2025 European Conference on Computing in Construction

CIB W78 Conference on IT in Construction

Porto, Portugal
July 14-17, 2025

DEEP NEURAL NETWORKS FOR OBJECT-DETECTION AND INSTANCE SEGMENTATION OF MECHANICAL, ELECTRICAL AND PLUMBING COMPONENTS: TRANSFER LEARNING ON RADIATORS AS AN EXAMPLE

Rainer Partl, Francois Veynandt, David Ziermann and Christian Heschl
University of Applied Sciences Burgenland, Pinkafeld, Austria

Abstract

Cost estimation and as-built documentation on construction sites require extensive manual work. Combining LiDAR with computer vision offers an effective solution for 5D BIM applications. Therefore, two deep neural networks are trained on a database of over 1000 images of radiators, as an example of MEP components. Transfer learning has been applied on SOLOv2 for instance segmentation and on YOLOx for object detection. Given the variety of the image database, the generated models achieve satisfying performance, suitable for the intended applications. The developed models will be used in a scan-to-BIM workflow, enabling the semi-automatic capture of as-built information.

Introduction

The numerous interconnected components of Mechanical, Electrical and Plumbing (MEP) systems across different trades significantly increase the complexity of construction projects, making coordination processes particularly challenging (Yarmohammadi & Ashuri, 2015). Insufficient integration of MEP systems poses high risks for disruptions, with for example in Austria 77% of disruptions related to MEP installations in construction projects, which leads to claims and additional costs (Monsberger, 2019). A major cause of claims is incomplete project documentation (Awed & Fini, 2024). Despite being time-consuming and costly, as-built documentation is therefore crucial during construction (Casini, 2022).

Building Information Modeling (BIM) tools are typically used for planning and managing new construction projects (Doukari et al., 2023), but are not effectively utilized so far for cost estimation during construction (Casini, 2022). Challenges include the time required to create detailed BIM models, their reliability, costs and lack of digital tool skills among workers (Disney et al., 2023). A promising approach is "scan-to-BIM", using laser scanning technologies (LiDAR) to create 3D models from point clouds. This would enable effective as-built documentation, offering time savings in 3D reconstruction and comparisons (Bosché et al., 2013). However, manual tasks remain in preparing and classifying point cloud objects (Mirzaei et al., 2022). Integrating deep learning techniques in computer vision can enhance automation in cost estimation (Abioye et al., 2021).

This paper contributes by developing computer vision of MEP components using deep neural networks, with two objectives: (i) object detection for counting MEP components on construction sites and (ii) instance segmentation to localize MEP components in images, supporting a scan-to-BIM workflow by associating pixels with BIM objects.

Research framework

Background

Several studies show the potential of LiDAR technologies for MEP systems, which explains the research effort in this field (Noichl et al., 2021; Wang et al., 2021). For example, a scan-to-BIM workflow for MEP components focused on pipes in Noichl et al. (2021) and valves in Wang et al. (2021) for generating MEP scenes in BIM by semantic segmentation in point clouds.

Three approaches of using computer vision methods can be found in the literature. First, single experiments of object detection in images: Tofan et al. (2020) and Kufuor et al. (2021) tested Faster Region-based Convolutional Neural Network (R-CNN) for object detection to recognize radiators in images. Tofan et al. (2020) used 123 images of sectional radiator. Kufuor et al. (2021) used two different datasets for model training and analyzed the performance of each dataset separately. 105 pictures of 360° images and 97 of phone cameras were taken for model training. Both experiments achieved satisfactory results despite the small number of images, but room for improvement is mentioned: better results could be achieved with more labelled training images.

Second, deep learning methods are used for semantic segmentation in point clouds. For example, Perez-Perez et al. (2021) segmented components like columns, floors, ceilings, pipes, walls and beams with an average accuracy of 86.13% by using a recurrent neural network and two convolutional neural networks. Yin et al. (2021) developed the ResPointNet++ network, which has ResNet blocks integrated, and used millions of data points for

training, reaching segmentation accuracy of more than 90% for beams, pipes, pumps and tanks. Nevertheless, a high performance for typical point cloud segmentation can primary be achieved for large-sized components, e.g., walls, beams, ceilings. For smaller types of components such as furniture and MEP components the application is much more difficult. Furthermore, it is easier to annotate small objects in 2D datasets than in 3D (Wang et al., 2022).

