of plant instances per drone capture. The legend matches in both diagrams, where sorghum plants are denoted in blue and weed plants in orange. Overlapping instances are counted as one.

Supplementary Text 1: Sharpness Indicators

UAV captures may suffer from several distortions during acquisition (and compression). The acquisition step is especially fragile when using a low-cost drone. It cannot stabilize sufficiently before capture, which leads to visible motion blur in the images. Visible motion blur is subjective and the extent is difficult to measure even by a human oracle [1]. Image Quality Assessment (IQA) techniques aim to measure the quality of images in an automated fashion [2]. They can be divided into three groups: Full-reference IQA compares the degraded image to a high-quality reference image. Reduced-reference IQA takes into account partial information of the reference image. Non-reference IQA does not compare the distorted image, as there is no high-quality reference image available. It calculates indices for image quality only based on the pixel values of the degraded image. For this task, the type of distortion must be known between the reference and degraded image [3]. In addition, the proposed methods are highly dependent on the image content. One major drawback of non-reference IQA is the variation in the image content, illumination, noise and spatial activity, which have an effect on the blur estimation of images [4, 5]. These methods may yield highly different scores for images with similar subjective blur estimation or inconsistent scores for subjectively differently blurred images. A current review classified and systematically compared 13 representative methods and provided some references [2]. They conclude that most existing methods perform well on gaussian blurred images but fail estimating the image quality of realistically blurred images.

Here we focus on non-reference IQA, as there were no high-quality reference images captured for our dataset. In addition, we focus on IQA that is specific to motion blur, e.g. camera motion or camera shake. In general, blur-specific non-reference IQA can be roughly categorized into edge-based and wavelet (Fourier-like) methods [1, 6]. In our experiments we selected several blur estimation methods, which are briefly defined next.

The Total Variance of Laplacian (TVoL) [7] is a quick and accurate edge-based method for scoring how sharp an image is. It assumes, that blurry images have fewer edges than sharp images. The Laplacian operator highlights well-defined edges in an image. Therefore, the calculated variance will be higher if there are more edges present.

The sharpness of natural scenes [8] can also be estimated using second order derivatives in grayscale images. The Difference of Medians (DoM) calculates the changes in the difference as an indicator for sharpness. The authors tested this method on images of documents with high contrast. It can also be applied on images of natural scenes and has been shown to give a better estimate than methods using edge-width.

The Discrete Fourier Transform (DFT) [9] is a wavelet-based method that can be used as indicator for image blur [2, 10]. High spatial frequencies in the Fourier domain are reduced when blur is present in an image [10].

Eventually, a method for scoring general image quality is called Blind/Referenceless Image Spatial QUality Evaluator (BRISQUE) [11]. It uses a natural image model (Support Vector Regression), calculates several statistics (i.e. pixel intensity) of the query image, and compares them with the distribution of natural images. The model is trained on a variety of known distortions, such as compression artifacts, noise and blur. The difference of the distributions is then used as a measure of image quality. In our dataset, the most dominant distortion is motion blur of the drone, so we estimate the sharpness of the image using the general image quality.

We conducted experiments using an additionally collected dataset of matching sharp and blurry image pairs to compare different blur indicators. This dataset was used to estimate a threshold for the subsequent blurriness analysis. Afterwards we selected the best performing measure and applied it to the blurry dataset described in this manuscript.

The evaluation of sharpness estimates was done on 256x256 $pixel^2$ patches using an additional dataset with sorghum in BBCH stages 11-13. Therefore, the blurry and sharp patch were matched spatially to ensure the same image content. This dataset contains only patches, where weed or sorghum is present; there are no patches of only background pixels. Example patches are shown in Figure S2.

Figure S2: Examples from the additional dataset used to evaluate different sharpness indicators. The growth stage of sorghum was measured at BBCH stage 11-13 a Sorghum and weed instances are visible in one patch. **b** Only sorghum instances are present. **c** Only weeds are visible.

The sharpness indicators used in this study have different output scales, so the relative score between the sharpest and most blurry patch in the dataset was calculated. This ensured comparable results. In addition, no absolute scoring about the sharpness of an image is needed, which is another subjective factor that might affect the results.

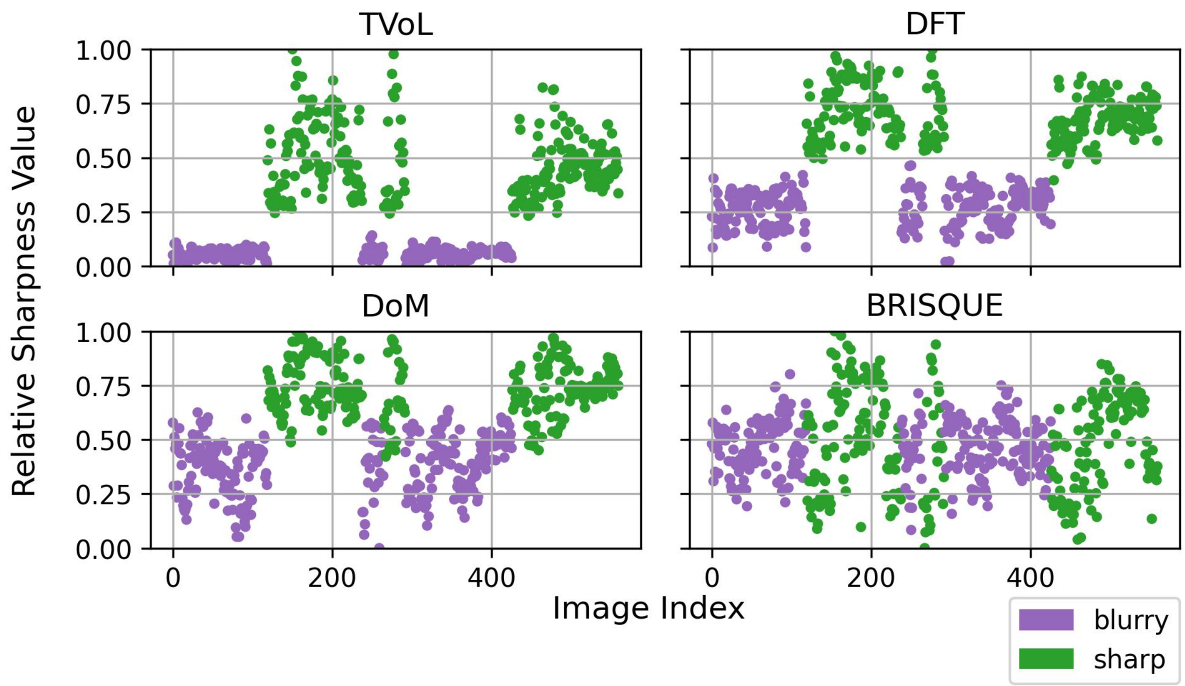

Figure S3: Relative Sharpness Value of different methods for non-reference IQA. High values indicate sharper images than lower values. A total of 560 images patches from an extra dataset were used, where the blurry and sharp patch contained the same content. The best separation could be achieved by using TVoL. DFT scores are also a viable option, as only a few patches cannot be separated. DoM and BRISQUE scores were not competitive in our experiments.

The TVoL method was then applied on all image patches of the dataset published with this work. The relative sharpness is shown in Figure S4a. Most captures had a small relative sharpness value of <0.2, indicating a severe degree of motion blur. The GPS coordinates of the captures showed, that relatively sharp patches were located at the edges of the flight plan (Figure S4b). There the drone needed to slow down to rotate before taking a new capture. In addition, one can see that some captures from the test-set were not located at the edges and thus were severely degraded by motion blur. Additionally, relatively sharp captures were present in the test-set. The relative sharpness calculated by TVOL matches the manual assessment of the images, as shown in Figure S4c.

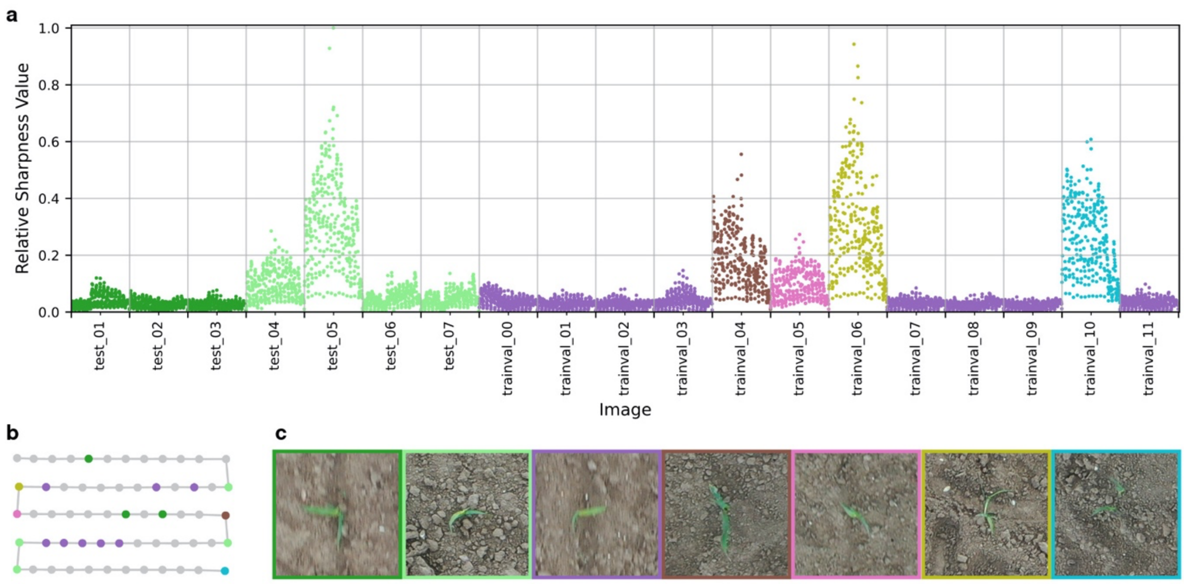

Figure S4: Sharpness assessment for the dataset used in this study. The calculated sharpness matches the manual assessment of this dataset, as images 04, 05, 06 and 10 from the trainval dataset were relatively sharper as the other captures. The test-set captures were degraded with different degrees of motion blur with 01, 02 and 03 being blurred the most. The color code matches in the sub-figures. **a** Relative Sharpness Value based on TVoL. Lower values indicate a higher degree of motion blur. Most captures in the dataset were blurred, but there were also some relatively sharper images present. **b** Flight plan of the used dataset showing the location of the captures. Gray dots indicate captures that were not annotated. Relatively sharp captures were located at the edges of the flight plan, where the drone was flying slower to rotate, as indicated by the limegreen, brown, pink, yellow and cyan dots. **c** Example patches from the captures, where color of the border indicates the corresponding capture.

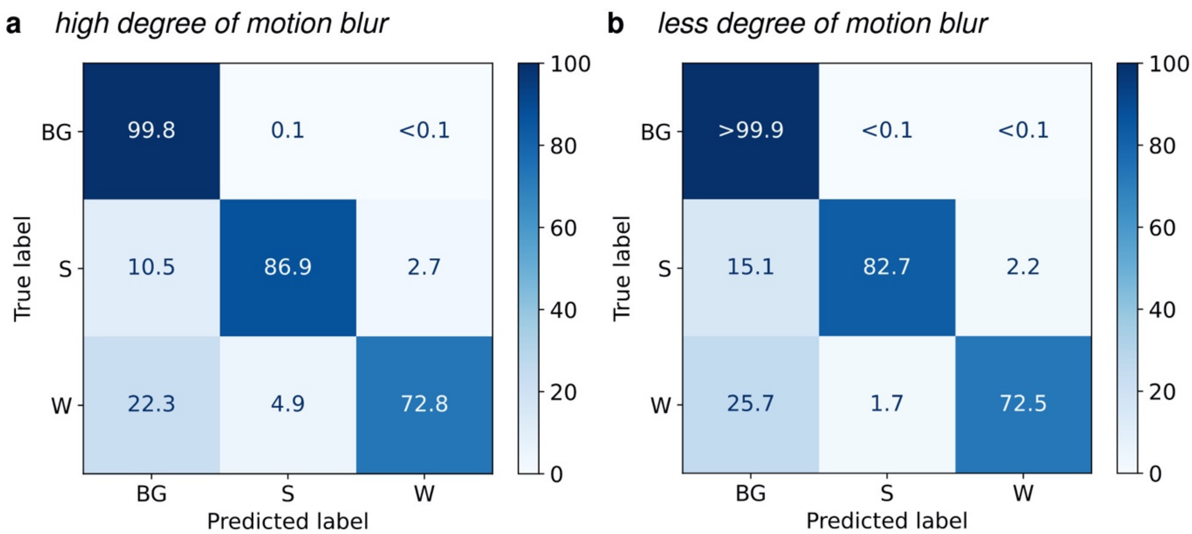

Figure S5: Normalized confusion matrix by support size in percent shows the pixel-based classification results on the hold-out test-set. Background/soil is denoted by BG, sorghum by S and weed by W. **a** Test images with a high degree of motion blur (test_01 to test_03, as indicated in Supplementary Figure S4). **b** Test images with less degree of motion blur (test_04 to test_07, as indicated in Supplementary Figure S4).

Additional Model Evaluation Statistics

Table S5 Per-class metrics of the hold-out test-set for *sorghum_17* in percent, the amount of pixels in the ground-truth is denoted as support.

class	precision	recall	f1-score	support
Background	99.80	99.93	99.86	137909280
Sorghum	91.58	86.10	88.76	1249145
Weed	87.64	72.71	79.48	574567
macro avg	93.01	86.25	89.37	139732992
weighted avg	99.68	99.69	99.68	139732992

References

1. Tiwari S, Shukla VP, Singh AK, Biradar SR. Review of Motion Blur Estimation Techniques. JOIG. 2014:176–84. doi:10.12720/JOIG.1.4.176-184.
2. Li D, Jiang T. Blur-Specific No-Reference Image Quality Assessment: A Classification and Review of Representative Methods. In: Jiang M, Ida N, Louis AK, Quinto ET, editors. The Proceedings of the International Conference on Sensing and Imaging. Cham: Springer International Publishing; 2019. p. 45–68. doi:10.1007/978-3-319-91659-0_4.
3. Wang Z, Simoncelli EP. Reduced-reference image quality assessment using a wavelet-domain natural image statistic model. In: Rogowitz BE, Pappas TN, Daly SJ, editors. Electronic Imaging 2005; Monday 17 January 2005; San Jose, CA: SPIE; 2005. p. 149. doi:10.1117/12.597306.
4. Vu CT, Phan TD, Chandler DM. A spectral and spatial measure of local perceived sharpness in natural images. IEEE Trans Image Process. 2012;21:934–45. doi:10.1109/TIP.2011.2169974.
5. Ciancio A, Da Costa ALNT, Da Silva EAB, Said A, Samadani R, Obrador P. No-reference blur assessment of digital pictures based on multifeature classifiers. IEEE Trans Image Process. 2011;20:64–75. doi:10.1109/TIP.2010.2053549.
6. Fang X, Wu H, Wu Z, Luo B. An Improved Method for Robust Blur Estimation. Information Technology J. 2011;10:1709–16. doi:10.3923/ITJ.2011.1709.1716.
7. De K, Masilamani V. Image Sharpness Measure for Blurred Images in Frequency Domain. Procedia Engineering. 2013;64:149–58. doi:10.1016/j.proeng.2013.09.086.
8. Kumar J, Chen F, David Doermann, editors. Sharpness Estimation for Document and Scene Images; 2012.
9. Cannon M. Blind deconvolution of spatially invariant image blurs with phase. IEEE Trans. Acoust., Speech, Signal Process. 1976;24:58–63. doi:10.1109/TASSP.1976.1162770.
10. Marziliano P, Dufaux F, Winkler S, Ebrahimi T. Perceptual blur and ringing metrics: application to JPEG2000. Signal Processing: Image Communication. 2004;19:163–72. doi:10.1016/j.image.2003.08.003.
11. Mittal A, Moorthy AK, Bovik AC. Blind/Referenceless Image Spatial Quality Evaluator. In: 2011 45th Asilomar Conference on Signals, Systems and Computers; 06.11.2011 - 09.11.2011; Pacific Grove, CA, USA: IEEE; 06.11.2011 - 09.11.2011. p. 723–727. doi:10.1109/ACSSC.2011.6190099.

B Publication B

The following pages show the full version of the published manuscript:

Publication B

Nikita Genze, Maximilian Wirth, Christian Schreiner, Raymond Ajekwe, Michael Grieb & Dominik G. Grimm (2023). Improved weed segmentation in UAV imagery of sorghum fields with a combined deblurring segmentation model. *Plant Methods*, 19(1), 87

Genze et al. Plant Methods (2023) 19:87
https://doi.org/10.1186/s13007-023-01060-8

Plant Methods

RESEARCH

Open Access

Improved weed segmentation in UAV imagery of sorghum fields with a combined deblurring segmentation model

Nikita Genze[1,2*], Maximilian Wirth[1,2], Christian Schreiner[1], Raymond Ajekwe[4], Michael Grieb[4] and Dominik G. Grimm[1,2,3*]

Abstract

Background Efficient and site-specific weed management is a critical step in many agricultural tasks. Image captures from drones and modern machine learning based computer vision methods can be used to assess weed infestation in agricultural fields more efficiently. However, the image quality of the captures can be affected by several factors, including motion blur. Image captures can be blurred because the drone moves during the image capturing process, e.g. due to wind pressure or camera settings. These influences complicate the annotation of training and test samples and can also lead to reduced predictive power in segmentation and classification tasks.

Results In this study, we propose `DeBlurWeedSeg`, a combined deblurring and segmentation model for weed and crop segmentation in motion blurred images. For this purpose, we first collected a new dataset of matching sharp and naturally blurred image pairs of real sorghum and weed plants from drone images of the same agricultural field. The data was used to train and evaluate the performance of `DeBlurWeedSeg` on both sharp and blurred images of a hold-out test-set. We show that `DeBlurWeedSeg` outperforms a standard segmentation model that does not include an integrated deblurring step, with a relative improvement of 13.4% in terms of the Sørensen-Dice coefficient.

Conclusion Our combined deblurring and segmentation model `DeBlurWeedSeg` is able to accurately segment weeds from sorghum and background, in both sharp as well as motion blurred drone captures. This has high practical implications, as lower error rates in weed and crop segmentation could lead to better weed control, e.g. when using robots for mechanical weed removal.

Keywords Weed detection, Segmentation, Machine learning, Computer vision, Deblurring, UAV

*Correspondence:
Nikita Genze
nikita.genze@hswt.de
Dominik G. Grimm
dominik.grimm@hswt.de
[1] Technical University of Munich, TUM Campus Straubing for Biotechnology and Sustainability, Bioinformatics, Schulgasse 22, 94315 Straubing, Germany
[2] Weihenstephan-Triesdorf University of Applied Sciences, Bioinformatics, Petersgasse 18, 94315 Straubing, Germany
[3] Technical University of Munich, TUM School of Computation, Information and Technology (CIT), Boltzmannstr. 3, 85748 Garching, Germany
[4] Technology and Support Centre in the Centre of Excellence for Renewable Resources (TFZ), Schulgasse 18, 94315 Straubing, Germany

Background

Weed control in agricultural fields is a time-sensitive and critical task. Depending on the quantity and distribution of weeds in agricultural fields, farmers must consider different strategic and economic options, ranging from chemical to mechanical or manual weed control. These decisions depend on several factors, including efficacy, cost, and regulations. The most common way to control weeds is with herbicides, which can have a negative impact on groundwater quality and thus cause public concern [1]. Mechanical weed control is a possible solution to deal with the resulting environmental degradation, although there are challenges such as efficiency, management and erosion effects. Therefore, automated image analysis combined with Unmanned Aerial Vehicle (UAV) imagery is a fast and effective method to reliably detect weeds in agricultural landscapes [2–7]. Automatic weed detection is difficult due to many factors, such as large variations in plant species, occlusions, or changing outdoor conditions. Therefore, modern deep learning based techniques have shown promising results in many agricultural tasks and are replacing conventional methods due to higher accuracy and flexibility [8–11]. In addition, UAVs come with their own difficulties, such as degraded image quality due to challenging illumination conditions on agricultural sites. In addition, the image quality of these images is prone to motion blur, as either the UAV or also possibly the plants may move due to wind pressure during the capture [12]. In addition, the drone and camera settings can affect the influence of motion blur due to the interdependence of flight speed and shutter speed.

Although motion blur is common when using UAVs to capture images in agricultural fields, its effect on the predictive power of weed segmentation models has not been widely studied. In our previous work, we focused on weed segmentation in motion blurred UAV captures [12]. The captures were degraded by different levels of motion blur, and we concluded that it is possible to train deep learning-based segmentation models on motion blurred captures. However, the annotation process of these degraded captures is more difficult and, more importantly, highly time consuming. The weed segmentation model from our recent study [12] consists of a feature extractor and a semantic segmentation architecture to decode the features. Here, residual networks [13] serve as feature extractors because they mitigate the vanishing gradient problem using identity skip connections and allow data to be passed from any layer directly to any subsequent layer. For our semantic segmentation architecture, we used UNet [14], which was originally developed for biomedical tasks. It has also been shown in a variety of non-medical domains that this architecture can achieve sufficient segmentation results even when little training data is available [15–17]. UNet uses skip connections between the encoder and decoder to link information from the encoder and decoder layers.

In general, computer vision tasks such as object detection or segmentation are often degraded by motion blur [18, 19], making motion deblurring an important task in image enhancement. Motion blur is a common form of image distortion. It depends on the magnitude of several overlapping effects [20] and can be described mathematically with respect to different sources of blur. However, most deblurring approaches are based on the simplified blur model [21–24], which is described by the following equation:

$$b = H \cdot s + n. \tag{1}$$

It follows that a blurred image b is the result of a convolution between the underlying sharp image s and the blur kernel H and an addition of noise n. Deblurring tasks can be divided into two categories: non-blind deblurring if the blur kernel is known, and blind deblurring otherwise. In addition, the spatial invariance of the blur kernel is often assumed to produce uniform blur [25]. However, in the agricultural domain, wind-induced shifts are not only possible for the drone (i.e., camera shake), but might also occur at the plant level. Therefore, these real-world images are often degraded by spatially varying and blind blur. Blind blur removal is a highly ill-posed inverse problem, as there are many possible outcomes of a deblurred image [26, 27]. Deep learning methods for image deblurring [28, 29] typically do not explicitly estimate the underlying blur kernel, but by using a Convolutional Neural Network (CNN), relevant image features are extracted and used to directly restore a sharp output image. Therefore, these models are trained on blurry-sharp image pairs. There are several types of deblurring methods already existing in the research community, ranging from encoder-decoder-based models [30–32] to transformer-based models [33–35] to generative models [36–38].

Recently, Chen et al. proposed a computationally efficient encoder-decoder model called NAFNet [39]. The authors used a single-stage UNet architecture with skip connections and concluded that nonlinear activation functions are not necessary for image deblurring. They trained different image restoration models on several different datasets. In particular, they conducted experiments on the REalistic and Diverse Scenes (REDS) dataset [40], where images were not only blurred, but also degraded by compression. Their results surpassed the previous state-of-the-art on several benchmark datasets while using only a fraction of the computational resources.

