An Advanced Dirichlet Prior Network for Out-of-Distribution Detection in Remote Sensing

Jakob Gawlikowski, *Student Member, IEEE*, Sudipan Saha, *Member, IEEE*, Anna Kruspe, and Xiao Xiang Zhu, *Fellow, IEEE*

***Abstract*— Remote sensing deals with a plethora of sensors, a large number of classes/categories, and a huge variation in geography. Due to the difficulty of collecting labeled data uniformly representing all scenarios, data-hungry deep learning models are often trained with labeled data in a source domain that is limited in the above-mentioned aspects. However, during the test/inference phase, such deep learning models are often subjected to a distributional shift, also called out-of-distribution (OOD) samples, in the form of unseen classes, geographic differences, and multisensor differences. Deep learning models can behave in an unexpected manner when subjected to such distributional uncertainties. Vulnerability to OOD data severely reduces the reliability of deep learning models and trusting on such predictions in the absence of any reliability indicator may lead to wrong policy decisions or mishaps in time-bound remote sensing applications. Motivated by this, in this work, we propose a Dirichlet prior network-based model to quantify the distributional uncertainty of deep learning-based remote sensing models. The approach seeks to maximize the representation gap between the in-domain and OOD examples for better segregation of OOD samples at test time. Extensive experiments on several remote sensing image classification datasets demonstrate that the proposed model can quantify distributional uncertainty. To the best of our knowledge, this is the first work to elaborately study distributional uncertainty in context of remote sensing. The codes are publicly available at https://gitlab.lrz.de/ai4eo/Uncertainty/-/tree/main/DPN-RS.**

***Index Terms*— Distributional uncertainty, open-set recognition, out-of-distribution (OOD), reliability, remote sensing, robustness, uncertainty.**

Manuscript received April 19, 2021; revised October 3, 2021; accepted November 24, 2021. Date of publication January 4, 2022; date of current version March 8, 2022. This work was supported in part by the German Federal Ministry of Education and Research (BMBF) in the Framework of the International Future AI Lab "AI4EO—Artificial Intelligence for Earth Observation: Reasoning, Uncertainties, Ethics and Beyond" under Grant 01DD20001; in part by the German Federal Ministry of Economics and Technology in the Framework of the "National Center of Excellence ML4Earth" under Grant 50EE2201C; in part by the Helmholtz Association through the Framework of Helmholtz AI (Local Unit "Munich Unit @Aeronautics, Space and Transport (MASTr)") under Grant ZT-I-PF-5-01; and in part by the Helmholtz Excellent Professorship "Data Science in Earth Observation—Big Data Fusion for Urban Research" under Grant W2-W3-100. *(Corresponding author: Xiao Xiang Zhu.)*

Jakob Gawlikowski is with the Institute of Data Science, German Aerospace Center (DLR), 07745 Jena, Germany, and also with the Data Science in Earth Observation, Technical University of Munich, 85521 Ottobrunn, Germany (e-mail: jakob.gawlikowski@dlr.de).

Sudipan Saha and Anna Kruspe are with the Data Science in Earth Observation, Technical University of Munich, 85521 Ottobrunn, Germany (e-mail: sudipan.saha@tum.de; anna.kruspe@tum.de).

Xiao Xiang Zhu is with the Remote Sensing Technology Institute, German Aerospace Center (DLR), 82234 Weßling, Germany, and also with the Data Science in Earth Observation, Technical University of Munich, 85521 Ottobrunn, Germany (e-mail: xiaoxiang.zhu@dlr.de).

Digital Object Identifier 10.1109/TGRS.2022.3140324

I. Introduction

DEEP learning has revolutionized the field of remote sensing in the last few years [1] and has been successfully applied in various remote sensing tasks, including classification [2], [3], hyperspectral image analysis [4]–[6], semantic segmentation [7]–[9], change detection [10]–[12], image retrieval [13], [14], target detection [15], [16], disaster management [17], [18], cloud detection and removal [19]–[21], and image fusion [22]–[24]. Most of these methods assume that the model is trained on a dataset that has similar geographical characteristics as the target area [25], i.e., the source data distribution is the same as the target data distribution. Moreover, they assume that the source and the target data have an identical set of classes. However, in practice, remote sensing deals with a large number of sensors, which operates across a significant variation of geographies and a large number of classes [26]. Considering this variation, the above assumptions often do not hold in remote sensing. There are a few works related to domain adaptation [27]–[29] that try to align the target distribution with the source distribution. However, such methods are only effective when the domain shift between the source and the target is small. Moreover, they do not consider the presence of unseen/open-set classes. Deep learning models are likely to fail or behave in an unexpected way when faced with open-set classes, e.g., when a deep model trained on images from forest area is applied on urban images consisting of residential complexes and parking lots. Similarly, deep models behave in an unexpected way when they are fed with data from seen classes but with considerable geographic variation, e.g., when a model trained on European urban area (where skyscrapers are rare) is used to predict test images from Asian urban areas. When deep learning models fail, they do not provide sufficient clue to the user, having unforeseeable impact on remote sensing applications, especially in time-bound and safety-critical applications. As an example, we can consider the scenario of disaster management after an earthquake where unreliable predictions may lead the rescue team to the wrong site, at the expense of human lives. Nevertheless, unreliable predictions may also negatively impact nontime-bound applications, e.g., a building detection model trained for Europe and used unreliably on Asia/Africa may lead to incorrect estimations of the building density and thus impacts the subsequent policy decisions.

Toward designing reliable deep learning models that are aware of different sources of uncertainty, predictive uncertainty

estimation has recently emerged as a research topic in the machine learning community [30]. Uncertainty estimation informs users about the confidence on a prediction, thus improving the reliability of such systems. Deep learning-based classification models are prone to predictive uncertainties from three different sources [31], model (also known as epistemic) uncertainty, data (also known as aleatoric) uncertainty, and distributional uncertainty. Epistemic uncertainty stems from a model's lack of knowledge (e.g., limited training data, limited complexity, and errors in the training process), while aleatoric uncertainty arises from complexities related to data distribution (e.g., class overlap in data). Distributional uncertainty is related to the mismatch between the training and the test data and can be seen as a special case of model uncertainty [32]. In remote sensing, distributional uncertainty may arise due to various reasons, e.g., unseen classes, geographic differences, and sensor differences. Considering its high relevance in remote sensing, in this work, we focus on distributional uncertainty [31].

The key contributions of this article are as follows:

1) introducing the concept of out-of-distribution (OOD) detection in remote sensing,
2) proposing a Dirichlet prior network (DPN)-based model that can quantify distributional uncertainty in context of different remote sensing uncertainty sources,
3) extensively experimenting on large-scale remote sensing datasets for open-set recognition, sensor shift, and region shift,
4) providing a benchmark that can facilitate further research on remote sensing distributional uncertainty.

Extensive experiments demonstrated that the proposed approach is able to detect OOD examples in remote sensing images, thus improving the reliability and robustness of deep learning-based models. To the best of our knowledge, this is the first work that extensively addresses OOD detection in remote sensing.[1]

The rest of this article is organized as follows. We briefly discuss the related works in Section II. In Section III, we detail the proposed method, and in Section IV, the datasets, experiments, and results are presented. A critical discussion on different distributional uncertainties in context of our results is presented in Section V. We conclude this article and discuss scope of future research in Section VI.

II. Related Works

Uncertainty quantification gained attention of the remote sensing community even before the emergence of deep learning [33], [34]. Despite this, there are only a few works that explore distributional uncertainty for remote sensing and topics closely related to it [35]. In Section II-A, we briefly discuss them. We also briefly discuss different existing Bayesian paradigms in the machine learning literature to handle uncertainty (Section II-B). Our work is not in contrast with the domain adaptation literature, as explained in Section II-C.

[1]The code for this work is available under https://gitlab.lrz.de/ai4eo/Uncertainty/-/tree/main/DPN-RS

A. Detecting Distributional Shifts in Remote Sensing

One common form of distributional shift is the presence of new classes in the target data. This problem has also been dealt as open-set recognition; da Silva *et al.* [36] proposed a method for open-set aerial image segmentation. They assign a pixel with a class confidence (given by the soft max) that exceeds a threshold as belonging to that class. However, if the pixelwise probability is inferior to the threshold, the pixel is classified as open set. Dang *et al.* [37] proposed an open-set incremental learning-based method for target recognition by exploiting an extreme value theory (EVT). Wu *et al.* [38] introduced open-set recognition to hyperspectral image classification.

A few works identified that the models may likely fail if applied to new geographic locations considerably different from the training data [39], [40]. To quantify the area of applicability, Meyer and Pebesma [39] proposed a dissimilarity index based on the minimum distance to the training data in multidimensional predictor space. In [25], an applicable model is learned by using unlabeled data from each geography of interest.

Contrary to the previous works, our work tackles all forms of distributional shift (e.g., open set, spatial shift, and sensor shift) in the same framework. Moreover, on the contrary to previous works [36] that employ trivial solutions, our work is based on DPNs, a well-founded theoretical framework for uncertainty estimation. Our work is also a step forward toward building explainable remote sensing model [41], [42].

B. Bayesian Frameworks for Uncertainty

Bayesian frameworks are traditionally used to model the predictive uncertainty of a classifier. The sources of uncertainty [31] can be broadly categorized into the following three categories.

1) Epistemic uncertainty characterizes the uncertainty caused by the lack of knowledge of the network, caused, for example, by insufficient training data, a shortage of model capabilities, or an insufficient training process.
2) Aleatoric uncertainty arises from the complexity in the data distribution, e.g., class overlap and label noise. For example, data having different values in label space may have very similar representation in the feature domain.
3) Distributional uncertainty arises from a mismatch in the distribution of the training and the test data. Distributional uncertainty is likely in remote sensing due to differences caused by new classes in the target data, geographic shift, and multisensor differences.

Data uncertainty is in general modeled as a confidence prediction by the neural network itself, e.g. by a soft-max probability vector [32], [43]. Bayesian neural network-based approaches capture the model uncertainty by modeling the network parameters as probability variables. A posterior distribution over the parameters is derived based on the given training data and predictions are realized by sampling different sets of parameters from this posterior. Different ways of approximating such a posterior are available, e.g., Monte Carlo dropout [43], Laplace approximation [44], or deep ensembles [32]. However,

it is computationally very expensive to produce such ensembles, thus limiting the application of existing ensemble and Bayesian approaches in such scenario. DPN and its variants are introduced in [31] and [45] as an efficient adaptation of the Bayesian networks. Our work directly derives from [31] and [45], thus exploiting the benefits of Bayesian modeling while still being computationally efficient.

C. Position in Reference to Domain Adaptation

Domain adaptation [28], [46] is a branch of multidomain learning. A model trained on a source domain is modulated by domain adaptation techniques to be applied to another target domain. However, domain adaptation assumes that either a few labeled data samples or a large unlabeled dataset from the target domain is available during the training of the model. If the target domain data are completely unseen during the training, the most domain adaptation methods do not have the capability to mitigate differences between domains and may eventually produce unreliable predictions. Thus, it is important to be able to identify the test samples that are drawn from a distribution unseen during the training. This is where the OOD detection comes into play, further pushing forward the paradigm of multidomain learning.

III. Proposed Method

Remote sensing deals with a vast set of data types, varying in geography, climate conditions, sensor properties, end applications, and target classes. It is expensive, both in terms of time and effort, to collect labeled data uniformly representing all scenarios. Thus, most deep learning models are trained with limited training samples in a source domain that is limited in the above-mentioned aspects. During test/inference, when the model is fed with data that do not follow the source domain distribution, the model predicts in unexpected fashion. Our goal is to propose a framework that handles the above-mentioned sources of uncertainty in the same framework without any adjustment being made for different sources of uncertainty. Toward this, we adopt an efficient adaptation of the DPN approach [45]. The Dirichlet distribution is popularly used as a prior distribution in Bayesian learning [47]. Motivated by this, Malinin and Gales [31] proposed DPNs for an improved detection of OOD samples. DPNs are deterministic neural networks that efficiently mimic the behavior of Bayesian neural networks by parameterizing a Dirichlet distribution over the categorical distribution given by a soft-max classification output. Convenient to remote sensing applications, any neural network with a soft-max activation can be considered as a DPN. Following the idea of Malinin and Gales, several other DPN-based methods for OOD detection were developed [45], [48], [49]. In this work, we take inspiration from the Dirichlet distribution-based approaches and propose DPN-RS that transfers DPNs to remote sensing settings.

Section III-A briefly introduces the Dirichlet distribution. In Section III-B, we detail the DPNs, and we briefly discuss its suitability for remote sensing data in Section III-C. Finally, we present DPN-RS in Section III-D.

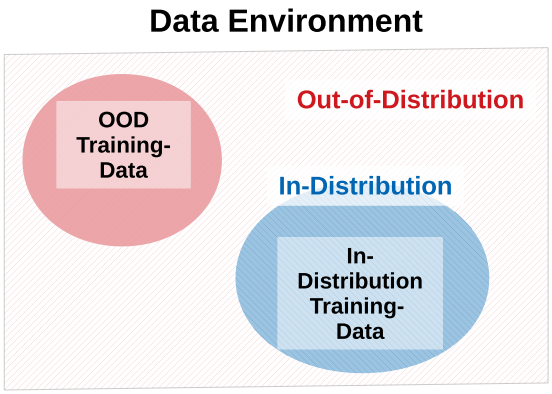

Fig. 1. Visualization of the data environment. The in-distribution represents the data on which the network is expected to deliver accurate predictions. OOD covers any other kind of data that are significantly different from the training data distribution. The OOD training dataset is used to train the network to handle OOD examples but can only cover a small portion of the OOD region. While trained, the network can handle any OOD data.

A. Dirichlet Distribution

In probability theory, a categorical distribution is a discrete probability distribution that describes the possible results of a random variable that can take on one of K possible categories [50]. In classification tasks, the popularly used soft-max activation function transforms the output of a neural network to a probability vector describing a categorical distribution. The Dirichlet distribution is the conjugate prior of the categorical distribution and can be interpreted as a distribution over categorical distributions. While the probability vector given by a soft-max function represents a single point on the underlying solution simplex, the Dirichlet distribution represents a distribution on this simplex. Following, it can be used to represent the uncertainty on a classification network's output vector. A Dirichlet distribution for K classes is described by class concentrations $\{\alpha_1, \ldots, \alpha_K\} > 0$ and a derived precision value $\alpha_0 = \sum_{c=1}^{K} \alpha_c$. With this, the density of a Dirichlet distribution is given by

$$\mathrm{Dir}(\mu|\alpha) = \frac{\Gamma(\alpha_0)}{\prod_{c=1}^{K} \Gamma(\alpha_c)} \prod_{c=1}^{K} \mu_c^{\alpha_c - 1} \tag{1}$$

where Γ is the gamma function. A higher class concentration α_i for the class c_i leads to more probability mass shifted toward the corner of class c_i. Also, the higher the resulting precision value, the sharper is the Dirichlet distribution, i.e., the lower is the variety in the plausible categorical probability vectors. In Fig. 2, this is visualized for a Dirichlet distribution based on three classes.

B. Dirichlet Prior Network

Since a Dirichlet distribution describes a distribution over categorical distributions, it can be used as a distribution over the outputs of a neural network with K outputs. For a neural network $f_\theta(x)$ with parameters θ and input x, the network outputs before the soft-max activation function are called the logits and are given by $f_\theta(x) = z(x) = (z_1(x), \ldots, z_K(x)) \in \mathbb{R}^K$. The logits are in general unbounded and can be both, positive and negative. A DPN uses the logit output to predict the log

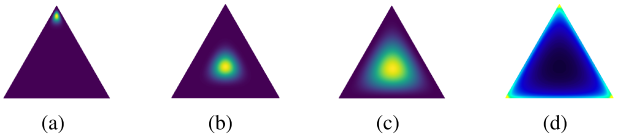

Fig. 2. Different desired Dirichlet distributions shown over the simplex (cf. [45]). (a) In-domain confident: high-class concentration α_i for a single class c_i. (b) In-domain aleatoric uncertainty: high-class concentrations for all three classes leading to a sharp and centered Dirichlet distribution. (c) OOD (with DPN): low-class concentrations $\alpha_i > 1$ for all classes leading to flat and centered Dirichlet distribution. (d) OOD (with DPN-RS): low-class concentrations $\alpha_i < 1$ for all classes leading to a degenerated and multimodal Dirichlet distribution.

concentration of a Dirichlet distribution. With predicted logit values $z_1, \ldots, z_K$, the network parameterizes a Dirichlet distribution with (positive) concentrations $\alpha_c = exp(z_c(x^*)),\ c = 1, \ldots, K$. Equivalently, the precision value is given by $\alpha_0 = \sum_{c=1}^{K} \alpha_c = \sum_{c=1}^{K} exp(z_c(x^*))$.

