Enabling Component-Based Progress Monitoring on Construction Sites Through Image-Based Computer Vision

Felix Friedl[1] and Fabian Pfitzner[1]

[1]Chair of Computational Modeling and Simulation, Technical University of Munich, Arcisstraße 21, 80333 Munich, Germany

E-mail(s): felix.friedl@tum.de, fabian.pfitzner@tum.de

Abstract: Vision-based construction monitoring methods have improved on-site transparency. However, many point cloud-based techniques are complex and often involve an image-dependent reconstruction step, making them prone to uncertainties. Additionally, few address productivity insights at the construction activity level. This paper presents a novel computer vision approach for automating construction progress monitoring, extracting information directly from image data enhanced through as-built details. A PIDNet Semantic segmentation model was trained to identify cast-in-place concrete walls, columns, and slabs during panel, rebar, and concrete phases. The detected components were processed using averaging techniques to monitor element-specific progress. The resulting data was integrated with as-built models through geometric projections, forming the basis for a digital twin construction. Our method was deployed on two-month construction data, providing detailed progress information and demonstrating its robustness. Compared to previous methods, this approach effectively merges existing as-built models with comprehensive as-performed image data.

Keywords: Construction Monitoring, Computer Vision, Semantic Segmentation, As-built geometry

Erschienen in Tagungsband 35. Forum Bauinformatik 2024, Hamburg, Deutschland, DOI: will/be/assigned/by/editor

1 Introduction

There have been few increases in productivity and digitization during the last decades in the construction sector despite the recent developments towards Industry 4.0 [1]. Specifically, project management and progress monitoring are often carried out manually, comparing the status of building elements against a construction schedule. This tedious and error-prone process results in frequent cost and time overruns [2]. Novel approaches towards site monitoring need to be developed that aim at the digital twin idea [3] supporting management in everyday construction projects.

However, recent approaches lack monitoring functionalities for cast-in-place concrete elements (e.g. columns, walls and slabs) [4]. The authors aim to close this gap by providing a new approach for semantic segmentation based progress data collection on these elements' panel, rebar, and concrete phases. The potential of our approach is highlighted by demonstrating improved workflows for extracting and coupling element-specific progress data with a BIM model.

2 Related Work

The monitoring approach proposed by this paper relies on automating the visual understanding of construction environments with the help of 2D semantic segmentation. This computer vision discipline allows the detection of specific objects in images by partitioning them into multiple segments with matching characteristics and assigning a categorical label from a predefined set of classes to each one. Convolutional Neural Networks (CNNs) use learnable convolutional kernels to efficiently process image data and perform this task autonomously [5]. However, as most models retain the problem of generality, curated datasets with annotated ground truths tailored to the specific research problem are required for training.

Many applications of computer vision and digital twinning pose significant benefits to the construction sector. In this context, a digital twin is a virtual replica of a building that uses real-time data to simulate, monitor, and optimize the performance of the structure throughout its lifecycle [6]. Computer Vision techniques are used to evaluate productivity and progress monitoring on an activity-level in construction operations [7] or in smaller areas of application like crack detection in reinforced concrete structures [8].

Progress monitoring is fundamental for digital twins in construction by extending as-planned data with as-performed data [3]. To fully leverage the digital twin construction, it's essential to gather and organize vast data for various assessments [6]. Pfitzner, Braun, and Borrmann [9], for example, enable extracting precise construction metrics through knowledge graph-based methods. Despite extensive research in vision-based construction monitoring, significant gaps remain. Existing work, such as that described in [10], primarily focuses on tracking single elements like pre-cast walls, which limits the scope of progress monitoring. Expanding existing datasets to include additional elements like slabs and pillars could provide a more comprehensive understanding of the construction process. Moreover, approaches as demonstrated by [8] are limited by uncertainties in unordered photographs, which can be mitigated through advanced averaging algorithms to enhance data reliability. Additionally, while [7] demonstrates detailed activity-level monitoring, it involves complex processing steps limiting reproducibility. Therefore, there is a need for a more efficient and accurate approach that can expand the range of elements monitored, enhance tracking precision, and simplify the monitoring process.