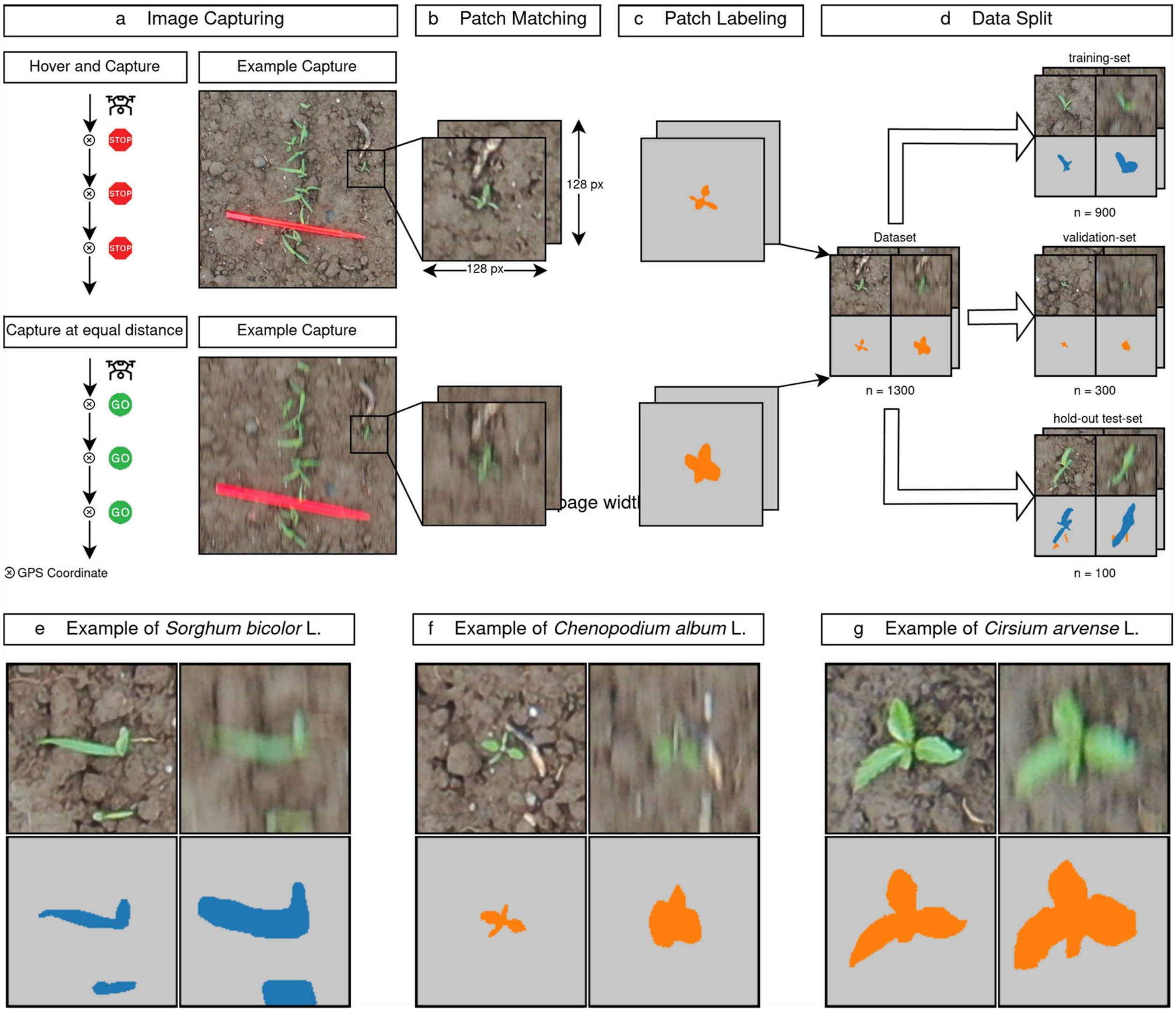

Fig. 1 Overview of the image acquisition and data processing pipeline. **a** Image acquisition using two different drone settings. **b** Matching patches of the same image content. **c** Pixel-wise annotation of the patches. **d** Data splitting into training, validation and a hold-out test-set. **e–g** Example patches of sorghum and the main weeds

In this work, we developed a novel weed segmentation model `DeBlurWeedSeg` to accurately segment weeds from sorghum and background in both sharp as well as motion blurred drone images. Therefore, we first conducted two consecutive UAV flights on the same agricultural field with different flight modes to capture sharp and motion blurred drone images and to enable an in-depth comparison of the effect of motion blur on weed segmentation. For this purpose, we trained a weed segmentation model using the easier to annotate sharp images (`WeedSeg`) and compared the segmentation behaviour of `WeedSeg` with a model combining NAFNet for deblurring with a subsequent segmentation step (`DeBlurWeedSeg`) on both, sharp and blurred images of a hold-out test-set. All data, containing blurry-sharp image pairs, as well as the corresponding expert generated semantic segmentation masks are published in our GitHub repository together with the code and the pre-trained segmentation models: https://github.com/grimmlab/DeBlurWeedSeg. The final model is available at Mendeley Data: https://data.mendeley.com/datasets/k4gvsjv4t3/1.

Materials and methods

In the following, we first outline the image acquisition process, followed by the data preparation and processing pipeline used for this study. The main aspects of the data acquisition are summarized in Fig. 1a–d and explained in detail below. This description is followed by a detailed summary of our deblurring and weed segmentation models, including an overview of the hyperparameter optimization and evaluation metrics.

Image acquisition

The images for this study were taken in an experimental agricultural sorghum field in southern Germany using a consumer-grade "DJI Mavic 2 Pro" drone equipped with a 20 MP Hasselblad camera (L1D-20c). The sorghum (*Sorghum bicolor* L.) crop was sown with a row spacing of 37.5 cm and a density of 25 seeds per m^2. The main weed species observed in this experimental field was *Chenopodium album* L. We also observed *Cirsium arvense* L. (Scop.) in small quantities (examples shown in Fig. 1e–g). We conducted automated drone missions at the end of September 2020, flying at an altitude of five meters with a drone velocity of 6.9 km h^{-1} and an ambient wind current of 6 km h^{-1}. This resulted in a Ground Sampling Distance (GSD) of one millimeter, which is accurate enough to detect sorghum and weeds in early growth stages. Here, the sorghum was at growth stage 13 on the BBCH scale [41].

We used two different UAV settings for the flights, i.e. (i) "Hover and Capture" and (ii) "Capture at Equal Distance", as shown in Fig. 1a. The first setting, "Hover and Capture" was used to stop and stabilize the UAV prior to image capture. This ensured sharp contours of the plants and mitigated the effects of motion blur. In addition, we repeated the flight on the same field and the same flight plan using the UAV's "Capture at Equal Distance" setting. This caused the UAV to capture images at predetermined points without stopping and stabilizing. This resulted in degraded image quality with visible motion blur because the camera shutter was open while the UAV was moving. The images were captured with a shutter speed of $\frac{1}{120}s$, an aperture of approx. 4.0, an ISO of 100 and a manually added exposure bias of -0.3.

Data processing

We first matched image pairs of sharp and motion blurred patches, as shown in Fig. 1b. This was necessary to ensure that each image pair contained the same or similar content, since flying with the "Capture at Equal Distance" setting resulted in images being captured at a slightly different location (difference of about 1 m) relative to the UAV's flight direction due to GPS inaccuracies. In addition, several difficulties in the image capture process were identified, e.g. differences in the flight altitude which resulted in objects of different sizes, or that several plants were connected and appeared as a single plant in the blurred image due to the lower image quality. Therefore, a 128×128 px^2 patch was extracted for each plant instance, with the plant in the center, resulting in 1300 non-overlapping blurry-sharp image pairs as our final dataset. Further dataset statistics are summarized in Additional file 1.

Next, the dataset was manually semantically annotated using the open source software GIMP 2.10,[1] as shown in Fig. 1c. This means that each image pair was separated into the three classes soil/background (gray), sorghum (blue), and weeds (orange).

For hyperparameter optimization and model selection, we sampled a distinct validation set. Therefore we split our dataset into three parts, as shown in Fig. 1d. The hold-out test-set for the final evaluation contains 100 image patches. From the remaining 1200 patches, we selected 25 % for the validation set. We stratified our dataset by the number of plants in each patch, as there was usually one plant instance present per patch. Additionally, we used the type of plants present in the patch (sorghum only, weeds only, both) as a second feature for stratification, since the majority of our dataset (about 70 %) consists of patches where only weeds are visible.

Model selection

We implemented two models for the task of semantic weed segmentation for our comparison, namely `WeedSeg` and `DeBlurWeedSeg`. The `WeedSeg` model follows a classical encoder-decoder based architecture. It consists of an encoder part, where features of images are extracted and encoded into a high-level representation. This representation is of low spatial resolution, and therefore is decoded by a separate model to restore the shape of the input image. As decoder, we chose a UNet-based [14] architecture, similar to our previous work [12]. For the encoder, we use four different residual neural networks, namely ResNet-18, 34, 50, and 101. They were initialized with weights trained on the ImageNet dateset [42] to ensure comparability and faster training convergence.

In the training stage, we evaluated two different scenarios, as shown in Fig. 2a: First, similar to our previous model in [12], we used sharp and motion blurred images to train `WeedSeg` (Scenario 1). Therefore we collected sharp and motion blurred images together with their corresponding semantic ground-truths before training the model. In the second scenario we assumed, that only sharp image patches were available when training `WeedSeg`. This is more realistic, as generating high quality segmentation masks for motion blurred images is time-consuming and error-prone.

[1] https://www.gimp.org.

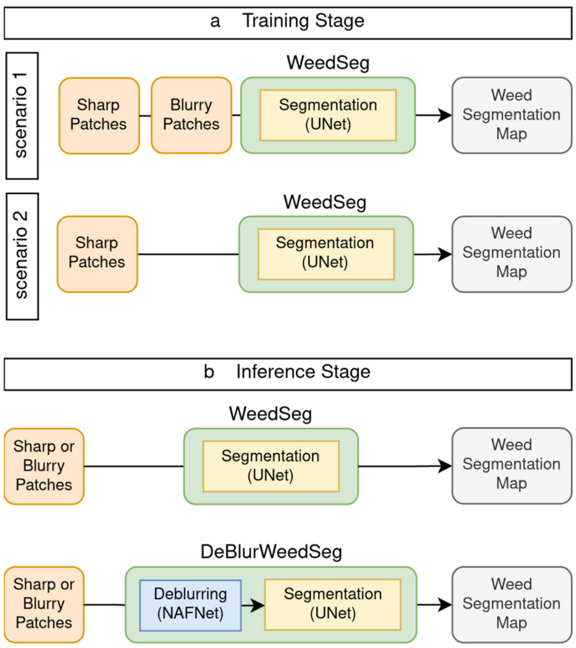

Fig. 2 Our proposed weed segmentation model called `DeBlurWeedSeg`. The model consists of an additional deblurring model prior to weed segmentation model and is able to segment weeds in blurry and sharp images in the inference stage

However, in a real-world scenario it cannot be assumed that the input images in the inference stage are all of the same quality (or distribution) as in the training stage (Fig. 2b). This can lead to a domain shift [43] and decrease prediction performance. In the case of weed segmentation using UAV captures, this shift could be caused by motion blur. Therefore, the classical `WeedSeg` model architecture may not generalize to motion blurred image captures. For this purpose, we propose `DeBlurWeedSeg`, a combined deblurring and segmentation model that can be used to detect weeds in both blurred and sharp image patches in production use. More importantly, the training phase is still performed only on sharp images. This has two advantages over `WeedSeg`: First, it eliminates the effort of semantically annotating motion-blurred UAV imagery, and second, training models on the new dataset is unnecessary, which might be resource intensive. `DeBlurWeedSeg` consists of two modules, a deblurring module based on the computationally efficient deblurring model NAFNet [39] and the segmentation model `WeedSeg` as described above.

Model training and hyperparameter optimization

The performance of a model is highly sensitive to the hyperparameter configuration, especially when limited data are available. Therefore, hyperparameter optimization is a crucial step to select the best model. In this study, we used grid-search to optimize the learning rate and batch size for each encoder. For this purpose, ten different learning rates were selected from a log-uniform distribution starting from 1e−4 to 1e−3. The batch size was optimized starting from 128 up to the maximum possible size, depending on the size of the network architecture and the available GPU memory, in steps of 128. In total, we sampled 160 different hyperparameter sets for each scenario, as summarized in Table 1. Adam [44] was used as the optimizer with different learning rates. In addition, early stopping [45] was used to avoid overfitting.

Table 1 Summary of all hyperparameter sets used for each scenario

Encoder	Possible batch sizes	Num hyperparameter sets
ResNet-18	128, 256, 384, 512, 640, 768	60
ResNet-34	128, 256, 384, 512, 6 40	50
ResNet-50	128, 256, 384	30
ResNet-101	128, 256	20

Evaluation metrics

Our proposed model `DeBlurWeedSeg` consists of two parts, as shown in Fig. 2. In the following, we summarize the metrics used to evaluate the deblurring and segmentation part of our model.

Deblurring metrics

To compare the output of a deblurring model, the quality of an image must be determined. Human evaluation is a reliable but expensive method. Alternatively, several metrics have been proposed for automatic Image Quality Assessment. The main goal is to imitate human perception with these metrics, which is a challenging task. We use the well-known Peak Signal-to-Noise Ratio (PSNR) [46] and Structural SIMilarity index (SSIM) [47] to evaluate the performance of the deblurring model. Although they are not suitable to measure the impact of artifacts that may have been introduced [48], they are still widely used in the literature.

Recent work [49] assessed the perceptual similarity between two images and evaluated different metrics. The authors concluded, that features on which deep learning-based networks are trained for classification tasks can be used to evaluate image quality. In this study, we use the metric called Learned Perceptual Image Patch Similarity (LPIPS) to evaluate our deblurring model.

Table 2 Best results on the validation set for each encoder and training strategy. The best performing combination is shown in bold

Scenario	Encoder name	Best of	Batch size	Step	Learning rate	DS ↑
1	ResNet-18	60	128	4900	$1.43 \cdot 10^{-4}$	0.8732
1	ResNet-34	50	384	4280	$2.37 \cdot 10^{-4}$	0.8655
1	ResNet-50	30	256	3540	$3.97 \cdot 10^{-4}$	0.8982
1	ResNet-101	20	256	4020	$5.40 \cdot 10^{-4}$	0.8995
2	ResNet-18	60	128	2160	$5.40 \cdot 10^{-4}$	0.8862
2	ResNet-34	50	384	1680	$5.11 \cdot 10^{-4}$	0.8837
2	**ResNet-50**	**30**	**128**	**2960**	$\mathbf{5.40 \cdot 10^{-4}}$	**0.9048**
2	ResNet-101	20	256	440	$7.35 \cdot 10^{-4}$	0.9011

Segmentation metrics

In supervised learning tasks, the confusion matrix is often used to evaluate the performance of different models. Considering a binary case, the classes are referred to as positive (P) and negative (N). A test example is defined as a true positive (TP), if it was correctly predicted to be positive. A true negative (TN) example is correctly predicted to be negative. Similarly, an example from the negative class that is misclassified as positive is called a false positive (FP), and a positive example that is misclassified as negative is called a false negative (FN). In the case of a multi-class classification problem, the values are calculated in a one-vs-all fashion. The confusion matrix is defined with a shape of NxN, where the N is the number of classes (three in our case). In addition, a set of quantitative metrics such as Accuracy (AC), Precision (PR), Recall (RE) and F1-Score (F1) can be derived from this matrix.

The weed segmentation task can be defined as classifying each pixel in an image. In our study, the dataset consists of three classes and we observed a high class imbalance especially for the majority class background (> 98% of pixels). This makes these evaluation metrics insufficient due to several reasons: Accuracy is dominated by the majority class. Precision does not provide insight into the number of samples from the FN. Also, Recall does not consider the number of samples from the FP. Additionally, a high F1-Value can be a result from the imbalance between PR and RE. To evaluate the segmentation performance of our models, we used the Sørensen-Dice coefficient [50], also called Dice-Score (DS). This function measures the similarity between two samples and is used in segmentation tasks with high class imbalance [51, 52]. It is mathematically defined as follows:

$$DS = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}. \tag{2}$$

Table 3 Evaluation of NAFNet on the hold-out test-set

	SSIM ↑	PSNR ↑	LPIPS ↓
Blurry	**0.39**	**19.61**	0.45
Deblurred	0.38	18.66	**0.32**

Dataset with the best performance is shown in bold

We selected the best performing hyperparameter set with respect to this metric to train a final model with the combination of the training and validation set.

Hardware and software

All models are implemented in Python 3.8.10 [53] using the packages numpy [54], pandas [55], pytorch [56], scikit-image [57], scikit-learn [58], albumentations [59] and kornia [60]. Our code is publicly available on GitHub.[2] All experiments were conducted under Ubuntu 20.04 LTS on a machine with 104 CPU cores, 756 GB of memory, and four NVIDIA GeForce RTX 3090 GPUs. Each model was trained and evaluated on a single GPU.

Results

In this section, we first give an overview of the training of `WeedSeg` and evaluate the deblurring model. Then, we evaluate the generalization performance of `WeedSeg` and compare it to `DeBlurWeedSeg`. We then analyze the predictions of both models in more detail. Finally, we compare the models qualitatively.

WeedSeg model training and selection

To select the best performing `WeedSeg` model, we first trained ResNet-based feature extractors of different sizes using different hyperparameter sets. We show the best performing hyperparameter set for each feature extractor in Table 2. In summary, the ResNet-50 encoder

[2] https://github.com/grimmlab/DeBlurWeedSeg.

Fig. 3 Qualitative examples of the deblurring step with NAFNet. This model was presented with motion blurred patches. Sharp patches only for reference

performed best with a DS of 0.9048 using a batch size of 128. This model was selected for all further evaluations and comparisons. The results for all hyperparameter sets and their evaluations can be found in Additional file 2. The training curve on the validation set is shown in Additional file 3 .

Deblurring evaluation

Next, the deblurring network (NAFNet) was evaluated on the hold-out test-set. The sharp image patches were considered as a reference. Here we can see, that SSIM and PSNR decreased slightly, as shown in Table 3. One possible reason could be, that these metrics do not correlate with human perception. Nevertheless, LPIPS was significantly reduced, indicating good deblurring performance.

In addition, during a qualitative assessment of the deblurring step we observed a significant improvement in perceived sharpness, as shown in Fig. 3. Also, the deblurred patches showed less camera noise.

In some rare cases, the deblurring step failed for tiny weeds, making them indistinguishable (see Additional file 4). However, these patches were not critical to the

Table 4 DS for different scenarios and datasets. The sum of the sharp and motion blurred dataset is denoted in the column "Combined"

	Sharp	Motion blurred	Combined
WeedSeg (scenario 1)	0.8966	0.5864	0.7327
WeedSeg (scenario 2)	**0.9055**	0.5881	0.7381
DeBlurWeedSeg	0.8741	**0.8011**	**0.8373**

The best performing scenario is shown in bold

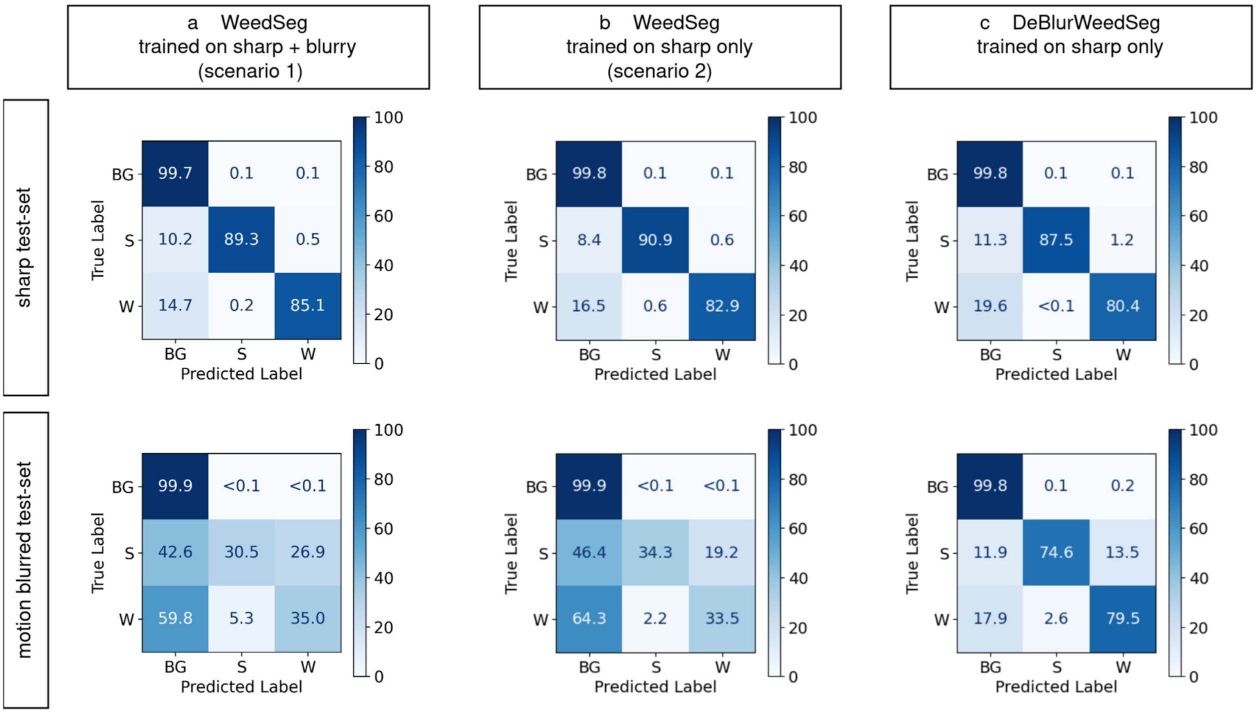

Fig. 4 Pixel-based classification results shown in a normalized confusion matrix in percent. Background is denoted by BG, sorghum by S, and weeds by W. **a + b** The performance of `WeedSeg` is degraded by motion blurred image patches. This is true for training a segmentation model on sharp and blurry image patches (scenario 1) or on sharp parches only (scenario 2) **c** `DeBlurWeedSeg` has a significantly better performance on motion blurred image patches and is only slightly worse on sharp image patches

average weed segmentation performance due to the tiny size of the plants.

Generalization performance of `WeedSeg` and `DeBlurWeedSeg`

Next, we estimated the generalization performance of `WeedSeg` on a hold-out test-set. This set contains both blurred and sharp image patches, as shown in Table 4.

There were only little differences on our hold-out test-set based on the DS when trained on motion blurred and sharp images (scenario 1) or on sharp images only (scenario 2), as shown in Table 4. This is similar to the results on the validation-set (compare Table 2). Both models were able to segment sharp image patches with a high DS, but failed to segment image patches with motion blur. Therefore, we focus on scenario 2 in further analysis, as no motion blurred images and segmentation masks are needed during the training process. Our model `DeBlurWeedSeg` has a high DS on sharp and motion blurred images, resulting in a relative improvement of 13.4 % for the combined dataset. This is not surprising, since `DeBlurWeedSeg` contains a prior deblurring step and is thus able to sharpen blurred images before segmentation. Therefore, `DeBlurWeedSeg` is able to better generalize to new images with unknown drone settings.

Furthermore, we provide a more detailed analysis of the hold-out test-set by showing a normalized confusion matrix of the accuracy calculated on a pixel basis (see Fig. 4). Here, the pixel-wise ground-truths and predictions were compared for each class, i.e., sorghum, weed and background.

We can see that the background class was predicted similarly well by all models, as indicated by an accuracy of more than 99.8 %. However, we clearly see a severe difference for sorghum and weed. We analyzed the segmentation capabilities for both sharp and motion blurred test images independently, as shown in Fig. 4. Here we can see that the performance of `WeedSeg` is highly accurate for sharp images. However, the performance drops severely for motion blurred images. This is to be expected in scenario2, since `WeedSeg` is trained only on sharp images. In particular, `DeBlurWeedSeg` performs well for both individual classes due to the prior deblurring step. We see a significant relative improvement in segmentation accuracy for blurred images of $\sim 117\%$ for the class sorghum and $\sim 137\%$ for the weed class. On sharp images, however, we observe a slight decrease in relative performance using `DeBlurWeedSeg`, i.e. 3.7% for sorghum and 3.1% for weed.

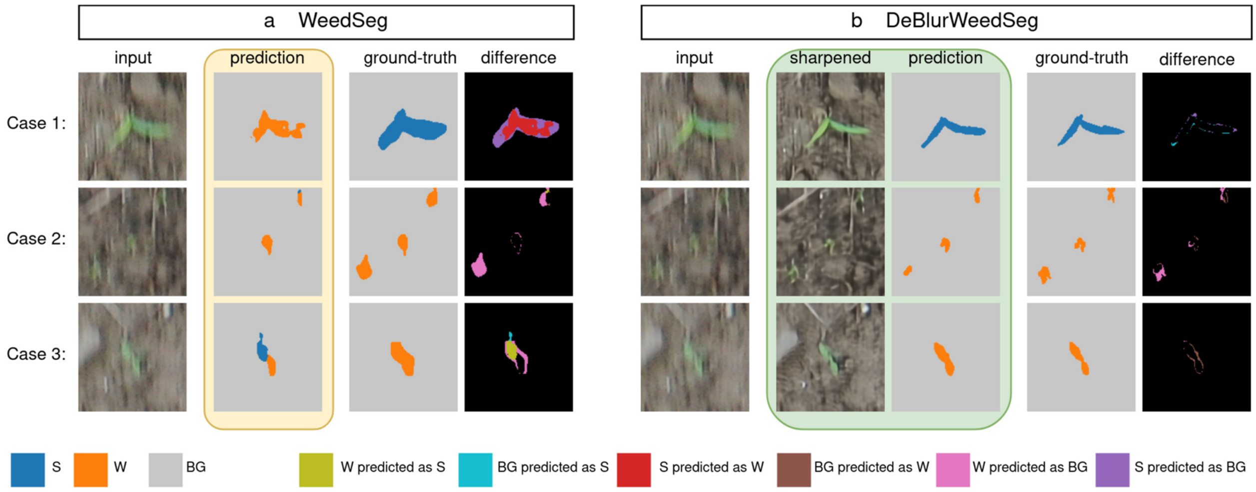

Fig. 5 Examples of the blurred test-set. Background is gray, sorghum is blue, and weeds are orange. **a** Predictions from `WeedSeg`. **b** Predictions from `DeBlurWeedSeg` including the sharpened image patch

Qualitative segmentation results

Finally, we analyze some example images from the test-set and their predictions in more detail. For this purpose, we generated segmentation difference maps of the prediction and the ground-truth and summarized them in Fig. 5.