With this formulation, the posterior distribution $p(\omega|x, \theta)$ over the possible class labels $\omega \in \{1, 2, \ldots, K\}$ is given by the expected value of the Dirichlet distribution

$$\begin{aligned} p(\omega|x, \theta) &= \mathbb{E}_{\mathrm{Dir}(\mu|\alpha)}[p] \\ &= \int p(\omega)\ \mathrm{Dir}(\mu|\alpha)\ d\mu \\ &= \frac{\alpha_c}{\alpha_0}. \end{aligned} \tag{2}$$

The posterior given in (2) is equivalent to applying the soft-max function to the logit values of the network.

The challenge in optimizing the posterior distribution using standard neural networks with a soft-max activation function and cross-entropy loss function lies in the scaling of the posterior. As evident from (2), the scaling of the concentrations (α_c) affects the precision (α_0). Thus, looking only at the soft-max value, one can not conclude on the precision of the Dirichlet distribution. Following, the network is optimized based on pointwise estimations of the posterior distribution instead of taking the uncertainty on the posterior into account. As a result, it is not possible to separate distributional and data uncertainty effectively, leading to difficulty in the detection of OOD samples.

The DPN tackles the above-mentioned challenge by designing a multitask learning paradigm. In order to separate in-distribution samples and OOD samples, the network is trained on a mixture of two sets, a set of in-distribution samples (D_{in}) and an additional set of OOD samples (D_{out}). Please note that the set D_{out} for training is not necessarily drawn from the same distribution as the OOD samples during test/evaluation (see Fig. 1). The OOD samples during training (D_{out}) are only used to learn a boundary on the in-distribution samples. Once trained, the network can be applied on any OOD samples, even those that have a completely different distribution than the OOD samples used during training.

The general purpose of DPNs is to predict different forms of Dirichlet distributions in order to separate the following three cases:

1) in-distribution examples where the network is certain in its prediction;
2) in-distribution examples where the network is uncertain;
3) OOD examples.

DPNs seek to differentiate between in-domain and OOD samples based on the predicted class concentrations. More explicitly, they aim to produce a unimodal distribution at the corner of the solution simplex with the correct class [Fig. 2(a)] [31]. For in-domain samples with high data uncertainty, DPNs aim to produce a sharp distribution at the center [Fig. 2(b)] and for OOD data, DPNs aim to produce a flat distribution [Fig. 2(c)].

The key architecture of the deep model remains unmodified with a DPN, except removing the soft-max activation after the final layer, i.e., outputting the logits. However, the key to achieve the desired behavior is the design of a multitask optimization loss function, i.e., a loss that simultaneously supports the network in learning the classification task for in-distribution samples and learning to predict very small class concentrations for OOD examples. For that, the loss has to differentiate whether a received prediction is based on a sample from D_{in} or D_{out} and hence should be of the form of

$$\mathcal{L}(\theta) = \mathcal{L}_{\mathrm{in}}(\theta) + \gamma \cdot \mathcal{L}_{\mathrm{out}}(\theta) \tag{3}$$

where $\gamma > 0$ is a scalar, balancing the impact of in-distribution and OOD samples. In order to achieve the desired behavior, Malinin and Gales [31] presented a loss function based on the Kullback–Leibler (KL) divergence between the target Dirichlet distribution $\mathrm{Dir}(\mu|\alpha^{\mathrm{in}})$ or $\mathrm{Dir}(\mu|\alpha^{\mathrm{out}})$ for some sample x, and the corresponding predicted Dirichlet distribution $p(\mu|x, \theta)$

$$\begin{aligned} \mathcal{L}^{\mathrm{kl}}\left(\theta; \alpha^{\mathrm{in}}, \alpha^{\mathrm{out}}\right) = \mathbb{E}_{P_{\mathrm{in}}(x)}\left[\mathrm{KL}\left[\mathrm{Dir}\left(\mu|\alpha^{\mathrm{in}}\right)||p(\mu|x, \theta)\right]\right] \\ + \mathbb{E}_{P_{\mathrm{out}}(x)}\left[\mathrm{KL}\left[\mathrm{Dir}\left(\mu|\alpha^{\mathrm{out}}\right)||p(\mu|x, \theta)\right]\right]. \end{aligned} \tag{4}$$

P_{in} and P_{out} describe the in- and out-distribution, respectively, and α^{in} and α^{out} represent the ground-truth target concentrations. Since the target concentrations cannot be derived from the one-hot encoding (due to the scaling described before), these values have to be chosen beforehand [31].

Based on further investigations, Malinin and Gales [48] also presented a loss function based on reverse KL divergence

$$\begin{aligned} \mathcal{L}^{\mathrm{rkl}}\left(\theta; \alpha^{\mathrm{in}}, \alpha^{\mathrm{out}}\right) = \mathbb{E}_{P_{\mathrm{in}}}\left[\mathrm{KL}\left[p(\mu|x, \theta)||\mathrm{Dir}\left(\mu|\alpha^{\mathrm{in}}\right)\right]\right] \\ + \mathbb{E}_{P_{\mathrm{out}}}\left[\mathrm{KL}\left[p(\mu|x, \theta)||\mathrm{Dir}\left(\mu|\alpha^{\mathrm{out}}\right)\right]\right]. \end{aligned} \tag{5}$$

The reverse KL divergence showed improvement in the numerical stability and OOD detection results in comparison to [31]. However, as shown by Nandy *et al.* [45], for in-domain examples with high aleatoric uncertainty among multiple classes, DPNs produce flat Dirichlet distributions [45]. In practice, this could easily lead to representations that are indistinguishable from OOD examples.

C. Suitability of Classical DPN for Remote Sensing

The DPN is a suitable framework for remote sensing image classification for the following reasons.

1) Considering the variety of remote sensing data, OOD data may come in many unforeseeable forms in remote

sensing. DPNs provide the flexibility that all samples from all possible distributions do not need to be seen during the training phase. For example, considering a spatially varying system, if the in-domain training data belong to Europe and OOD training data belong to Africa, the DPN model is capable of handling OOD test data from Asia.

2) DPNs can be used without altering the key architecture of the models already used in remote sensing classification.
3) A DPN is a single deterministic neural network where only one forward pass per evaluation has to be performed. This leads to less computation than for other approaches as ensembles or Bayesian neural networks. This is an important advantage, especially for very-large-scale Earth Observation (EO) applications.

Due to the large number of classes in remote sensing with strong interclass similarity, it is common in remote sensing for in-domain samples to have high aleatoric uncertainty among multiple classes. In such cases, DPNs produce a flatter Dirichlet distribution [45]. This leads to representations that are harder to distinguish from the OOD samples. In other words, for remote sensing applications, DPN may confuse between aleatoric uncertainty and distributional uncertainty. This limits the practical application of traditional DPNs [48] in remote sensing. Hence, to alleviate this problem, inspired by [45], we propose DPN-RS that can effectively segregate the OOD samples from in-domain data.

D. DPN-RS

To overcome the challenges introduced in Section III-C, our approach aims at learning a sharp multimodal distribution ($\alpha_0 \ll 1$) [see Fig. 2(d)] instead of a flat unimodal distribution for OOD examples. The precision regularization is achieved by introducing a bounded regularization term given by the sigmoid function on the logits

$$\alpha_0' = \frac{1}{K}\sum_{c=1}^{K} \text{sigmoid}(z_c(x)).$$

α_0' is used as a regularizer along with the cross-entropy loss. This gives the following two loss formulations for in-domain and OOD examples:

$$\mathcal{L}_{\text{in}}(\theta;\lambda_{\text{in}}) := \mathbb{E}_{P_{\text{in}}(x,y)}\left[-\log p(y|x,\theta) - \lambda_{\text{in}}\alpha_0'\right] \quad (6)$$

and

$$\mathcal{L}_{\text{out}}(\theta;\lambda_{\text{out}}) := \mathbb{E}_{P_{\text{out}}(x,y)}\left[\mathcal{H}_{\text{ce}}(\mathcal{U};\, p(y|x,\theta)) - \lambda_{\text{out}}\alpha_0'\right] \quad (7)$$

where $\mathcal{U}$ denotes the uniform distribution over all classes and $\mathcal{H}_{\text{ce}}$ denotes the cross-entropy function. With this approach, the ground truth is given by a probability vector and can be therefore directly derived from the class labels and no target concentrations have to be chosen. The precision is controlled by two hyperparameters λ_{in} and λ_{out} [45] and the combined loss function is given by

$$\mathcal{L}^{\text{DPN-RS}}(\theta;\gamma,\lambda_{\text{in}},\lambda_{\text{out}}) = \mathcal{L}_{\text{in}}(\theta,\lambda_{\text{in}}) + \gamma\,\mathcal{L}_{\text{out}}(\theta,\lambda_{\text{out}}) \quad (8)$$

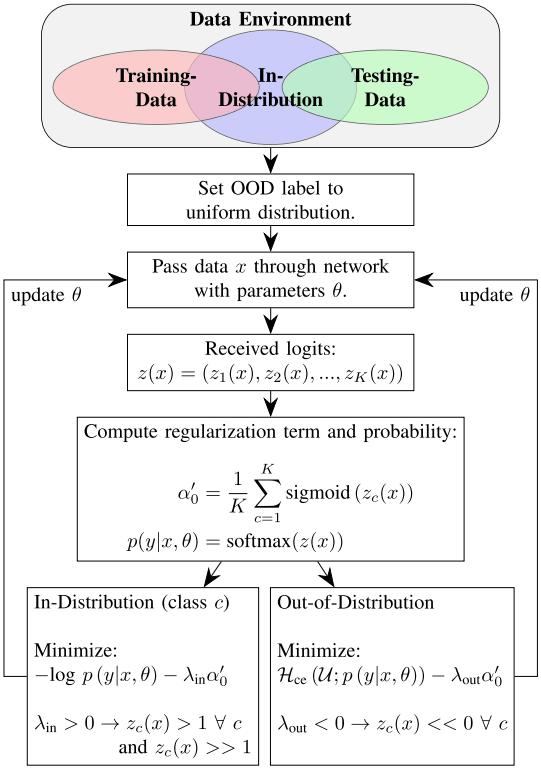

Fig. 3. Visualization of the training procedure for the considered DPN-RS network.

where again in-domain and OOD samples are balanced by a hyperparameter $\gamma > 0$. For the proposed approach, we use $\lambda_{\text{in}} > 0$, while $\lambda_{\text{out}} < 0$. For in-domain examples that are confidently predicted, the cross-entropy loss maximizes the logit value of the correct class. However, for in-domain samples with aleatoric uncertainty, the optimizer maximizes sigmoid($z_c(x)$) for all classes, thus yielding a sharp distribution centered on the solution simplex. By choosing $\lambda_{\text{out}} < 0$, DPN-RS produces negative values for $z_c(x^*)$ for an OOD example x^*. This leads to $\alpha_c \ll 1$ for all $c = 1, \ldots, K$, and thus, an OOD sample yields a sharp multimodal Dirichlet distribution with probability mass at each corner of the simplex [Fig. 2(d)]. Fig. 2(b) and (d) shows more distinct over the simplex, making the OOD samples easily distinguishable from the in-domain ones. In Fig. 3, a visualization of the training process of the proposed approach is given. In Fig. 5, an example for a certain in-distribution prediction, an uncertain in-distribution prediction, and an OOD prediction is presented together with different derived measures, which can be used to separate between in-distribution and OOD.

IV. Experimental Validation

In Section IV-A, we briefly present the datasets used in our experiments. Experimental settings are discussed in Section IV-B. The rest of this section presents the results and the analyses for each of the experiments.

A. Datasets

We perform our experiments on three different datasets, namely, the Aerial Image dataset (AID) [52], the UC-Merced

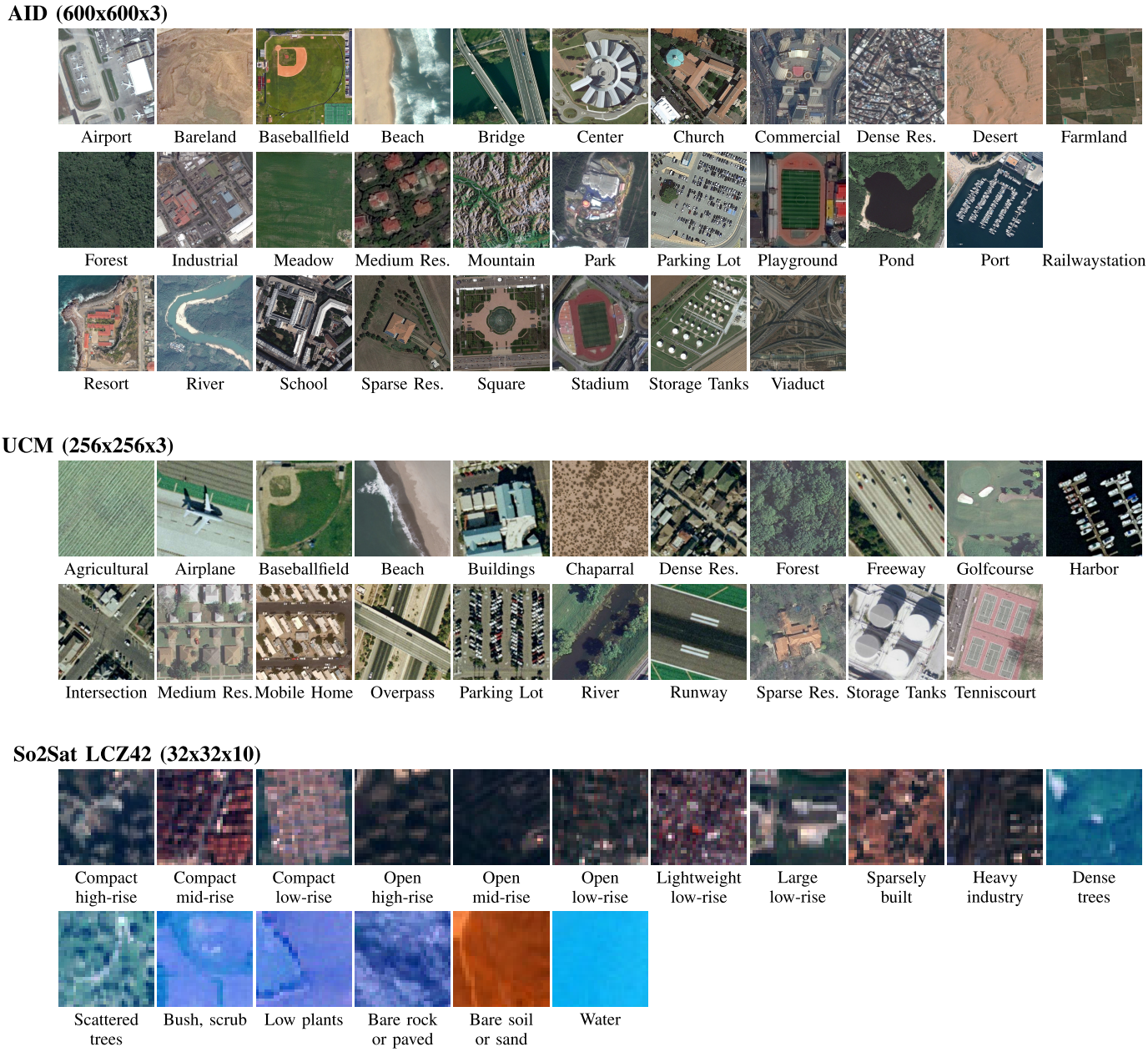

Fig. 4. Defined classes and corresponding example patches of the UCM dataset [51], the AID [52], and the So2Sat LCZ42 dataset [53]. For LCZ42, only the three bands representing red, green, and blue are visualized.

Land Use (UCM) dataset [51], and the So2Sat Local Climate Zone 42 (LCZ42) dataset [53]. In the following, the datasets are briefly described. An overview over the classes contained in the different datasets is given in Fig. 4.