This study aims to enhance vision-based techniques for monitoring construction progress by addressing existing limitations. Specifically, it focuses on expanding the scope of monitored elements, enabling in-depth progress tracking, and streamlining the overall monitoring process. By leveraging state-of-the-art semantic segmentation networks, the proposed method intends to reduce the time and effort required for effective monitoring while maintaining the accuracy and comprehensiveness of the tracking.

3 Method

Our approach provides a semantic segmentation-based pipeline towards automating construction progress monitoring by detecting element-specific construction progress timestamps. The generated information is coupled with an as-built model.

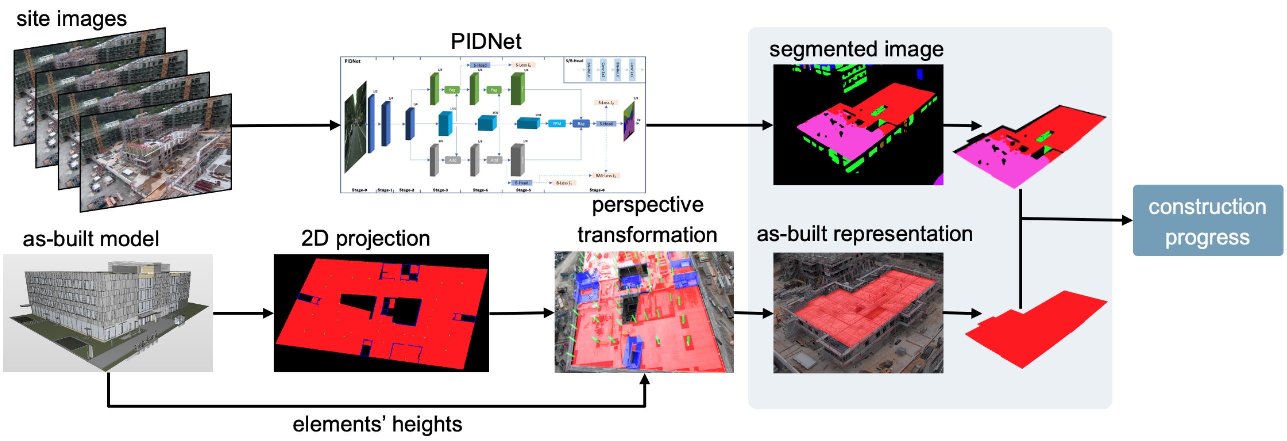

Figure 1: Method overview

Pretrained semantic segmentation models [11] are used to generate predictions from the site images. Our work focuses on cast-in-place slabs, walls and columns. The detected classes include three different construction stages: panel, rebar and concrete.

Figure 1 shows the proposed pipeline: In the first step, crane-mounted cameras with a fixed viewpoint capture a continuous stream of site images. These images are then processed by the semantic segmentation model PIDNet [12]. Finally, element-wise construction progress is extracted from the resulting prediction files.

The integration of the as-built data is done as follows: First, the elements' geometry is extracted from the BIM model and then plotted element-wise in a top-down perspective. These geometries are then represented in a simplified, top-down view image where each element's ground floor area is highlighted in red against a black background, as illustrated Fig. 1. The orientation of each element is aligned with the actual building directions using rotation based on the as-built model. To match the camera's percepective, the as-built representations are converted using perspective transformation. Per building floor and camera, a transformation matrix is computed using a direct linear transformation algorithm [13] and point picking. The resulting matrix is applied to each pixel of an element's as-built representation. A vector computation determining the elements' heights was applied to overcome the 2D limitation of the perspective transformation approach. A vanishing point, predefined on two parallel lines, determines the z-direction of a building component. The element height extracted from the as-built representation and the vector computation is converted into pixels and projected onto the image. The element-wise as-built representations are ultimately passed back to the progress monitoring pipeline, fusing the resulting data.