Third, the development of a holistic scan-to-BIM workflow for MEP, using computer vision by combining image and point cloud data, is mentioned in only a few studies. Wang et al. (2022, 2024) set a focus on recognizing objects in images and map the shape of the object to the point cloud for spatial clustering. Wang et al. (2022) used 2000 images of MEP components (e.g. electrical pipes, ducts, valves, pumps etc.) for training, validation and testing by using DeepLabv3+. An overall segmentation accuracy of 77.7% was achieved. Wang et al. (2024) developed a workflow called Omni-Scan2BIM for reconstruction of a pipe system based on visual foundation models (DINOv2, Segment Anything Model). The two studies focused on individual sub-components of MEP and do not cover the whole range.

However, automated segmentation based on images and point clouds of a laser-scanned scene in relation to MEP components is not possible yet. The maturity of automated segmentation of MEP components as integral part in a scan-to-BIM workflow is not yet achieved. In the construction industry, robust and trustworthy artificial intelligence solutions are essential, because, in practice, companies bear the liability risk and responsibility for incorrect models. As a consequence, full automation of the scan-to-BIM process lacks user acceptance, as BIM models fully generated by computer vision would need to be verified by a domain expert anyway.

Proposed approach

For this purpose, a workflow for object recognition of MEP components with deep neural networks is developed within the research project *LiDAR4HVAC*. This project aims to develop a semi-automatic scan-to-BIM workflow integrated as an application for BIM software. Figure 1 shows the framework of this workflow. The functional scope encompasses advanced image recognition to support counting MEP components and the automated generation of an as-built BIM model. Terrestrial laser scanners, equipped with high-resolution cameras and LiDAR sensors, provide images and a 3D point cloud as raw data. Object detection enables the recognition of MEP components within georeferenced panoramic imagery and can be used for object counting. In addition, geometric information is transferred by overlaying image material with the point cloud with instance segmentation. The segmented pixels in the images are projected onto the point cloud and used for geometric pre-segmentation and semantic enrichment of the object. A significant advantage of this workflow is the preliminary filtering of point cloud data through instance segmentation, which significantly reduces the manual effort required for geometric segmentation. The pre-segmented elements are the basis for the user's decision to transfer them as 3D object instances into a BIM model. For process transparency and quality assurance, the application implements a user validation interface where the final authorization of the transfer of 3D object instances is done by the user –a domain expert. This workflow is scalable and can be used for any type of reality capture based on terrestrial laser scanning.

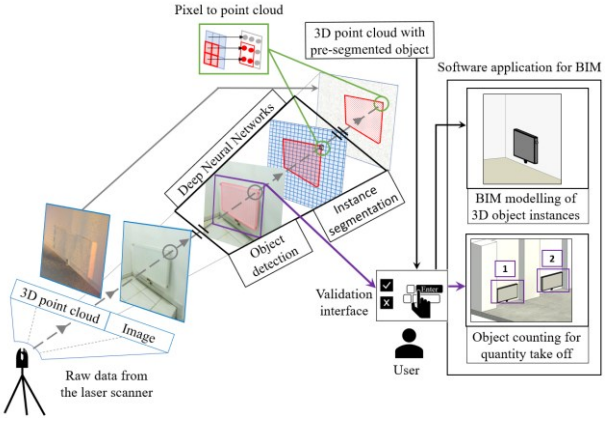

Figure 1: Framework of a software application for a scan-to-BIM workflow with deep neural networks

Objectives of the study

To support the proposed approach, this study aims to develop computer vision models for object detection and instance segmentation for MEP components. The objective is to demonstrate the feasibility, starting with an example. Radiators are chosen, as they are common in MEP systems and can usually be recorded with a laser scanner thanks to their size and visibility. More specifically, the following questions will be addressed:

- Which performance can be achieved for object detection and instance segmentation depending on the number of training images?
- What is the best balance between precision and recall, in terms of detection threshold for object detection and for instance segmentation, considering the intended object counting and scan-to-BIM applications?
- How suitable are the models for the proposed applications, based on a detailed analysis of the quality of the results, considering the characteristics of the images and the objects?

The results will support the selection of a relevant model for object detection and instance segmentation in the perspective of the proposed applications, respectively object counting and scan-to-BIM process.

Methodology

Overview

The workflow to recognize radiators with a deep neural network is illustrated in Figure 2. The basis for training a neural network is the acquisition of images with radiators. Each radiator needs to be labelled to constitute the ground truth information that the learning algorithm will use. A key part is the pretrained Deep Neural Network on which

transfer learning is applied. Part of the database is used for training, including validation. A separate part of the database is kept for testing the obtained model and evaluating its performance. The details of the methodology are described in the following subsections.