1) AID Dataset: The AID dataset [52] contains very high-resolution aerial RGB images with 600×600 pixels size. The dataset covers 30 different classes, each represented by more than 300 samples in the dataset. We split the dataset randomly into 70% for training and 30% for testing. Furthermore, the images are cropped and resized to 256×256 pixels. All experiments are based on a ResNet50 neural network pretrained on imagenet.

2) UC Merced Landcover Dataset: The UC Merced (UCM) dataset [51] contains high-resolution aerial RGB images with 1-ft ground sampling distance and 256×256 pixels size. The dataset covers 21 different classes, each represented by 100 samples in the dataset. Again, we split the dataset randomly into 70% for training and 30% for testing. All experiments are based on a ResNet50 neural network pretrained on Imagenet.

3) So2Sat LCZ42 Dataset: The So2Sat LCZ42 dataset [53] provides about half a million coregistered Sentinel-1 and Sentinel-2 patches. For our experiments, we only use the optical Sentinel-2 images. The 32×32 patches are taken from 42 different regions worldwide, and for each sample, a local climate zone (LCZ) label is provided. The data are split into a training set of 352 366 patches and a validation and test set containing 24 188 and 24 119 patches, respectively, sampled from regions different from the regions of the training set. An overview over the LCZs can be seen in Fig. 4. We build

TABLE I
OVERVIEW OVER THE COMPARED METHODS AND THEIR FUNDAMENTAL PROPERTIES AND DIFFERENCES

Abbreviation	Description	OOD Examples at training time	Targets for in-distribution	Targets for out-of-distribution	Loss function
DPN-RS	Aims to predict high class concentrations for in-distribution samples and low and uniform class concentrations for OOD samples.	✓	One-hot vector of the form $(0, .., 1, .., 0)$.	Uniform probability vector of the form $(\frac{1}{K}, .., \frac{1}{K}, .., \frac{1}{K})$.	Cross-entropy with a precision regularization.
DPN$^+$	Aims to predict high class concentrations for in-distribution samples and high and and uniform class concentrations for OOD samples.	✓	One-hot vector of the form $(0, .., 1, .., 0)$.	Uniform probability vector of the form $(\frac{1}{K}, .., \frac{1}{K}, .., \frac{1}{K})$.	Cross-Entropy with a precision regularization.
DPN$_{\text{rev}}$	Aims to predict high class concentrations for in-distribution samples and low and uniform class concentrations for out-of-distribution samples.	✓	Target concentrations of the form $(1, .., 100, .., 1)$.	Target concentrations of the form $(1, .., 1, .., 1)$	KL-Divergence with the predicted Dirichlet distribution as first input.
DPN$_{\text{for}}$	Aims to predict high class concentrations for in-distribution samples and low and uniform class concentrations for out-of-distribution samples.	✓	Target concentrations of the form $(1, .., 100, .., 1)$	Target concentrations of the form $(1, .., 1, .., 1)$	KL-Divergence with the predicted Dirichlet distribution as second input.
ENN	Predicts 'an evidence' for in-distribution classes and identifies OOD examples based on missing evidence.	✗	One-hot vector of the form $(0, .., 1, .., 0)$.	-	Expected Cross-Entropy with an evidence regularization.

our networks based on the network structure proposed in [54] but without multilevel fusion. For the experiments related to open-set recognition and sensor shifts, we want to avoid a region shift and therefore work only on the training set of the original dataset which we split into 70%–30% for our training and testing.

B. Experimental Settings

We evaluate the performance of the presented methods on three different remote sensing tasks.

1) *Open-Set Recognition:* Where the test set contains classes unseen during training.
2) *Channel Separation:* Where the test set contains images from different channels than the training images. This simulates a multisensor scenario.
3) *Location Separation:* Where the test set contains images from different spatial locations than the training images.

We run the experiments within single datasets and without mixing different datasets. Intuitively, it is clear that when working with different datasets, the similarity between the dataset used for in-distribution and the dataset used for OOD during training builds a crucial point for the OOD detection performance.

We compare the proposed method to the following paradigms in which main properties are also summarized in Table I.

1) *DPN$^+$ [45]:* A DPN-based approach with precision regularizing factors $\lambda_{\text{in}} > 0$ and $\lambda_{\text{out}} > 0$.
2) *DPN^{rev} [48]:* A DPN that uses the reverse KL divergence as in (5) to compare the predicted and the ground-truth Dirichlet distribution.
3) *DPN^{forw} [31]:* A DPN that uses the KL divergence as in (4) to compare the predicted and the ground-truth Dirichlet distribution.
4) *Evidential neural network (ENN) [49]:* The ENN does not require any OOD training data. ENN is motivated by subjective logic and also interprets the logits as a parameterization of a Dirichlet distribution. However, in contrast to DPNs, ENNs set the class concentrations in relation to an additional constant concentration that is interpreted as an unknown class. For ENNs, different loss functions are presented in [49]. For our analysis, we use the expected cross-entropy loss.

The receiver operator characteristic (ROC) is popularly used to present the results for binary decision problems in machine learning [55]. Conforming to this, we use the area under ROC (AUROC) to present the OOD detection performance based on four popularly indicators, namely, maximum probability, entropy, mutual information, and α_0 [45].

For the approaches that make use of OOD samples at training time, we generated batches that contain 50% in-distribution and 50% OOD samples. Based on preliminary experiments, we have chosen the hyperparameters $\lambda_{\text{in}}, \lambda_{\text{out}}$ and γ for the losses defined in (6)–(8) as

$$\lambda_{\text{in}} = 0.5, \quad \lambda_{\text{out}} = \frac{1}{K} - 0.5, \quad \gamma = 1. \tag{9}$$

The targets are chosen, as shown in Table I.

We show all results as mean and standard deviation based on seven different runs. Even though the objective of the proposed method is OOD detection, we also show in-domain classification performances. We show them as accuracy computed over all in-domain samples (denoted as "accuracy" in the tables) and accuracy computed separately over all classes and then averaged (denoted as "average accuracy" in the tables) and as Cohen's kappa value.

TABLE II

DIFFERENT SPLITS OF THE 30 CLASSES CONTAINED IN AID INTO IN-DOMAIN, OOD FOR TRAINING, AND OOD FOR TESTING SETS

	Simple Setting	Random Setting 1	Random Setting 2	Random Setting 3	Random Setting 4	Random Setting 5
In-Domain Classes	Farmland, Bare Land, River, Forest, Dessert, Meadow, Beach, Mountain, Park, Pond	Port, Commercial, Center, Baseball Field, Beach, Railway Station, Meadow, Stadium, Industrial, Bare Land	Mountain, Dense Residential, Church, Square, Beach, Forest, Medium Residential, Railway Station, Desert, Center	Commercial, Resort, Playground, Medium Residential, Bare Land, Railway Station, Bridge, Mountain, River, Dessert	Resort, Mountain, Square, Bare Land, Sparse Residential, Dense Residential, Commercial, Viaduct, Parking, Railway Station	Baseball Field, Playground, Mountain, Beach, Sparse Residential, Bare Land, Port, Stadium, Dessert, Center
Out-of-Distribution Classes Training	Airport, Industrial, Baseball Field, Bridge, Center, Church, Dense Residential, Medium Residential, Playground, Parking	Pond, Medium Residential, Playground, Church, Forest, River, School, Storage Tanks, Resort, Farmland	Farmland, Resort, Baseball Field, Bridge, Port, Viaduct, Pond, Airport, Storage Tanks, School	Port, Dense Residential, Pond, Airport, Farmland, Storage Tanks, Viaduct, Sparse Residential, Meadow, Center	Airport, Pond, Bridge, Forest, Dessert, Stadium, Meadow, Center, Medium Residential, School	Railway Station, Storage Tanks, Parking, Medium Residential, Square, Pond, Airport, Viaduct, Meadow
Out-of-Distribution Classes Testing	Commercial, Port, Railway Station, Resort, School, Sparse Residential, Square, Stadium, Storage Tanks, Viaduct	Square, Viaduct, Dense Residential, Sparse Residential, Park, Parking, Airport, Mountain, Bridge, Dessert	Playground, Commercial, Bare Land, Industrial, Meadow, River, Parking, Stadium, Park, Sparse Residential	Square, Park, Industrial, School, Beach, Baseball Field, Parking, Church, Stadium, Forest	River, Beach, Industrial, Port, Baseball Field, Church, Storage Tanks, Farmland, Park, Playground	Dense Residential, Park, River, Farmland, Church, Commercial, Resort, Industrial, School, Forest

TABLE III

OOD DETECTION RESULTS ON THE OPEN-SET RECOGNITION TASK ON THE AID DATASET. THE PERFORMANCE IS MEASURED BY 100× THE AUROC. THE RESULTS ARE GIVEN AS MEAN AND STANDARD DEVIATION COMPUTED BASED ON SEVEN RUNS. THE BEST RESULTS PER APPROACH ARE GIVEN BOLDFACED AND THE BEST RESULTS ON THE SINGLE SETTINGS OVER ALL APPROACHES ARE ITALICIZED

		Simple Setting	Random Setting 1	Random Setting 2	Random Setting 3	Random Setting 4	Random Setting 5
DPN-RS	Max. Prob.	98.28±0.70	94.55±0.97	93.24±1.34	93.65±1.68	94.89±0.73	97.87±1.57
	Entropy	98.24±0.67	**94.62±0.92**	93.37±1.30	**93.78±1.47**	**95.04±0.60**	98.09±1.36
	Mutual Info	98.43±0.60	93.58±0.83	93.50±1.01	92.25±1.34	94.09±0.89	98.19±1.10
	α_0	**98.50±0.60**	93.89±0.89	**93.83±1.09**	92.83±1.18	94.62±0.65	**98.23±1.09**
DPN$^+$	Max. Prob.	98.37±0.39	***95.42±0.74***	93.92±0.80	93.85±1.47	***95.97±0.52***	***98.90±0.38***
	Entropy	98.46±0.37	95.39±0.71	***93.96±0.85***	**93.86±1.43**	95.96±0.54	***98.90±0.39***
	Mutual Info	98.71±0.38	93.60±1.31	91.81±1.49	92.17±1.09	94.37±0.53	98.27±0.47
	α_0	***98.76±0.35***	94.55±0.81	92.75±1.30	92.95±1.27	95.04±0.48	98.32±0.44
DPN^{rev}	Max. Prob.	98.13±0.31	**95.37±0.26**	93.30±1.14	94.92±0.60	95.78±0.50	98.44±0.59
	Entropy	98.26±0.28	**95.37±0.23**	93.46±1.04	**94.97±0.60**	**95.83±0.52**	**98.60±0.51**
	Mutual Info	**98.44±0.29**	95.01±0.28	**93.57±0.99**	94.72±0.74	95.73±0.64	98.55±0.57
	α_0	**98.44±0.28**	95.01±0.28	93.57±0.99	94.72±0.73	95.73±0.63	98.55±0.57
DPN^{forw}	Max. Prob.	98.18±0.26	94.97±0.60	93.42±0.77	94.4±0.71	95.48±0.55	98.83±0.14
	Entropy	98.24±0.28	**95.01±0.59**	93.53±0.84	94.45±0.71	**95.53±0.57**	**98.87±0.14**
	Mutual Info	**98.31±0.21**	94.62±0.62	93.57±0.89	94.40±1.05	94.70±1.54	98.76±0.16
	α_0	98.27±0.20	94.52±0.65	**93.55±0.96**	***94.32±1.19***	94.17±1.81	98.73±0.15
ENN	Max. Prob.	80.46±4.55	89.13±2.78	82.64±2.39	85.21±4.25	88.53±1.68	91.36±2.31
	Entropy	80.47±4.59	**89.28±2.74**	**82.70±2.41**	85.30±4.25	**88.63±1.68**	**91.44±2.36**
	Mutual Info	81.13±5.53	89.00±2.79	82.49±2.70	**85.96±4.56**	88.35±2.06	90.83±2.76
	α_0	**83.60±4.63**	87.31±2.96	81.16±3.37	83.93±5.25	87.17±2.91	88.90±3.04

C. Open-Set Recognition

Open-set recognition is an important problem in computer vision [56] and remote sensing [36]. To simulate open-set behavior in a remote sensing dataset, we split the given datasets into three subsets. The sets of classes in each subset are disjoint, i.e., each class is part of exactly one subset. For the open-set recognition problem, we use one of the subsets as in-distribution samples, one as OOD samples that are given at training time and the third one as OOD samples reserved for testing the OOD detection performance.

1) Open-Set Recognition on AID: For the open-set recognition, we split the 30 classes of the AID into three groups

TABLE IV

CLASSIFICATION ACCURACY AND AVERAGE ACCURACY ON THE OPEN-SET RECOGNITION TASK ON THE AID DATASET. THE RESULTS ARE GIVEN AS MEAN AND STANDARD DEVIATION COMPUTED BASED ON SEVEN RUNS. THE BEST AVERAGE ACCURACY FOR EACH SETTING IS HIGHLIGHTED

		Simple Setting	Random Setting 1	Random Setting 2	Random Setting 3	Random Setting 4	Random Setting 5
DPN-RS	Accuracy	97.29±0.63	96.77±0.59	96.05±0.62	97.25±0.78	96.86±0.34	97.27±1.39
	Avg. Acc.	97.35±0.56	96.70±0.62	95.85±0.65	97.03±0.86	96.63±0.44	97.21±1.43
	Kappa	97.28±0.64	96.72±0.59	96.02±0.63	97.21±0.78	96.84±0.34	97.23±1.40
DPN^{+}	Accuracy	97.43±0.84	97.19±0.37	96.64±0.66	97.36±0.94	97.21±0.83	98.01±0.40
	Avg. Acc.	97.42±0.89	97.12±0.42	96.44±0.74	97.18±1.03	97.01±0.78	97.93±0.46
	Kappa	97.42±0.84	97.15±0.04	96.61±0.01	97.33±0.01	97.19±0.01	97.99±0.41
DPN^{rev}	Accuracy	98.04±0.30	98.01±0.51	97.17±0.93	98.06±0.26	97.92±0.32	98.18±0.69
	Avg. Acc.	98.04±0.35	**97.97±0.53**	97.04±0.88	97.89±0.29	**97.79±0.32**	98.07±0.75
	Kappa	98.02±0.31	97.98±0.51	97.15±0.94	98.04±0.26	97.90±0.32	98.16±0.70
DPN^{forw}	Accuracy	98.31±0.33	97.66±0.47	97.29±0.52	98.72±0.53	97.14±0.80	98.75±0.08
	Avg. Acc.	**98.34±0.38**	97.59±0.43	**97.13±0.52**	**98.21±0.42**	96.93±0.87	**98.68±0.07**
	Kappa	98.30±0.34	97.62±0.41	97.27±0.53	98.71±0.53	97.12±0.81	98.74±0.08
ENN	Accuracy	96.84±0.49	96.11±0.95	92.17±4.18	95.46±1.84	95.20±1.23	96.23±1.33
	Avg. Acc.	96.87±0.53	96.10±1.07	91.99±3.97	95.29±1.83	95.02±1.24	96.10±1.45
	Kappa	96.82±0.50	96.05±0.97	92.12±4.21	95.41±1.87	95.17±1.24	96.18±1.34

TABLE V

DIFFERENT SPLITS OF THE 21 CLASSES CONTAINED IN THE UCM DATASET INTO IN-DOMAIN, OOD FOR TRAINING, AND OOD FOR TESTING SETS

	Simple Setting	Random Setting 1	Random Setting 2	Random Setting 3	Random Setting 4	Random Setting 5
In-Domain Classes	Beach, Forest, Harbor, River, Agriculture, Golf Course, Chaparral	Overpass, Chaparral, Baseball Diamond, Intersection, Forest, Sparse Residential, Beach	Beach, Chaparral, Harbor, Dense Residential, Parking Lot, Sparse Residential, Tennis Court	Forest, Golf Course, Harbor, Tennis Court, Mobile Home Park, Freeway, Overpass	Intersection, Mobile Home Park, Agricultural, Airplane, Freeway, Chaparral, Forest	Airplane, Tennis Court, Parking Lot, Chaparral, Baseball Diamond, Buildings, Mobile Home Park
Out-of-Distribution Classes Training	Airplane, Baseball Diamond, Intersection, Medium Residential, Mobile Home Park, Buildings, Freeway	Storage Tank, Dense Residential, Medium Residential, Runway, Tennis Court, Freeway, Harbor	Golf Course, Overpass, Buildings, Intersection, Agricultural, Storage Tanks, Freeway	Parking Lot, Sparse Residential, Chaparral, Buildings, Airplane, Storage Tanks, Agriculture	Runway, Parking Lot, Overpass, Beach, Tennis Court, Medium Residential, Golf Course	Beach, Runway, Sparse Residential, Golf Course, Overpass, Freeway, Agricultural
Out-of-Distribution Classes Testing	Overpass, Parking Lot, Runway, Sparse Residential, Storage Tanks, Tennis Court, Dense Residential	Airplane, Mobile Home Park, River, Agricultural, Parking Lot, Buildings, Golf Course	Baseball Diamond, Forest, Airplane, Medium Residential, Mobile Home Park, Runway, River	Medium Residential, River, Beach, Baseball Diamond, Runway, Dense Residential, Intersection	Buildings, Baseball Diamond, Dense Residential, Storage Tanks, Harbor, River, Sparse Residential	River, Storage Tanks, Harbor, Dense Residential, Forest, Medium Residential, Intersection

of ten classes each. In order to evaluate the robustness of the considered approaches, we consider a handcrafted split into human built scenes and nonhuman built scenes. Furthermore, we consider five random splits of classes. The resulting in-domain and OOD datasets are described in Table II. We tabulate the open-set recognition accuracy and classification accuracy in Tables III and IV, respectively. All methods perform relatively well for this dataset. While the DPN-based approaches (DPN-RS, DPN^{+}, DPN^{forw}, and DPN^{rev}) receive AUROC values above 0.9 for the OOD detection task, the ENN achieves at least 0.80 in average. Over all test cases, all DPN-based approaches are among the best performances with not more than 1% difference.