A percentage value indicating the components progress is computed based on the overlapping area. Two thresholds per construction state are assumed: 10%: the start of a panel, rebar, or concreting phase and 80%: the end. To reduce the number of computations and minimize outliers, all entries within a specified *averaging interval* are aggregated to compute the mean pixel count for each class.

4 Case Study

4.1 Data and Setup

A custom dataset of site images [9] containing cast-in-place concrete elements of real-world site environments was prepared to showcase our method. The annotated dataset contains 390 samples. The segmentation model *PIDNet-S* was used within the computation pipeline.

Two examples were investigated thoroughly to determine the in-depth accuracy of the proposed method. All cast-in-place types (*rebar, slab, concrete*) were included. The gap between both timestamps was used to measure the framework's accuracy. In the following tables, a negative error indicates the derived timestamp is earlier than the actual, while a positive error means it's later.

Various averaging techniques were tested by adjusting starting and ending thresholds from 0% to 50% and from 50% to 100%, respectively, in 5% increments.

In total, 121 combinations were analyzed for averaging intervals from 0 to 30 minutes in one-minute intervals for each camera. The total error over all evaluated timestamps was recorded. However, only the best-scoring combinations of threshold and averaging interval parameters were further evaluated to enable precise monitoring.

All as-built representations were compared against manually drawn annotations as shown in Figure 3, 4 and 5. Common pixels are coloured in red, pixels missing from model-generated ones in green and pixels missing from the model-generated pictures in blue. The combined error value was calculated as a fraction of wrong to correct pixels in percent.

Lastly, the entire pipeline was tested on one construction site using images from two cameras over two months. The as-built representations were generated for all building elements from the first and second floors, resulting in 290 elements. Considering only concrete slabs, walls and columns, all elements were monitored over the whole two-month timespan using a parameter configuration of an averaging interval of 7 minutes and thresholds of 75% and 10%, respectively. The detected construction progress was analyzed further by comparing it to manually derived as-built timestamps for 8 elements.

4.2 Results

The segmentation model *PIDNet-S* was trained on a specific dataset consisting of 390 annotated images from two different construction sites and achieved a mean average precision (mAP) of 74.20%, demonstrating sufficient precision and performance. The test results shown in Table 1 indicate sufficient performance across all classes. Yet, a lower mAP was achieved for wall or column elements in the rebar phase presumably due to lacking features. However, to achieve higher generalisability of the model, additional training data is required. A comprehensive investigation of the semantic training and testing process was carried out in the authors' prior work [11].

The monitoring results for a specific wall element are shown in Figure 2 with a negative error indicating, that the derived timestamp is earlier than the actual, while a positive error means it's later. These results illustrate, that precise results can be achieved when using different averaging intervals for singular cameras and by averaging the monitoring results over multiple cameras.

Table 1: Results for the PIDNet-S model - mAP [%]

class	background	column			slab			wall		
		concrete	panel	rebar	concrete	panel	rebar	concrete	panel	rebar
mAP [%]	98.34	63.60	72.60	22.59	66.88	93.95	94.00	83.02	79.90	67.12