For this analysis, we focus only on cases where `WeedSeg` predicted the motion blurred patches worse than the sharp counterpart. As shown in Fig. 5a, the failure cases of `WeedSeg` can be summarized in three cases: First, blurry sorghum plants were predicted as weeds (see Case 1). Second, small weeds could not be detected and were predicted as background (see Case 2). And third, parts of weed plants were misclassified as sorghum (see Case 3). In addition, we show the difference maps between ground-truth and prediction to highlight the areas of misclassification. All of these cases were successfully corrected by `DeBlurWeedSeg`, as shown in Fig. 5b. The remaining errors can be attributed to inaccuracies at the plant boundaries and tiny errors due to incorrect predictions of the entire plant instance, as shown in the difference map.

Discussion

In this work, we trained a semantic segmentation model called `WeedSeg`, based on a UNet-shaped architecture with residual networks as encoders, to segment weeds from sorghum and background using only sharp training images. Selecting sharp images for training has the advantage that the annotation process is less time-consuming and error-prone compared to blurred and degraded images. Training and evaluating different segmentation models directly on blurred images has been studied extensively by Genze et al. in [12]. In the current study study, we aimed to investigate the generalization abilities of models trained under idealized conditions and then deployed in productive environments. We observed a significant drop in performance when applying `WeedSeg` to naturally motion blurred images, i.e. motion blurred images due to non-ideal flight settings. Also, training a model on sharp and motion blurred image patches (scenario 1) yielded inferior results. We identified motion blur as a major bottleneck for semantic weed segmentation. Therefore, we generated a dataset containing matching blurry-sharp image pairs of sorghum and weed plants and their corresponding semantic ground-truths.

In this study, we proposed a combined deblurring and semantic segmentation model `DeBlurWeedSeg` that is able to segment sorghum and weeds from the background in sharp and motion blurred images. Here, NAFNet [39], a computationally efficient deblurring model is used as a prior step to produce a sharpened version from the blurred input images, which is then segmented by our weed segmentation model. Our proposed model achieves significantly better performance with motion blurred and sharp image patches. Nevertheless, `DeBlurWeedSeg` still misclassified 13.5% of the sorghum pixels as weeds in motion blurred patches (see Fig. 4c), indicating that there is room for improvement in classifying the correct plant species. These errors could be resolved by training the weed segmentation model with additional sorghum images, as our dataset contains more images of weeds. Also, there was a slight drop in performance when using `DeBlurWeedSeg` on sharp image patches. This might be an indication, that the segmentation model is slightly dependent on low-level noise that is present in the sharp image patches (i.e. ISO noise) and is subject of a future study.

Although DeBlurWeedSeg performed well on our hold-out test-set, there are a number of factors that were not evaluated in this study. First, our dataset was generated from a single UAV mission over a specific agricultural field and the sorghum plants were at a low growth stage of BBCH 13. However, the weed flora may be different in other regions and for different growth stages of sorghum. Second, different weather conditions and sampling times could affect the illumination of the images and thus the segmentation performance. Here, we focused on one UAV flight where *Chenopodium album* L. was the main weed present in the field. As future work, we would like to evaluate this method on a variety of growth stages and weed species.

This research could also be integrated into agricultural robots to deal with motion blur on the fly, which is the subject of another study.

Conclusion

Accurate detection and segmentation of weeds in the early growth stages of sorghum is critical for effective weed management. However, UAVs are prone to motion blur, which is a major problem for real-world applications and use of deep-learning based weed segmentation models. In this study, we propose a combined deblurring and weed segmentation model DeBlurWeedSeg. We demonstrate that we can efficiently mitigate the performance loss that was caused by motion blur. In addition, this method could be used to segment already sharp image patches without a substantial drop in performance. Finally, this model could lead to better weed control due to lower error rates in weed detection and help to enforce agricultural robots in combination with mechanical weed control.

Abbreviations

CNN	Convolutional neural network
GSD	Ground sampling distance
UAV	Unmanned aerial vehicle
PSNR	Peak signal to noise ratio
SSIM	Structural Similarity Index
LPIPS	Learned perceptual image patch similarity

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13007-023-01060-8.

Additional file 1. Dataset Statistics.

Additional file 2. Results of hyperparameter optimisation.

Additional file 3. Training Curve of the best performing hyperparameter set.

Additional file 4. Insufficient cases of the deblurring step.

Acknowledgements

This article is funded by the Open Access Publication Fund of Weihenstephan-Triesdorf University of Applied Sciences.

Author contributions

DGG and NG conceived and designed the study. RA conducted the drone flights and the image acquisition. CS and MW labeled the dataset with supervision of NG. RA and MG guided the labeling process with domain expertise. NG implemented the machine-learning pipeline and conducted all computational experiments. NG and DGG analyzed the results. NG and DGG wrote the manuscript with contributions from all authors.

Funding

Open Access funding enabled and organized by Projekt DEAL. Funding for the research presented in this paper is provided by the Bavarian State Ministry for Food, Agriculture and Forests within the EWIS project (Funding ID: G2/N/19/13).

Availability of data and materials

All data, annotations and code is publicly available on GitHub under https://github.com/grimmlab/DeBlurWeedSeg and Mendeley Data under https://data.mendeley.com/datasets/k4gvsjv4t3/1.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 27 February 2023 Accepted: 22 July 2023
Published online: 22 August 2023

References

1. Kudsk P, Streibig JC. Herbicides–a two-edged sword*. Weed Res. 2003;43(2):90–102. https://doi.org/10.1046/j.1365-3180.2003.00328.x.
2. Kim J, Kim S, Ju C, Son HI. Unmanned aerial vehicles in agriculture: a review of perspective of platform, control, and applications. IEEE Access. 2019;7:105100–15. https://doi.org/10.1109/ACCESS.2019.2932119.
3. Lottes P, Hörferlin M, Sander S, Stachniss C. Effective vision-based classification for separating sugar beets and weeds for precision farming. J Field Robot. 2017;34(6):1160–78.
4. Lottes P, Khanna R, Pfeifer J, Siegwart R, Stachniss C. Uav-based crop and weed classification for smart farming. In: 2017 IEEE International Conference on Robotics and Automation (ICRA), 2017; pp. 3024–3031. IEEE.
5. Lottes P, Behley J, Milioto A, Stachniss C. Fully convolutional networks with sequential information for robust crop and weed detection in precision farming. IEEE Robot Automation Lett. 2018;3(4):2870–7.
6. Sa I, Chen Z, Popović M, Khanna R, Liebisch F, Nieto J, Siegwart R. weednet: Dense semantic weed classification using multispectral images and mav for smart farming. IEEE Robotics and Automation Letters. 2017;3(1):588–95.
7. Sa I, Popović M, Khanna R, Chen Z, Lottes P, Liebisch F, Nieto J, Stachniss C, Walter A, Siegwart R. Weedmap: a large-scale semantic weed mapping framework using aerial multispectral imaging and deep neural network for precision farming. Remote Sens. 2018;10(9):1423.
8. Genze N, Bharti R, Grieb M, Schultheiss SJ, Grimm DG. Accurate machine learning-based germination detection, prediction and quality assessment of three grain crops. Plant Methods. 2020;16(1):1–11.

9. Wu Z, Chen Y, Zhao B, Kang X-B, Ding Y. Review of weed detection methods based on computer vision. Sensors (Basel, Switzerland). 2021;21.
10. Veeragandham S, Santhi H. A detailed review on challenges and imperatives of various cnn algorithms in weed detection. In: 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), 2021; pp. 1068–1073. https://doi.org/10.1109/ICAIS50930.2021.9395986
11. Zhang Y, Wang M, Zhao D, Liu C, Liu Z. Early weed identification based on deep learning: a review. Smart Agric Technol. 2023;3: 100123. https://doi.org/10.1016/j.atech.2022.100123.
12. Genze N, Ajekwe R, Güreli Z, Haselbeck F, Grieb M, Grimm DG. Deep learning-based early weed segmentation using motion blurred uav images of sorghum fields. Comput Electron Agric. 2022;202: 107388.
13. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2015;1512:03385.
14. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-assisted Intervention, Springer; 2015, pp. 234–241.
15. Iglovikov V, Shvets A. Ternausnet: U-net with VGG11 encoder pre-trained on imagenet for image segmentation. CoRR abs/1801.05746; 2018.
16. Boyina L, Sandhya G, Vasavi S, Koneru L, Koushik V. Weed detection in broad leaves using invariant u-net model. In: 2021 International Conference on Communication, Control and Information Sciences (ICCISc), 2021; 1:1–4. https://doi.org/10.1109/ICCISc52257.2021.9485001
17. Siam M, Gamal M, Abdel-Razek M, Yogamani S, Jagersand M, Zhang H. A comparative study of real-time semantic segmentation for autonomous driving. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. 2018.
18. Guo Q, Juefei-Xu F, Xie X, Ma L, Wang J, Yu B, Feng W, Liu Y. Watch out! motion is blurring the vision of your deep neural networks. Adv Neural Inf Process Syst. 2020;33:975–85.
19. Sayed M, Brostow G. Improved handling of motion blur in online object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021; pp. 1706–1716.
20. Potmesil M, Chakravarty I. Modeling motion blur in computer-generated images. SIGGRAPH Comput Graph. 1983;17(3):389–99. https://doi.org/10.1145/964967.801169.
21. Whyte O, Sivic J, Zisserman A, Ponce J. Non-uniform deblurring for shaken images. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2010; pp. 491–498. https://doi.org/10.1109/CVPR.2010.5540175
22. Gupta A, Joshi N, Zitnick C.L, Cohen M, Curless B. Single image deblurring using motion density functions. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 6311 LNCS. 2010; pp. 171–184.
23. Harmeling S, Michael H, Schölkopf B. Space-variant single-image blind deconvolution for removing camera shake. Adv Neural Informat Process Syst. 2010; 23
24. Hirsch M, Schuler C.J, Harmeling S, Schölkopf B. Fast removal of non-uniform camera shake. In: Proceedings of the IEEE International Conference on Computer Vision. 2011; pp. 463–470. https://doi.org/10.1109/ICCV.2011.6126276
25. Cho S, Matsushita Y, Lee S. Removing non-uniform motion blur from images. In: 2007 IEEE 11th International Conference on Computer Vision. 2007; pp. 1–8. https://doi.org/10.1109/ICCV.2007.4408904.
26. Xu R, Xiao Z, Huang J, Zhang Y, Xiong Z. Edpn: enhanced deep pyramid network for blurry image restoration. CVPR. 2021.
27. Liu S, Qiao P, Dou Y. Multi-Outputs Is All You Need For Deblur. arXiv; 2022. https://doi.org/10.48550/ARXIV.2208.13029. https://arxiv.org/abs/2208.13029.
28. Lai W-S, Huang J-B, Hu Z, Ahuja N, Yang M-H. A comparative study for single image blind deblurring. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016; pp. 1701–1709.
29. Su J, Xu B, Yin H. A survey of deep learning approaches to image restoration. Neurocomputing. 2022;487:46–65.
30. Zhang H, Dai Y, Li H, Koniusz P. Deep stacked hierarchical multi-patch network for image deblurring. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2019.
31. Ji S-W, Lee J, Kim S-W, Hong JP, Baek S-J, Jung S-W, Ko S-J. Xydeblur: Divide and conquer for single image deblurring. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022; pp. 17400–17409.
32. Ye M, Lyu D, Chen G. Scale-iterative upscaling network for image deblurring. IEEE Access. 2020;8:18316–25.
33. Wang Z, Cun X, Jianmin Zhou BW, Liu J, Li H. Uformer: a general u-shaped transformer for image restoration. CVPR. 2022.
34. Tsai F-J, Peng Y-T, Lin Y-Y, Tsai C-C, Lin C-W. Stripformer: Strip transformer for fast image deblurring.
35. Zamir SW, Arora A, Khan S, Hayat M, Khan FS, Yang M-H. Restormer: efficient transformer for high-resolution image restoration. CVPR. 2022.
36. Kupyn O, Martyniuk T, Wu J, Wang Z. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. ICCV 2019-October, 8877–8886; 2019. https://doi.org/10.1109/ICCV.2019.00897.
37. Zhang K, Luo W, Zhong Y, Ma L, Stenger B, Liu W, Li H. Deblurring by realistic blurring. 2020.
38. Hexin X, Li Z, Yan J. Motion blur image restoration by multi-scale residual neural network. Int J Adv Netw Monit Controls 2021;6:57–67. https://doi.org/10.21307/IJANMC-2021-009
39. Chen L, Chu X, Zhang X, Sun J. Simple baselines for image restoration. arXiv preprint arXiv:2204.04676. 2022.
40. Nah S, Baik S, Hong S, Moon G, Son S, Timofte R, Lee KM. Ntire 2019 challenge on video deblurring and super-resolution: Dataset and study. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2019; pp. 1996–2005.
41. Hess M, Barralis G, Bleiholder H, Buhr L, Eggers T, Hack H, Stauss R. Use of the extended bbch scale-general for the descriptions of the growth stages of mono; and dicotyledonous weed species. Weed Res. 1997;37(6):433–41.
42. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. Imagenet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. 2009; pp. 248–255. https://doi.org/10.1109/CVPR.2009.5206848.
43. Zhou K, Liu Z, Qiao Y, Xiang T, Loy C.C. Domain generalization: a survey. In: IEEE Transactions on Pattern Analysis and Machine Intelligence. 2022;1–20. https://doi.org/10.1109/TPAMI.2022.3195549
44. Kingma DP, Ba JL. Adam: A method for stochastic optimization. In: 3rd International Conference on Learning Representations, ICLR 2015—Conference Track Proceedings. 2014. https://doi.org/10.48550/arxiv.1412.6980
45. Prechelt L. Early stopping-but when? In: Neural Networks: Tricks of the trade. 2002;55–69.
46. Teo PC, Heeger DJ. Perceptual image distortion. In: Proceedings—International Conference on Image Processing, ICIP. 1994;2:982–6. https://doi.org/10.1109/ICIP.1994.413502.
47. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process. 2004;13:600–12. https://doi.org/10.1109/TIP.2003.819861.
48. Liu Y, Wang J, Cho S, Finkelstein A, Rusinkiewicz S. A no-reference metric for evaluating the quality of motion deblurring. ACM Trans Graphics (TOG). 2013;32. https://doi.org/10.1145/2508363.2508391
49. Zhang R, Isola P, Efros AA, Shechtman E, Wang O. The unreasonable effectiveness of deep features as a perceptual metric. 2018.
50. Bertels J, Eelbode T, Berman M, Vandermeulen D, Maes F, Bisschops R, Blaschko MB. Optimizing the dice score and jaccard index for medical image segmentation: Theory and practice. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 11765 LNCS, 2019;92–100. https://doi.org/10.1007/978-3-030-32245-8_11
51. Yao AD, Cheng DL, Pan I, Kitamura F. Deep learning in neuroradiology: a systematic review of current algorithms and approaches for the new wave of imaging technology. Radiol Artif Intell. 2020;2(2):190026.https://doi.org/10.1148/ryai.2020190026 .
52. Muhammad K, Hussain T, Ullah H, Ser JD, Rezaei M, Kumar N, Hijji M, Bellavista P, de Albuquerque VHC. Vision-based semantic segmentation in scene understanding for autonomous driving: recent achievements, challenges, and outlooks. IEEE Trans Intell Transp Syst. 2022;23(12):22694–715. https://doi.org/10.1109/TITS.2022.3207665.
53. Van Rossum G, Drake FL. Python 3 Reference Manual. Scotts Valley, CA: CreateSpace; 2009.
54. Harris RC, Millman KJ, van der Walt JS, et al. Array programming with NumPy. Nature. 2020;585(7825):357–62. https://doi.org/10.1038/s41586-020-2649-2.

55. Wes McKinney: data structures for statistical computing in Python. In: Stéfan van der Walt, Jarrod Millman, editors Proceedings of the 9th Python in Science Conference. 2010; 56–61. https://doi.org/10.25080/Majora-92bf1922-00a.
56. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Kopf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J, Chintala S. Pytorch: an imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems 32: 8024–8035. Curran Associates, Inc., 2019. http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.
57. Van der Walt S, Schönberger JL, Nunez-Iglesias J, Boulogne F, Warner JD, Yager N, Gouillart E, Yu T. scikit-image: image processing in python. PeerJ. 2014;2:453.
58. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825–30.
59. Buslaev A, Iglovikov VI, Khvedchenya E, et al. Albumentations: fast and flexible image augmentations. ArXiv e-prints; 2018. arXiv.1809.06839.
60. Riba E, Mishkin D, Ponsa D. Kornia: an open source differentiable computer vision library for pytorch. In: Winter conference on applications of computer vision. 2020. https://arxiv.org/pdf/1910.02190.pdf.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

C Publication C

The following pages show the full version of the published manuscript:

Publication C

Nikita Genze, Wouter K. Vahl, Jennifer Groth, Maximilian Wirth, Michael Grieb & Dominik G. Grimm (2024). Manually annotated and curated Dataset of diverse Weed Species in Maize and Sorghum for Computer Vision. *Scientific Data*, 11(1), 109

scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Manually annotated and curated Dataset of diverse Weed Species in Maize and Sorghum for Computer Vision

Nikita Genze[1,2], Wouter K. Vahl[3], Jennifer Groth[3], Maximilian Wirth[1,2], Michael Grieb[4] & Dominik G. Grimm[1,2,5] ✉

Sustainable weed management strategies are critical to feeding the world's population while preserving ecosystems and biodiversity. Therefore, site-specific weed control strategies based on automation are needed to reduce the additional time and effort required for weeding. Machine vision-based methods appear to be a promising approach for weed detection, but require high quality data on the species in a specific agricultural area. Here we present a dataset, the Moving Fields Weed Dataset (MFWD), which captures the growth of 28 weed species commonly found in sorghum and maize fields in Germany. A total of 94,321 images were acquired in a fully automated, high-throughput phenotyping facility to track over 5,000 individual plants at high spatial and temporal resolution. A rich set of manually curated ground truth information is also provided, which can be used not only for plant species classification, object detection and instance segmentation tasks, but also for multiple object tracking.

Background & Summary

Weeds are plants that, although not specifically cultivated, are adapted to grow on arable land. Typically, weeds are considered to be an undesirable element in crop production. Their negative impact on crop development can be described in terms of competition with the crop for resources (nutrients, sunlight, space and water), reduction in productivity, increased challenges during harvesting and an overall increase in the cost of agricultural production. In addition, weeds can be hosts for insects and diseases[1,2], which might further increase the necessity for control strategies. Nevertheless, weeds might also have positive effects on biodiversity[3] and soil structure[4]. Therefore, only highly competitive and invasive weed species should be removed which might lead to more sustainable agriculture[5].

Over centuries, several crop management strategies were established to mitigate the negative impact of weeds, which can be divided into five main categories[6]: 'preventative' (preventing weeds from establishing), 'cultural' (maintaining field hygiene with low weed seed bank), 'mechanical' (removing weeds by mowing, mulching or tilling), 'biological' (using natural enemies such as insects or animals), and 'chemical' (applying herbicides). Disadvantages of these approaches include financial burden, additional time and effort to varying degrees. In addition, control treatments may impact the health of people, plants, soil, animals, and the environment[7–9].

Sustainable strategies for managing weeds are critical to feed the world's population while conserving the ecosystems and biodiversity[9]. The limited and rational use of herbicides is an important principle of sustainable farming, as spraying of herbicides leads to waste and can pollute soil and water sources. Furthermore, agrochemical residues are one of the most important food-related concerns. Therefore, additional non-chemical and site-specific weed management (SSWM) strategies[10] are needed, which should be linked with the farm

[1]Technical University of Munich, TUM Campus Straubing for Biotechnology and Sustainability, Bioinformatics, Schulgasse 22, 94315, Straubing, Germany. [2]Weihenstephan-Triesdorf University of Applied Sciences, Bioinformatics, Petersgasse 18, 94315, Straubing, Germany. [3]Institute for Crop Science and Plant Breeding, Bavarian State Research Center for Agriculture, Am Gereuth 6, 85354, Freising, Germany. [4]Technology and Support Centre in the Centre of Excellence for Renewable Resources (TFZ), Schulgasse 18, 94315, Straubing, Germany. [5]Technical University of Munich, TUM School of Computation, Information and Technology (CIT), Boltzmannstr. 3, 85748, Garching, Germany. ✉e-mail: dominik.grimm@hswt.de

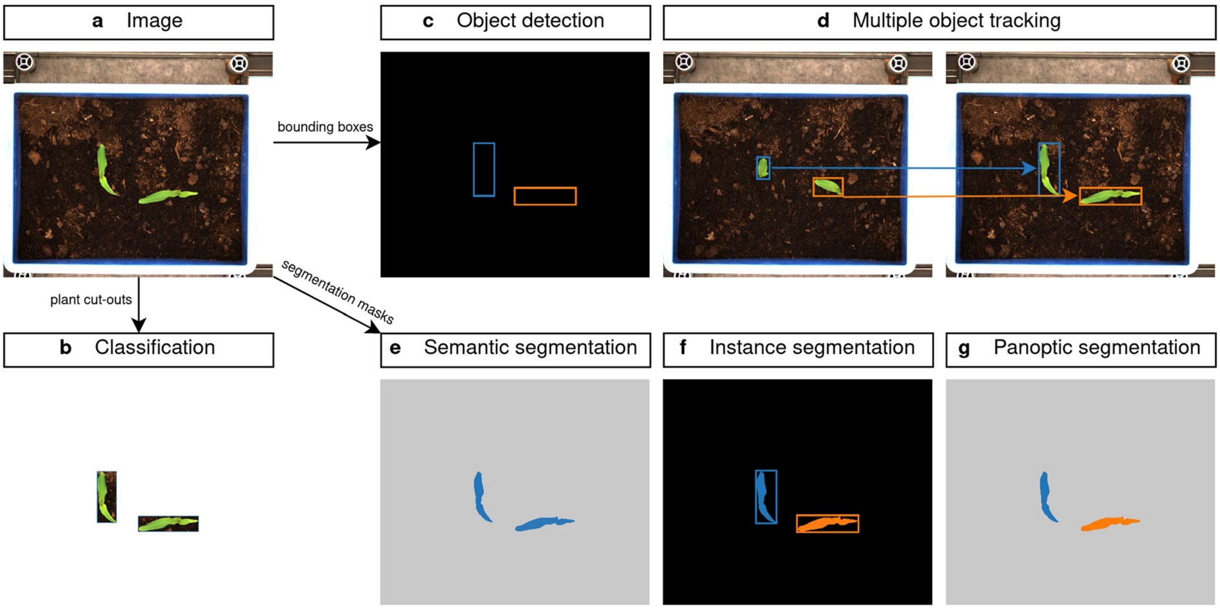

Fig. 1 Comparison of different computer vision tasks used in plant phenotyping. Plants (**a**) can be classified (**b**) as individual cut-outs, detected by bounding boxes (**c**), tracked through time (**d**), or segmented containing pixel-wise information. This can be done either semantically (**e**), by instance (**f**) or by the combination of both called panoptic segmentation (**g**).

management system. One key aspect is to automatically and precisely detect weeds to mitigate the additional time and effort for either site-specific or weed-specific herbicide application or mechanical weed control. Numerous studies demonstrated methods to automatically detect weeds on the field or in greenhouses, where computer vision-based methods seem the most promising[11–19]. These methods can be grouped into different tasks with varying ground truth information, as shown in Fig. 1. Starting from an image as shown in Fig. 1a, image classification is the least accurate task (Fig. 1b). It is applied to a single plant cut-out without location and thus cannot be applied in SSWM tasks, where the simultaneous detection and localization of multiple plants is desired. Therefore, the object detection[20] task might be utilized. However, this task detects rectangular bounding boxes (Fig. 1c) which is not satisfactory due to the complex and irregular shape of the plants. Also, these models are prone to occlusion[21] which diminishes their performance in areas with high weed infestation[22]. Nevertheless, the analysis of a plant's growth dynamics can be achieved, where multiple objects are tracked through time (see Fig. 1d). Moreover, by using segmentation masks, additional tasks can be performed. Here, by convention any countable entity (i.e. plant, person, etc.) is named '*thing*' and any uncountable region (i.e. soil, sky) is called '*stuff*': The semantic segmentation task[23] provides a precise delineation of *stuff*, which separates every pixel in the image by class label, but cannot separate different plants of the same class[24] (see Fig. 1e). This is crucial in selective weed management[25] to conserve biodiversity by removing only competitive weeds. Consequently, instance segmentation[24] (Fig. 1f) can generate accurate detections of *things* individually, which can be used in many downstream tasks such as weed density assessment or biomass estimation[26]. Nevertheless, this type of model is only able to detect countable objects (*things*) and does not consider *stuff* regions. Finally, panoptic segmentation[27] combines the concept of both semantic and instance segmentation and assigns two labels (semantic label and instance id) to each pixel in an image (Fig. 1g).