All approaches perform satisfactorily in regard to the classification accuracy on the in-distribution samples. All DPN-based approaches obtain an average accuracy higher than 95% for all test cases.

2) Open-Set Recognition on UCM: We split the 21 classes of the UCM dataset into three groups of seven classes each. In order to evaluate the robustness of the considered approaches, we consider a handcrafted split into human built scenes and nonhuman built scenes. Furthermore, we consider five random splits of classes. The resulting in-domain and OOD datasets are described in Table V. The OOD detection performance and classification results on the UCM dataset are presented in Tables VI and VII, respectively. Regarding

TABLE VI

OOD DETECTION RESULTS ON THE OPEN-SET RECOGNITION TASK ON THE UCM DATASET. THE PERFORMANCE IS MEASURED BY 100× THE AUROC. THE RESULTS ARE GIVEN AS MEAN AND STANDARD DEVIATION COMPUTED BASED ON SEVEN RUNS. THE BEST RESULTS PER APPROACH ARE GIVEN BOLDFACED AND THE BEST RESULTS ON THE SINGLE SETTINGS OVER ALL APPROACHES ARE ITALICIZED

		Simple Setting	Random Setting 1	Random Setting 2	Random Setting 3	Random Setting 4	Random Setting 5
DPN-RS	Max. Prob.	99.81±0.11	**98.40±0.67**	***95.91±1.35***	97.41±0.57	99.24±0.33	95.38±2.88
	Entropy	99.84±0.10	98.37±0.64	***95.91±1.21***	***97.48±0.63***	***99.26±0.30***	**95.64±2.46**
	Mutual Info	99.86±0.11	96.42±1.12	94.98±0.98	96.71±0.94	99.09±0.45	94.70±1.44
	α_0	***99.87±0.10***	96.57±1.07	95.24±1.01	96.88±0.83	98.98±0.54	94.98±1.37
DPN^{+}	Max. Prob.	99.66±0.47	***98.47±0.93***	**95.76±0.65**	97.26±0.87	98.94±0.74	***97.38±0.69***
	Entropy	99.71±0.38	98.42±0.90	95.75±0.67	**97.38±0.86**	99.12±0.42	97.33±0.64
	Mutual Info	99.70±0.28	96.86±1.12	94.92±1.19	96.67±0.71	**99.21±0.37**	95.85±1.17
	α_0	**99.72±0.23**	97.21±0.94	95.35±0.88	96.74±0.73	**99.21±0.37**	96.22±0.96
DPN^{rev}	Max. Prob.	98.91±0.67	**97.94±1.26**	**93.98±2.10**	95.83±1.28	**97.73±0.81**	**95.85±1.39**
	Entropy	98.99±0.74	97.82±1.24	93.96±1.95	**96.09±1.19**	97.70±0.82	**95.83±1.31**
	Mutual Info	**99.13±0.74**	96.49±1.70	93.31±2.15	95.87±1.27	97.55±0.83	94.72±1.61
	α_0	**99.13±0.74**	96.26±1.80	93.20±2.20	95.78±1.30	97.50±0.84	94.58±1.66
DPN^{forw}	Max. Prob.	98.53±1.45	98.16±1.35	**93.95±1.07**	95.27±1.11	98.06±2.10	**95.08±1.76**
	Entropy	**98.67±1.40**	**98.25±1.03**	**93.95±1.09**	**95.46±1.09**	**98.11±1.98**	94.99±1.95
	Mutual Info	95.00±8.73	96.90±2.99	93.33±1.04	95.41±1.19	96.57±4.81	93.80±3.56
	α_0	93.70±9.71	96.55±3.15	92.98±1.13	95.26±1.24	96.25±5.00	93.37±4.11
ENN	Max. Prob.	88.74±3.68	86.32±6.71	85.50±7.49	90.53±3.42	90.99±3.37	90.35±4.91
	Entropy	**88.80±3.62**	86.61±6.49	**85.56±7.50**	90.63±3.37	91.09±3.35	**90.42±4.94**
	Mutual Info	88.62±3.35	86.98±6.11	85.48±7.18	**90.74±3.43**	91.22±3.38	90.20±4.97
	α_0	86.99±3.13	**87.80±5.05**	85.06±6.32	90.05±3.65	**92.18±2.96**	89.17±5.23

TABLE VII

CLASSIFICATION ACCURACY AND AVERAGE ACCURACY ON THE OPEN-SET RECOGNITION TASK ON THE UCM DATASET. THE RESULTS ARE GIVEN AS MEAN AND STANDARD DEVIATION COMPUTED BASED ON SEVEN RUNS. THE BEST AVERAGE ACCURACY FOR EACH SETTING IS HIGHLIGHTED

		Simple Setting	Random Setting 1	Random Setting 2	Random Setting 3	Random Setting 4	Random Setting 5
DPN-RS	Accuracy	98.38±0.93	99.22±0.37	99.31±0.65	98.78±0.77	99.48±0.52	98.96±0.68
	Avg. Acc.	98.41±0.90	99.18±0.39	99.33±0.62	98.78±0.79	99.47±0.47	98.95±0.65
	Kappa	98.34±0.96	99.20±0.37	99.29±0.66	98.76±0.80	99.48±0.48	98.94±0.63
DPN^{+}	Accuracy	97.81±1.90	99.06±1.04	99.74±0.26	99.17±0.63	99.32±0.57	99.27±0.48
	Avg. Acc.	97.81±1.94	99.04±1.09	**99.75±0.25**	**99.16±0.63**	99.31±0.59	**99.27±0.48**
	Kappa	97.76±1.94	99.04±1.05	99.73±0.27	99.15±0.64	99.30±0.59	99.26±0.49
DPN^{rev}	Accuracy	99.22±0.75	99.41±0.71	99.19±1.35	99.11±0.60	99.69±0.35	98.96±1.09
	Avg. Acc.	**99.23±0.72**	**99.43±0.69**	99.16±1.39	99.10±0.61	99.70±0.34	98.91±0.91
	Kappa	99.17±0.78	99.37±0.73	99.15±1.37	99.07±0.63	99.65±0.35	98.92±0.89
DPN^{forw}	Accuracy	99.21±2.26	98.75±2.06	99.28±0.36	98.85±0.76	98.83±2.72	98.26±1.32
	Avg. Acc.	98.19±2.33	98.69±2.27	99.29±0.37	98.84±0.79	**99.85±0.23**	98.22±1.38
	Kappa	98.13±2.31	98.68±2.10	99.23±0.39	98.80±0.80	99.80±0.26	98.19±1.34
ENN	Accuracy	96.21±2.71	96.50±5.67	94.84±4.96	97.86±1.24	99.07±0.54	95.52±4.35
	Avg. Acc.	96.24±2.71	96.61±5.51	94.81±4.95	97.85±1.28	99.07±0.53	95.41±4.54
	Kappa	96.11±2.77	96.43±5.81	94.74±5.02	97.80±1.27	99.05±0.55	95.42±4.46

the OOD detection task, the DPN-based approaches perform satisfactorily with AUROC values of at least 0.95. Even though DPN-RS gives the highest average AUROC score in four out of six cases, the results are very close to each other considering the stated standard deviations. The performances of all DPN-based approaches do not differ more than 2% in almost all cases. The ENN results are worse compared to the DPN-based methods.

The average in-domain classification accuracy is satisfactory for all approaches and all settings. The best average accuracy is above 99% with only small deviations between the different approaches.

3) Open-Set Recognition on LCZ42: In comparison to the AID and UCM datasets, the interclass similarity is much stronger in the low spatial-resolution LCZ42 dataset, making it a more challenging dataset. For our experiments, we split the classes into urban (classes 1–10), vegetation (classes A–F), and water (class G). First, we test the performance with urban as in-domain and vegetation and water as OOD data. Second, we test the performance with vegetation as in-domain and urban and water as OOD data. The OOD detection performance and the classification results on the LCZ42 dataset are presented in Tables VIII and IX, respectively.

The proposed DPN-RS performs best on all test settings based on the LCZ42 dataset. Not only the average separation performance is better than for the other approaches but also the accuracy on the in-distribution classification task is larger in three out of four settings and the variances in the results are smaller. The setting with urban classes as in-domain and vegetation as OOD during training and water as OOD for testing leads to the best results over all test settings with all considered AUROC values above 0.99 for DPN-RS.

In Table X, we show the results of DPN-RS on the open-set problem when using UCM, AID, and LCZ42 at the same time. One can clearly see that the similar resolution of AID and UCM has a significant influence on the OOD detection performance.

TABLE VIII

OOD DETECTION RESULTS ON THE OPEN-SET RECOGNITION TASK ON THE SO2SAT LCZ42 DATASET. THE PERFORMANCE IS MEASURED BY 100× THE AUROC. THE RESULTS ARE GIVEN AS MEAN AND STANDARD DEVIATION COMPUTED BASED ON SEVEN RUNS. THE BEST RESULTS PER APPROACH ARE GIVEN BOLDFACED AND THE BEST RESULTS ON THE SINGLE SETTINGS OVER ALL APPROACHES ARE ITALICIZED

		Urban Vegetation Water	Urban Water Vegetation	Vegetation Urban Water	Vegetation Water Urban
DPN-RS	Max. Prob.	98.18±1.03	75.52±2.49	***90.28±2.01***	90.01±1.89
	Entropy	98.68±0.89	77.38±2.88	89.31±2.48	91.09±1.78
	Mutual Info	***99.24±0.28***	87.74±1.96	86.13±3.37	***94.19±1.02***
	α_0	99.22±0.28	***87.81±2.01***	86.15±3.38	94.11±0.99
DPN^+	Max. Prob.	96.33±1.54	74.81±1.66	**86.03±2.29**	89.14±3.82
	Entropy	**97.56±1.46**	**76.94±1.70**	**86.03±2.83**	90.18±3.63
	Mutual Info	90.35±9.10	74.63±8.85	78.97±8.75	90.89±2.65
	α_0	90.35±9.08	74.75±9.02	79.47±8.62	**91.02±2.82**
DPN^{rev}	Max. Prob.	**89.09±6.41**	70.53±3.45	82.78±5.88	72.73±6.89
	Entropy	88.41±6.99	71.02±4.72	**83.71±6.03**	71.97±5.87
	Mutual Info	85.00±1.83	**73.29±5.73**	82.69±6.28	**77.99±1.35**
	α_0	85.19±4.69	72.14±4.97	82.97±7.27	77.01±1.04
DPN^{forw}	Max. Prob.	89.91±3.08	**41.34±7.23**	**73.08±8.79**	84.82±2.70
	Entropy	**91.42±2.07**	39.69±7.73	71.97±9.28	**88.38±3.79**
	Mutual Info	59.74±8.79	35.21±7.13	67.81±10.95	84.14±2.16
	α_0	53.61±9.75	30.51±7.87	65.87±9.36	83.78±2.26
ENN	Max. Prob.	73.01±6.99	**58.51±8.54**	**82.38±4.55**	86.34±1.72
	Entropy	70.59±7.80	57.66±9.42	82.07±4.93	86.21±1.93
	Mutual Info	**78.17±5.07**	56.89±7.93	81.45±6.05	**86.65±1.62**
	α_0	73.12±3.20	55.34±8.43	81.83±6.28	81.48±2.49

TABLE IX

CLASSIFICATION ACCURACY AND AVERAGE ACCURACY ON THE OPEN-SET RECOGNITION TASK ON THE SO2SAT LCZ42 DATASET. THE RESULTS ARE GIVEN AS MEAN AND STANDARD DEVIATION COMPUTED BASED ON SEVEN RUNS. THE BEST AVERAGE ACCURACY FOR EACH SETTING IS HIGHLIGHTED

		Urban Vegetation Water	Urban Water Vegetation	Vegetation Urban Water	Vegetation Water Urban
DPN-RS	Accuracy	88.59±0.90	89.13±1.27	97.23±0.80	95.87±0.76
	Avg. Acc.	**86.18±0.72**	87.29±0.33	**95.52±1.42**	**93.41±1.09**
	Kappa	88.18±0.91	88.74±1.30	97.22±0.80	95.52±0.56
DPN^+	Accuracy	88.66±0.95	89.35±1.13	95.41±4.13	93.53±0.25
	Avg. Acc.	85.91±1.58	**87.60±0.47**	93.61±5.14	91.02±3.08
	Kappa	88.26±0.99	88.97±1.16	95.39±4.15	93.50±0.25
DPN^{rev}	Accuracy	69.62±6.8	89.40±0.70	94.58±3.66	81.23±6.95
	Avg. Acc.	55.64±21.3	86.25±0.83	90.27±5.47	66.71±16.22
	Kappa	68.24±16.6	88.99±0.75	94.59±3.63	80.00±7.46
DPN^{forw}	Accuracy	78.49±4.20	64.80±8.01	84.19±4.99	87.56±6.13
	Avg. Acc.	71.10±5.43	52.14±11.84	66.53±12.44	75.40±13.21
	Kappa	78.40±4.42	64.80±8.01	84.19±4.99	87.56±6.13
ENN	Accuracy	82.53±3.49	81.57±3.54	93.55±1.21	93.62±1.51
	Avg. Acc.	78.03±4.74	76.89±4.82	89.07±2.30	89.50±2.81
	Kappa	81.89±3.65	80.89±3.69	93.51±1.22	93.59±1.52

TABLE X

OOD DETECTION RESULTS OF THE PROPOSED METHOD ON THE OPEN-SET RECOGNITION TASK ON THE MIXTURE OF THE UCM, THE AID, AND THE LCZ42 DATASET. UCM IS USED AS IN-DISTRIBUTION AND THE OTHER TWO DATASETS FOR OOD FOR TRAINING AND TESTING. ONE CAN CLEARLY SEE THAT THE LCZ42 DATASET WITH A MUCH LOWER RESOLUTION THAN UCM IS DETECTED SIGNIFICANTLY BETTER AS OOD

	UCM-AID-LCZ42	UCM-LCZ42-AID
Max. Prob.	100 ± 0.0	89.66 ± 0.51
Entropy	100 ± 0.0	90.07 ± 0.60
Mutual Info	100 ± 0.0	81.32 ± 1.01
α_0	100 ± 0.0	87.17 ± 0.78

D. Channel Separation

For the channel separation, we use the R-, G-, and B-channels of the samples of the three datasets. All classes are considered, but each sample is separated into an in-domain channel, an OOD channel for training, and an OOD channel for testing. The in-domain classification results and OOD detection indices for the AID, UCM, and LCZ42 datasets are tabulated in Tables XI–XVI. For all datasets, the DPN-RS and DPN^+ provide the best OOD detection performance, with DPN-RS reaching four out of six top scores. For the UCM and the AID dataset, DPN^{rev}performs worse on the OOD detection than DPN-RS and DPN^+ but better than DPN^{forw} and the ENN approach. The two settings where the blue channel