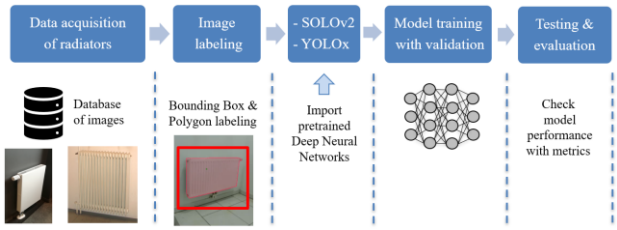

Figure 2: Workflow for object detection and instance segmentation of radiators

Deep neural network selection

In the domain of computer vision, tasks such as object detection and instance segmentation stand as exemplary processes that can be facilitated by a range of deep neural networks, which differ in their architecture and purpose. The following key aspects have been considered for choosing a deep neural network:

- recognized quality of the results obtained with the models, according to references
- capacity of dealing with the variety of image sizes and the number of instances, without preprocessing (e.g. encoding, transformation etc.)
- availability in the MATLAB library: in this work, the MATLAB environment is used for image processing and computer vision tasks; this brings practical benefits in data handling and fast implementation of transfer learning.

For object detection, one of the most commonly used deep neural networks is the YOLO (You Only Look Once) series (Jiang et al., 2022). In this study, the YOLOx model is selected, as the latest and best version of the YOLO series available in Matlab. Compared to YOLO v2, v3 and v4, YOLOx can readily handle different input image sizes, making real-time applications easier.

For instance segmentation, the Mask R-CNN and SOLOv2 (Segmenting Objects by Locations) methods are good candidates in Matlab. SOLOv2 is preferred for its better suitability to real-time applications. Indeed, SOLOv2 performs instance segmentation in a single stage, identifying pixel masks corresponding to each instance, without the need for bounding boxes. It supports any image input size. SOLOv2 has been used in combination with BIM in Wei et al. (2023) for mapping components in soil-foundation construction, as well as in Hart et al. (2023) for reconstructing pipes and fittings.

The advanced design of the selected methods combines accuracy and speed (Ge et al., 2021; Wang et al., 2020). This is especially relevant for real-time applications, such as object counting tasks in a dynamic building environment, like virtual walkthroughs in BIM, established by images from a laser scan (acquiring 2D images and 3D point clouds of the place).

Database for transfer learning

The developed models build upon pretrained deep neural networks, using transfer learning. For this, a database of more than 1000 images with radiators has been constituted. Targeting this intermediate number of images, sufficiently accurate results are expected. The influence of the number of training images on the model performance will be investigated to evaluate if more data is required.

The database gathers self-taken images of panel radiators and sectional radiators in different installation situations. The database consists of 1004 images, with at least one radiator in each image, summing up 1179 radiators. The image sizes range from 200 to 9200 pixels of equivalent square side length $\sqrt{width \times height}$, with $width$ the number of pixels on the x-axis of the image and $height$ the number of pixels of the y-axis of the image. The median equivalent square side length is 2096 pixels. The full distribution of image sizes can be seen in Figure 3.

This figure also provides an overview of the size of the radiators in equivalent square side length $\sqrt{A_{radiator}}$, with $A_{radiator}$ the surface area of the radiator mask in number of pixels. There are up to three radiators (mask 1, mask 2 and mask 3) in a single image. As Figure 4 shows, the image collection covers a wide range of radiator sizes, relatively to the image size. It can be noticed that many radiators are very small in the images: 40% of the radiators cover less than 4% of the image surface area and 18% less than 1%. The database includes several types of radiators, in various contexts. The radiators are often well