The basis for the development, validation and assessment of such systems is the availability of high quality data on weed diversity in a particular area of interest. Several datasets are publicly available, but lack several aspects for precise plant phenotyping, as summarized in Table 1.

Most datasets available lack variability in the data (low number of individuals or plant species) limiting their usability in different studies. Only a few datasets are larger than 100,000 annotated plant samples, including Open Plant Phenotype Database (OPPD)[28] and Pl@ntNet-300k[29]. However, Pl@ntNet-300k can only be used for classification tasks without tracking plant growth stages and OPPD is missing semantic and instance segmentation masks, which are important for precise phenotyping. Also, the bounding box information of a plant is often not sufficient, as it is too coarse for most weed management applications. Therefore, semantic segmentation or even more accurate instance segmentation masks are required. Finally, tracking a plant over time provides valuable insight into growth dynamics.

In this work we have created a high-quality dataset of different plant species with a high temporal and spatial resolution. We added manually curated semantic and instance segmentation masks of a subset to make this dataset suitable for weed management tasks. For this purpose, we used a high throughput phenotyping system to ensure a high degree of automation, as this system was equipped with controlled illumination and an automatic irrigation system. In our dataset, we included images of plants captured multiple times per day. This included captures in the evening, when the appearance of some species changes due to their dependency on sunlight. In

Dataset name	Labeled samples	Classification	Object detection	Semantic segmentation	Instance segmentation	Multiple object tracking (growth)
Open Plant Phenotype Database (OPPD)[28]	315,038	✓	✓			✓
Pl@ntNet-300k[29]	306,146	✓				
MFWD (ours)[30]	**200,148**	✓	✓	✓	✓	✓
Plant Village Dataset[41]	54,303	✓				
Deep Seedling[42]	33,444	✓	✓			
DeepWeeds[43]	17,509	✓				
Weed25[44]	14,023	✓	✓			
WE3DS[45]	11,544	✓	✓	✓		
Sudars *et al.*[46]	7,853	✓	✓			
Plant Seedling[47]	5,539	✓	✓			✓
Champ *et al.*[24]	2,489	✓	✓	✓	✓	
Ladybird Cobbitty 2017 Brassica Dataset[48]	2,245	✓	✓	✓		
CWFID[49]	494	✓	✓			
Leaf Segmentation and Counting Challenge[50]	284		✓	✓	✓	✓

Table 1. Comparison of ground-based image datasets for plant detection.

addition, we generated data from different varieties of sorghum and maize, focusing on a wide range of seedling weeds that are also common in agricultural sites where these crops are grown.

Methods

The methodology can be summarized into three steps, as shown in Fig. 2. First, we will describe the experimental setup (Fig. 2a). Second, we will illustrate the image generation (Fig. 2b) and conclude with the labeling process (Fig. 2c).

Experimental Setup. To generate a dataset consisting of high-quality images that capture the initial growth dynamics of individual plants of several weed species, a greenhouse experiment was performed at the Moving Fields facility (https://www.lfl.bayern.de/verschiedenes/ueberuns/272457/index.php) of the Bavarian State Research Center for Agriculture in Freising, Germany. In the experiment, plants were grown in micro-plots that were watered and photographed automatically on at least a daily basis from the day of sowing until harvest, which took place at around shooting. Built into a greenhouse, the Moving Fields facility (LemnaTec GmbH) consists of a conveyor belt system, three irrigation stations and four 'Scanalyzer 3d' photo cabins, which together enable experimental units consisting of plants growing on micro-plots to be automatically moved, watered, weighted and photographed. The greenhouse can be climatically controlled with regard to humidity and temperature and can be illuminated by 48 sodium-vapor lamps (Philips Son-T AGRO). The conveyor belt system (Bosch Rexroth TS2plus) accommodates and enables the movement of 390 carriers (micro-plots). At the three measuring stations, digital scales (Bizerba ST) and high-pressure pumps (Wartson-Marlow) enable carriers, together with any plants transported by them, to be weighed and to be watered to a unit-specific target weight. The plant species included in the dataset, listed in Table 2 and Table 3, were selected as weed species common to fields of sorghum grown in Germany. Additional selection criteria were 1) commercial availability and 2) the ability to be grown at the climatically controlled conditions of a greenhouse. Seeds of the species involved were acquired from commercial breeders in Germany, the Netherlands and France. All plants were grown in boxes of size $40 \times 30 \times 22$ cm (outer dimensions). The color of these boxes was blue, to facilitate image analysis afterwards. Each box was filled to about half height (roughly 11 cm after compression) with a commercial peat-free substrate (Höfter GmbH), primarily consisting of coconut fibers. Plants were grown as monocultures; each box contained plants of one species only. To yield data for enough individual plants per species, the number of boxes varied between species due to different germination rates.

The number of seeds planted in each box was made dependent on the expected germination rate, which was adjusted throughout the experiment. Thus, seed density varied both within species over time and between species. Following breeder recommendations, some seeds were kept in a vernalization room (at 4 °C) or treated with gibberellin acid (GA_3) to ensure germination success. Units were sent to an automatic watering station as often as frequent imaging allowed, in practice at least twice a day. Each unit was watered to its unique target weight. This target weight initially corresponded to the unit's weight at sowing. Throughout the experiment, target weights were adjusted repeatedly to prevent boxes from becoming either too dry or too wet. Twice a week, all boxes were examined to score seedling emergence, to thin the standing stock in order to minimize overlap between individual plants and to harvest plants that either started shooting or that became too big for the box they were growing in.

In addition, different varieties of maize (*Zea Mays*) and sorghum (*Sorghum bicolor*) were grown and captured, as shown in Table 3.

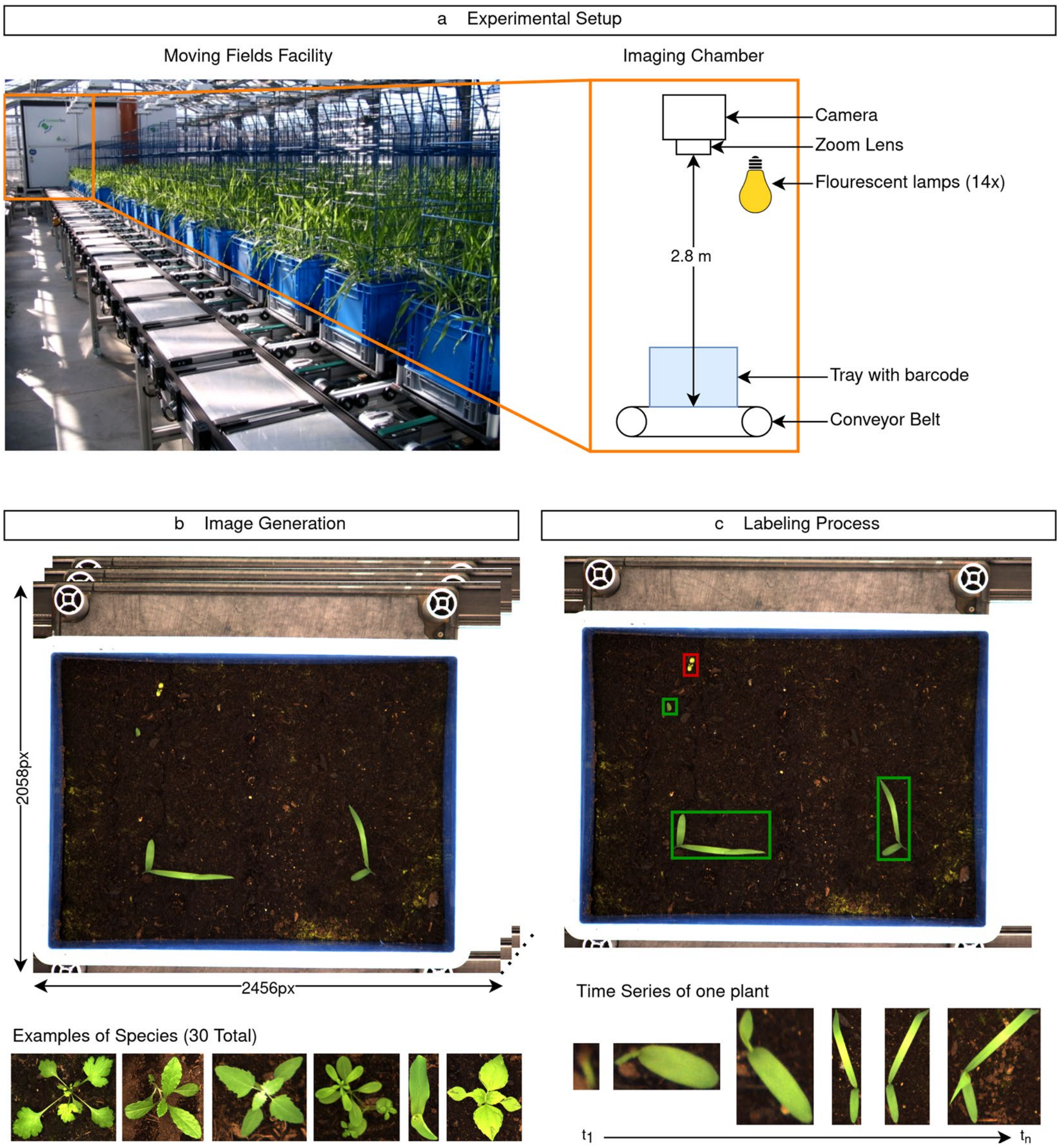

Fig. 2 Outline of the experimental workflow. (**a**) Experimental setup using an automated phenotyping facility. (**b**) Generation of images with high temporal and spatial resolution of 30 different plant species. (**c**) Labeling process illustrated by bounding box information for the object detection task (weed illustrated in red and sorghum plants in green).

Image Generation. To generate well illuminated, high-resolution top-down images of the experimental units, one of the Scanalyzer 3D imaging cabins of the Moving Fields facility was used. In this cabin, one RGB camera (Basler piA2400-17gm) is mounted 2.8 m perpendicular above the conveyor band. This camera takes images with 2456×2058 pixels. This camera is equipped with a motorized zoom lens (Pentax C6Z1218M3-5); throughout the experiment, however, this lens was set fixed at a single position, which resulted in a ground resolution of ~ 0.17 mm per pixel. The micro-plots were illuminated by 14 fluorescent lamps (Osram HE 28 W/865) that were also mounted perpendicular above the conveyor band. Units were imaged as often as possible, in practice at least once a day on the two days per week on which plant maintenance took place and at least twice a day on all other days. The units were tracked over a development period from sowing till the last plant either started shooting or became too big for the setup, representing the relevant stages for weed control on the field for sorghum. The images collected were stored in a LemnaTec-specific raw format, after which they were converted to PNG format. Each experimental unit was marked with a unique numerical barcode composed of identity codes for 1) the species, 2) the treatment (with or without GA3) and 3) the replicate at issue. Each image was saved with the barcode of the unit that was on the image as well as the date and time of image acquisition.

Labeling Process. The complete dataset was labeled using the open-source software CVAT (Computer Vision Annotation Tool; https://www.cvat.ai/) as a self-hosted solution. This software made the labeling time-efficient, as it provided an easy interface for multiple object tracking by adding time series information

EPPO Code	Dicot	Botanical Name	# Images	Captured Trays	Labeled Trays for classification	detection	segmentation
ACHMI	✓	*Achillea millefolium*	2,893	25	25	5	0
AGRRE		*Elymus repens*	2,848	29	12	12	0
ALOMY		*Alopecurus myosuroides*	3,595	28	11	11	0
ARTVU	✓	*Artemisia vulgaris*	2,881	34	34	18	1
CHEAL	✓	*Chenopodium album*	1,652	23	23	9	1
CIRAR	✓	*Cirsium arvense*	2,546	15	15	15	1
CONAR	✓	*Convolvulus arvensis*	4,914	27	27	27	1
ECHCG		*Echinochloa crus-galli*	2,622	27	25	25	0
GALAP	✓	*Galium aparine*	1,513	23	23	14	1
GASPA	✓	*Galinsoga parviflora*	847	13	13	13	1
LAMAL	✓	*Lamium album*	5,276	36	36	32	0
MATCH	✓	*Matricaria chamomilla*	1,681	21	21	8	0
PLAMA	✓	*Plantago major*	7,856	43	43	37	0
POAAN		*Poa annua*	2,574	23	11	11	0
POLCO	✓	*Fallopia convolvulus*	330	9	9	9	1
POROL	✓	*Portulaca oleracea*	4,435	21	21	5	1
PULDY	✓	*Pulicaria dysenterica*	1,745	21	21	10	1
SOLNI	✓	*Solanum nigrum*	1,472	22	22	12	1
SSYOF	✓	*Sisymbrium officinale*	2,477	14	14	10	0
STEME	✓	*Stellaria media*	1,887	22	22	8	0
THLAR	✓	*Thlaspi arvense*	4,044	38	38	36	1
VEROF	✓	*Veronica officinalis*	3,987	22	22	5	0
VIOAR	✓	*Viola arvensis*	1,261	17	17	11	1

Table 2. Selected weed species with corresponding amount of data captured and labeled.

Abbreviation	EPPO Code	Variety	# Images	Captured Trays	Labeled Trays for classification	detection	segmentation
SORFR	SORVU	KWS Freya	2,881	27	27	27	0
SORHA	SORVU	KWS Hannibal	552	27	5	5	1
SORKM	SORVU	KWS Merlin	511	27	5	5	0
SORSA	SORVU	KWS Sammos	1,279	27	12	12	0
SORKS	SORVU	KWS Sole	604	27	5	5	0
SORRS	SORVU	RAGT Swingg	447	27	5	5	0
ZEALP	ZEAMX	Lidea Palladium	930	38	38	38	0
ZEAKJ	ZEAMX	KWS Johaninio	1,660	38	38	38	1

Table 3. Selected crop varieties with corresponding amount of data captured and labeled.

EPPO code	Botanical name	# Plant individuals	# Plant samples
AETCY	*Aethusa cynapium*	4	110
GERMO	*Geranium molle*	4	139
POLAM	*Persicaria amphibia*	55	2237
POLAV	*Polygonum aviculare*	2	62
VICVI	*Vicia villosa*	1	32

Table 4. Additional germinating weed species (dicots) that were not sown explicitly.

per plant. Each species was labeled individually, providing the EPPO code as a label. Although only one species was seeded per tray, more weed species germinated during the experiment (compare Table 4), as the seed assortment was not completely pure. Therefore, the additional label "weed" was used to annotate these plants and the unknown species was identified by an expert in a second step. The correct species could not be specified for all plants due to several reasons (i.e., little germination resulting in small plants, occlusion with other plants, etc.), especially when they were not part of our initial assortment.

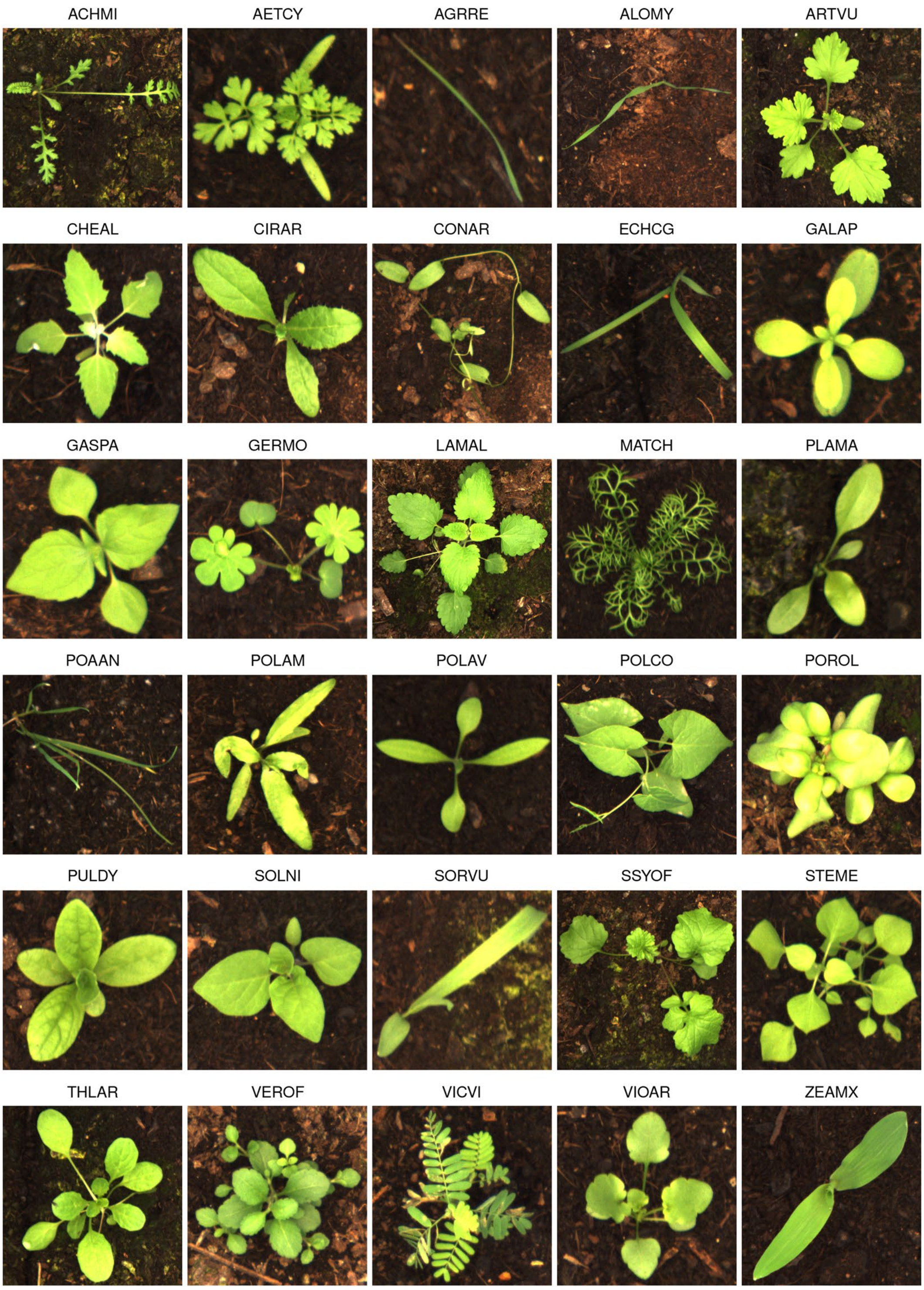

Fig. 3 Example of each plant species with corresponding EPPO code.

The instance segmentation masks were manually drawn using another open-source software, GIMP (GNU Image Manipulation Program), which provides pixel-level information. Therefore, we selected one tray from each of 14 plant species, as not all species could be labeled due to complexity and time constraints (see Fig. 1g).

Data Records

The Moving Fields Weed Dataset[30] (MFWD) is deposited at the digital library of the Technical University of Munich (https://mediatum.ub.tum.de/1717366). The dataset consists of 94,321 high temporal and spatial resolution images of 30 different plant species (see Fig. 3).

Data Content	Data Statistics
High Resolution Tray Images	94,321
Individual Plant Samples (Object Detection)	200,148
Individual Plant Samples (Instance Segmentation)	2,025
Time Series of individual Plants	5,068
Weed Species	28
Sorghum Varieties	6
Maize Varieties	2
All PNG images [GB]	814
All JPEG images [GB]	114

Table 5. Summary of the current version of our dataset.

Column	Explanation
track_id	ID of the individual time series connecting bbox_id of different images in one tray
label_id	ID of the plant species, saved as EPPO code
bbox_id	ID of the individual sample
xmin	left coordinate of the individual sample
ymin	top coordinate of the individual sample
xmax	right coordinate of the individual sample
ymax	bottom coordinate of the individual sample
filename	relative file path to the image
tray_id	ID of the tray

Table 6. Description of the csv file containing ground truth information.

Additional ground truth data is provided, consisting of the plant species, a bounding box per plant, and time series information to track the same plant individual through growth. We labeled a large subset of these images, resulting in 200,148 records for 5,068 plants in the current version. Additional information is shown in Table 5.

Image data is stored in PNG format to ensure the highest possible quality without compression artifacts. Compressed (JPEG) images are also stored to ensure accessibility with lower Internet bandwidth. All object detection and object tracking information are stored in a separate CSV file named "gt.csv". The segmentation masks are stored in the folder "masks". An additional folder is provided containing all images without ground truth annotations. The contents of the CSV file are explained in Table 6:

Technical Validation

The growth experiments were conducted using multiple trays, resulting in different numbers of replicates per species. Here, a minimum of nine replicates were used. Seeds were treated according to breeders' recommendations. Some weeds germinated only when treated with gibberellin acid. Therefore, the optimal procedure was evaluated in a prior experiment.

The quality of the bounding boxes and labels was ensured twofold. First, a valid bounding box could be evaluated during the labeling process by using tools in CVAT directly and by an additional human inspector doing quality control. Second, using the time-series information, all plant cut-outs of one plant individual could be plotted in a series of images to visually inspect the bounding boxes. Finally, plants of additional not sown species could be assessed and classified in an additional step. Remaining instances were labeled as class "Weed", as they were tiny and thus could not be labeled by species.

The high variability in the seed germination rate of different weed species resulted in a very diverse data set of different weeds with different germination rates, as shown in Fig. 4.

Usage Notes

The dataset can be downloaded via custom Python scripts (https://github.com/grimmlab/MFWD) and used as a resource for precision agriculture, smart farming, and computer vision related tasks. In addition, we encourage the development of algorithms that take plant phenotyping data into account. Since the dataset consists of high-resolution images, we also provide a custom Python script to easily resize the images. The dataset could be a useful resource for the computer science community in general to develop novel machine learning and computer vision algorithms for automatic weed detection. Here, the data could be used in classification as well as object detection and segmentation tasks. Furthermore, the additional time series information makes our dataset suitable for multiple object tracking. Finally, the inherent inter- and intra-class variance can be used to evaluate new algorithms that address class imbalance, which is a challenge in many machine learning tasks[31–33].

To demonstrate the usefulness of the dataset, we provide a simple baseline experiment on the image classification task. For this purpose, we use plant cut-outs from the jpeg-compressed images and rescale them to an image size of 224×224 pixels^2. Here we focused on the multi-species classification for sorghum, i.e. all sorghum

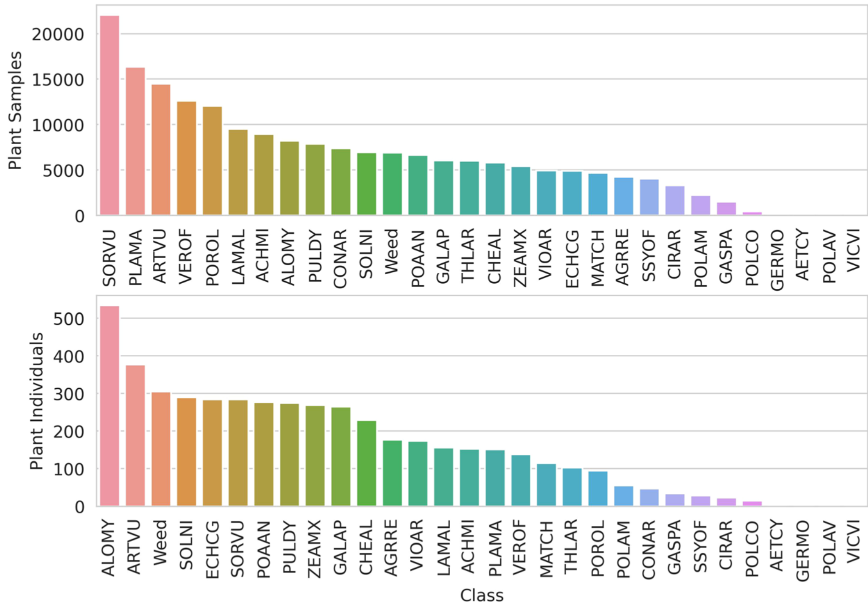

Fig. 4 Distribution of plant samples (separate images) and plant individuals (multiple images of one plant through time) per class.

model	learning rate ($*10^{-4}$)	best epoch	f1-score (%)
EfficientNet_b0	2.37	8	88.75
EfficientNet_b0	8.93	1	83.93
EfficientNet_b0	**5.40**	**32**	**90.00**
EfficientNet_b0	3.97	8	89.35
EfficientNet_b0	1.43	26	88.88
ResNet-10	2.37	20	88.24
ResNet-10	8.93	1	81.28
ResNet-10	5.40	21	88.74
ResNet-10	3.97	20	88.70
ResNet-10	1.43	20	87.75

Table 7. Summary of the hyperparameter optimization calculated on the validation set.

varieties were labeled as "SORVU" and excluded all maize images. Additionally, the generic weed class was excluded for this experiment, as these were mostly small plants which could not even be classified by the human eye. We also excluded POLAV and VICVI from the experiment because they contained less than three plant individuals and thus could not be separated into training, validation, and test sets. We strongly recommend stratifying the data by plant individuals, as the temporal resolution is high, and models may overfit if the data were randomly split. The final dataset of 27 plant species contained 167,505 images and was split into a training- (~60%), validation- (~20%), and test-set (~20%).