TABLE XI

OOD DETECTION UNDER SENSOR SHIFT IN THE AID DATASET. THE PERFORMANCE IS MEASURED BY 100× THE AUROC. THE RESULTS ARE GIVEN AS MEAN AND STANDARD DEVIATION COMPUTED BASED ON SEVEN RUNS. THE BEST RESULTS PER APPROACH ARE GIVEN BOLDFACED AND THE BEST RESULTS ON THE SINGLE SETTINGS OVER ALL APPROACHES ARE ITALICIZED

		R-G-B Split	R-B-G Split	G-R-B Split	G-B-R Split	B-R-G Split	B-G-R Split
DPN-RS	Max. Prob.	67.24±0.60	63.51±1.21	70.22±0.69	***69.29±0.59***	77.63±0.61	74.84±1.01
	Entropy	67.31±0.62	63.61±1.21	70.38±0.73	68.44±0.59	***77.98±0.66***	74.84±1.53
	Mutual Info	67.53±0.90	53.41±1.21	70.55±1.08	68.25±1.63	77.13±1.49	74.21±1.66
	α_0	***67.58±0.92***	***63.69±1.16***	**70.61±1.11**	68.29±1.69	77.15±1.51	***75.86±1.69***
DPN$^+$	Max. Prob.	67.08±0.76	63.32±1.53	70.52±0.72	68.35±0.52	77.51±0.71	74.13±0.57
	Entropy	**67.09±0.65**	63.52±1.39	***70.81±0.79***	**68.56±0.55**	**77.89±0.79**	74.13±0.58
	Mutual Info	66.11±0.91	63.31±1.54	69.93±1.24	67.47±1.35	76.81±1.53	73.12±0.65
	α_0	66.21±0.88	**63.26±1.45**	70.08±1.41	67.51±1.49	76.88±1.74	**75.46±1.97**
DPN$^{\text{rev}}$	Max. Prob.	65.25±0.64	61.69±2.97	66.30±1.21	63.16±2.09	71.18±0.37	71.89±1.24
	Entropy	65.64±0.63	61.94±3.05	66.58±1.17	63.52±2.07	71.77±0.31	72.29±1.17
	Mutual Info	66.12±0.70	62.24±3.16	66.88±1.17	64.04±1.95	**72.33±0.17**	72.83±1.06
	α_0	**66.13±0.72**	**62.25±3.15**	**66.89±1.75**	**64.07±1.93**	72.31±0.17	**72.84±1.05**
DPN$^{\text{forw}}$	Max. Prob.	60.84±2.28	59.31±2.58	63.41±2.69	62.79±1.93	66.45±4.65	69.72±1.74
	Entropy	**61.25±2.10**	**59.83±2.47**	**63.82±2.36**	**63.53±1.42**	**67.20±4.43**	**69.94±1.75**
	Mutual Info	45.70±3.39	51.03±5.23	46.73±4.97	51.58±8.09	44.76±6.85	46.75±10.50
	α_0	44.45±3.08	50.05±5.22	45.37±4.66	50.53±7.86	42.87±6.47	45.06±9.81
ENN	Max. Prob.	52.41±0.58	51.97±0.51	51.17±0.83	51.19±0.33	51.99±0.41	**52.61±0.38**
	Entropy	**52.43±0.58**	51.97±0.49	**51.19±0.81**	51.19±0.35	52.00±0.43	52.58±0.37
	Mutual Info	52.42±0.52	51.98±0.49	51.15±0.78	**51.40±0.39**	**52.02±0.50**	52.57±0.36
	α_0	52.41±0.46	**51.99±0.54**	51.00±0.75	51.33±0.45	52.00±0.58	52.59±0.37

TABLE XII

CLASSIFICATION ACCURACY AND AVERAGE ACCURACY ON THE SENSOR SHIFT TASKS ON THE AID DATASET. THE RESULTS ARE GIVEN AS MEAN AND STANDARD DEVIATION COMPUTED BASED ON SEVEN RUNS. THE BEST AVERAGE ACCURACY FOR EACH SETTING IS HIGHLIGHTED

		R-G-B	R-B-G	G-R-B	G-B-R	B-R-G	B-G-R
DPN-RS	Accuracy	87.90±1.41	88.63±0.95	86.85±0.98	87.51±1.23	85.64±1.51	84.25±2.10
	Avg. Acc.	87.71±1.46	88.30±1.10	86.39±1.05	87.20±1.37	85.31±1.62	83.87±2.15
	Kappa	87.88±1.41	88.61±0.95	86.84±0.98	87.50±1.23	85.62±1.52	84.22±2.10
DPN$^+$	Accuracy	87.95±0.71	87.36±1.82	85.01±2.14	85.01±2.29	83.10±4.41	84.44±3.86
	Avg. Acc.	87.56±0.72	87.02±1.84	84.65±2.28	84.65±2.24	82.81±4.14	84.27±2.86
	Kappa	87.93±0.71	87.34±1.82	84.99±2.14	84.98±2.30	80.74±4.42	84.41±3.89
DPN$^{\text{rev}}$	Accuracy	90.67±1.59	89.50±4.69	90.75±1.07	91.09±0.39	89.75±2.10	71.19±0.69
	Avg. Acc.	**90.25±1.63**	**89.13±4.83**	**90.41±0.97**	**90.73±0.36**	**89.34±2.21**	**90.81±0.74**
	Kappa	90.66±1.60	89.48±4.71	90.74±1.07	91.08±0.39	89.75±2.10	91.18±0.69
DPN$^{\text{forw}}$	Accuracy	82.39±3.22	82.65±2.53	77.53±9.62	83.77±2.74	57.33±7.24	76.27±8.70
	Avg. Acc.	81.65±3.41	81.77±2.71	76.72±9.73	82.99±3.00	74.09±7.91	75.67±9.10
	Kappa	82.38±3.22	82.64±2.53	77.51±9.63	83.76±2.74	75.31±7.24	76.25±8.70
ENN	Accuracy	85.13±4.14	85.69±3.69	87.54±1.24	87.28±1.09	85.56±4.05	86.69±2.51
	Avg. Acc.	84.65±4.49	85.10±3.61	87.18±1.24	86.74±1.71	85.00±4.25	86.25±2.40
	Kappa	85.12±4.14	85.68±3.69	87.53±1.24	87.27±1.09	85.56±4.05	86.69±2.51

TABLE XIII

OOD DETECTION UNDER SENSOR SHIFT IN THE UCM DATASET. THE PERFORMANCE IS MEASURED BY 100× THE AUROC. THE RESULTS ARE GIVEN AS MEAN AND STANDARD DEVIATION COMPUTED BASED ON SEVEN RUNS. THE BEST RESULTS PER APPROACH ARE GIVEN BOLDFACED AND THE BEST RESULTS ON THE SINGLE SETTINGS OVER ALL APPROACHES ARE ITALICIZED

		R-G-B Split	R-B-G Split	G-R-B Split	G-B-R Split	B-R-G Split	B-G-R Split
DPN-RS	Max. Prob.	78.07±0.93	66.83±1.32	**57.29±0.93**	***56.06±0.82***	79.29±1.12	81.58±2.01
	Entropy	***78.08±0.89***	67.00±1.40	57.21±0.89	***56.06±0.80***	79.58±1.19	**81.88±2.14**
	Mutual Info	77.54±0.85	66.84±2.04	56.71±1.02	54.67±0.79	***80.76±1.19***	81.47±2.87
	α_0	77.51±0.84	**67.11±2.01**	56.71±0.98	54.83±0.71	80.63±1.20	81.40±2.87
DPN$^+$	Max. Prob.	77.85±1.42	67.96±0.99	***58.39±1.15***	**55.80±1.62**	79.00±1.22	82.32±1.62
	Entropy	**77.97±1.45**	***68.06±0.97***	58.35±1.11	55.76±1.68	**79.24±1.22**	***82.50±1.68***
	Mutual Info	75.99±1.55	67.44±1.11	56.87±1.35	54.77±1.72	79.06±1.16	81.89±1.58
	α_0	76.04±1.53	67.68±1.13	56.90±1.33	55.04±1.75	78.98±1.19	81.92±1.63
DPN$^{\text{rev}}$	Max. Prob.	71.43±2.36	65.02±1.18	57.61±1.60	**54.75±1.02**	75.07±1.87	79.87±1.97
	Entropy	71.81±2.29	65.37±1.17	**57.68±1.59**	54.73±1.04	75.47±1.91	80.37±1.94
	Mutual Info	**71.95±2.45**	**65.77±1.15**	57.55±1.67	54.64±0.98	**76.11±1.93**	**80.89±1.96**
	α_0	71.90±2.51	65.75±1.15	57.53±1.68	54.64±0.97	**76.11±1.93**	80.85±1.97
DPN$^{\text{forw}}$	Max. Prob.	62.97±2.80	60.23±1.69	**54.39±2.31**	53.65±1.11	63.65±4.45	66.41±2.76
	Entropy	**63.00±2.95**	**60.92±1.65**	54.26±2.32	**53.67±0.99**	**64.63±4.61**	**67.00±2.89**
	Mutual Info	48.54±2.39	52.38±2.34	49.72±3.87	50.37±0.61	47.60±3.94	43.21±3.32
	α_0	46.98±2.38	51.13±2.22	49.40±4.02	49.94±0.51	46.72±3.59	41.34±3.03
ENN	Max. Prob.	52.31±0.73	50.25±0.45	52.02±0.83	51.32±0.55	**51.31±0.53**	**52.66±1.38**
	Entropy	52.35±0.75	50.25±0.46	**52.04±0.87**	51.36±0.53	51.29±0.53	52.59±1.39
	Mutual Info	52.36±0.85	50.40±0.40	52.01±0.85	**51.37±0.56**	51.12±0.73	52.43±1.36
	α_0	**52.47±0.81**	**50.62±0.28**	51.99±0.79	51.24±0.65	50.91±0.97	52.48±1.31

TABLE XIV

CLASSIFICATION ACCURACY AND AVERAGE ACCURACY ON THE SENSOR SHIFT TASKS ON THE UCM DATASET. THE RESULTS ARE GIVEN AS MEAN AND STANDARD DEVIATION COMPUTED BASED ON SEVEN RUNS. THE BEST AVERAGE ACCURACY FOR EACH SETTING IS HIGHLIGHTED

		R-G-B	R-B-G	G-R-B	G-B-R	B-R-G	B-G-R
DPN-RS	Accuracy	87.85±1.75	88.65±1.91	85.86±2.09	87.62±2.65	88.72±1.92	86.70±2.28
	Avg. Acc.	87.85±1.74	88.00±2.04	85.83±2.09	87.16±2.70	88.73±1.93	86.69±2.33
	Kappa	87.82±1.76	88.62±1.91	85.82±2.09	87.59±2.66	88.69±1.92	86.67±2.28
DPN^{+}	Accuracy	87.83±1.46	88.49±1.99	86.98±1.99	89.43±2.16	88.79±1.61	87.95±2.10
	Avg. Acc.	87.79±1.50	88.48±1.94	86.99±1.95	89.46±2.16	88.81±1.61	87.92±2.12
	Kappa	87.80±1.46	88.46±1.95	86.95±1.99	89.40±2.16	88.76±1.62	87.92±2.11
DPN^{rev}	Accuracy	94.76±0.64	94.92±0.62	95.61±0.71	95.56±1.07	95.07±1.42	94.95±0.89
	Avg. Acc.	**94.80±0.66**	**94.94±0.62**	**95.60±0.72**	**95.57±0.97**	**95.07±1.41**	**94.96±0.88**
	Kappa	94.75±0.64	94.91±0.63	95.60±0.71	95.54±0.99	95.05±1.42	94.94±0.89
DPN^{forw}	Accuracy	84.61±2.46	84.36±4.45	83.74±2.75	83.83±2.13	83.60±4.63	81.56±3.36
	Avg. Acc.	82.33±3.92	83.10±2.46	83.78±2.70	83.84±2.17	82.30±1.84	81.55±3.40
	Kappa	84.57±2.47	84.32±4.46	83.69±2.77	83.80±2.14	83.56±4.64	81.50±3.38
ENN	Accuracy	87.29±4.31	86.42±6.29	85.31±6.05	86.72±5.37	86.05±7.22	83.19±6.95
	Avg. Acc.	87.30±4.30	86.38±6.37	85.29±6.09	86.76±5.40	96.14±6.17	83.18±5.91
	Kappa	87.27±4.31	86.39±6.03	85.28±6.07	86.71±5.37	86.03±7.23	83.16±6.95

TABLE XV

OOD DETECTION UNDER SENSOR SHIFT IN THE SO2SAT LCZ42 DATASET. THE PERFORMANCE IS MEASURED BY 100× THE AUROC. THE RESULTS ARE GIVEN AS MEAN AND STANDARD DEVIATION COMPUTED BASED ON SEVEN RUNS. THE BEST RESULTS PER APPROACH ARE GIVEN BOLDFACED AND THE BEST RESULTS ON THE SINGLE SETTINGS OVER ALL APPROACHES ARE ITALICIZED

		R-G-B Split	R-B-G Split	G-R-B Split	G-B-R Split	B-R-G Split	B-G-R Split
DPN-RS	Max. Prob.	91.79±0.20	78.34±0.09	**56.38±0.53**	**43.41±0.77**	80.57±0.21	96.19±0.12
	Entropy	93.81±0.17	79.67±0.09	54.49±0.52	39.26±0.66	82.85±0.22	96.78±0.10
	Mutual Info	***95.52±0.29***	***82.27±0.08***	35.59±0.32	26.80±2.02	***86.63±0.24***	***96.83±0.05***
	α_0	***95.52±0.38***	82.00±0.13	35.59±0.32	26.72±2.04	86.55±0.23	96.18±0.03
DPN^{+}	Max. Prob.	91.97±0.19	78.13±0.16	**55.44±0.72**	**43.82±0.40**	80.65±0.47	96.17±0.15
	Entropy	**94.00±0.12**	79.48±0.14	53.38±0.79	39.42±0.36	83.01±0.40	**96.73±0.15**
	Mutual Info	92.94±0.34	**80.59±0.46**	41.80±1.52	24.00±0.84	**83.19±0.43**	88.46±2.12
	α_0	92.94±0.34	80.59±0.45	41.80±1.52	24.10±0.82	83.19±0.43	88.46±2.12
DPN^{rev}	Max. Prob.	81.51±0.92	56.85±6.28	**57.17±0.41**	**46.62±1.69**	74.68±1.30	92.14±0.81
	Entropy	86.76±0.85	**59.16±8.79**	55.58±0.67	41.84±3.23	77.74±1.15	93.61±0.67
	Mutual Info	93.71±0.34	48.49±16.57	49.15±0.936	32.65±4.70	82.34±1.00	95.34±0.51
	α_0	**93.88±0.32**	47.76±16.63	48.83±0.90	32.68±4.90	**82.47±1.06**	**95.39±0.50**
DPN^{forw}	Max. Prob.	71.89±4.53	70.14±1.86	51.48±1.02	41.32±0.13	70.25±1.57	**90.32±1.87**
	Entropy	**76.22±4.67**	**72.05±0.99**	50.81±0.92	38.72±0.83	**71.72±1.36**	90.29±2.02
	Mutual Info	12.26±3.10	30.70±5.94	59.40±0.64	53.61±1.30	25.58±4.19	9.45±3.50
	α_0	11.80±2.71	29.25±5.45	**59.90±0.67**	**54.63±1.00**	24.29±3.29	8.00±2.62
ENN	Max. Prob.	58.89±0.70	**59.96±0.22**	***59.98±0.46***	53.40±1.24	49.62±0.85	37.07±1.58
	Entropy	**58.97±0.73**	59.90±0.26	59.88±0.48	52.81±1.31	49.13±0.84	36.13±1.56
	Mutual Info	58.17±1.23	59.40±0.18	59.46±0.42	53.15±1.36	49.83±0.50	35.26±0.44
	α_0	56.83±2.02	58.42±0.33	58.80±0.37	***56.85±0.94***	**53.93±0.88**	**40.23±1.13**