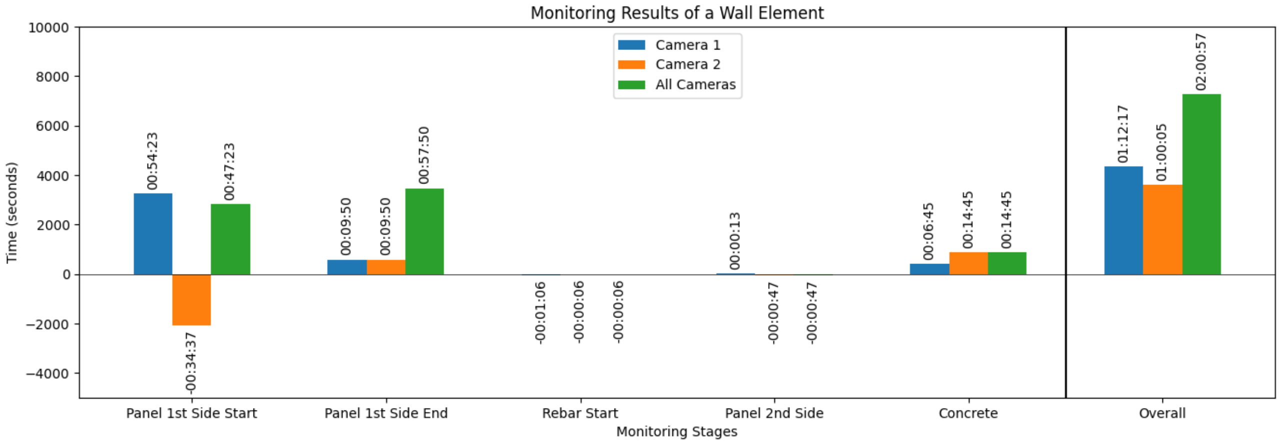

Figure 2: In-depth monitoring results of a wall element

Including this wall example, six elements were closer investigated to establish a best suited combination of top-level parameters being the length of the averaging interval and the upper and lower threshold used for detecting element specific progress milestones. The element-wise best suited upper threshold varied highly: Especially for partially occluded elements, a conservative threshold like the median of all examples of 75% seemed practical. However, most studied examples generated precise results with a minor lower threshold of 5%-10%. Over all samples, a smaller averaging interval of 5 to 10 minutes generated the best results as errors could be compensated, and a certain level of accuracy was maintained. This configuration of parameters poses a likely combination to work for monitoring arbitrary construction elements and is used in the following studies.

The generated as-built representations for slabs showed very high accuracy of 96.47% correctly colored pixels. It must be pointed out that one stairwell platform rendered insufficient results because of its different base height.

An tolerable error of 18.51% was achieved for the wall elements. This was primarily due to the fact that in the model-generated images, some elements that should be at the bottom appeared higher due to the limitations in pixel precision.

Higher errors of 61.12% for column elements were encountered as a result of the height conversion problem and the column's smaller mask area.

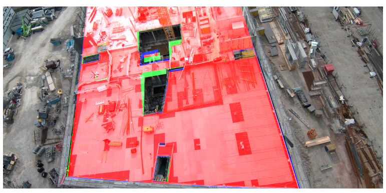

Figure 3: Slab elements

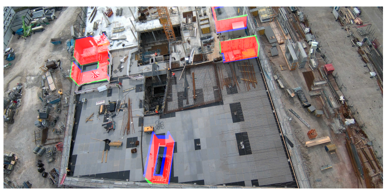

Figure 4: Wall elements

Figure 5: Column elements

The resulting differences between detected and as-built timestamps for the eight closer investigated elements passed through the end-end pipeline are shown in Table 2. A positive value indicates the detected event occurring earlier than the actual one. Additionally, all samples are color-coded with entries below 15 minutes appearing in green, below 24 hours in orange and higher than 24 hours in red. Unsuccessful tests are shown with a grey background.