We selected two different deep learning-based model architectures, ResNet-10[34] and EfficientNet_b0[35], to evaluate the classification performance. For the hyperparameter optimization we used grid-search and five different learning rates (sampled from a log uniform distribution in the range between 1e-3 and 1e-4). We sampled 512 plant cut-outs in a batch by oversampling the minority classes, due to high class imbalances. The networks were initialized with weights from ImageNet. The Adam[36] optimizer with a learning rate scheduler and the cross-entropy loss[37] was used to train the models. We validated the models using the validation-set by calculating the weighted f1-score[38,39], due to the high class imbalance. Each model was trained for a maximum of 50 epochs. Early stopping[40] was used as a regularization technique to avoid overfitting. After training the models, EfficientNet_b0 with a learning rate of ~$5.4*10^{-4}$ gave the best results on the validation set with an f1-score of 90.00%. The summary of the hyperparameter optimization is shown in Table 7.

Finally, the best performing model was applied to the hold-out test set to evaluate the generalization abilities on an unseen dataset. Here, the model achieved a weighted f1-score of 90.57%, indicating good generalization performance within the MFWD dataset. The complete code for training and testing the model is publicly available in our GitHub repository.

However, deep learning models trained on this dataset may not be applicable to out-of-context data, such as weed detection in drone imagery. Here, pre-training a model on our dataset and fine-tuning it to the target task might be a feasible strategy to scale up weed detection in agricultural landscapes. However, the main target application of our dataset is to encourage the research community to develop new computer vision algorithms on a unified dataset, thus increasing the reproducibility of the results.

Code availability

The code to download the dataset is publicly available for download on GitHub: https://github.com/grimmlab/MFWD.

Received: 29 August 2023; Accepted: 10 January 2024;
Published online: 23 January 2024

References

1. Dentika, P., Ozier-Lafontaine, H. & Penet, L. Weeds as Pathogen Hosts and Disease Risk for Crops in the Wake of a Reduced Use of Herbicides: Evidence from Yam (Dioscorea alata) Fields and Colletotrichum Pathogens in the Tropics. *J Fungi* **7**, 283, https://doi.org/10.3390/jof7040283 (2021).
2. Norris, R. F. & Kogan, M. Interactions between weeds, arthropod pests, and their natural enemies in managed ecosystems. *Weed Sci* **48**, 94–158 (2000).
3. Schumacher, M., Dieterich, M. & Gerhards, R. Effects of weed biodiversity on the ecosystem service of weed seed predation along a farming intensity gradient. *Glob Ecol Conserv;* **24**, e01316, https://doi.org/10.1016/j.gecco.2020.e01316 (2020).
4. Logsdon, S. D. Root effects on soil properties and processes: Synthesis and future research needs. In: *Enhancing Understanding and Quantification of Soil-Root Growth Interactions*. John Wiley & Sons, Ltd, pp 173–196, https://doi.org/10.2134/advagricsystmodel4.c8 (2015).
5. Harker, K. N., Clayton, G. W. & O'Donovan, J. T. Reducing agroecosystem vulnerability to weed invasion. In: *Invasive Plants: Ecological and Agricultural Aspects*. Birkhäuser Basel, pp 195–207, https://doi.org/10.1007/3-7643-7380-6_12 (2005).
6. Harker, K. N. & O'Donovan, J. T. Recent Weed Control, Weed Management, and Integrated Weed Management. *Weed Technol;* **27**, 1–11, https://doi.org/10.1614/WT-D-12-00109.1 (2013).
7. Myers, J. P. *et al*. Concerns over use of glyphosate-based herbicides and risks associated with exposures: A consensus statement. *Environ. Heal. A Glob. Access Sci. Source.* **15**, 1–13, https://doi.org/10.1186/s12940-016-0117-0 (2016).
8. Steinmetz, Z. *et al*. Plastic mulching in agriculture. Trading short-term agronomic benefits for long-term soil degradation? *Sci. Total Environ.;* **550**, 690–705, https://doi.org/10.1016/j.scitotenv.2016.01.153 (2016).
9. MacLaren, C., Storkey, J., Menegat, A., Metcalfe, H. & Dehnen-Schmutz, K. An ecological future for weed science to sustain crop production and the environment. A review. *Agron. Sustain. Dev.* **40**, 1–29, https://doi.org/10.1007/s13593-020-00631-6 (2020).
10. Christensen, S. *et al*. Site-specific weed control technologies. *Weed Res;* **49**, 233–241, https://doi.org/10.1111/j.1365-3180.2009.00696.x (2009).
11. Hasan, A. S. M. M., Sohel, F., Diepeveen, D., Laga, H. & Jones, M. G. K. A survey of deep learning techniques for weed detection from images. Comput. Electron. Agric.; **184**, https://doi.org/10.1016/j.compag.2021.106067 (2021).
12. Wang, A., Zhang, W. & Wei, X. A review on weed detection using ground-based machine vision and image processing techniques. *Comput. Electron. Agric.;* **158**, 226–240, https://doi.org/10.1016/j.compag.2019.02.005 (2019).
13. Dian Bah, M., Hafiane, A. & Canals, R. Deep learning with unsupervised data labeling for weed detection in line crops in UAV images. *Remote Sens;* **10**, https://doi.org/10.3390/rs10111690 (2018).
14. Bakhshipour, A. & Jafari, A. Evaluation of support vector machine and artificial neural networks in weed detection using shape features. *Comput. Electron. Agric.;* **145**, 153–160, https://doi.org/10.1016/j.compag.2017.12.032 (2018).
15. Sivakumar, A. N. V. *et al*. Comparison of object detection and patch-based classification deep learning models on mid-to late-season weed detection in UAV imagery. *Remote Sens;* **12**, https://doi.org/10.3390/rs12132136 (2020).
16. Genze, N. *et al*. Improved weed segmentation in UAV imagery of sorghum fields with a combined deblurring segmentation model. *Plant Methods*; **19**. https://doi.org/10.1186/s13007-023-01060-8 (2023).
17. Genze, N. *et al*. Deep learning-based early weed segmentation using motion blurred UAV images of sorghum fields. *Comput. Electron. Agric.;* **202**, https://doi.org/10.1016/j.compag.2022.107388 (2022).
18. Milioto, A., Lottes, P. & Stachniss, C. Real-Time Semantic Segmentation of Crop and Weed for Precision Agriculture Robots Leveraging Background Knowledge in CNNs. In: *Proceedings - IEEE International Conference on Robotics and Automation*. Institute of Electrical and Electronics Engineers Inc., pp 2229–2235, https://doi.org/10.1109/ICRA.2018.8460962 (2018).
19. Lottes, P., Khanna, R., Pfeifer, J., Siegwart, R. & Stachniss, C. UAV-based crop and weed classification for smart farming. In: *Proceedings - IEEE International Conference on Robotics and Automation*. Institute of Electrical and Electronics Engineers Inc., pp 3024–3031, https://doi.org/10.1109/ICRA.2017.7989347 (2017).
20. Zhao, Z. Q., Zheng, P., Xu, S. T. & Wu, X. Object Detection with Deep Learning: A Review. *IEEE Trans. Neural Networks Learn. Syst.;* **30**, 3212–3232, https://doi.org/10.1109/TNNLS.2018.2876865 (2019).
21. Genze, N., Bharti, R., Grieb, M., Schultheiss, S. J. & Grimm, D. G. Accurate machine learning-based germination detection, prediction and quality assessment of three grain crops. *Plant Methods*; **16**. https://doi.org/10.1186/s13007-020-00699-x (2020).
22. Janneh, L. L., Zhang, Y., Cui, Z. & Yang, Y. Multi-level feature re-weighted fusion for the semantic segmentation of crops and weeds. *J King Saud Univ - Comput Inf Sci*; **35**, https://doi.org/10.1016/j.jksuci.2023.03.023 (2023).
23. Long, J., Shelhamer, E. & Darrell, T. Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, pp 431–440, https://doi.org/10.1109/CVPR.2015.7298965 (2015).
24. Champ, J. *et al*. Instance segmentation for the fine detection of crop and weed plants by precision agricultural robots. *Appl Plant Sci*; **8**. https://doi.org/10.1002/aps3.11373 (2020).
25. von Redwitz, C. *et al*. Better-Weeds – Next generation weed management. *Tagungsband 30 Dtsch Arbeitsbesprechung über Frag der Unkrautbiologie und -bekämpfung*; 432–437, https://doi.org/10.5073/20220124-075254 (2022).
26. Sapkota, B. B. *et al*. Use of synthetic images for training a deep learning model for weed detection and biomass estimation in cotton. *Sci Rep*; **12**. https://doi.org/10.1038/s41598-022-23399-z (2022).
27. Kirillov, A., He, K., Girshick, R., Rother, C. & Dollar, P. Panoptic segmentation. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, pp 9396–9405, https://doi.org/10.1109/CVPR.2019.00963 (2019).
28. Madsen, S. L. *et al*. Open plant phenotype database of common weeds in Denmark. *Remote Sens;* **12**. https://doi.org/10.3390/RS12081246 (2020).
29. Garcin, C., Joly, A., Bonnet, P., Chouet, M. & Servajean, M. Pl@ntNet-300K: a plant image dataset with high label ambiguity and a long-tailed distribution. *NeurIPS 2021-35th Conf Neural Inf Process Syst* 2021.
30. Genze, N. *et al*. Manually annotated and curated dataset of diverse weed species in maize and sorghum for computer vision. *Technical University of Munich, mediaTUM,* https://doi.org/10.14459/2023mp1717366 (2023).

31. Chou, H. P., Chang, S. C., Pan, J. Y., Wei, W. & Juan, D. C. Remix: Rebalanced Mixup. In: *Lecture Notes in Computer Science*. Springer Science and Business Media Deutschland GmbH, pp 95–110, https://doi.org/10.1007/978-3-030-65414-6_9 (2020).
32. Zhou, F., Yang, S., Fujita, H., Chen, D. & Wen, C. Deep learning fault diagnosis method based on global optimization GAN for unbalanced data. *Knowledge-Based Syst*; **187**. https://doi.org/10.1016/j.knosys.2019.07.008 (2020).
33. Kaur, H., Pannu, H. S. & Malhi, A. K. A systematic review on imbalanced data challenges in machine learning: Applications and solutions. ACM Comput. Surv.; **52**. https://doi.org/10.1145/3343440 (2019).
34. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp 770–778, https://doi.org/10.1109/CVPR.2016.90 (2016).
35. Tan, M. & Le, Q. V. EfficientNet: Rethinking model scaling for convolutional neural networks. In: *36th International Conference on Machine Learning, ICML* 2019. 2019, pp 10691–10700.
36. Kingma, D. & Ba, J. Adam: A Method for Stochastic Optimization. In: *International Conference on Learning Representations (ICLR)*. San Diega, CA, USA, 2015.
37. Cox, D. R. The Regression Analysis of Binary Sequences. *J R Stat Soc Ser B* **20**, 215–232, https://doi.org/10.1111/j.2517-6161.1958.tb00292.x (1958).
38. Chinchor, N. MUC-4 evaluation metrics. *4th Messag Underst Conf MUC 1992 - Proc*: 22–29, https://doi.org/10.3115/1072064.1072067 (1992).
39. He, H. & Garcia, E. A. Learning from Imbalanced Data. *IEEE Trans Knowl Data Eng* **21**, 1263–1284, https://doi.org/10.1109/TKDE.2008.239 (2009).
40. Yao, Y., Rosasco, L. & Caponnetto, A. On early stopping in gradient descent learning. *Constr Approx* **26**, 289–315, https://doi.org/10.1007/s00365-006-0663-2 (2007).
41. Hughes, D. P. & Salathe, M. An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv.org* 2015.http://arxiv.org/abs/1511.08060 (accessed 27 Oct2023).
42. Jiang, Y., Li, C., Paterson, A. H. & Robertson, J. S. DeepSeedling: Deep convolutional network and Kalman filter for plant seedling detection and counting in the field. *Plant Methods*; **15**. https://doi.org/10.1186/s13007-019-0528-3 (2019).
43. Olsen, A. *et al*. DeepWeeds: A Multiclass Weed Species Image Dataset for Deep Learning. *Sci Rep*; **9**. https://doi.org/10.1038/s41598-018-38343-3 (2019).
44. Wang, P. *et al*. Weed25: A deep learning dataset for weed identification. *Front Plant Sci*; **13**. https://doi.org/10.3389/fpls.2022.1053329 (2022).
45. Kitzler, F., Barta, N., Neugschwandtner, R. W., Gronauer, A. & Motsch, V. WE3DS: An RGB-D Image Dataset for Semantic Segmentation in Agriculture. *Sensors*; **23**. https://doi.org/10.3390/s23052713 (2023).
46. Sudars, K., Jasko, J., Namatevs, I., Ozola, L. & Badaukis, N. Dataset of annotated food crops and weed images for robotic computer vision control. *Data Br*; **31**. https://doi.org/10.1016/j.dib.2020.105833 (2020).
47. Giselsson, T. M., Jørgensen, R. N., Jensen, P. K., Dyrmann, M. & Midtiby, H. S. A Public Image Database for Benchmark of Plant Seedling Classification Algorithms. *arXiv.org* 2017.http://arxiv.org/abs/1711.05458 (accessed 27 Oct2023).
48. Bender, A., Whelan, B. & Sukkarieh, S. A high-resolution, multimodal data set for agricultural robotics: A Ladybird's-eye view of Brassica. *J F Robot;* **37**, 73–96, https://doi.org/10.1002/rob.21877 (2020).
49. Haug, S. & Ostermann, J. A crop/weed field image dataset for the evaluation of computer vision based precision agriculture tasks. In: *Lecture Notes in Computer Science*. Springer Verlag, pp 105–116, https://doi.org/10.1007/978-3-319-16220-1_8 (2015).
50. Minervini, M., Fischbach, A., Scharr, H. & Tsaftaris, S. A. Finely-grained annotated datasets for image-based plant phenotyping. *Pattern Recognit Lett;* **81**, 80–89, https://doi.org/10.1016/j.patrec.2015.10.013 (2016).

Acknowledgements
We thank Marlene Hanly and Agnes Wappler for their assistance in labeling the dataset. Funding for the research presented in this paper is provided by the Bavarian State Ministry for Food, Agriculture and Forests within the EWIS project (Funding ID: G2/N/19/13).

Author contributions
D.G.G., N.G., M.G. conceived and designed the study. J.G. and W.K.V. acquired the plants' seeds. J.G. and W.K.V. designed the experiments and were responsible for the image acquisition. N.G. and M.W. labeled the dataset. M.G. guided the labeling process with domain expertise. N.G. was responsible for quality control of the data. N.G. and D.G.G. wrote the manuscript with contributions from all authors. All authors read and approved the final manuscript.

Funding
Open Access funding enabled and organized by Projekt DEAL.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to D.G.G.

Acronyms

AI	Artificial Intelligence
ASPP	Atrous Spatial Pyramid Pooling
CNN	Convolutional Neural Network
CVAT	Computer Vision Annotation Tool
DC	Sørensen-Dice Coefficient
DL	Deep Learning
FCN	Fully Convolutional Network
FN	False Negatives
FP	False Positives
GPS	Global Positioning System
GSD	Ground Sampling Distance
LfL	Bavarian State Research Center for Agriculture
LPIPS	Learned Perceptual Image Patch Similarity
MFWD	Moving Fields Weed Dataset
MSE	Mean Squared Error
NAFNet	Nonlinear Activation Free Network
OOD	Out-of-distribution
PSF	point-spread function
PSNR	Peak Signal-to-Noise Ratio
ReLU	Rectified Linear Unit
ResNet	Residual Network
SSIM	Structural SIMilarity Index
SSWM	Site-Specific Weed Management
TFZ	Technology and Support Centre in the Centre of Excellence for Renewable Resources
TP	True Positives
UAV	Unmanned Aerial Vehicle
UGV	Unmanned Ground Vehicle
YOLO	You Only Look Once

Bibliography

[1] Samuel Knapp and Marcel G. A. van der Heijden. "A Global Meta-Analysis of Yield Stability in Organic and Conservation Agriculture". In: *Nature Communications* 9.1 (Sept. 7, 2018), p. 3632. ISSN: 2041-1723. DOI: 10.1038/s41467-018-05956-1. (Visited on 01/09/2025).

[2] Deepthi G Pai, Radhika Kamath, and Mamatha Balachandra. "Deep Learning Techniques for Weed Detection in Agricultural Environments: A Comprehensive Review". In: *IEEE Access* (2024), pp. 1–1. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2024.3418454. (Visited on 08/19/2024).

[3] Marco Esposito et al. "Drone and Sensor Technology for Sustainable Weed Management: A Review". In: *Chemical and Biological Technologies in Agriculture* 8.1 (Mar. 25, 2021), p. 18. ISSN: 2196-5641. DOI: 10.1186/s40538-021-00217-8. (Visited on 08/21/2024).

[4] K. K. Barman et al. "Challenges and Opportunities in Weed Management Under a Changing Agricultural Scenario". In: *Recent Advances in Weed Management*. Ed. by Bhagirath S. Chauhan and Gulshan Mahajan. New York, NY: Springer, 2014, pp. 365–390. ISBN: 978-1-4939-1019-9. DOI: 10.1007/978-1-4939-1019-9_16. (Visited on 09/27/2024).

[5] Philippe Jeanneret et al. "An Increase in Food Production in Europe Could Dramatically Affect Farmland Biodiversity". In: *Communications Earth & Environment* 2.1 (Sept. 2, 2021), pp. 1–8. ISSN: 2662-4435. DOI: 10.1038/s43247-021-00256-x. (Visited on 09/24/2024).

[6] Clement A. Okia. *Global Perspectives on Sustainable Forest Management*. BoD - Books on Demand, 2012. ISBN: 978-953-51-0569-5 978-953-51-5283-5. (Visited on 03/17/2025).

[7] Norman Myers. *The Primary Source: Tropical Forests & Our Future*. 1992. URL: https://www.cabidigitallibrary.org/doi/full/10.5555/19920661903 (visited on 01/17/2025).

[8] Harald Sioli. "The Effects of Deforestation in Amazonia". In: *The Geographical Journal* 151.2 (1985), pp. 197–203. ISSN: 0016-7398. DOI: 10.2307/633533. JSTOR: 633533. (Visited on 01/17/2025).

[9] Cassiano D'Almeida et al. "The Effects of Deforestation on the Hydrological Cycle in Amazonia: A Review on Scale and Resolution". In: *International Journal of Climatology* 27.5 (2007), pp. 633–647. ISSN: 1097-0088. DOI: 10.1002/joc.1475. (Visited on 01/17/2025).

[10] Sarah Carter et al. "Agriculture-Driven Deforestation in the Tropics from 1990–2015: Emissions, Trends and Uncertainties". In: *Environmental Research Letters* 13.1 (Dec. 2017), p. 014002. ISSN: 1748-9326. DOI: 10.1088/1748-9326/aa9ea4. (Visited on 01/17/2025).

[11] Ryan Abman et al. *Agricultural Productivity and Deforestation.* Sept. 14, 2020. DOI: 10.2139/ssrn.3692682. Social Science Research Network: 3692682. (Visited on 01/17/2025). Pre-published.

[12] *Cash for Carbon: A Randomized Trial of Payments for Ecosystem Services to Reduce Deforestation.* DOI: 10.1126/science.aan0568. (Visited on 01/17/2025).

[13] Ashley E. Larsen, L. Claire Powers, and Sofie McComb. "Identifying and Characterizing Pesticide Use on 9,000 Fields of Organic Agriculture". In: *Nature Communications* 12.1 (Sept. 15, 2021), p. 5461. ISSN: 2041-1723. DOI: 10.1038/s41467-021-25502-w. (Visited on 01/09/2025).

[14] Julian Cribb. *The Coming Famine: The Global Food Crisis and What We Can Do to Avoid It.* Univ of California Press, 2010. (Visited on 01/17/2025).

[15] E. -C. Oerke and H. -W. Dehne. "Safeguarding Production - Losses in Major Crops and the Role of Crop Protection". In: *Crop Protection* 23.4 (Apr. 1, 2004), pp. 275–285. ISSN: 0261-2194. DOI: 10.1016/j.cropro.2003.10.001. (Visited on 07/25/2024).

[16] Forrest Nutter, Paul Teng, and Matthew Royer. "Terms and Concepts for Yield, Crop Loss, and Disease Thresholds". In: (1993).

[17] Juan I Rattalino Edreira et al. "Beyond the Plot: Technology Extrapolation Domains for Scaling out Agronomic Science". In: *Environmental Research Letters* 13.5 (May 1, 2018), p. 054027. ISSN: 1748-9326. DOI: 10.1088/1748-9326/aac092. (Visited on 02/07/2025).

[18] L. G. Firbank and A. R. Watkinson. "Modelling the Population Dynamics of an Arable Weed and Its Effects Upon Crop Yield". In: *The Journal of Applied Ecology* 23.1 (Apr. 1986), p. 147. ISSN: 00218901. DOI: 10.2307/2403088. JSTOR: 2403088. (Visited on 01/17/2025).

[19] Robert L. Zimdahl and Nicholas T. Basinger. *Fundamentals of Weed Science.* Elsevier, Feb. 14, 2024. 578 pp. ISBN: 978-0-443-15724-0. Google Books: NOTPEAAAQBAJ.

[20] K. F. Egbadzor and E. K. Sakyi. "Effects of Synthetic Agricultural Chemicals on Health: Views of Smallholder Farmers in the Ho West District". In: *Journal of Agriculture and Food Sciences* 18.2 (2 2020), pp. 73–84. ISSN: 1597-1074. DOI: 10.4314/jafs.v18i2.5. (Visited on 01/17/2025).

[21] Aaron Baum et al. "A Fresh Look at Energy, Materials, and Labor in Agriculture". In: *Nature Precedings* (Mar. 19, 2010), pp. 1–1. ISSN: 1756-0357. DOI: 10.1038/npre.2010.4291.1. (Visited on 09/26/2024).

[22] D. C. Slaughter, D. K. Giles, and D. Downey. "Autonomous Robotic Weed Control Systems: A Review". In: *Computers and Electronics in Agriculture.* Emerging Technologies For Real-time and Integrated Agriculture Decisions 61.1 (Apr. 1, 2008), pp. 63–78. ISSN: 0168-1699. DOI: 10.1016/j.compag.2007.05.008. (Visited on 08/21/2024).

[23] Nikolaos Antonopoulos et al. “Hot Foam: Evaluation of a New, Non-Chemical Weed Control Option in Perennial Crops”. In: *Smart Agricultural Technology* 3 (Feb. 1, 2023), p. 100063. ISSN: 2772-3755. DOI: `10.1016/j.atech.2022.100063`. (Visited on 08/21/2024).

[24] Graham Brodie et al. “Microwave Soil Treatment and Plant Growth”. In: *Sustainable Crop Production*. IntechOpen, Oct. 15, 2019. ISBN: 978-1-78985-318-6. DOI: `10.5772/intechopen.89684`. (Visited on 08/21/2024).

[25] Anja Löbmann et al. “Electrical Weed Control and Its Effect on Soil Organisms”. In: *Tagungsband: 30. Deutsche Arbeitsbesprechung über Fragen der Unkrautbiologie und -bekämpfung* (2022), pp. 277–282. DOI: `10.5073/20220124-064756`. (Visited on 08/21/2024).

[26] Clément Vigneault and Diane L. Benoît. “Electrical Weed Control: Theory and Applications”. In: *Physical Control Methods in Plant Protection*. Ed. by Charles Vincent, B. Panneton, and F. Fleurat-Lessard. Berlin, Heidelberg: Springer, 2001, pp. 174–188. ISBN: 978-3-662-04584-8. DOI: `10.1007/978-3-662-04584-8_12`. (Visited on 08/21/2024).

[27] Ian Heap. “Herbicide Resistant Weeds”. In: *Integrated Pest Management: Pesticide Problems, Vol.3*. Ed. by David Pimentel and Rajinder Peshin. Dordrecht: Springer Netherlands, 2014, pp. 281–301. ISBN: 978-94-007-7796-5. DOI: `10.1007/978-94-007-7796-5_12`. (Visited on 08/20/2024).