TABLE XVI

CLASSIFICATION ACCURACY AND AVERAGE ACCURACY ON THE SENSOR SHIFT TASKS ON THE LCZ42 DATASET. THE RESULTS ARE GIVEN AS MEAN AND STANDARD DEVIATION COMPUTED BASED ON SEVEN RUNS. THE BEST AVERAGE ACCURACY FOR EACH SETTING IS HIGHLIGHTED

		R-G-B	R-B-G	G-R-B	G-B-R	B-R-G	B-G-R
DPN-RS	Accuracy	77.23±0.38	77.80±0.48	78.07±0.24	78.54±0.21	76.72±0.35	76.73±0.43
	Avg. Acc.	65.99±0.59	67.22±0.97	67.70±0.56	**68.36±0.74**	67.16±0.13	66.70±0.66
	Kappa	77.01±0.38	77.58±0.49	77.85±0.25	78.32±0.21	76.48±0.35	76.50±0.42
DPN^{+}	Accuracy	77.79±0.12	77.91±0.29	78.43±0.43	78.57±0.15	76.79±0.19	77.03±0.08
	Avg. Acc.	67.03±0.17	**67.83±0.29**	**68.23±0.91**	68.00±0.26	**67.62±0.49**	**66.94±0.03**
	Kappa	77.56±0.12	77.69±0.29	78.20±0.44	78.36±0.15	76.56±0.20	76.80±0.07
DPN^{rev}	Accuracy	78.27±0.70	43.13±6.17	76.71±0.12	62.85±7.04	74.81±1.77	75.86±1.74
	Avg. Acc.	**67.17±1.31**	26.98±4.48	65.20±0.08	43.64±10.70	63.29±3.14	64.66±2.78
	Kappa	78.04±0.70	42.13±6.31	76.46±0.12	62.39±7.16	68.59±8.54	64.91±9.19
DPN^{forw}	Accuracy	45.24±9.30	56.99±6.49	65.50±1.72	62.48±5.70	53.86±8.96	62.00±3.31
	Avg. Acc.	28.53±5.18	36.64±7.26	46.45±3.30	42.96±7.44	34.68±8.23	41.32±4.58
	Kappa	44.57±9.17	56.39±6.64	65.13±1.76	62.01±5.85	53.17±9.23	61.49±3.41
ENN	Accuracy	75.85±0.15	76.00±0.27	76.71±0.36	76.44±0.17	75.43±0.19	75.37±0.45
	Avg. Acc.	64.94±0.65	65.09±0.65	66.12±0.59	65.82±0.66	65.15±0.49	65.02±1.01
	Kappa	76.84±0.15	0.76±0.27	76.11±0.36	76.43±0.17	75.43±0.19	75.19±0.51

TABLE XVII

OOD DETECTION UNDER REGION SHIFT IN ALL CLASSES OF THE SO2SAT LCZ42 DATASET. THE PERFORMANCE IS MEASURED BY 100× THE AUROC. THE RESULTS ARE GIVEN AS MEAN AND STANDARD DEVIATION COMPUTED BASED ON SEVEN RUNS. THE BEST RESULTS PER APPROACH ARE GIVEN BOLDFACED AND THE BEST RESULTS ON THE SINGLE SETTINGS OVER ALL APPROACHES ARE ITALICIZED

		R1-R2-R3	R1-R3-R2	R2-R1-R3	R2-R3-R1	R3-R1-R2	R3-R2-R1
DPN-RS	Max. Prob.	83.57±1.92	86.81±1.39	91.76±4.11	82.33±2.20	95.33±1.12	88.58±1.22
	Entropy	84.56±1.87	***87.76±1.24***	93.84±2.16	***82.71±2.24***	95.80±1.09	88.95±1.22
	Mutual Info	83.59±2.53	83.63±1.25	94.97±2.16	79.64±4.13	96.71±1.20	90.82±1.01
	α_0	***85.32±2.38***	85.77±1.22	**95.10±2.05**	80.88±4.35	**96.76±1.19**	***91.02±0.98***
DPN$^+$	Max. Prob.	83.37±1.95	85.70±2.39	94.90±0.44	80.86±2.43	96.70±0.92	88.04±2.43
	Entropy	**84.02±2.02**	**86.52±2.13**	***95.72±0.48***	**81.24±2.51**	97.27±0.83	88.46±2.47
	Mutual Info	80.63±4.66	80.18±1.73	95.32±0.90	72.21±3.77	97.36±0.44	88.44±2.22
	α_0	82.53±3.81	82.59±1.68	95.41±0.90	73.77±3.12	***97.39±0.44***	**89.83±2.38**
DPN$^{\text{rev}}$	Max. Prob.	76.14±3.62	**83.81±1.91**	64.48±7.57	78.22±3.75	88.48±1.73	88.00±2.51
	Entropy	**77.68±2.91**	83.59±1.71	66.72±9.97	**78.24±4.02**	90.58±1.71	89.97±1.97
	Mutual Info	77.51±1.90	81.51±2.31	**71.81±6.79**	76.28±3.07	**92.97±1.25**	**89.99±2.03**
	α_0	77.46±1.94	81.69±2.34	70.44±7.89	76.34±2.98	92.81±1.29	89.95±2.33
DPN$^{\text{forw}}$	Max. Prob.	82.32±1.85	86.88±0.67	**91.02±0.96**	73.71±2.09	92.69±2.32	**87.10±0.99**
	Entropy	**82.40±1.91**	**87.07±0.63**	90.91±1.63	**75.17±2.10**	**93.91±2.05**	86.87±0.78
	Mutual Info	75.59±1.09	82.08±1.03	69.81±9.19	62.57±5.19	75.58±5.97	77.21±1.37
	α_0	74.93±1.14	81.61±1.17	66.46±8.71	60.11±5.39	75.21±6.82	75.88±1.50
ENN	Max. Prob.	67.89±2.54	70.60±3.62	63.67±3.78	61.84±2.88	69.50±1.74	70.86±1.87
	Entropy	67.93±2.56	**70.71±3.59**	63.64±3.82	61.76±2.98	69.58±1.70	**70.92±1.86**
	Mutual Info	67.95±2.33	70.60±3.47	63.98±3.93	61.82±3.37	**69.63±1.54**	70.76±1.74
	α_0	**68.47±2.02**	69.73±3.42	**65.97±4.03**	**63.60±3.42**	69.06±1.86	69.68±1.99

4

TABLE XVIII

CLASSIFICATION ACCURACY AND AVERAGE ACCURACY ON THE REGION SHIFT TASKS ON ALL CLASSES OF THE LCZ42 DATASET. THE RESULTS ARE GIVEN AS MEAN AND STANDARD DEVIATION COMPUTED BASED ON SEVEN RUNS. THE BEST AVERAGE ACCURACY ON IN-DISTRIBUTION SAMPLES IS HIGHLIGHTED FOR EACH SETTING

		R1-R2-R3	R1-R3-R2	R2-R1-R3	R2-R3-R1	R3-R1-R2	R3-R2-R1
DPN-RS	Accuracy	94.99±0.82	94.47±0.61	92.06±0.88	92.64±1.07	94.87±1.48	95.50±0.78
	Avg. Acc.	**92.34±0.82**	91.52±0.94	88.05±1.17	**89.60±1.33**	93.07±1.65	**93.48±1.05**
	Kappa	94.68±0.51	64.10±0.64	90.40±0.88	92.62±1.08	94.76±1.51	95.41±0.79
	Accuracy (Out)	44.62±2.15	39.12±3.20	25.90±4.15	34.94±3.25	29.56±4.00	31.70±1.92
	Avg. Acc. (Out)	30.84±2.36	32.68±4.21	18.96±2.93	26.11±4.42	18.12±2.06	20.33±0.85
	Kappa (Out)	43.80±2.12	38.03±3.23	24.91±4.58	33.09±4.53	29.52±4.08	25.98±4.41
DPN$^+$	Accuracy	95.05±0.35	95.51±0.45	93.47±0.83	92.74±0.81	94.93±0.33	94.99±0.76
	Avg. Acc.	91.54±1.00	**93.10±1.09**	**88.94±1.49**	88.62±0.88	**93.62±0.79**	92.72±0.66
	Kappa	94.72±0.37	95.23±0.46	93.45±0.83	92.13±0.81	94.82±0.34	94.49±0.76
	Accuracy (Out)	42.60±1.31	38.79±2.35	26.80±3.42	33.22±3.52	29.16±3.96	32.02±1.76
	Avg. Acc. (Out)	29.00±1.20	25.06±2.43	18.51±1.70	23.46±3.46	18.14±2.00	21.24±1.11
	Kappa (Out)	41.45±0.95	38.45±2.24	25.93±0.03	31.97±3.77	28.66±4.15	29.87±2.04
DPN$^{\text{rev}}$	Accuracy	88.35±3.36	93.77±1.56	65.92±9.14	92.00±3.05	90.99±1.28	95.06±1.71
	Avg. Acc.	75.23±6.95	88.08±2.64	40.99±11.50	84.71±4.77	83.72±2.09	91.11±3.12
	Kappa	87.52±3.61	93.34±1.67	65.88±9.12	91.98±3.52	90.76±1.30	94.45±1.76
	Accuracy (Out)	53.57±3.47	47.31±0.87	48.62±0.73	40.10±2.42	46.56±2.06	37.19±2.48
	Avg. Acc. (Out)	34.04±2.10	30.04±0.67	28.54±3.00	28.05±3.51	30.88±1.70	25.94±2.06
	Kappa (Out)	52.04±3.71	46.95±0.83	48.28±0.59	38.34±1.88	46.15±2.09	34.38±2.85
DPN$^{\text{forw}}$	Accuracy	89.47±1.67	92.22±0.73	84.77±5.98	81.14±6.08	89.89±3.65	92.52±0.12
	Avg. Acc.	79.47±4.93	83.51±1.23	69.48±7.61	64.97±8.35	83.00±5.28	86.42±0.74
	Kappa	90.89±1.73	92.19±0.59	84.43±6.40	81.70±7.44	80.72±3.53	92.44±0.82
	Accuracy (Out)	46.22±3.24	41.67±0.92	26.71±5.26	38.01±2.47	32.22±1.63	33.64±1.73
	Avg. Acc. (Out)	31.68±2.27	26.55±1.42	19.96±2.08	27.77±3.31	20.24±1.52	22.11±2.31
	Kappa (Out)	46.37±2.94	41.54±0.78	25.84±5.86	37.81±1.97	32.09±1.51	30.56±2.34
ENN	Accuracy	93.56±1.12	91.47±1.99	88.25±3.43	89.77±1.79	92.83±1.62	90.56±3.10
	Avg. Acc.	88.50±1.40	84.93±4.27	79.94±3.92	82.89±2.76	88.41±2.61	85.63±4.42
	Kappa	93.15±1.18	90.91±2.12	88.21±3.43	89.74±1.78	92.66±1.65	90.33±3.20
	Accuracy (Out)	53.75±2.81	47.84±2.22	49.86±3.92	45.87±3.55	44.87±6.38	42.38±4.45
	Avg. Acc. (Out)	35.62±2.10	30.96±2.66	33.66±2.84	31.66±4.12	31.53±5.13	28.19±2.17
	Kappa (Out)	52.46±3.00	47.57±2.23	49.47±3.96	43.94±3.98	42.91±6.82	36.87±3.56

is considered as in-domain provide the best OOD separation performance, while other settings provide less satisfactory OOD separation results. Remarkably, the performance of DPN-RS is superior for the more challenging LCZ42 dataset in comparison to the other datasets. At the same time, DPN-RS provides competitive in-domain prediction accuracies for the LCZ42 dataset by performing best or at most 1% below the best performing approach. In general, the performance is lower for the open-set recognition experiments.

E. Location Separation

The location separation experiments are conducted only on the LCZ42 dataset as any location information for the other two datasets is not available. We form three sets of regions contained in the LCZ42 dataset.

1) *Europe and North America:* Amsterdam, Cologne, London, Zurich, Los Angeles, Melbourne,

TABLE XIX

OOD DETECTION UNDER REGION SHIFT IN THE URBAN CLASSES OF THE SO2SAT LCZ42 DATASET. THE PERFORMANCE IS MEASURED BY 100× THE AUROC. THE RESULTS ARE GIVEN AS MEAN AND STANDARD DEVIATION COMPUTED BASED ON SEVEN RUNS. THE BEST RESULTS PER APPROACH ARE GIVEN BOLDFACED AND THE BEST RESULTS ON THE SINGLE SETTINGS OVER ALL APPROACHES ARE ITALICIZED

		R1-R2-R3	R1-R3-R2	R2-R1-R3	R2-R3-R1	R3-R1-R2	R3-R2-R1
DPN-RS	Max. Prob.	82.33±3.18	87.69±2.21	93.39±1.72	77.54±2.57	97.20±0.88	92.73±1.33
	Entropy	**83.16±3.18**	***88.95±2.16***	94.29±1.67	**77.91±2.50**	97.59±0.80	93.20±1.31
	Mutual Info	79.65±6.93	87.56±4.59	96.71±1.39	73.24±3.67	***98.39±0.55***	94.65±0.91
	α_0	80.83±5.69	88.25±4.04	***96.72±1.40***	73.39±3.78	***98.39±0.54***	**94.69±0.94**
DPN$^+$	Max. Prob.	84.28±2.33	86.10±2.11	93.63±1.97	79.88±4.15	97.24±0.77	93.65±0.79
	Entropy	85.11±2.39	87.27±2.02	94.55±1.87	***80.50±4.10***	**97.75±0.61**	94.18±0.73
	Mutual Info	85.54±3.31	87.57±2.57	95.13±2.47	71.80±5.33	97.49±1.30	95.52±0.87
	α_0	***85.97±3.18***	**88.12±2.35**	**95.17±2.45**	72.56±4.85	97.49±1.30	***95.55±0.88***
DPN$^{\text{rev}}$	Max. Prob.	80.51±1.33	79.60±1.14	65.64±8.28	67.45±2.29	95.24±1.27	91.71±0.92
	Entropy	**81.72±1.00**	**80.67±1.08**	66.97±9.75	**67.61±2.57**	95.82±1.21	92.34±0.91
	Mutual Info	80.52±1.27	75.87±1.31	72.89±7.31	63.14±3.17	96.27±1.09	**93.11±0.99**
	α_0	80.54±1.18	75.65±0.91	**76.62±7.02**	63.18±3.18	**96.31±1.23**	93.04±1.11
DPN$^{\text{forw}}$	Max. Prob.	**77.80±4.67**	79.24±2.16	60.57±6.63	**66.81±6.93**	93.01±0.94	**88.71±1.67**
	Entropy	77.07±5.28	**79.59±2.04**	**62.55±5.15**	64.88±7.83	**93.83±0.76**	88.65±1.97
	Mutual Info	62.28±1.035	71.02±2.11	29.00±8.18	43.99±5.53	52.81±5.18	58.91±4.70
	α_0	60.66±1.73	69.50±2.59	28.29±8.61	43.23±4.92	49.40±4.98	55.54±4.44
ENN	Max. Prob.	73.51±1.50	72.97±2.60	56.65±2.76	64.40±3.11	75.41±2.56	72.61±3.81
	Entropy	**73.59±1.53**	**73.10±2.64**	56.55±2.74	64.34±3.11	**75.59±2.55**	**72.74±3.86**
	Mutual Info	72.95±1.46	72.85±2.66	56.81±2.74	63.31±10.01	75.52±3.21	72.39±3.80
	α_0	71.40±1.78	72.22±2.58	**58.85±3.26**	**65.20±2.21**	74.45±5.32	71.11±3.63

TABLE XX

CLASSIFICATION ACCURACY AND AVERAGE ACCURACY ON THE SAMPLES OF THE REGION SHIFT TASK ON THE URBAN CLASSES OF THE LCZ42 DATASET. THE RESULTS ARE GIVEN AS MEAN AND STANDARD DEVIATION COMPUTED BASED ON SEVEN RUNS. THE BEST AVERAGE ACCURACY ON IN-DISTRIBUTION SAMPLES IS HIGHLIGHTED FOR EACH SETTING