Table 2: Examples for determining the accuracy

Event	Slab 1	Slab 2	Wall 1	Wall 2	Wall 3	Column 1	Column 2	Column3
Panel Start	00:11:13	00:00:00	>1day	>1day	03:50:08	>1day	-23:57:45	-01:43:36
Panel End	-14:36:13	>1day	00:01:06	-00:08:45	00:57:50	-	00:15:35	>1day
Rebar Start	>1day	03:07:34	-	-	-00:01:06	-00:00:13	00:06:02	02:06:20
Rebar End	-00:08:40	-22:39:57	-	-	-20:26:46	-	-	-
Concrete Start	>1day	01:40:40	>1day	>1day	>1day	-00:00:41	-00:00:44	-00:01:48
Concrete End	00:00:10	-	00:00:40	00:07:43	00:14:15	-	-	-

Unlike the positive outcomes depicted in Figure 2, challenges were encountered in identifying the completion times for the rebar stages of wall elements and both rebar and concrete completion events for column elements. Timestamps for panel phases, however, were detected more accurately. Similar observations were made for all examined elements despite using fixed parameters based on identical thresholds and averaging intervals. The underlying reasons and possible fixes for these inaccuracies are discussed below.

5 Discussion

Our method demonstrated sufficient results in precisely measuring fine-granular phases of construction progress using sequential on-site images and fusing it with as-built data. While the overall performance of this approach proved to be promising, some limitations were noted:

The elaborated monitoring methodology could not detect the completion of rebar works correctly. As the segmentation-based approach identifies an area as 100% completed once a single layer has been installed, the two necessary directional layers per side could not be monitored. Additionally, the completion of concreting for walls and columns is only detected once the formwork is removed and concrete-labeled pixels are visible in the proposed camera setup. Prior limitations can be addressed by exploring additional monitoring techniques increasing the level of detail [4], [14]. The perspective transformation method's inaccuracies can be approached using referenced on-site markers. A future investigation of this method sourced from laser-scanning [15] and its application to semantic segmentation would be desirable.

Figure 6 highlights noisy results stemming from a slab with detected *panel*, *rebar*, or *concrete* timestamps at the same time in different sections of the element. These noisy results were caused by a lack of details in the as-built model, e.g., missing reinforcement sections. Specific modeling rules are necessary to represent the element accurately in all phases.

Finally, inaccuracies coming from bad lighting conditions and occlusions could not be compensated for small elements. Moreover, camera sway from crane movements affected detection accuracy. On-site markers for dynamic perspective transformation could mitigate this issue.

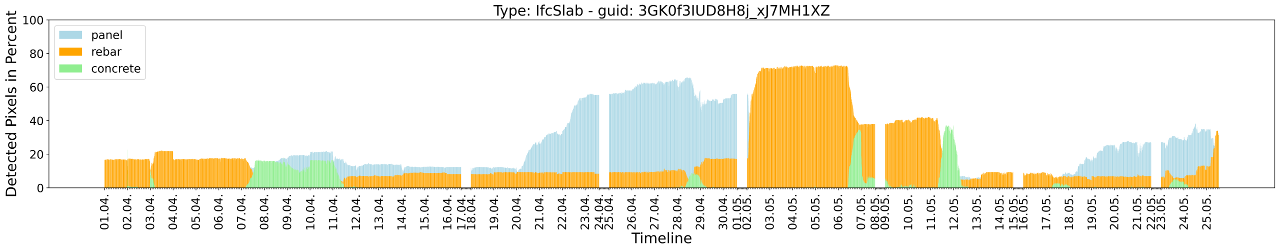

Figure 6: Evaluation of a slab in the second building floor with noisy results

While there has been exhaustive work in construction progress monitoring, our approach differs from existing work as follows: In contrast to [10], progress tracking beyond one element (pre-cast walls) is demonstrated here. The in-depth analysis of slabs and pillars highlighted additional challenges and requirements for precise monitoring. Fine-granular monitoring allowed a comprehensive understanding of the construction process, necessary for efficient and effective on-site resource allocation. This approach overcomes the uncertainties that arise from unordered photographs [8] through the use of an averaging algorithm. This algorithm not only smooths out the variations in the data but also provides a more accurate representation of the overall progress, enhancing the reliability of our results. While [7] has shown an innovative approach to monitor not only the construction progress but also the activity-level progress details, a lot of processing steps and computational resources were necessary to achieve these details. We demonstrate that similar results can be gained using a much simpler prediction pipeline based on image level. This underscores the need for further exploration of diverse monitoring methodologies to identify practical approaches for the task at hand.