[28] James L. Griffin, Joseph M. Boudreaux, and Donnie K. Miller. “Herbicides As Harvest Aids”. In: *Weed Science* 58.3 (Sept. 2010), pp. 355–358. ISSN: 0043-1745, 1550-2759. DOI: `10.1614/WS-09-108.1`. (Visited on 03/10/2025).

[29] Ti Zhang et al. “Early Application of Harvest Aid Herbicides Adversely Impacts Lentil”. In: *Agronomy Journal* 109.1 (2017), pp. 239–248. ISSN: 1435-0645. DOI: `10.2134/agronj2016.07.0419`. (Visited on 03/10/2025).

[30] H. Charles J. Godfray et al. “Food Security: The Challenge of Feeding 9 Billion People”. In: *Science* 327.5967 (Feb. 12, 2010), pp. 812–818. DOI: `10.1126/science.1185383`. (Visited on 01/17/2025).

[31] Jonathan A. Foley et al. “Solutions for a Cultivated Planet”. In: *Nature* 478.7369 (Oct. 2011), pp. 337–342. ISSN: 1476-4687. DOI: `10.1038/nature10452`. (Visited on 01/10/2025).

[32] Leonard P Gianessi. “The Increasing Importance of Herbicides in Worldwide Crop Production: The Increasing Importance of Herbicides”. In: *Pest Management Science* 69.10 (Oct. 2013), pp. 1099–1105. ISSN: 1526498X. DOI: `10.1002/ps.3598`. (Visited on 02/24/2025).

[33] I. Rao and T. Madhulety. “Role of Herbicides in Improving Crop Yields”. In: *Dev Physiol, Biochem and Mol Biol Plants* 1 (2005), pp. 203–287. (Visited on 02/24/2025).

[34] James H. Westwood et al. "Weed Management in 2050: Perspectives on the Future of Weed Science". In: *Weed Science* 66.3 (May 2018), pp. 275–285. ISSN: 1550-2759. DOI: 10.1017/wsc.2017.78. (Visited on 08/20/2024).

[35] Maor Matzrafi et al. "Evolution of Herbicide Resistance Mechanisms in Grass Weeds". In: *Plant Science* 229 (Dec. 1, 2014), pp. 43–52. ISSN: 0168-9452. DOI: 10.1016/j.plantsci.2014.08.013. (Visited on 08/21/2024).

[36] Ian Heap. "Global Perspective of Herbicide-Resistant Weeds". In: *Pest Management Science* 70.9 (2014), pp. 1306–1315. ISSN: 1526-4998. DOI: 10.1002/ps.3696. (Visited on 02/28/2025).

[37] Sahand Jorfi et al. "Herbicide Residues in Water Resources: A Scoping Review". In: *Avicenna Journal of Environmental Health Engineering* 8.2 (2021), pp. 126–133. URL: https://ajehe.umsha.ac.ir/Article/ajehe-4215 (visited on 03/11/2025).

[38] Vera Silva et al. "Pesticide Residues in European Agricultural Soils – A Hidden Reality Unfolded". In: *Science of The Total Environment* 653 (Feb. 25, 2019), pp. 1532–1545. ISSN: 0048-9697. DOI: 10.1016/j.scitotenv.2018.10.441. (Visited on 01/20/2025).

[39] Heinz-R. Köhler and Rita Triebskorn. "Wildlife Ecotoxicology of Pesticides: Can We Track Effects to the Population Level and Beyond?" In: *Science* 341.6147 (Aug. 16, 2013), pp. 759–765. DOI: 10.1126/science.1237591. (Visited on 01/20/2025).

[40] Rozidaini Mohd Ghazi et al. "Health Effects of Herbicides and Its Current Removal Strategies". In: *Bioengineered* 14.1 (Dec. 31, 2023), p. 2259526. ISSN: 2165-5979. DOI: 10.1080/21655979.2023.2259526. pmid: 37747278. (Visited on 03/11/2025).

[41] R. Hokanson et al. "Alteration of Estrogen-Regulated Gene Expression in Human Cells Induced by the Agricultural and Horticultural Herbicide Glyphosate". In: *Human & Experimental Toxicology* 26.9 (Sept. 1, 2007), pp. 747–752. ISSN: 0960-3271. DOI: 10.1177/0960327107083453. (Visited on 03/13/2025).

[42] Wen-Hao Su. "Crop Plant Signaling for Real-Time Plant Identification in Smart Farm: A Systematic Review and New Concept in Artificial Intelligence for Automated Weed Control". In: *Artificial Intelligence in Agriculture* 4 (Jan. 1, 2020), pp. 262–271. ISSN: 2589-7217. DOI: 10.1016/j.aiia.2020.11.001. (Visited on 08/21/2024).

[43] Ali A. Bajwa, Gulshan Mahajan, and Bhagirath S. Chauhan. "Nonconventional Weed Management Strategies for Modern Agriculture". In: *Weed science* 63.4 (2015), pp. 723–747. (Visited on 02/20/2025).

[44] María Sol Balbuena et al. "Effects of Sublethal Doses of Glyphosate on Honeybee Navigation". In: *Journal of Experimental Biology* 218.17 (Sept. 1, 2015), pp. 2799–2805. ISSN: 0022-0949. DOI: 10.1242/jeb.117291. (Visited on 03/13/2025).

[45] Gilles-Eric Séralini et al. "Republished Study: Long-Term Toxicity of a Roundup Herbicide and a Roundup-tolerantgenetically Modified Maize". In: *Environmental Sciences Europe* 26.1 (June 24, 2014), p. 14. ISSN: 2190-4715. DOI: `10.1186/s12302-014-0014-5`. (Visited on 03/13/2025).

[46] International Agency for Research on Cancer. "IARC Monographs Volume 112: Evaluation of Five Organophosphate Insecticides and Herbicides". In: *World Health Organization, Lyon* (2015).

[47] James L. Maino et al. "Estimating Rates of Pesticide Usage from Trends in Herbicide, Insecticide, and Fungicide Product Registrations". In: *Crop Protection* 163 (Jan. 1, 2023), p. 106125. ISSN: 0261-2194. DOI: `10.1016/j.cropro.2022.106125`. (Visited on 01/10/2025).

[48] S. European Commission. *The European Green Deal*. European Commission Brussels, Belgium, 2019.

[49] John P. Reganold and Jonathan M. Wachter. "Organic Agriculture in the Twenty-First Century". In: *Nature Plants* 2.2 (Feb. 3, 2016), pp. 1–8. ISSN: 2055-0278. DOI: `10.1038/nplants.2015.221`. (Visited on 01/20/2025).

[50] Verena Seufert, Navin Ramankutty, and Jonathan A. Foley. "Comparing the Yields of Organic and Conventional Agriculture". In: *Nature* 485.7397 (May 2012), pp. 229–232. ISSN: 1476-4687. DOI: `10.1038/nature11069`. (Visited on 09/26/2024).

[51] D. W. Lotter, R. Seidel, and W. Liebhardt. "The Performance of Organic and Conventional Cropping Systems in an Extreme Climate Year". In: *American Journal of Alternative Agriculture* 18.3 (Sept. 2003), pp. 146–154. ISSN: 1478-5498, 0889-1893. DOI: `10.1079/AJAA200345`. (Visited on 01/20/2025).

[52] Giuseppe Colla et al. "Soil Physical Properties and Tomato Yield and Quality in Alternative Cropping Systems". In: *Agronomy Journal* 92.5 (2000), pp. 924–932. ISSN: 1435-0645. DOI: `10.2134/agronj2000.925924x`. (Visited on 01/20/2025).

[53] H. Schebesta and J. J. L. Candel. "Game-Changing Potential of the EU's Farm to Fork Strategy". In: *Nature Food* 1 (2020), pp. 586–588. ISSN: 2662-1355. DOI: `10.1038/s43016-020-00166-9`. (Visited on 01/20/2025).

[54] Nicolas Lampkin and Jürn Sanders. *Policy Support for Organic Farming in the European Union 2010-2020*. DE: Johann Heinrich von Thünen-Institut, 2022. 132 pp. URL: `https://doi.org/10.3220/WP1663067402000` (visited on 01/20/2025).

[55] Elin Röös et al. "Risks and Opportunities of Increasing Yields in Organic Farming. A Review". In: *Agronomy for Sustainable Development* 38.2 (Feb. 27, 2018), p. 14. ISSN: 1773-0155. DOI: `10.1007/s13593-018-0489-3`. (Visited on 02/28/2025).

[56] Shanxing Gong et al. "Biodiversity and Yield Trade-Offs for Organic Farming". In: *Ecology Letters* 25.7 (2022), pp. 1699–1710. ISSN: 1461-0248. DOI: `10.1111/ele.14017`. (Visited on 02/28/2025).

[57] Markus Larsson et al. “Institutional Barriers to Organic Farming in Central and Eastern European Countries of the Baltic Sea Region”. In: *Agricultural and Food Economics* 1.1 (May 21, 2013), p. 5. ISSN: 2193-7532. DOI: 10.1186/2193-7532-1-5. (Visited on 02/28/2025).

[58] Peter R Hobbs, Ken Sayre, and Raj Gupta. “The Role of Conservation Agriculture in Sustainable Agriculture”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 363.1491 (July 24, 2007), pp. 543–555. DOI: 10.1098/rstb.2007.2169. (Visited on 01/20/2025).

[59] Agniva Mandal et al. “Chapter 2 - Conservation Agricultural Practices under Organic Farming”. In: *Advances in Organic Farming*. Ed. by Vijay Singh Meena et al. Woodhead Publishing, Jan. 1, 2021, pp. 17–37. ISBN: 978-0-12-822358-1. DOI: 10.1016/B978-0-12-822358-1.00014-6. (Visited on 01/20/2025).

[60] Stephen L. Young and Francis J. Pierce, eds. *Automation: The Future of Weed Control in Cropping Systems*. Dordrecht: Springer Netherlands, 2014. ISBN: 978-94-007-7511-4 978-94-007-7512-1. DOI: 10.1007/978-94-007-7512-1. (Visited on 02/21/2025).

[61] R. E. Garrett. “Development of a Synchronous Thinner”. In: *J Am Soc Sugar Beet Technol* 14.3 (1966), pp. 206–213.

[62] Dale L Shaner and Hugh J Beckie. “The Future for Weed Control and Technology”. In: *Pest Management Science* 70.9 (2014), pp. 1329–1339. ISSN: 1526-4998. DOI: 10.1002/ps.3706. (Visited on 02/20/2025).

[63] Francis J. Pierce and Peter Nowak. “Aspects of Precision Agriculture”. In: *Advances in Agronomy*. Ed. by Donald L. Sparks. Vol. 67. Academic Press, Jan. 1, 1999, pp. 1–85. DOI: 10.1016/S0065-2113(08)60513-1. (Visited on 01/20/2025).

[64] Stavros G. Vougioukas. “Agricultural Robotics”. In: *Annual Review of Control, Robotics, and Autonomous Systems* 2 (Volume 2, 2019 May 3, 2019), pp. 365–392. ISSN: 2573-5144. DOI: 10.1146/annurev-control-053018-023617. (Visited on 02/24/2025).

[65] Subhajit Basu et al. “Legal Framework for Small Autonomous Agricultural Robots”. In: *AI & SOCIETY* 35.1 (Mar. 2020), pp. 113–134. ISSN: 0951-5666, 1435-5655. DOI: 10.1007/s00146-018-0846-4. (Visited on 03/11/2025).

[66] D. Kent Shannon, David E. Clay, and Kenneth A. Sudduth. “An Introduction to Precision Agriculture”. In: *Precision Agriculture Basics*. John Wiley & Sons, Ltd, 2018, pp. 1–12. ISBN: 978-0-89118-367-9. DOI: 10.2134/precisionagbasics.2016.0084. (Visited on 01/17/2025).

[67] Victoria Gonzalez-Dugo et al. “Site-Specific Agriculture”. In: *Principles of Agronomy for Sustainable Agriculture*. Ed. by Francisco J. Villalobos and Elias Fereres. Cham: Springer International Publishing, 2024, pp. 583–597. ISBN: 978-3-031-69150-8. DOI: 10.1007/978-3-031-69150-8_39. (Visited on 02/03/2025).

[68] S Christensen et al. “Site-Specific Weed Control Technologies”. In: *Weed Research* 49.3 (2009), pp. 233–241. ISSN: 1365-3180. DOI: 10.1111/j.1365-3180.2009.00696.x. (Visited on 02/03/2025).

[69] Arjun Upadhyay et al. “Advances in Ground Robotic Technologies for Site-Specific Weed Management in Precision Agriculture: A Review”. In: *Computers and Electronics in Agriculture* 225 (Oct. 1, 2024), p. 109363. ISSN: 0168-1699. DOI: 10.1016/j.compag.2024.109363. (Visited on 02/03/2025).

[70] F López-Granados. “Weed Detection for Site-Specific Weed Management: Mapping and Real-Time Approaches”. In: *Weed Research* 51.1 (2011), pp. 1–11. ISSN: 1365-3180. DOI: 10.1111/j.1365-3180.2010.00829.x. (Visited on 07/25/2024).

[71] Edward C. Luschei et al. “Implementing and Conducting On-Farm Weed Research with the Use of GPS”. In: *Weed Science* 49.4 (2001), pp. 536–542. (Visited on 02/23/2025).

[72] Lee R. Van Wychen et al. “Accuracy and Cost Effectiveness of GPS-assisted Wild Oat Mapping in Spring Cereal Crops”. In: *Weed Science* 50.1 (2002), pp. 120–129. (Visited on 02/23/2025).

[73] H. J. Beckie and S. Shirriff. “Site-Specific Wild Oat (*Avena Fatua* L.) Management”. In: *Canadian Journal of Plant Science* 92.5 (Sept. 2012), pp. 923–931. ISSN: 0008-4220, 1918-1833. DOI: 10.4141/cjps2012-007. (Visited on 02/23/2025).

[74] Hugh J. Beckie, Linda M. Hall, and Barclay Schuba. “Patch Management of Herbicide-Resistant Wild Oat (Avena Fatua)”. In: *Weed technology* 19.3 (2005), pp. 697–705. (Visited on 02/23/2025).

[75] Guy R. Y. Coleman et al. “Weed Detection to Weed Recognition: Reviewing 50 Years of Research to Identify Constraints and Opportunities for Large-Scale Cropping Systems”. In: *Weed Technology* 36.6 (Dec. 2022), pp. 741–757. ISSN: 0890-037X, 1550-2740. DOI: 10.1017/wet.2022.84. (Visited on 09/26/2024).

[76] R. J. Haggar, C. J. Stent, and S. Isaac. “A Prototype Hand-Held Patch Sprayer for Killing Weeds, Activated by Spectral Differences in Crop/Weed Canopies”. In: *Journal of Agricultural Engineering Research* 28.4 (July 1, 1983), pp. 349–358. ISSN: 0021-8634. DOI: 10.1016/0021-8634(83)90066-5. (Visited on 01/20/2025).

[77] D. M. Woebbecke et al. “Color Indices for Weed Identification under Various Soil, Residue, and Lighting Conditions”. In: *Transactions of the American Society of Agricultural Engineers* 38.1 (Jan. 1995), pp. 259–269. ISSN: 0001-2351. URL: http://www.scopus.com/inward/record.url?scp=0029110322&partnerID=8YFLogxK (visited on 01/20/2025).

[78] Hiroshi Okamoto et al. “Plant Classification for Weed Detection Using Hyperspectral Imaging with Wavelet Analysis”. In: *Weed Biology and Management* 7.1 (2007), pp. 31–37. ISSN: 1445-6664. DOI: 10.1111/j.1445-6664.2006.00234.x. (Visited on 01/20/2025).

[79] Chunying Wang et al. "A Review of Deep Learning Used in the Hyperspectral Image Analysis for Agriculture". In: *Artificial Intelligence Review* 54.7 (Oct. 1, 2021), pp. 5205–5253. ISSN: 1573-7462. DOI: 10.1007/s10462-021-10018-y. (Visited on 01/20/2025).

[80] Won Suk Lee, D. C. Slaughter, and D. K. Giles. "Robotic Weed Control System for Tomatoes". In: *Precision Agriculture* 1 (1999), pp. 95–113. (Visited on 02/23/2025).

[81] Steven A. Fennimore et al. "Evaluation and Economics of a Rotating Cultivator in Bok Choy, Celery, Lettuce, and Radicchio". In: *Weed Technology* 28.1 (2014), pp. 176–188. (Visited on 02/23/2025).

[82] David C. Slaughter. "The Biological Engineer: Sensing the Difference Between Crops and Weeds". In: *Automation: The Future of Weed Control in Cropping Systems*. Ed. by Stephen L. Young and Francis J. Pierce. Dordrecht: Springer Netherlands, 2014, pp. 71–95. ISBN: 978-94-007-7512-1. DOI: 10.1007/978-94-007-7512-1_5. (Visited on 01/20/2025).

[83] Ralph B. Brown and Scott D. Noble. "Site-Specific Weed Management: Sensing Requirements - What Do We Need to See?" In: *Weed Science* 53.2 (Apr. 2005), pp. 252–258. ISSN: 0043-1745, 1550-2759. DOI: 10.1614/WS-04-068R1. (Visited on 08/16/2024).

[84] Tasawar Abbas et al. "Chapter Five - Limitations of Existing Weed Control Practices Necessitate Development of Alternative Techniques Based on Biological Approaches". In: *Advances in Agronomy*. Ed. by Donald L. Sparks. Vol. 147. Academic Press, Jan. 1, 2018, pp. 239–280. DOI: 10.1016/bs.agron.2017.10.005. (Visited on 02/28/2025).

[85] Xavier P. Burgos-Artizzu et al. "Real-Time Image Processing for Crop/Weed Discrimination in Maize Fields". In: *Computers and Electronics in Agriculture* 75.2 (Feb. 1, 2011), pp. 337–346. ISSN: 0168-1699. DOI: 10.1016/j.compag.2010.12.011. (Visited on 01/20/2025).

[86] Xiuhao Quan and Reiner Doluschitz. "Unmanned Aerial Vehicle (UAV) Technical Applications, Standard Workflow, and Future Developments in Maize Production - Water Stress Detection, Weed Mapping, Nutritional Status Monitoring and Yield Prediction". In: *agricultural engineering.eu* 76.1 (1 Mar. 24, 2021). ISSN: 2943-5641. DOI: 10.15150/lt.2021.3263. (Visited on 08/19/2024).

[87] R Gerhards and S Christensen. "Real-Time Weed Detection, Decision Making and Patch Spraying in Maize, Sugarbeet, Winter Wheat and Winter Barley". In: *Weed Research* 43.6 (2003), pp. 385–392. ISSN: 1365-3180. DOI: 10.1046/j.1365-3180.2003.00349.x. (Visited on 08/19/2024).

[88] C.P. Lo. "Applied Remote Sensing". In: *Geocarto International* 1.4 (Jan. 1986), pp. 60–60. ISSN: 1010-6049, 1752-0762. DOI: 10.1080/10106048609354071. (Visited on 03/06/2025).

[89] M. D. Steven and Jeremy Austin Clark. *Applications of Remote Sensing in Agriculture*. Elsevier, 2013. (Visited on 02/23/2025).

[90] Wouter H. Maes and Kathy Steppe. "Perspectives for Remote Sensing with Unmanned Aerial Vehicles in Precision Agriculture". In: *Trends in Plant Science* 24.2 (Feb. 1, 2019), pp. 152–164. ISSN: 1360-1385. DOI: 10.1016/j.tplants.2018.11.007. (Visited on 02/05/2025).

[91] Chad DeChant et al. "Automated Identification of Northern Leaf Blight-Infected Maize Plants from Field Imagery Using Deep Learning". In: *Phytopathology®* 107.11 (Nov. 2017), pp. 1426–1432. ISSN: 0031-949X. DOI: 10.1094/PHYTO-11-16-0417-R. (Visited on 02/27/2025).

[92] Jiang Lu et al. "An In-Field Automatic Wheat Disease Diagnosis System". In: *Computers and Electronics in Agriculture* 142 (Nov. 1, 2017), pp. 369–379. ISSN: 0168-1699. DOI: 10.1016/j.compag.2017.09.012. (Visited on 02/27/2025).

[93] Amanda Ramcharan et al. "Deep Learning for Image-Based Cassava Disease Detection". In: *Frontiers in Plant Science* 8 (Oct. 27, 2017). ISSN: 1664-462X. DOI: 10.3389/fpls.2017.01852. (Visited on 02/27/2025).

[94] Bilal Hazim Younus Alsalam et al. "Autonomous UAV with Vision Based On-Board Decision Making for Remote Sensing and Precision Agriculture". In: *2017 IEEE Aerospace Conference*. 2017 IEEE Aerospace Conference. Mar. 2017, pp. 1–12. DOI: 10.1109/AERO.2017.7943593. (Visited on 02/27/2025).

[95] Eugenio Vocaturo et al. "AI-Driven Agriculture: Opportunities and Challenges". In: *2023 IEEE International Conference on Big Data (BigData)*. 2023 IEEE International Conference on Big Data (BigData). Dec. 2023, pp. 3530–3537. DOI: 10.1109/BigData59044.2023.10386314. (Visited on 02/25/2025).

[96] Kamila Dilmurat, Vasit Sagan, and Stephen Moose. "AI-driven Maize Yield Forecasting Using Unmanned Aerial Vehicle-Based Hyperspectral and Lidar Data Fusion". In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 3 (2022), pp. 193–199. URL: https://isprs-annals.copernicus.org/articles/V-3-2022/193/2022/isprs-annals-V-3-2022-193-2022.html (visited on 02/25/2025).

[97] Sameer Qazi, Bilal A. Khawaja, and Qazi Umar Farooq. "IoT-Equipped and AI-Enabled Next Generation Smart Agriculture: A Critical Review, Current Challenges and Future Trends". In: *IEEE Access* 10 (2022), pp. 21219–21235. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2022.3152544. (Visited on 02/25/2025).

[98] Avital Bechar. "Agricultural Robotics for Precision Agriculture Tasks: Concepts and Principles". In: *Innovation in Agricultural Robotics for Precision Agriculture: A Roadmap for Integrating Robots in Precision Agriculture*. Ed. by Avital Bechar. Cham: Springer International Publishing, 2021, pp. 17–30. ISBN: 978-3-030-77036-5. DOI: 10.1007/978-3-030-77036-5_2. (Visited on 02/25/2025).

[99] Ivan Beloev et al. “Artificial Intelligence-Driven Autonomous Robot for Precision Agriculture”. In: *Acta Technologica Agriculturae* 24.1 (Mar. 1, 2021), pp. 48–54. ISSN: 1338-5267. DOI: 10.2478/ata-2021-0008. (Visited on 02/25/2025).

[100] Qianyu Chen et al. “AI-enhanced Soil Management and Smart Farming”. In: *Soil Use and Management* 38.1 (Jan. 2022), pp. 7–13. ISSN: 0266-0032, 1475-2743. DOI: 10.1111/sum.12771. (Visited on 02/25/2025).

[101] Parveen Kumar et al. “Intelligent Irrigation Systems”. In: *AI for Big Data-Based Engineering Applications from Security Perspectives*. CRC Press, 2023, pp. 187–213. (Visited on 02/25/2025).

[102] Syeda Iqra Hassan et al. “A Systematic Review on Monitoring and Advanced Control Strategies in Smart Agriculture”. In: *Ieee Access* 9 (2021), pp. 32517–32548. URL: https://ieeexplore.ieee.org/abstract/document/9350242/ (visited on 02/25/2025).

[103] Patrizia Vizza et al. “Tracking Agricultural Products for Wellness Care”. In: *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2018, pp. 2064–2067. URL: https://ieeexplore.ieee.org/abstract/document/8621161/ (visited on 02/25/2025).

[104] Giuseppe Tradigo et al. “SISTABENE: An Information System for the Traceability of Agricultural Food Production”. In: *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2019, pp. 2304–2309. URL: https://ieeexplore.ieee.org/abstract/document/8983039/ (visited on 02/25/2025).

[105] Asifullah Khan et al. “A Survey of the Recent Architectures of Deep Convolutional Neural Networks”. In: *Artificial Intelligence Review* 53.8 (Dec. 1, 2020), pp. 5455–5516. ISSN: 1573-7462. DOI: 10.1007/s10462-020-09825-6. (Visited on 01/20/2025).

[106] C Fernández-Quintanilla et al. “Is the Current State of the Art of Weed Monitoring Suitable for Site-Specific Weed Management in Arable Crops?” In: *Weed Research* 58.4 (2018), pp. 259–272. ISSN: 1365-3180. DOI: 10.1111/wre.12307. (Visited on 01/20/2025).

[107] Milan Šulc and Jiří Matas. “Fine-Grained Recognition of Plants from Images”. In: *Plant Methods* 13 (2017), p. 115. ISSN: 1746-4811. DOI: 10.1186/s13007-017-0265-4. pmid: 29299049.