		R1-R2-R3	R1-R3-R2	R2-R1-R3	R2-R3-R1	R3-R1-R2	R3-R2-R1
DPN-RS	Accuracy	92.40±1.61	92.65±0.74	91.00±1.01	89.83±0.85	92.75±1.74	92.03±2.62
	Avg. Acc.	88.12±1.65	**89.50±2.18**	**89.28±1.31**	**86.06±0.82**	91.17±2.26	90.05±2.72
	Kappa	91.91±1.71	92.17±0.79	90.67±1.01	89.81±0.86	92.61±1.77	91.85±2.68
	Accuracy (Out)	37.57±3.31	28.83±2.85	21.17±3.51	24.71±0.99	23.38±2.19	27.62±4.49
	Avg. Acc. (Out)	27.94±1.54	23.56±2.31	19.39±2.31	26.48±2.34	20.72±1.17	21.21±3.42
	Kappa (Out)	35.58±3.34	28.93±2.96	20.26±3.41	23.85±0.95	23.89±2.69	23.91±4.15
DPN$^+$	Accuracy	92.65±0.82	92.53±1.74	90.08±1.30	88.28±1.81	92.39±1.35	92.49±2.28
	Avg. Acc.	**88.79±1.88**	88.81±1.80	89.22±1.24	84.92±2.58	91.00±1.48	**91.20±1.47**
	Kappa	92.16±0.92	92.09±1.82	90.05±1.30	88.24±1.81	92.23±1.37	92.33±2.32
	Accuracy (Out)	38.42±3.88	29.98±1.91	24.29±3.19	27.22±2.76	23.62±2.95	27.89±4.12
	Avg. Acc. (Out)	29.73±2.27	23.85±1.38	19.68±3.06	28.25±2.33	20.81±1.68	22.60±3.19
	Kappa (Out)	37.14±4.07	29.71±2.00	24.10±3.30	25.23±2.88	23.77±2.93	24.60±3.15
DPN$^{\text{rev}}$	Accuracy	91.03±2.99	84.87±1.76	69.98±7.21	79.73±7.85	95.16±0.48	94.05±1.31
	Avg. Acc.	85.25±5.93	74.82±1.28	55.17±8.93	70.53±13.31	**92.45±0.63**	90.92±1.60
	Kappa	90.45±3.20	83.83±2.35	69.86±7.25	79.67±7.88	95.07±0.51	94.38±1.30
	Accuracy (Out)	39.43±2.01	33.08±1.87	38.69±0.89	37.77±4.07	31.75±2.81	32.67±2.03
	Avg. Acc. (Out)	28.93±2.60	25.03±2.21	30.73±1.91	34.71±3.81	29.37±2.01	24.72±1.44
	Kappa (Out)	37.65±2.02	32.33±2.55	37.27±1.02	37.39±4.13	31.29±3.10	28.89±1.75
DPN$^{\text{forw}}$	Accuracy	87.23±4.08	87.31±2.37	61.55±11.40	66.64±14.02	81.93±4.88	79.02±2.13
	Avg. Acc.	77.09±4.03	76.93±2.52	43.69±12.46	49.47±13.69	74.57±7.97	66.72±1.72
	Kappa	86.43±4.34	86.46±2.51	61.45±11.42	66.55±14.04	81.58±0.05	78.72±2.12
	Accuracy (Out)	35.30±3.50	27.32±2.17	30.53±1.54	28.71±2.65	27.16±2.13	25.92±1.67
	Avg. Acc. (Out)	24.60±2.00	21.39±1.41	21.27±1.40	26.01±4.95	25.46±3.63	22.10±1.79
	Kappa (Out)	35.08±3.92	26.02±2.37	29.30±0.99	26.19±3.46	26.28±1.97	24.34±1.99
ENN	Accuracy	91.20±1.31	90.14±1.80	87.51±2.28	86.91±2.89	89.73±1.44	88.33±4.19
	Avg. Acc.	84.37±2.66	84.89±1.45	83.53±3.97	80.31±5.59	86.57±1.69	84.50±6.03
	Kappa	90.62±1.42	89.51±1.91	87.47±2.29	86.89±2.89	89.50±1.48	88.08±4.28
	Accuracy (Out)	41.63±3.81	35.51±1.36	44.31±4.71	39.98±4.27	35.06±4.31	43.17±7.38
	Avg. Acc. (Out)	30.66±3.47	27.31±1.60	31.65±0.98	38.62±3.25	32.95±2.29	37.06±2.41
	Kappa (Out)	39.54±30.10	34.24±1.80	44.24±4.44	38.25±3.22	34.81±3.37	39.79±5.04

Madrid, Paris, Milan, Rome, Philadelphia, New and York.

2) *China and Japan:* Beijing, Changsha, Dongying, Hongkong, Wuhan, Tokyo, Shenzhen, Shanghai, Qingdao, Nanjing, and Kyoto.
3) *South America, Africa, and Middle East:* Cairo, Capetown, Islamabad, Istanbul, Dhaka, Lima, Orangitown, Caracas, Bogota, Sao Paulo, Salvador, and Rio de Janeiro.

The three groups exhibit distinct characteristics. While group 1 contains less high-rise buildings, group 2 contains many high-rise buildings. In contrast to this, group 3 contains many disorganized crowded settlements.

In order to evaluate the performance with a special focus on the urban and the vegetation classes, we apply three different types of experiments. For the first setting, we consider all classes, for the second setting, we only use the urban classes 1-10, and for the third setting, we only use the

TABLE XXI

OOD DETECTION UNDER REGION SHIFT IN THE VEGETATION CLASSES OF THE SO2SAT LCZ42 DATASET. THE PERFORMANCE IS MEASURED BY 100× THE AUROC. THE RESULTS ARE GIVEN AS MEAN AND STANDARD DEVIATION COMPUTED BASED ON SEVEN RUNS. THE BEST RESULTS PER APPROACH ARE GIVEN BOLDFACED AND THE BEST RESULTS ON THE SINGLE SETTINGS OVER ALL APPROACHES ARE ITALICIZED

		R1-R2-R3	R1-R3-R2	R2-R1-R3	R2-R3-R1	R3-R1-R2	R3-R2-R1
DPN-RS	Max. Prob.	87.79±3.86	92.11±1.58	93.77±1.92	72.22±1.14	95.44±2.66	91.52±3.09
	Entropy	***89.66±2.95***	92.41±1.60	94.41±1.64	***72.26±0.89***	**95.57±2.65**	***91.59±3.05***
	Mutual Info	83.21±3.12	**92.97±2.82**	94.51±0.83	69.14±1.02	95.04±2.43	88.05±2.81
	α_0	83.75±3.44	92.20±1.03	***94.53±0.86***	70.47±1.01	95.14±2.47	89.75±2.86
DPN$^+$	Max. Prob.	89.21±3.15	93.10±2.26	90.68±2.93	69.71±3.29	96.29±1.62	89.31±5.18
	Entropy	**89.36±3.14**	93.38±2.19	**91.72±2.58**	**70.01±3.28**	***96.43±1.55***	**89.40±5.16**
	Mutual Info	87.09±3.20	93.47±1.68	89.24±4.04	67.79±3.83	94.53±2.22	87.77±4.80
	α_0	87.94±2.56	***93.60±1.66***	90.66±3.64	67.94±3.80	94.54±2.22	88.48±4.72
DPN$^{\text{rev}}$	Max. Prob.	**81.15±6.13**	87.42±1.38	**66.98±7.35**	**62.11±4.88**	88.94±2.81	87.24±1.78
	Entropy	75.99±6.54	88.14±0.97	64.75±7.51	61.02±3.81	88.67±2.62	**87.38±2.31**
	Mutual Info	80.07±6.85	88.05±0.68	54.87±7.17	60.94±3.02	90.77±2.17	84.61±3.51
	α_0	79.87±7.78	**88.21±1.24**	55.95±9.84	60.74±3.71	**91.01±2.12**	83.61±4.08
DPN$^{\text{forw}}$	Max. Prob.	87.20±3.37	83.94±5.22	**90.97±2.00**	**71.62±5.17**	92.94±4.04	**89.15±3.83**
	Entropy	**87.21±3.42**	**84.02±5.13**	90.67±1.70	70.77±5.62	**93.12±3.81**	89.13±3.78
	Mutual Info	85.02±3.58	81.15±5.54	80.97±1.20	62.54±7.92	86.99±9.07	84.70±2.60
	α_0	84.56±3.64	80.37±5.70	80.39±1.10	61.94±7.76	85.83±9.96	84.06±2.63
ENN	Max. Prob.	78.45±3.20	81.51±2.64	60.26±8.15	57.98±5.74	82.23±1.65	78.22±1.22
	Entropy	78.47±3.19	**81.56±2.62**	60.82±8.16	57.90±5.72	82.25±1.66	**78.24±1.23**
	Mutual Info	78.72±3.34	81.46±2.81	60.41±8.15	57.91±5.60	82.31±1.53	77.96±1.04
	α_0	**80.08±3.87**	81.14±3.61	**62.81±8.34**	**58.69±5.19**	**83.10±1.38**	78.04±1.35

TABLE XXII

CLASSIFICATION ACCURACY AND AVERAGE ACCURACY ON THE SAMPLES OF THE REGION SHIFT TASK ON THE VEGETATION CLASSES OF THE LCZ42 DATASET. THE RESULTS ARE GIVEN AS MEAN AND STANDARD DEVIATION COMPUTED BASED ON SEVEN RUNS. THE BEST AVERAGE ACCURACY ON IN-DISTRIBUTION SAMPLES IS HIGHLIGHTED FOR EACH SETTING

		R1-R2-R3	R1-R3-R2	R2-R1-R3	R2-R3-R1	R3-R1-R2	R3-R2-R1
DPN-RS	Accuracy	98.23±0.67	98.59±0.50	93.92±1.82	91.68±4.47	98.11±0.24	96.92±0.54
	Avg. Acc.	96.58±1.37	97.81±0.57	**87.95±2.23**	85.56±3.01	96.17±0.47	**96.47±0.39**
	Kappa	98.23±0.67	98.59±0.50	93.59±2.00	91.67±4.48	98.03±1.21	96.69±3.52
	Accuracy (Out)	42.52±4.11	44.10±6.73	37.06±11.34	48.70±3.75	37.94±5.45	36.89±6.11
	Avg. Acc. (Out)	31.32±2.87	36.03±1.24	28.16±6.67	42.60±4.82	27.30±3.35	32.49±2.79
	Kappa (Out)	51.45±4.27	43.09±10.14	29.83±12.36	47.87±5.79	38.01±5.54	36.70±6.27
DPN$^+$	Accuracy	98.35±0.79	98.49±0.65	90.70±3.95	91.74±1.99	98.24±0.24	98.11±0.88
	Avg. Acc.	**97.46±1.29**	**98.02±0.67**	87.11±3.72	**87.07±2.41**	**96.92±0.96**	95.96±1.53
	Kappa	98.35±0.79	98.49±0.65	90.68±3.66	91.73±2.00	98.17±0.25	98.04±0.91
	Accuracy (Out)	52.25±9.17	81.38±3.24	43.27±9.42	46.74±3.16	31.07±6.08	39.97±4.34
	Avg. Acc. (Out)	40.72±6.84	40.72±2.41	29.11±7.00	46.84±4.77	24.31±2.63	34.40±2.81
	Kappa (Out)	54.16±9.81	52.05±3.14	37.47±10.19	47.11±3.26	31.25±6.13	40.25±4.66
DPN$^{\text{rev}}$	Accuracy	96.91±1.57	98.01±0.62	80.61±4.34	75.39±10.07	97.42±0.59	96.64±0.58
	Avg. Acc.	93.51±2.87	96.81±0.37	62.35±5.79	71.56±8.72	93.41±2.33	91.41±1.52
	Kappa	97.86±1.07	97.85±0.0	80.12±4.83	75.47±12.84	97.15±0.55	96.54±0.56
	Accuracy (Out)	49.99±0.92	56.87±0.96	60.62±8.61	55.18±6.34	53.71±3.40	47.69±4.33
	Avg. Acc. (Out)	33.56±1.78	42.97±0.98	40.91±2.04	44.14±8.65	35.71±1.74	43.73±5.01
	Kappa (Out)	49.49±0.79	56.79±0.97	59.62±8.41	55.15±6.34	53.45±3.85	47.65±4.34
DPN$^{\text{forw}}$	Accuracy	96.18±1.11	94.75±3.31	92.50±2.00	87.98±6.21	94.08±6.21	96.72±0.27
	Avg. Acc.	90.39±3.36	89.32±6.94	83.06±4.79	72.46±17.58	90.94±6.29	91.39±1.48
	Kappa	96.17±1.11	94.75±3.31	92.48±2.00	87.97±6.22	93.90±6.37	96.60±0.28
	Accuracy (Out)	53.11±6.95	60.36±3.16	32.43±12.54	50.82±4.33	35.63±8.43	39.96±7.76
	Avg. Acc. (Out)	38.01±4.42	41.32±3.01	29.40±4.65	32.63±5.56	25.09±5.66	35.10±3.20
	Kappa (Out)	53.31±6.66	57.68±2.58	30.41±11.28	51.34±4.38	36.89±8.59	39.75±7.84
ENN	Accuracy	97.88±0.33	97.70±1.40	91.17±0.72	91.14±3.21	97.11±1.02	97.52±0.55
	Avg. Acc.	96.03±1.04	95.96±3.14	84.35±1.62	85.95±5.44	93.43±2.54	94.96±0.83
	Kappa	97.88±0.33	97.70±1.40	91.15±0.72	91.12±0.32	96.99±1.07	97.42±0.57
	Accuracy (Out)	61.77±6.56	61.29±3.56	62.96±3.71	60.22±4.32	49.06±3.42	59.12±19.70
	Avg. Acc. (Out)	42.79±7.85	48.59±3.76	44.83±2.04	47.50±2.19	35.41±3.49	39.27±4.33
	Kappa (Out)	58.69±6.92	61.12±3.57	61.17±3.67	59.00±4.49	48.74±3.43	58.07±19.74

vegetation classes A–F. The OOD detection and in-domain classification results for the experiments with all classes, urban classes only, and vegetation classes only are presented in Tables XVII–XXII. For all three cases, DPN-RS and DPN$^+$ give the best results in almost all cases regarding both the OOD detection and the in-domain classification. They specially perform very good for the top four settings. Depending on the definition of in- and out-domain, the prior networks based on forward and reverse KL divergence perform poorly. While the best separation performance increases from all classes setting to the urban only setting, it decreases for the vegetation-only setting.

V. DISCUSSION

A. Open-Set Recognition

The experiments on the open-set recognition clearly demonstrate that the proposed method as well as the compared methods are capable of differentiating between in-domain and OOD samples for high-resolution images with clear differences in the class representations, as seen for the AID and the UCM dataset. On the low-resolution So2Sat LCZ42 dataset which contains multiple very similar classes, the proposed method clearly outperforms the other methods and is the only method that still delivers a separation between in-domain and

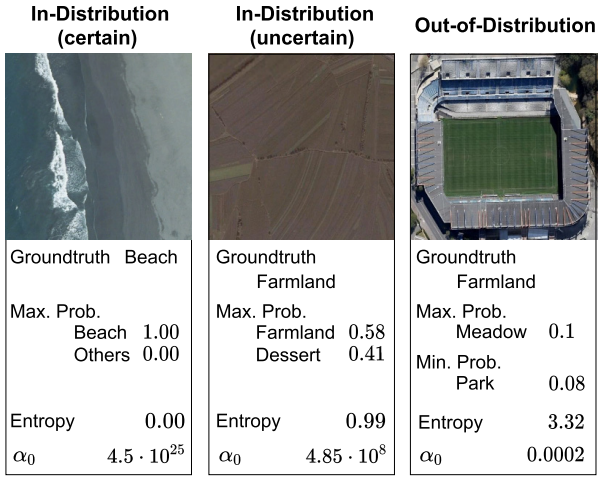

Fig. 5. Examples of DPN-RS predictions with low uncertainty, data uncertainty, and distributional uncertainty. One can clearly see that the differences in the stated maximum probability (Max. Prob), the entropy, and the precision.

OOD samples with the best AUROC scores between 0.88 and 0.99 on all considered test cases. This result underlines the main motivation of this method to derive a better separation between aleatoric in-distribution uncertainty and distributional uncertainty. However, the performance is lower in this dataset compared to the other two datasets, caused by its lower resolution and poor interclass separability. Small variations in such low-resolution data may lead to completely different predictions. Therefore, maximizing the gap between the in-distribution and OOD data is challenging on such datasets.