6 Conclusion

This paper presented a novel approach to automating progress monitoring of cast-in-place concrete slabs, walls, and columns, extracting information directly from image data and enhancing it through as-built details. For this task, algorithms for evaluating element-specific construction progress timestamps and procedures for coupling the obtained information with as-built models have been proposed. With the integration of different averaging approaches, robust performance with high precision was proven in case studies concerning the specific parts of the elaborated framework and the whole end-to-end approach. In contrast to many other studies, the suitability of the proposed methodology on real-world construction sites has been thoroughly analyzed. In conclusion, the approach demonstrated practical applicability and potential for supporting on-site management toward higher productivity.

Acknowledgements

We thankfully acknowledge Innovationsmanagement Bau GmbH for their financial support and for providing us access to multiple construction sites to collect valuable data.

References

[1] R. Agarwal, S. Chandrasekaran, and M. Sridhar, *Imagining construcion's digital future*, Singapur, Jun. 2016.

[2] K. Mostafa and T. Hegazy, "Review of image-based analysis and applications in construction", *Automation in Construction*, 2021.

[3] R. Sacks, I. Brilakis, E. Pikas, H. S. Xie, and M. Girolami, "Construction with digital twin information systems", *Data-Centric Engineering*, 2020.

[4] J. Yang, M.-W. Park, P. A. Vela, and M. Golparvar-Fard, "Construction performance monitoring via still images, time-lapse photos, and video streams: Now, tomorrow, and the future", *Advanced Engineering Informatics*, 2015.

[5] S. Ghosh, N. Das, I. Das, and U. Maulik, "Understanding deep learning techniques for image segmentation", *ACM Computing Surveys*, 2020.

[6] J. Schlenger, F. Pfitzner, A. Braun, S. Vilgertshofer, and A. Borrmann, "Digitaler zwilling baustelle – baustellenüberwachung zur automatisierten zeit– und kostenkontrolle", *Bautechnik*, 2023.

[7] A. Pal, J. J. Lin, S.-H. Hsieh, and M. Golparvar-Fard, "Activity-level construction progress monitoring through semantic segmentation of 3d-informed orthographic images", *Automation in Construction*, 2024.

[8] K. C. Laxman, N. Tabassum, L. Ai, C. Cole, and P. Ziehl, "Automated crack detection and crack depth prediction for reinforced concrete structures using deep learning", *Construction and Building Materials*, 2023.

[9] F. Pfitzner, A. Braun, and A. Borrmann, "From data to knowledge: Construction process analysis through continuous image capturing, object detection, and knowledge graph creation", *Automation in Construction*, Aug. 2024.

[10] Z. Wang, Q. Zhang, B. Yang, *et al.*, "Vision-based framework for automatic progress monitoring of precast walls by using surveillance videos during the construction phase", *Journal of Computing in Civil Engineering*, 2021.

[11] F. Friedl, "Bim-based progress monitoring using 2d semantic segmentation", M.S. thesis, Technische Universität München, Apr. 2024.

[12] J. Xu, Z. Xiong, and S. P. Bhattacharyya, *Pidnet: A real-time semantic segmentation network inspired by pid controllers*, 2023. arXiv: 2206.02066.

[13] D. A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*, 2. Auflage. Harlow: Pearson Education, 2012.

[14] F. Pfitzner, A. Braun, and A. Borrmann, "Object detection-based knowledge graph creation: Enabling insight into construction processes", in *Computing in Civil Engineering 2023*, Y. Turkan, J. Louis, F. Leite, and S. Ergan, editors, Reston: American Society of Civil Engineers, 2024.

[15] L. Cheng, S. Chen, X. Liu, *et al.*, "Registration of laser scanning point clouds: A review", *Sensors (Basel, Switzerland)*, 2018.