[108] I. Gogul and V. Sathiesh Kumar. “Flower Species Recognition System Using Convolution Neural Networks and Transfer Learning”. In: *2017 Fourth International Conference on Signal Processing, Communication and Networking (ICSCN)*. 2017 Fourth International Conference on Signal Processing, Communication and Networking (ICSCN). Mar. 2017, pp. 1–6. DOI: 10.1109/ICSCN.2017.8085675. (Visited on 01/20/2025).

[109] Chen Sun et al. “Revisiting Unreasonable Effectiveness of Data in Deep Learning Era”. In: Proceedings of the IEEE International Conference on Computer Vision. 2017, pp. 843–852. URL: `https://openaccess.thecvf.com/content_iccv_2017/html/Sun_Revisiting_Unreasonable_Effectiveness_ICCV_2017_paper.html` (visited on 02/28/2025).

[110] Xue-Wen Chen and Xiaotong Lin. “Big Data Deep Learning: Challenges and Perspectives”. In: *IEEE Access* 2 (2014), pp. 514–525. ISSN: 2169-3536. DOI: `10.1109/ACCESS.2014.2325029`. (Visited on 02/28/2025).

[111] Jayme Garcia Arnal Barbedo. “Impact of Dataset Size and Variety on the Effectiveness of Deep Learning and Transfer Learning for Plant Disease Classification”. In: *Computers and Electronics in Agriculture* 153 (Oct. 1, 2018), pp. 46–53. ISSN: 0168-1699. DOI: `10.1016/j.compag.2018.08.013`. (Visited on 02/28/2025).

[112] Amandalynne Paullada et al. “Data and Its (Dis)Contents: A Survey of Dataset Development and Use in Machine Learning Research”. In: *Patterns* 2.11 (Nov. 12, 2021). ISSN: 2666-3899. DOI: `10.1016/j.patter.2021.100336`. (Visited on 02/28/2025).

[113] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. “Imagenet Classification with Deep Convolutional Neural Networks”. In: *Advances in neural information processing systems* 25 (2012). URL: `https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html` (visited on 03/11/2025).

[114] Hansel J. Otero et al. “Twenty Years of Cost-effectiveness Analysis in Medical Imaging: Are We Improving?” In: *Radiology* 249.3 (Dec. 2008), pp. 917–925. ISSN: 0033-8419. DOI: `10.1148/radiol.2493080237`. pmid: `19011188`. (Visited on 02/27/2025).

[115] Boyang Deng, Yuzhen Lu, and Jiajun Xu. “Weed Database Development: An Updated Survey of Public Weed Datasets and Cross-Season Weed Detection Adaptation”. In: *Ecological Informatics* 81 (July 1, 2024), p. 102546. ISSN: 1574-9541. DOI: `10.1016/j.ecoinf.2024.102546`. (Visited on 03/15/2025).

[116] Yuzhen Lu and Sierra Young. “A Survey of Public Datasets for Computer Vision Tasks in Precision Agriculture”. In: *Computers and Electronics in Agriculture* 178 (Nov. 1, 2020), p. 105760. ISSN: 0168-1699. DOI: `10.1016/j.compag.2020.105760`. (Visited on 02/03/2025).

[117] Mark D. Wilkinson et al. “The FAIR Guiding Principles for Scientific Data Management and Stewardship”. In: *Scientific Data* 3.1 (Mar. 15, 2016), p. 160018. ISSN: 2052-4463. DOI: `10.1038/sdata.2016.18`. (Visited on 02/27/2025).

[118] Nikita Genze et al. “Deep Learning-Based Early Weed Segmentation Using Motion Blurred UAV Images of Sorghum Fields”. In: *Computers and Electronics in Agriculture* 202 (Nov. 1, 2022), p. 107388. ISSN: 0168-1699. DOI: `10.1016/j.compag.2022.107388`. (Visited on 02/14/2025).

[119] Mohamed Alloghani et al. “A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science”. In: *Supervised and Unsupervised Learning for Data Science*. Ed. by Michael W. Berry, Azlinah Mohamed, and Bee Wah Yap. Cham: Springer International Publishing, 2020, pp. 3–21. ISBN: 978-3-030-22475-2. DOI: 10.1007/978-3-030-22475-2_1. (Visited on 02/28/2025).

[120] Krupa Patel and Hiren B. Patel. “A Comparative Analysis of Supervised Machine Learning Algorithm for Agriculture Crop Prediction”. In: *2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT)*. 2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT). Sept. 2021, pp. 1–5. DOI: 10.1109/ICECCT52121.2021.9616731. (Visited on 02/28/2025).

[121] S. V. Jansi Rani et al. “Automated Weed Detection System in Smart Farming for Developing Sustainable Agriculture”. In: *International Journal of Environmental Science and Technology* 19.9 (Sept. 1, 2022), pp. 9083–9094. ISSN: 1735-2630. DOI: 10.1007/s13762-021-03606-6. (Visited on 01/20/2025).

[122] Ildar Rakhmatulin, Andreas Kamilaris, and Christian Andreasen. “Deep Neural Networks to Detect Weeds from Crops in Agricultural Environments in Real-Time: A Review”. In: *Remote Sensing* 13.21 (21 Jan. 2021), p. 4486. ISSN: 2072-4292. DOI: 10.3390/rs13214486. (Visited on 08/21/2024).

[123] Pallab Bharman et al. “Deep Learning in Agriculture: A Review”. In: *Asian Journal of Research in Computer Science* (Feb. 15, 2022), pp. 28–47. ISSN: 2581-8260. DOI: 10.9734/ajrcos/2022/v13i230311. (Visited on 08/19/2024).

[124] Muhammad Hammad Saleem, Johan Potgieter, and Khalid Mahmood Arif. “Automation in Agriculture by Machine and Deep Learning Techniques: A Review of Recent Developments”. In: *Precision Agriculture* 22.6 (Dec. 1, 2021), pp. 2053–2091. ISSN: 1573-1618. DOI: 10.1007/s11119-021-09806-x. (Visited on 08/19/2024).

[125] Nikita Genze et al. “Manually Annotated and Curated Dataset of Diverse Weed Species in Maize and Sorghum for Computer Vision”. In: *Scientific Data* 11.1 (Jan. 23, 2024), p. 109. ISSN: 2052-4463. DOI: 10.1038/s41597-024-02945-6. (Visited on 02/14/2025).

[126] Jonathan Long, Evan Shelhamer, and Trevor Darrell. “Fully Convolutional Networks for Semantic Segmentation”. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015, pp. 3431–3440. URL: https://openaccess.thecvf.com/content_cvpr_2015/html/Long_Fully_Convolutional_Networks_2015_CVPR_paper.html (visited on 02/28/2025).

[127] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Ed. by Nassir Navab et al. Vol. 9351. Cham: Springer International Publishing, 2015, pp. 234–241. ISBN: 978-3-319-24573-7 978-3-319-24574-4. DOI: 10.1007/978-3-319-24574-4_28. (Visited on 02/17/2025).

[128] Liang-Chieh Chen et al. “DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40.4 (Apr. 2018), pp. 834–848. ISSN: 1939-3539. DOI: 10.1109/TPAMI.2017.2699184. (Visited on 02/28/2025).

[129] Kevin Trebing, Tomasz Stanczyk, and Siamak Mehrkanoon. “SmaAt-UNet: Precipitation Nowcasting Using a Small Attention-UNet Architecture”. In: *Pattern Recognition Letters* 145 (May 1, 2021), pp. 178–186. ISSN: 0167-8655. DOI: 10.1016/j.patrec.2021.01.036. (Visited on 03/07/2025).

[130] Hu Cao et al. “Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation”. In: *Computer Vision – ECCV 2022 Workshops.* Ed. by Leonid Karlinsky, Tomer Michaeli, and Ko Nishino. Cham: Springer Nature Switzerland, 2023, pp. 205–218. ISBN: 978-3-031-25066-8. DOI: 10.1007/978-3-031-25066-8_9.

[131] Mulham Fawakherji et al. “Crop and Weeds Classification for Precision Agriculture Using Context-Independent Pixel-Wise Segmentation”. In: *2019 Third IEEE International Conference on Robotic Computing (IRC).* 2019 Third IEEE International Conference on Robotic Computing (IRC). Feb. 2019, pp. 146–152. DOI: 10.1109/IRC.2019.00029. (Visited on 02/28/2025).

[132] Fangyu Liu and Linbing Wang. “UNet-based Model for Crack Detection Integrating Visual Explanations”. In: *Construction and Building Materials* 322 (Mar. 7, 2022), p. 126265. ISSN: 0950-0618. DOI: 10.1016/j.conbuildmat.2021.126265. (Visited on 02/28/2025).

[133] Junfeng Jing et al. “Mobile-Unet: An Efficient Convolutional Neural Network for Fabric Defect Detection”. In: *Textile Research Journal* 92.1–2 (Jan. 1, 2022), pp. 30–42. DOI: 10.1177/0040517520928604. (Visited on 02/28/2025).

[134] Liang-Chieh Chen et al. “Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation”. In: *Proceedings of the European Conference on Computer Vision (ECCV).* 2018, pp. 801–818. URL: http://openaccess.thecvf.com/content_ECCV_2018/html/Liang-Chieh_Chen_Encoder-Decoder_with_Atrous_ECCV_2018_paper.html (visited on 03/10/2025).

[135] Liang-Chieh Chen et al. *Rethinking Atrous Convolution for Semantic Image Segmentation.* Dec. 5, 2017. DOI: 10.48550/arXiv.1706.05587. arXiv: 1706.05587 [cs]. (Visited on 03/10/2025). Pre-published.

[136] Kaiming He et al. “Deep Residual Learning for Image Recognition”. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, June 2016, pp. 770–778. ISBN: 978-1-4673-8851-1. DOI: 10.1109/CVPR.2016.90. (Visited on 02/17/2025).

[137] Kaixuan Huang et al. "Why Do Deep Residual Networks Generalize Better than Deep Feedforward Networks? — A Neural Tangent Kernel Perspective". In: *Advances in Neural Information Processing Systems*. Vol. 33. Curran Associates, Inc., 2020, pp. 2698–2709. URL: https://proceedings.neurips.cc/paper/2020/hash/1c336b8080f82bcc2cd2499b4c57261d-Abstract.html (visited on 03/15/2025).

[138] Mingxing Tan and Quoc Le. "Efficientnet: Rethinking Model Scaling for Convolutional Neural Networks". In: *International Conference on Machine Learning*. PMLR, 2019, pp. 6105–6114. URL: http://proceedings.mlr.press/v97/tan19a.html?ref=jina-ai-gmbh.ghost.io (visited on 03/10/2025).

[139] Irwan Bello et al. "Revisiting ResNets: Improved Training and Scaling Strategies". In: *Advances in Neural Information Processing Systems*. Vol. 34. Curran Associates, Inc., 2021, pp. 22614–22627. URL: https://proceedings.neurips.cc/paper/2021/hash/bef4d169d8bddd17d68303877a3ea945-Abstract.html (visited on 02/17/2025).

[140] Seungjun Nah et al. "NTIRE 2021 Challenge on Image Deblurring". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 149–165. URL: http://openaccess.thecvf.com/content/CVPR2021W/NTIRE/html/Nah_NTIRE_2021_Challenge_on_Image_Deblurring_CVPRW_2021_paper.html (visited on 03/10/2025).

[141] Zohair Al-Ameen et al. "A Comprehensive Study on Fast Image Deblurring Techniques". In: *International Journal of Advanced Science and Technology* 44 (2012). URL: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=c3477382086585b11dfc879e006985ad47722c04 (visited on 03/10/2025).

[142] D. Kundur and D. Hatzinakos. "Blind Image Deconvolution". In: *IEEE Signal Processing Magazine* 13.3 (May 1996), pp. 43–64. ISSN: 1558-0792. DOI: 10.1109/79.489268. (Visited on 03/10/2025).

[143] Norbert Wiener. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. The MIT press, 1964. URL: https://dl.acm.org/doi/abs/10.5555/1097023 (visited on 03/10/2025).

[144] Jiangxin Dong, Stefan Roth, and Bernt Schiele. "Deep Wiener Deconvolution: Wiener Meets Deep Learning for Image Deblurring". In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 1048–1059. URL: https://proceedings.neurips.cc/paper/2020/hash/0b8aff0438617c055eb55f0ba5d226fa-Abstract.html (visited on 03/10/2025).

[145] Jiangxin Dong, Stefan Roth, and Bernt Schiele. "DWDN: Deep Wiener Deconvolution Network for Non-Blind Image Deblurring". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.12 (2021), pp. 9960–9976. URL: https://ieeexplore.ieee.org/abstract/document/9664009 (visited on 03/10/2025).

[146] Esmaeil Faramarzi, Dinesh Rajan, and Marc P. Christensen. "Unified Blind Method for Multi-Image Super-Resolution and Single/Multi-Image Blur Deconvolution". In: *IEEE Transactions on Image Processing* 22.6 (2013), pp. 2101–2114. URL: https://ieeexplore.ieee.org/abstract/document/6408136/ (visited on 03/10/2025).

[147] Christian J. Schuler et al. "Learning to Deblur". In: *IEEE transactions on pattern analysis and machine intelligence* 38.7 (2015), pp. 1439–1451. URL: https://ieeexplore.ieee.org/abstract/document/7274732 (visited on 03/10/2025).

[148] Jinshan Pan et al. "Learning to Deblur Images with Exemplars". In: *IEEE transactions on pattern analysis and machine intelligence* 41.6 (2018), pp. 1412–1425. URL: https://ieeexplore.ieee.org/abstract/document/8353166 (visited on 03/10/2025).

[149] Lingyan Ruan et al. "Learning to Deblur Using Light Field Generated and Real Defocus Images". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 16304–16313. URL: http://openaccess.thecvf.com/content/CVPR2022/html/Ruan_Learning_to_Deblur_Using_Light_Field_Generated_and_Real_Defocus_CVPR_2022_paper.html (visited on 03/10/2025).

[150] Kaihao Zhang et al. "Deep Image Deblurring: A Survey". In: *International Journal of Computer Vision* 130.9 (2022), pp. 2103–2130. (Visited on 03/10/2025).

[151] Liangyu Chen et al. "Simple Baselines for Image Restoration". In: *Computer Vision – ECCV 2022*. Ed. by Shai Avidan et al. Cham: Springer Nature Switzerland, 2022, pp. 17–33. ISBN: 978-3-031-20071-7. DOI: 10.1007/978-3-031-20071-7_2.

[152] M. Stone. "Cross-Validatory Choice and Assessment of Statistical Predictions". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 36.2 (Jan. 1, 1974), pp. 111–133. ISSN: 1369-7412, 1467-9868. DOI: 10.1111/j.2517-6161.1974.tb00994.x. (Visited on 03/07/2025).

[153] N. A. Diamantidis, D. Karlis, and E. A. Giakoumakis. "Unsupervised Stratification of Cross-Validation for Accuracy Estimation". In: *Artificial Intelligence* 116.1 (Jan. 1, 2000), pp. 1–16. ISSN: 0004-3702. DOI: 10.1016/S0004-3702(99)00094-6. (Visited on 03/15/2025).

[154] Michael Yeung et al. "Unified Focal Loss: Generalising Dice and Cross Entropy-Based Losses to Handle Class Imbalanced Medical Image Segmentation". In: *Computerized Medical Imaging and Graphics* 95 (Jan. 1, 2022), p. 102026. ISSN: 0895-6111. DOI: 10.1016/j.compmedimag.2021.102026. (Visited on 01/28/2025).

[155] Carole H. Sudre et al. "Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations". In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Ed. by M. Jorge Cardoso et al. Cham: Springer International Publishing, 2017, pp. 240–248. ISBN: 978-3-319-67558-9. DOI: 10.1007/978-3-319-67558-9_28.

[156] Lutz Prechelt. "Automatic Early Stopping Using Cross Validation: Quantifying the Criteria". In: *Neural Networks* 11.4 (June 1, 1998), pp. 761–767. ISSN: 0893-6080. DOI: 10.1016/S0893-6080(98)00010-0. (Visited on 10/28/2024).

[157] Diego Vidaurre, Concha Bielza, and Pedro Larrañaga. "A Survey of L1 Regression". In: *International Statistical Review* 81.3 (2013), pp. 361–387. ISSN: 1751-5823. DOI: 10.1111/insr.12023. (Visited on 03/07/2025).

[158] ZongBen Xu et al. "L1/2 Regularization". In: *Science China Information Sciences* 53.6 (June 1, 2010), pp. 1159–1169. ISSN: 1869-1919. DOI: 10.1007/s11432-010-0090-0. (Visited on 03/07/2025).

[159] Geoffrey E. Hinton et al. *Improving Neural Networks by Preventing Co-Adaptation of Feature Detectors*. July 3, 2012. DOI: 10.48550/arXiv.1207.0580. arXiv: 1207.0580 [cs]. (Visited on 03/07/2025). Pre-published.

[160] Lutz Prechelt. "Early Stopping - But When?" In: *Neural Networks: Tricks of the Trade*. Ed. by Genevieve B. Orr and Klaus-Robert Müller. Red. by Gerhard Goos, Juris Hartmanis, and Jan Van Leeuwen. Vol. 1524. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, pp. 55–69. ISBN: 978-3-540-65311-0 978-3-540-49430-0. DOI: 10.1007/3-540-49430-8_3. (Visited on 03/07/2025).

[161] Christian Darken, Joseph Chang, and John Moody. "Learning Rate Schedules for Faster Stochastic Gradient Search". In: *Neural Networks for Signal Processing*. Vol. 2. Citeseer, 1992, pp. 3–12. URL: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=9db554243d7588589569aea127d676c9644d069a (visited on 03/07/2025).

[162] Andrew Senior et al. "An Empirical Study of Learning Rates in Deep Neural Networks for Speech Recognition". In: *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 6724–6728. URL: https://ieeexplore.ieee.org/abstract/document/6638963 (visited on 03/15/2025).

[163] Sebastian Thrun and Lorien Pratt. "Learning to Learn: Introduction and Overview". In: *Learning to Learn*. Ed. by Sebastian Thrun and Lorien Pratt. Boston, MA: Springer US, 1998, pp. 3–17. ISBN: 978-1-4615-5529-2. DOI: 10.1007/978-1-4615-5529-2_1. (Visited on 01/30/2025).

[164] Stevo Bozinovski and Ante Fulgosi. "The Influence of Pattern Similarity and Transfer Learning upon Training of a Base Perceptron B2". In: *Proceedings of Symposium Informatica*. Vol. 3. 1976, pp. 121–126.

[165] Rémi Bardenet et al. "Collaborative Hyperparameter Tuning". In: *Proceedings of the 30th International Conference on Machine Learning*. International Conference on Machine Learning. PMLR, May 13, 2013, pp. 199–207. URL: https://proceedings.mlr.press/v28/bardenet13.html (visited on 10/25/2024).

[166] Matthias Feurer et al. *Practical Transfer Learning for Bayesian Optimization*. Oct. 24, 2022. DOI: 10.48550/arXiv.1802.02219. arXiv: 1802.02219. (Visited on 10/25/2024). Pre-published.

[167] Martin Wistuba, Nicolas Schilling, and Lars Schmidt-Thieme. "Scalable Gaussian Process-Based Transfer Surrogates for Hyperparameter Optimization". In: *Machine Learning* 107.1 (Jan. 2018), pp. 43–78. ISSN: 0885-6125, 1573-0565. DOI: 10.1007/s10994-017-5684-y. (Visited on 10/25/2024).

[168] James Bergstra and Yoshua Bengio. "Random Search for Hyper-Parameter Optimization". In: ().

[169] Sigrún Andradóttir. "A Review of Random Search Methods". In: *Handbook of Simulation Optimization*. Ed. by Michael C Fu. Vol. 216. New York, NY: Springer New York, 2015, pp. 277–292. ISBN: 978-1-4939-1383-1 978-1-4939-1384-8. DOI: 10.1007/978-1-4939-1384-8_10. (Visited on 03/07/2025).

[170] James Bergstra, Daniel Yamins, and David Cox. "Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures". In: *International Conference on Machine Learning*. PMLR, 2013, pp. 115–123. URL: https://proceedings.mlr.press/v28/bergstra13.html (visited on 03/07/2025).

[171] William R. Crum, Oscar Camara, and Derek LG Hill. "Generalized Overlap Measures for Evaluation and Validation in Medical Image Analysis". In: *IEEE transactions on medical imaging* 25.11 (2006), pp. 1451–1461. URL: https://ieeexplore.ieee.org/abstract/document/1717643 (visited on 03/06/2025).

[172] Aaron Carass et al. "Evaluating White Matter Lesion Segmentations with Refined Sørensen-Dice Analysis". In: *Scientific Reports* 10.1 (May 19, 2020), p. 8242. ISSN: 2045-2322. DOI: 10.1038/s41598-020-64803-w. (Visited on 10/24/2024).

[173] J.C. Nascimento and J.S. Marques. "Performance Evaluation of Object Detection Algorithms for Video Surveillance". In: *IEEE Transactions on Multimedia* 8.4 (Aug. 2006), pp. 761–774. ISSN: 1520-9210. DOI: 10.1109/TMM.2006.876287. (Visited on 10/24/2024).

[174] Cedric Scheerlinck et al. "Fast Image Reconstruction with an Event Camera". In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2020, pp. 156–163. URL: https://openaccess.thecvf.com/content_WACV_2020/html/Scheerlinck_Fast_Image_Reconstruction_with_an_Event_Camera_WACV_2020_paper.html (visited on 03/15/2025).

[175] Younghyun Jo, Sejong Yang, and Seon Joo Kim. "Investigating Loss Functions for Extreme Super-Resolution". In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020, pp. 424–425. URL: https://openaccess.thecvf.com/content_CVPRW_2020/html/w31/Jo_Investigating_Loss_Functions_for_Extreme_Super-Resolution_CVPRW_2020_paper.html (visited on 03/15/2025).

[176] M. Hess et al. "Use of the Extended BBCH Scale - General for the Descriptions of the Growth Stages of Mono; and Dicotyledonous Weed Species". In: *Weed Research* 37.6 (1997), pp. 433–441. ISSN: 1365-3180. DOI: 10.1046/j.1365-3180.1997.d01-70.x. (Visited on 03/07/2025).

[177] Nikita Genze et al. "Improved Weed Segmentation in UAV Imagery of Sorghum Fields with a Combined Deblurring Segmentation Model". In: *Plant Methods* 19.1 (Aug. 22, 2023), p. 87. ISSN: 1746-4811. DOI: 10.1186/s13007-023-01060-8. (Visited on 02/14/2025).

[178] Huan Wang, Luping Zhou, and Lei Wang. "Miss Detection vs. False Alarm: Adversarial Learning for Small Object Segmentation in Infrared Images". In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019, pp. 8509–8518. URL: https://openaccess.thecvf.com/content_ICCV_2019/html/Wang_Miss_Detection_vs._False_Alarm_Adversarial_Learning_for_Small_Object_ICCV_2019_paper.html (visited on 02/28/2025).

[179] Zhengeng Yang et al. "Small Object Augmentation of Urban Scenes for Real-Time Semantic Segmentation". In: *IEEE Transactions on Image Processing* 29 (2020), pp. 5175–5190. ISSN: 1941-0042. DOI: 10.1109/TIP.2020.2976856. (Visited on 02/28/2025).

[180] Ailong Ma et al. "FactSeg: Foreground Activation-Driven Small Object Semantic Segmentation in Large-Scale Remote Sensing Imagery". In: *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), pp. 1–16. ISSN: 1558-0644. DOI: 10.1109/TGRS.2021.3097148. (Visited on 01/30/2025).

[181] Yang Liu et al. "A Survey and Performance Evaluation of Deep Learning Methods for Small Object Detection". In: *Expert Systems with Applications* 172 (June 15, 2021), p. 114602. ISSN: 0957-4174. DOI: 10.1016/j.eswa.2021.114602. (Visited on 02/28/2025).

[182] Bartosz Krawczyk. "Learning from Imbalanced Data: Open Challenges and Future Directions". In: *Progress in Artificial Intelligence* 5.4 (Nov. 1, 2016), pp. 221–232. ISSN: 2192-6360. DOI: 10.1007/s13748-016-0094-0. (Visited on 02/12/2025).

[183] Guo Haixiang et al. "Learning from Class-Imbalanced Data: Review of Methods and Applications". In: *Expert Systems with Applications* 73 (May 1, 2017), pp. 220–239. ISSN: 0957-4174. DOI: 10.1016/j.eswa.2016.12.035. (Visited on 02/12/2025).

[184] Haibo He and Edwardo A. Garcia. "Learning from Imbalanced Data". In: *IEEE Transactions on Knowledge and Data Engineering* 21.9 (Sept. 2009), pp. 1263–1284. ISSN: 1558-2191. DOI: 10.1109/TKDE.2008.239. (Visited on 02/12/2025).