B. Sensor Shift

Contrary to the open-set recognition, the results indicate that the OOD detection under sensor shift is easier with lower resolution images and more challenging with higher resolution images. Furthermore, the similarity of the different sensors highly affects the OOD detection performance. It can be clearly observed that separating the blue channel from the green and the red channel gives the best results. Furthermore, the results on the LCZ42 dataset indicate that using a band more similar to the in-distribution as OOD data for training leads to a better separation. This underlines the obvious assumption that the similarity of the sensors highly affects the performance of such approaches and has to be considered for further research in this direction. The classification performance on the in-distribution data is pretty similar among the different experiments.

C. Region Shift

The region shift shows that DPN-RS is in general capable of detecting unknown city structures from other regions. Moreover, such a regionwise shift is almost similarly prevalent in both urban classes and vegetation classes. Poor OOD detection performance is obtained when using group 2 as in-domain data, group 3 as OOD training data, and group 1 as OOD test data. This shows that group 3 has a more diverse distribution than the other two, and thus, a boundary learned on group 2 using group 3 as OOD training data is less efficient to detect group 1 as OOD.

In contrast to the experiments in sensor shift experiments, the accuracy on the in-distribution samples is competitive, even though it might change significantly between different regions. Taking the OOD detection into account is therefore a promising way to improve the classification performance by rejecting uncertain samples from new regions.

VI. Conclusion

In this article, we proposed a method for OOD detection for remote sensing data. While deep learning is currently being applied to almost all remote sensing problems, their reliability is still questionable when the test data have a distributional shift from the training data. OOD detection is a crucial step for improving the trustworthiness of deep learning models. Toward this, we propose a DPN-based model that can effectively increase the gap between in-domain data and OOD data. The proposed method is tested extensively on three remote sensing datasets and three different tasks, namely, open-set recognition, sensor shift, and region shift. The proposed method shows satisfactory performance in all of the above settings. In general, DPN-based methods perform very well on OOD-detection and outperform the compared ENN approach and other baselines. Successful detection of OOD samples is a stride forward for building reliable, trustworthy deep learning-based remote sensing models. To the best of our knowledge, our work is the first extensive study on remote sensing data for this topic. Our future work will aim toward extending the OOD detection in the context of multitemporal analysis and multimodal fusion.

References

[1] X. X. Zhu *et al.*, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2018.

[2] K. Nogueira, O. A. B. Penatti, and J. A. dos Santos, "Towards better exploiting convolutional neural networks for remote sensing scene classification," *Pattern Recognit.*, vol. 61, pp. 539–556, Jan. 2017.

[3] G. Cheng, C. Yang, X. Yao, L. Guo, and J. Han, "When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2811–2821, May 2018.

[4] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," *J. Sensors*, vol. 2015, pp. 1–12, Jan. 2015.

[5] L. Mou, P. Ghamisi, and X. X. Zhu, "Deep recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3639–3655, Jul. 2017.

[6] L. Mou, S. Saha, Y. Hua, F. Bovolo, L. Bruzzone, and X. Xiang Zhu, "Deep reinforcement learning for band selection in hyperspectral image classification," 2021, *arXiv:2103.08741*.

[7] M. Volpi and D. Tuia, "Deep multi-task learning for a geographically-regularized semantic segmentation of aerial images," *ISPRS J. Photogramm. Remote Sens.*, vol. 144, pp. 48–60, Oct. 2018.

[8] R. Kemker, C. Salvaggio, and C. Kanan, "Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning," *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 60–77, Nov. 2018.

[9] L. Mou, Y. Hua, and X. X. Zhu, "A relation-augmented fully convolutional network for semantic segmentation in aerial scenes," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 12416–12425.

[10] Y. Zhan, K. Fu, M. Yan, X. Sun, H. Wang, and X. Qiu, "Change detection based on deep Siamese convolutional network for optical aerial images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1845–1849, Oct. 2017.

[11] F. Rahman, B. Vasu, J. V. Cor, J. Kerekes, and A. Savakis, "Siamese network with multi-level features for patch-based change detection in satellite imagery," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Nov. 2018, pp. 958–962.

[12] S. Saha, F. Bovolo, and L. Brurzone, "Unsupervised multiple-change detection in VHR optical images using deep features," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2018, pp. 1902–1905.

[13] Y. Li, Y. Zhang, X. Huang, H. Zhu, and J. Ma, "Large-scale remote sensing image retrieval by deep hashing neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 950–965, Feb. 2018.

[14] P. Li *et al.*, "Hashing nets for hashing: A quantized deep learning to hash framework for remote sensing image retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7331–7345, Oct. 2020.

[15] P. Zhou, G. Cheng, Z. Liu, S. Bu, and X. Hu, "Weakly supervised target detection in remote sensing images based on transferred deep features and negative bootstrapping," *Multidimensional Syst. Signal Process.*, vol. 27, no. 4, pp. 925–944, 2016.

[16] Z. Chen, D. Chen, Y. Zhang, X. Cheng, M. Zhang, and C. Wu, "Deep learning for autonomous ship-oriented small ship detection," *Saf. Sci.*, vol. 130, Oct. 2020, Art. no. 104812.

[17] O. Ghorbanzadeh, T. Blaschke, K. Gholamnia, S. R. Meena, D. Tiede, and J. Aryal, "Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection," *Remote Sens.*, vol. 11, no. 2, p. 196, Jan. 2019.

[18] S. Saha, F. Bovolo, and L. Bruzzone, "Building change detection in VHR SAR images via unsupervised deep transcoding," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 1917–1929, Mar. 2021.

[19] Z. Wu, J. Li, Y. Wang, Z. Hu, and M. Molinier, "Self-attentive generative adversarial network for cloud detection in high resolution remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 10, pp. 1792–1796, Oct. 2020.

[20] A. Meraner, P. Ebel, X. X. Zhu, and M. Schmitt, "Cloud removal in Sentinel-2 imagery using a deep residual neural network and SAR-optical data fusion," *ISPRS J. Photogramm. Remote Sens.*, vol. 166, pp. 333–346, Aug. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271620301398

[21] J. Li *et al.*, "A lightweight deep learning-based cloud detection method for Sentinel-2A imagery fusing multiscale spectral and spatial features," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–19, 2022.

[22] J. Zhong, B. Yang, G. Huang, F. Zhong, and Z. Chen, "Remote sensing image fusion with convolutional neural network," *Sens. Imag.*, vol. 17, no. 1, pp. 1–16, Dec. 2016.

[23] Z. Shao and J. Cai, "Remote sensing image fusion with deep convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1656–1669, May 2018.

[24] J. Gawlikowski, M. Schmitt, A. Kruspe, and X. X. Zhu, "On the fusion strategies of Sentinel-1 and Sentinel-2 data for local climate zone classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Sep. 2020, pp. 2081–2084.

[25] S. Saha, F. Bovolo, and L. Bruzzone, "Change detection in image time-series using unsupervised LSTM," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[26] J. Li, X. Huang, and J. Gong, "Deep neural network for remote-sensing image interpretation: Status and perspectives," *Nat. Sci. Rev.*, vol. 6, no. 6, pp. 1082–1086, May 2019.

[27] Z. Zhang, K. Doi, A. Iwasaki, and G. Xu, "Unsupervised domain adaptation of high-resolution aerial images via correlation alignment and self training," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 4, pp. 746–750, Apr. 2021.

[28] S. Saha, B. Banerjee, and S. N. Merchant, "Unsupervised domain adaptation without source domain training samples: A maximum margin clustering based approach," in *Proc. 10th Indian Conf. Comput. Vis., Graph. Image Process. (ICVGIP)*, 2016, pp. 1–8.

[29] O. Tasar, S. L. Happy, Y. Tarabalka, and P. Alliez, "SemI2I: Semantically consistent Image-to-Image translation for domain adaptation of remote sensing data," 2020, *arXiv:2002.05925*.

[30] J. Gawlikowski *et al.*, "A survey of uncertainty in deep neural networks," 2021, *arXiv:2107.03342*.

[31] A. Malinin and M. Gales, "Predictive uncertainty estimation via prior networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 7047–7058.

[32] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 6402–6413.

[33] C. E. Woodcock, "Uncertainty in remote sensing," in *Uncertainty in Remote Sensing and GIS*. Hoboken, NJ, USA: Wiley, Dec. 2002, pp. 19–24. [Online]. Available: https://onlinelibrary.wiley.com/doi/book/10.1002/0470035269

[34] M. A. Ibrahim, M. K. Arora, and S. K. Ghosh, "Estimating and accommodating uncertainty through the soft classification of remote sensing data," *Int. J. Remote Sens.*, vol. 26, no. 14, pp. 2995–3007, Jul. 2005.

[35] Q. Zhang and P. Zhang, "An uncertainty descriptor for quantitative measurement of the uncertainty of remote sensing images," *Remote Sens.*, vol. 11, no. 13, p. 1560, Jul. 2019.

[36] C. C. V. da Silva, K. Nogueira, H. N. Oliveira, and J. A. D. Santos, "Towards open-set semantic segmentation of aerial images," in *Proc. IEEE Latin Amer. GRSS ISPRS Remote Sens. Conf. (LAGIRS)*, Mar. 2020, pp. 16–21.

[37] S. Dang, Z. Cao, Z. Cui, Y. Pi, and N. Liu, "Open set incremental learning for automatic target recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4445–4456, Jul. 2019.

[38] L. Wu, Y. Peng, and C. Li, "Hyperspectral image open set recognition based on the extreme value machine," *Proc. SPIE*, vol. 11565, Nov. 2020, Art. no. 115650Q.

[39] H. Meyer and E. Pebesma, "Predicting into unknown space? Estimating the area of applicability of spatial prediction models," 2020, *arXiv:2005.07939*.

[40] T. Stark, M. Wurm, X. X. Zhu, and H. Taubenbock, "Satellite-based mapping of urban poverty with transfer-learned slum morphologies," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5251–5263, 2020.

[41] L. Bergamasco, S. Saha, F. Bovolo, and L. Bruzzone, "An explainable convolutional autoencoder model for unsupervised change detection," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. XLIII-B2-2020, pp. 1513–1519, Aug. 2020.

[42] R. Roscher, B. Bohn, M. F. Duarte, and J. Garcke, "Explainable machine learning for scientific insights and discoveries," *IEEE Access*, vol. 8, pp. 42200–42216, 2020.

[43] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1050–1059.

[44] J. Lee, M. Humt, J. Feng, and R. Triebel, "Estimating model uncertainty of neural networks in sparse information form," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 5702–5713.

[45] J. Nandy, W. Hsu, and M. L. Lee, "Towards maximizing the representation gap between in-domain & out-of-distribution examples," 2020, *arXiv:2010.10474*.

[46] D. Tuia, C. Persello, and L. Bruzzone, "Domain adaptation for the classification of remote sensing data: An overview of recent advances," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 41–57, Jun. 2016.

[47] D. Geiger and D. Heckerman, "A characterization of the Dirichlet distribution with application to learning Bayesian networks," in *Maximum entropy Bayesian Methods*. Dordrecht, The Netherlands: Springer, 1996, pp. 61–68.

[48] A. Malinin and M. Gales, "Reverse KL-divergence training of prior networks: Improved uncertainty and adversarial robustness," 2019, *arXiv:1905.13472*.

[49] M. Sensoy, L. Kaplan, and M. Kandemir, "Evidential deep learning to quantify classification uncertainty," in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst.*, 2018, pp. 3183–3193.

[50] W. Mendenhall, R. J. Beaver, and B. M. Beaver, *Introduction to Probability and Statistics*. Boston, MA, USA: Cengage Learning, 2012.

[51] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. 18th SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst. (GIS)*, 2010, pp. 270–279.

[52] G.-S. Xia *et al.*, "AID: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3965–3981, Jul. 2017.

[53] X. X. Zhu *et al.*, "So2Sat LCZ42: A benchmark data set for the classification of global local climate zones [software and data sets]," *IEEE Geosci. Remote Sens. Mag.*, vol. 8, no. 3, pp. 76–89, Sep. 2020.

[54] C. Qiu, X. Tong, M. Schmitt, B. Bechtel, and X. X. Zhu, "Multilevel feature fusion-based CNN for local climate zone classification from Sentinel-2 images: Benchmark results on the So2Sat LCZ42 dataset," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2793–2806, 2020.

[55] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," in *Proc. 23rd Int. Conf. Mach. Learn. (ICML)*, 2006, pp. 233–240.

[56] P. Perera *et al.*, "Generative-discriminative feature representations for open-set recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11814–11823.

Jakob Gawlikowski (Student Member, IEEE) received the bachelor's and master's degrees in mathematics from the Technical University of Munich, Ottobrunn, Germany, in 2015 and 2019, respectively. He is currently pursuing the Ph.D. degree with the Technical University of Munich, Ottobrunn, and the German Aerospace Center, Munich.

He is currently a Researcher at the Institute of Data Science, German Aerospace Center (DLR), Jena, Germany. His research interests are related to multimodal machine learning, uncertainty quantification, and robustness in deep learning models.

Sudipan Saha (Member, IEEE) received the M.Tech. degree in electrical engineering from IIT Bombay, Mumbai, India, in 2014, and the Ph.D. degree in information and communication technologies from the University of Trento, Trento, Italy, and Fondazione Bruno Kessler, Trento, in 2020.

He worked as an Engineer with TSMC Ltd., Hsinchu, Taiwan, from 2015 to 2016. In 2019, he was a Guest Researcher with the Technical University of Munich (TUM), Ottobrunn, Germany, where he is currently a Post-Doctoral Researcher. His research interests are related to multitemporal remote sensing image analysis, domain adaptation, time-series analysis, image segmentation, deep learning, image processing, and pattern recognition.

Dr. Saha was a recipient of the Fondazione Bruno Kessler Best Student Award 2020. He is a reviewer for several international journals and served as a Guest Editor for *Remote Sensing* (MDPI)—Special Issue on Advanced Artificial Intelligence for Remote Sensing: Methodology and Application.

Anna Kruspe received the Diploma degree in media technology and the Ph.D. degree from Technische Universität Ilmenau, Ilmenau, Germany, in 2011 and 2018, respectively.

She was the Head of the Machine Learning Group, German Aerospace Center (DLR), Weßling, Germany, and a Guest Researcher at Johns Hopkins University Baltimore, MD, USA, and AIST Japan, Tsukuba, Japan. She is currently the Deputy Head of the Data Science in Earth Observation (EO), Technical University of Munich, Ottobrunn, Germany, and co-leading a group on social media research at DLR.

Xiao Xiang Zhu (Fellow, IEEE) received the M.Sc., Dr.Ing., and Habilitation degrees in signal processing from the Technical University of Munich (TUM), Ottobrunn, Germany, in 2008, 2011, and 2013, respectively.

Since 2019, she has been a Co-Coordinator of the Munich Data Science Research School. Since 2019, she has also been heading the Helmholtz Artificial Intelligence—Research Field "Aeronautics, Space and Transport." Since May 2020, she has been the Director of the International Future AI Lab "AI4EO—Artificial Intelligence for Earth Observation (EO): Reasoning, Uncertainties, Ethics and Beyond," Munich. Since October 2020, she has been the Co-Director of the Munich Data Science Institute (MDSI), TUM. She was a Guest Scientist or a Visiting Professor at the Italian National Research Council (CNR-IREA), Naples, Italy, Fudan University, Shanghai, China, The University of Tokyo, Tokyo, Japan, and the University of California at Los Angeles, Los Angeles, CA, USA, in 2009, 2014, 2015, and 2016, respectively. She is currently the Professor with the Data Science in EO (former: Signal Processing in EO), TUM, and the Head of the Department "EO Data Science," Remote Sensing Technology Institute, German Aerospace Center (DLR), Weßling, Germany. She is also a Visiting AI Professor at ESA's Phi-lab. Her main research interests are remote sensing and EO, signal processing, machine learning, and data science, with a special application focus on global urban mapping.

Dr. Zhu is a member of the young academy (Junge Akademie/Junges Kolleg) at the Berlin-Brandenburg Academy of Sciences and Humanities, the German National Academy of Sciences Leopoldina, and the Bavarian Academy of Sciences and Humanities. She serves on the Scientific Advisory Board in several research organizations, including the German Research Center for Geosciences (GFZ) and the Potsdam Institute for Climate Impact Research (PIK). She is an Associate Editor of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. She serves as an Area Editor for special issues of *IEEE Signal Processing Magazine*.