[185] Roweida Mohammed, Jumanah Rawashdeh, and Malak Abdullah. "Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results". In: *2020 11th International Conference on Information and Communication Systems (ICICS)*. 2020 11th International Conference on Information and Communication Systems (ICICS). Apr. 2020, pp. 243–248. DOI: 10.1109/ICICS49469.2020.239556. (Visited on 02/12/2025).

[186] Seyda Ertekin, Jian Huang, and C. Lee Giles. "Active Learning for Class Imbalance Problem". In: *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.* SIGIR07: The 30th Annual International SIGIR Conference. Amsterdam The Netherlands: ACM, July 23, 2007, pp. 823–824. ISBN: 978-1-59593-597-7. DOI: 10.1145/1277741.1277927. (Visited on 02/12/2025).

[187] Hoel Kervadec et al. "Boundary Loss for Highly Unbalanced Segmentation". In: *Medical Image Analysis* 67 (Jan. 1, 2021), p. 101851. ISSN: 1361-8415. DOI: 10.1016/j.media.2020.101851. (Visited on 01/28/2025).

[188] T.-YLPG Ross and GKHP Dollár. "Focal Loss for Dense Object Detection". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 2017, pp. 2980–2988. URL: http://openaccess.thecvf.com/content_ICCV_2017/papers/Lin_Focal_Loss_for_ICCV_2017_paper (visited on 02/12/2025).

[189] Alexey Bokhovkin and Evgeny Burnaev. "Boundary Loss for Remote Sensing Imagery Semantic Segmentation". In: *Advances in Neural Networks - ISNN 2019.* Ed. by Huchuan Lu, Huajin Tang, and Zhanshan Wang. Vol. 11555. Cham: Springer International Publishing, 2019, pp. 388–401. ISBN: 978-3-030-22807-1 978-3-030-22808-8. DOI: 10.1007/978-3-030-22808-8_38. (Visited on 02/12/2025).

[190] Adrian Celaya, Beatrice Riviere, and David Fuentes. *A Generalized Surface Loss for Reducing the Hausdorff Distance in Medical Imaging Segmentation.* Jan. 24, 2024. DOI: 10.48550/arXiv.2302.03868. arXiv: 2302.03868 [eess]. (Visited on 02/12/2025). Pre-published.

[191] Da Li et al. "Deeper, Broader and Artier Domain Generalization". In: Proceedings of the IEEE International Conference on Computer Vision. 2017, pp. 5542–5550. URL: https://openaccess.thecvf.com/content_iccv_2017/html/Li_Deeper_Broader_and_ICCV_2017_paper.html (visited on 03/15/2025).

[192] Jooyoung Moon et al. "Confidence-Aware Learning for Deep Neural Networks". In: *Proceedings of the 37th International Conference on Machine Learning.* International Conference on Machine Learning. PMLR, Nov. 21, 2020, pp. 7034–7044. URL: https://proceedings.mlr.press/v119/moon20a.html (visited on 03/03/2025).

[193] Christian Szegedy et al. *Intriguing Properties of Neural Networks.* Feb. 19, 2014. DOI: 10.48550/arXiv.1312.6199. arXiv: 1312.6199 [cs]. (Visited on 03/03/2025). Pre-published.

[194] Anh Nguyen, Jason Yosinski, and Jeff Clune. "Deep Neural Networks Are Easily Fooled: High Confidence Predictions for Unrecognizable Images". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 2015, pp. 427–436. (Visited on 03/03/2025).

[195] Davood Karimi and Ali Gholipour. "Improving Calibration and Out-of-Distribution Detection in Deep Models for Medical Image Segmentation". In: *IEEE Transactions on Artificial Intelligence* 4.2 (Apr. 2023), pp. 383–397. ISSN: 2691-4581. DOI: 10.1109/TAI.2022.3159510. (Visited on 03/03/2025).

[196] Dokwan Oh et al. "Boosting Out-of-Distribution Image Detection with Epistemic Uncertainty". In: *IEEE Access* 10 (2022), pp. 109289–109298. URL: https://ieeexplore.ieee.org/abstract/document/9915595/ (visited on 03/12/2025).

[197] Bartłomiej Olber et al. "Detection of Out-of-Distribution Samples Using Binary Neuron Activation Patterns". In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023, pp. 3378–3387. (Visited on 03/15/2025).

[198] Christoph Berger et al. "Confidence-Based Out-of-Distribution Detection: A Comparative Study and Analysis". In: *Uncertainty for Safe Utilization of Machine Learning in Medical Imaging, and Perinatal Imaging, Placental and Preterm Image Analysis*. Ed. by Carole H. Sudre et al. Cham: Springer International Publishing, 2021, pp. 122–132. ISBN: 978-3-030-87735-4. DOI: 10.1007/978-3-030-87735-4_12.

[199] Mingyang Yi et al. "Improved Ood Generalization via Adversarial Training and Pretraing". In: *International Conference on Machine Learning*. PMLR, 2021, pp. 11987–11997. URL: https://proceedings.mlr.press/v139/yi21a.html (visited on 03/15/2025).

[200] Ting Chen et al. "A Simple Framework for Contrastive Learning of Visual Representations". In: *Proceedings of the 37th International Conference on Machine Learning*. Vol. 119. ICML'20. JMLR.org, July 13, 2020, pp. 1597–1607.

[201] Jingkang Yang et al. "Generalized Out-of-Distribution Detection: A Survey". In: *International Journal of Computer Vision* 132.12 (Dec. 1, 2024), pp. 5635–5662. ISSN: 1573-1405. DOI: 10.1007/s11263-024-02117-4. (Visited on 03/12/2025).

[202] Guy Hacohen and Daphna Weinshall. "On The Power of Curriculum Learning in Training Deep Networks". In: *Proceedings of the 36th International Conference on Machine Learning*. International Conference on Machine Learning. PMLR, May 24, 2019, pp. 2535–2544. URL: https://proceedings.mlr.press/v97/hacohen19a.html (visited on 02/12/2025).

[203] Chaeyeong Yun et al. "WRA-Net: Wide Receptive Field Attention Network for Motion Deblurring in Crop and Weed Image". In: *Plant Phenomics* 5 (Apr. 5, 2023), p. 0031. DOI: 10.34133/plantphenomics.0031. (Visited on 03/12/2025).

[204] Bo Zhao et al. "A Survey on Deep Learning-Based Fine-Grained Object Classification and Semantic Segmentation". In: *International Journal of Automation and Computing* 14.2 (Apr. 2017), pp. 119–135. ISSN: 1476-8186, 1751-8520. DOI: 10.1007/s11633-017-1053-3. (Visited on 03/12/2025).

[205] Uduak Idio Akpan and Andrew Starkey. “Review of Classification Algorithms with Changing Inter-Class Distances”. In: *Machine Learning with Applications* 4 (June 15, 2021), p. 100031. ISSN: 2666-8270. DOI: 10.1016/j.mlwa.2021.100031. (Visited on 03/12/2025).

[206] Beibei Xu et al. “Instance Segmentation Method for Weed Detection Using UAV Imagery in Soybean Fields”. In: *Computers and Electronics in Agriculture* 211 (Aug. 1, 2023), p. 107994. ISSN: 0168-1699. DOI: 10.1016/j.compag.2023.107994. (Visited on 03/15/2025).

[207] Rafał Pilarczyk and Władysław Skarbek. “On Intra-Class Variance for Deep Learning of Classifiers”. In: *Foundations of Computing and Decision Sciences* 44.3 (Sept. 1, 2019), pp. 285–301. ISSN: 2300-3405. DOI: 10.2478/fcds-2019-0015. (Visited on 03/10/2025).

[208] Zeeshan Ahmad et al. “Weed Species Composition and Distribution Pattern in the Maize Crop under the Influence of Edaphic Factors and Farming Practices: A Case Study from Mardan, Pakistan”. In: *Saudi Journal of Biological Sciences* 23.6 (Nov. 1, 2016), pp. 741–748. ISSN: 1319-562X. DOI: 10.1016/j.sjbs.2016.07.001. (Visited on 03/13/2025).

[209] Brandon T. Bestelmeyer et al. “Scaling Up Agricultural Research With Artificial Intelligence”. In: *IT Professional* 22.3 (May 2020), pp. 33–38. ISSN: 1941-045X. DOI: 10.1109/MITP.2020.2986062. (Visited on 01/29/2025).

[210] Liu Li et al. “Attention Based Glaucoma Detection: A Large-Scale Database and CNN Model”. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019, pp. 10571–10580. (Visited on 03/12/2025).

[211] Alexey Dosovitskiy et al. “An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale”. In: International Conference on Learning Representations. Oct. 2, 2020. URL: https://openreview.net/forum?id=YicbFdNTTy (visited on 03/12/2025).

[212] Shai Ben-David et al. “A Theory of Learning from Different Domains”. In: *Machine Learning* 79.1–2 (May 2010), pp. 151–175. ISSN: 0885-6125, 1573-0565. DOI: 10.1007/s10994-009-5152-4. (Visited on 03/04/2025).

[213] Gabriela Csurka, ed. *Domain Adaptation in Computer Vision Applications*. Advances in Computer Vision and Pattern Recognition. Cham: Springer International Publishing, 2017. ISBN: 978-3-319-58346-4 978-3-319-58347-1. DOI: 10.1007/978-3-319-58347-1. (Visited on 03/04/2025).

[214] Poojan Oza et al. “Unsupervised Domain Adaptation of Object Detectors: A Survey”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46.6 (June 2024), pp. 4018–4040. ISSN: 1939-3539. DOI: 10.1109/TPAMI.2022.3217046. (Visited on 03/04/2025).

[215] Riccardo Bertoglio et al. "A Comparative Study of Fourier Transform and CycleGAN as Domain Adaptation Techniques for Weed Segmentation". In: *Smart Agricultural Technology* 4 (Aug. 1, 2023), p. 100188. ISSN: 2772-3755. DOI: 10.1016/j.atech.2023.100188. (Visited on 01/30/2025).

[216] Yingchao Huang and Abdul Bais. "Unsupervised Domain Adaptation for Weed Segmentation Using Greedy Pseudo-labelling". In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024, pp. 2484–2494. (Visited on 01/30/2025).

[217] Eric Tzeng et al. "Adversarial Discriminative Domain Adaptation". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, pp. 7167–7176. (Visited on 03/15/2025).

[218] Chaoqi Chen et al. "Progressive Feature Alignment for Unsupervised Domain Adaptation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 627–636. (Visited on 03/15/2025).

[219] Hao-Wei Yeh et al. "Sofa: Source-data-free Feature Alignment for Unsupervised Domain Adaptation". In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2021, pp. 474–483. (Visited on 03/15/2025).

[220] Barret Zoph et al. "Rethinking Pre-training and Self-training". In: *Advances in Neural Information Processing Systems*. Vol. 33. Curran Associates, Inc., 2020, pp. 3833–3845. (Visited on 02/13/2025).

[221] Kaiming He et al. "Momentum Contrast for Unsupervised Visual Representation Learning". In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020, pp. 9729–9738. (Visited on 02/13/2025).

[222] Tushar Kataria, Beatrice Knudsen, and Shireen Elhabian. "To Pretrain or Not to Pretrain? A Case Study of Domain-Specific Pretraining for Semantic Segmentation in Histopathology". In: *Medical Image Learning with Limited and Noisy Data*. Ed. by Zhiyun Xue et al. Cham: Springer Nature Switzerland, 2023, pp. 246–256. ISBN: 978-3-031-44917-8. DOI: 10.1007/978-3-031-44917-8_24.

[223] Yixuan Qiu et al. "Pre-Training in Medical Data: A Survey". In: *Machine Intelligence Research* 20.2 (Apr. 2023), pp. 147–179. ISSN: 2731-538X, 2731-5398. DOI: 10.1007/s11633-022-1382-8. (Visited on 02/13/2025).

[224] Di Wang et al. "An Empirical Study of Remote Sensing Pretraining". In: *IEEE Transactions on Geoscience and Remote Sensing* 61 (2023), pp. 1–20. ISSN: 1558-0644. DOI: 10.1109/TGRS.2022.3176603. (Visited on 02/13/2025).

[225] M. Victoria Lantschner, Gerardo de la Vega, and Juan C. Corley. "Predicting the Distribution of Harmful Species and Their Natural Enemies in Agricultural, Livestock and Forestry Systems: An Overview". In: *International Journal of Pest Management* 65.3 (July 3, 2019), pp. 190–206. ISSN: 0967-0874. DOI: 10.1080/09670874.2018.1533664. (Visited on 02/12/2025).

[226] Javier López-Tirado and Jose L. Gonzalez-Andújar. "Spatial Weed Distribution Models under Climate Change: A Short Review". In: *PeerJ* 11 (Apr. 10, 2023), e15220. ISSN: 2167-8359. DOI: 10.7717/peerj.15220. (Visited on 02/12/2025).

[227] Ke Sun et al. "Deep High-Resolution Representation Learning for Human Pose Estimation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 2019, pp. 5693–5703. (Visited on 03/13/2025).

[228] Huisi Wu et al. "Optimized HRNet for Image Semantic Segmentation". In: *Expert Systems with Applications* 174 (July 15, 2021), p. 114532. ISSN: 0957-4174. DOI: 10.1016/j.eswa.2020.114532. (Visited on 03/13/2025).

[229] Fatih Cagatay Akyon, Sinan Onur Altinuc, and Alptekin Temizel. "Slicing Aided Hyper Inference and Fine-Tuning for Small Object Detection". In: *2022 IEEE International Conference on Image Processing (ICIP).* 2022 IEEE International Conference on Image Processing (ICIP). Oct. 2022, pp. 966–970. DOI: 10.1109/ICIP46576.2022.9897990. (Visited on 03/13/2025).

[230] Jia Deng et al. "Imagenet: A Large-Scale Hierarchical Image Database". In: *2009 IEEE Conference on Computer Vision and Pattern Recognition.* Ieee, 2009, pp. 248–255. URL: https://ieeexplore.ieee.org/abstract/document/5206848/ (visited on 03/13/2025).

[231] Jason Yosinski et al. "How Transferable Are Features in Deep Neural Networks?" In: *Advances in Neural Information Processing Systems.* Vol. 27. Curran Associates, Inc., 2014. (Visited on 02/13/2025).

[232] Ricardo Ribani and Mauricio Marengoni. "A Survey of Transfer Learning for Convolutional Neural Networks". In: *2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T).* 2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T). Oct. 2019, pp. 47–57. DOI: 10.1109/SIBGRAPI-T.2019.00010. (Visited on 01/30/2025).

[233] Ross Girshick et al. "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation". In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014, pp. 580–587. (Visited on 02/13/2025).

[234] D. Matthew Zeiler and Fergus Rob. "Visualizing and Understanding Convolutional Neural Networks". In: *Proceedings of the European Conference on Computer Vision, Zurich, Switzerland.* 2014, pp. 6–12.

[235] W. Schickier and Anthony Thorpe. "Operational Procedure for Automatic True Orthophoto Generation". In: *International Archives of Photogrammetry and Remote Sensing* 32 (1998), pp. 527–532. (Visited on 03/13/2025).

[236] Ayman F. Habib, Eui-Myoung Kim, and Chang-Jae Kim. "New Methodologies for True Orthophoto Generation". In: *Photogrammetric Engineering & Remote Sensing* 73.1 (2007), pp. 25–36. (Visited on 03/13/2025).

[237] Joseph L. Awange and John B. Kyalo Kiema. “Fundamentals of Photogrammetry”. In: *Environmental Geoinformatics: Monitoring and Management*. Ed. by Joseph L. Awange and John B. Kyalo Kiema. Berlin, Heidelberg: Springer, 2013, pp. 157–174. ISBN: 978-3-642-34085-7. DOI: 10.1007/978-3-642-34085-7_11. (Visited on 03/13/2025).

[238] Norman J. W. Thrower and John R. Jensen. “The Orthophoto and Orthophotomap: Characteristics, Development and Application”. In: *The American Cartographer* 3.1 (Jan. 1976), pp. 39–56. DOI: 10.1559/152304076784080249. (Visited on 03/13/2025).

[239] A. P. Barnes et al. “Influencing Incentives for Precision Agricultural Technologies within European Arable Farming Systems”. In: *Environmental Science & Policy* 93 (Mar. 1, 2019), pp. 66–74. ISSN: 1462-9011. DOI: 10.1016/j.envsci.2018.12.014. (Visited on 02/03/2025).

[240] Sharon A. Clay and J. Anita Dille. “Precision Weed Management”. In: *Women in Precision Agriculture: Technological Breakthroughs, Challenges and Aspirations for a Prosperous and Sustainable Future*. Ed. by Takoi Khemais Hamrita. Cham: Springer International Publishing, 2021, pp. 85–106. ISBN: 978-3-030-49244-1. DOI: 10.1007/978-3-030-49244-1_5. (Visited on 02/03/2025).

[241] Scott M. Swinton. “Economics of Site-Specific Weed Management”. In: *Weed Science* 53.2 (2005), pp. 259–263. ISSN: 0043-1745. JSTOR: 4046966. URL: https://www.jstor.org/stable/4046966 (visited on 02/03/2025).

[242] David R. Shaw. “Remote Sensing and Site-Specific Weed Management”. In: *Frontiers in Ecology and the Environment* 3.10 (2005), pp. 526–532. ISSN: 1540-9295. DOI: 10.2307/3868608. JSTOR: 3868608. (Visited on 02/03/2025).

[243] Roland Gerhards et al. “Advances in Site-Specific Weed Management in Agriculture - A Review”. In: *Weed Research* 62.2 (2022), pp. 123–133. ISSN: 1365-3180. DOI: 10.1111/wre.12526. (Visited on 02/03/2025).

[244] Joseph Redmon et al. “You Only Look Once: Unified, Real-Time Object Detection”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 779–788. (Visited on 03/14/2025).

[245] Luiz Carlos M. Junior and José Alfredo C. Ulson. “Real Time Weed Detection Using Computer Vision and Deep Learning”. In: *2021 14th IEEE International Conference on Industry Applications (INDUSCON)*. 2021 14th IEEE International Conference on Industry Applications (INDUSCON). Aug. 2021, pp. 1131–1137. DOI: 10.1109/INDUSCON51756.2021.9529761. (Visited on 03/14/2025).

[246] Ch Lakshmi Narayana and Kondapalli Venkata Ramana. “An Efficient Real-Time Weed Detection Technique Using YOLOv7”. In: *International Journal of Advanced Computer Science and Applications (IJACSA)* 14.2 (2 Dec. 28, 2023). ISSN: 2156-5570. DOI: 10.14569/IJACSA.2023.0140265. (Visited on 03/14/2025).

[247] Fengying Dang et al. "YOLOWeeds: A Novel Benchmark of YOLO Object Detectors for Multi-Class Weed Detection in Cotton Production Systems". In: *Computers and Electronics in Agriculture* 205 (Feb. 1, 2023), p. 107655. ISSN: 0168-1699. DOI: 10.1016/j.compag.2023.107655. (Visited on 03/14/2025).

[248] Rui Hu et al. "Real-Time Lettuce-Weed Localization and Weed Severity Classification Based on Lightweight YOLO Convolutional Neural Networks for Intelligent Intra-Row Weed Control". In: *Computers and Electronics in Agriculture* 226 (Nov. 1, 2024), p. 109404. ISSN: 0168-1699. DOI: 10.1016/j.compag.2024.109404. (Visited on 03/14/2025).

[249] Qifan Wang et al. "A Deep Learning Approach Incorporating YOLO v5 and Attention Mechanisms for Field Real-Time Detection of the Invasive Weed *Solanum Rostratum* Dunal Seedlings". In: *Computers and Electronics in Agriculture* 199 (Aug. 1, 2022), p. 107194. ISSN: 0168-1699. DOI: 10.1016/j.compag.2022.107194. (Visited on 03/14/2025).

[250] Yeong Sheng Tey and Mark Brindal. "Factors Influencing the Adoption of Precision Agricultural Technologies: A Review for Policy Implications". In: *Precision Agriculture* 13.6 (Dec. 1, 2012), pp. 713–730. ISSN: 1573-1618. DOI: 10.1007/s11119-012-9273-6. (Visited on 02/03/2025).

[251] Margit Paustian and Ludwig Theuvsen. "Adoption of Precision Agriculture Technologies by German Crop Farmers". In: *Precision Agriculture* 18.5 (Oct. 1, 2017), pp. 701–716. ISSN: 1573-1618. DOI: 10.1007/s11119-016-9482-5. (Visited on 02/03/2025).

[252] Isabel Cisternas et al. "Systematic Literature Review of Implementations of Precision Agriculture". In: *Computers and Electronics in Agriculture* 176 (Sept. 1, 2020), p. 105626. ISSN: 0168-1699. DOI: 10.1016/j.compag.2020.105626. (Visited on 02/03/2025).

[253] Mark Ryan, Gohar Isakhanyan, and Bedir Tekinerdogan. "An Interdisciplinary Approach to Artificial Intelligence in Agriculture". In: *NJAS: Impact in Agricultural and Life Sciences* 95.1 (Dec. 31, 2023), p. 2168568. ISSN: 2768-5241. DOI: 10.1080/27685241.2023.2168568. (Visited on 02/25/2025).

[254] Sai Sree Laya Chukkapalli et al. "Ontologies and Artificial Intelligence Systems for the Cooperative Smart Farming Ecosystem". In: *IEEE Access* 8 (2020), pp. 164045–164064. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2020.3022763. (Visited on 02/25/2025).

[255] M. J. Robertson et al. "Adoption of Variable Rate Fertiliser Application in the Australian Grains Industry: Status, Issues and Prospects". In: *Precision Agriculture* 13.2 (Apr. 1, 2012), pp. 181–199. ISSN: 1573-1618. DOI: 10.1007/s11119-011-9236-3. (Visited on 02/05/2025).

[256] Carrie S. Alexander, Mark Yarborough, and Aaron Smith. "Who Is Responsible for 'Responsible AI'?: Navigating Challenges to Build Trust in AI Agriculture and Food System Technology". In: *Precision Agriculture* 25.1 (Feb. 1, 2024), pp. 146–185. ISSN: 1573-1618. DOI: `10.1007/s11119-023-10063-3`. (Visited on 02/25/2025).

[257] Rayda Ben Ayed and Mohsen Hanana. "Artificial Intelligence to Improve the Food and Agriculture Sector". In: *Journal of Food Quality* 2021 (Apr. 22, 2021). Ed. by Rijwan Khan, pp. 1–7. ISSN: 1745-4557, 0146-9428. DOI: `10.1155/2021/5584754`. (Visited on 02/25/2025).

[258] Jussi Nikander, Onni Manninen, and Mikko Laajalahti. "Requirements for Cybersecurity in Agricultural Communication Networks". In: *Computers and electronics in agriculture* 179 (2020), p. 105776. URL: `https://www.sciencedirect.com/science/article/pii/S0168169920314812` (visited on 02/25/2025).

[259] Willard Downs and Kelley Ann Newton. "Legal Implications in Development and Use of Expert Systems in Agriculture". In: *Journal of agricultural ethics* 2.1 (Mar. 1, 1989), pp. 53–58. ISSN: 1573-322X. DOI: `10.1007/BF01831548`. (Visited on 02/10/2025).

[260] Rozita Dara, Seyed Mehdi Hazrati Fard, and Jasmin Kaur. "Recommendations for Ethical and Responsible Use of Artificial Intelligence in Digital Agriculture". In: *Frontiers in Artificial Intelligence* 5 (July 29, 2022). ISSN: 2624-8212. DOI: `10.3389/frai.2022.884192`. (Visited on 02/10/2025).

[261] Mrinalini Kochupillai and Julia Köninger. "Creating a Digital Marketplace for Agrobiodiversity and Plant Genetic Sequence Data: Legal and Ethical Considerations of an Ai and Blockchain Based Solution". In: *Towards Responsible Plant Data Linkage: Data Challenges for Agricultural Research and Development* (2023), p. 223. (Visited on 02/25/2025).

[262] Asaf Tzachor et al. "Responsible Artificial Intelligence in Agriculture Requires Systemic Understanding of Risks and Externalities". In: *Nature Machine Intelligence* 4.2 (2022), pp. 104–109. URL: `https://www.nature.com/articles/s42256-022-00440-4` (visited on 02/25/2025).

[263] Ismaguil Hanadé Houmma et al. "Modelling Agricultural Drought: A Review of Latest Advances in Big Data Technologies". In: *Geomatics, Natural Hazards and Risk* 13.1 (Dec. 31, 2022), pp. 2737–2776. ISSN: 1947-5705, 1947-5713. DOI: `10.1080/19475705.2022.2131471`. (Visited on 02/25/2025).

[264] Remco Schrijver. *Precision Agriculture and the Future of Farming in Europe: Scientific Foresight Study: Study*. European Parliament, 2016.