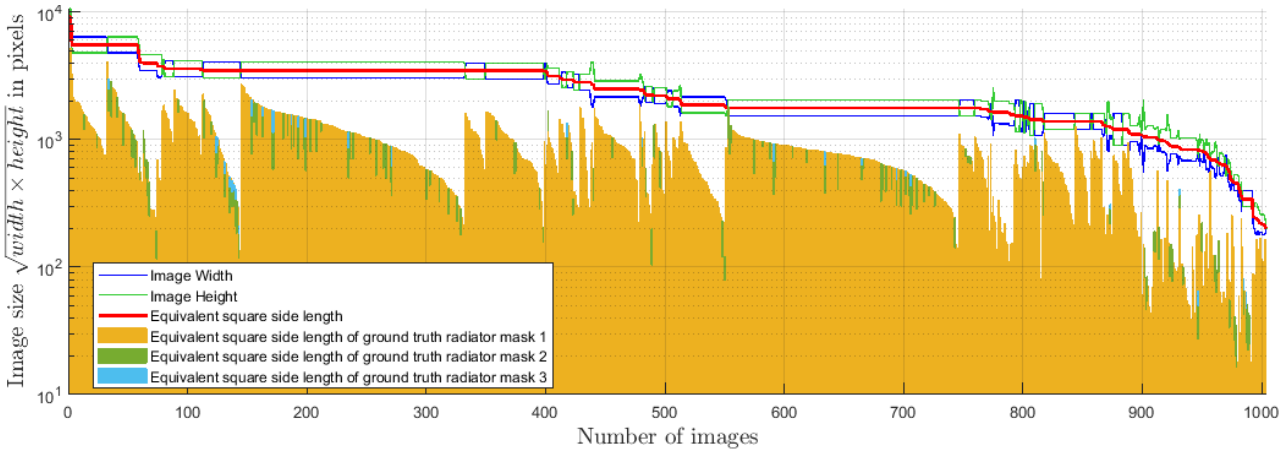

Figure 3: Distribution of the image sizes and radiator pixel mask sizes, with up to three instances per image (mask, 1, 2, and 3)

visible, but in 15% of the cases, the radiator mask covers less than 50% of the corresponding bounding box. This indicates that the radiator is significantly obstructed by other objects or pieces of furniture.

Each image has been labelled manually by a single annotator –a domain expert–, using the "image labeler"

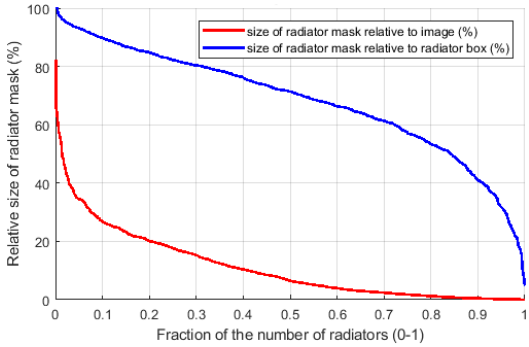

Figure 4: Distribution of the surface area of the 1179 radiator masks, relative to the image or box surface areas

application of Matlab (version R2022b), available in the computer vision toolbox. Depending on the task, different kind of labels are necessary. For object detection, rectangle boundary boxes are drawn around the radiator, while for instance segmentation, separate pixel masks have been drawn on each radiator in the images.

Applying transfer learning

To avoid training a model from scratch, the custom database is used for transfer-learning on pretrained deep neural network in MATLAB (R2024b). For object detection, the basis used is the "medium-coco" pretrained YOLOx network. For instance segmentation, the "resnet50-coco" pretrained SOLOv2 deep neural network is used as basis. As mentioned, with YOLOx and SOLOv2, the labelled data can be used directly, regardless of the image size, saving time in data pre-processing.

The labelled dataset represents the ground truth for training, validation and testing of the deep neural networks. Validation has been applied during the training process, to enable early stopping and to control the training progress. In particular, possible over- or underfitting can be detected in this way. A common practice is to divide the dataset into 70-80% for training, 10-20% for validation and 10-20% for testing (Rosebrock, 2017).

Depending on the training run, the number of images selected for the respective purposes varies, as summed up in Table 1. The images are selected randomly among the entire database and split into the three purposes: training, validation and testing. For each variant when varying the number of training images, the same test dataset is used. To obtain the best variant with each model, the training hyperparameters have been fine-tuned.

Evaluating the trained models' performance

The performance of the trained models is analyzed with typical evaluation metrics: the Precision-Recall (PR) curve, the Average Precision (AP) and the mean Average Precision (mAP). The PR curves are plot by default for an overlap threshold of 50% –also called Intersection over Union (IoU) threshold. This means that the ratio of the intersection over the union of the surface areas of the predicted and ground truth radiators must be at least of 50%. For object detection, the surface area of the bounding box is used, while for instance segmentation, the surface area of the pixel mask is taken. The Average Precision (AP) is given for the overlap threshold 50%, noted AP50. The mean Average Precision AP@[0.5:0.05:0.95] is calculated by varying the overlap threshold from 50% to 95% in 5 percentage points steps.

For accurate object counting, as well as for the development of an input mask for a BIM plug-in to perform semi-automatic segmentation of MEP components, it is important to determine the best balance between recall and precision in both models. The user will be assisted in the selection of MEP components on the basis of the automatic identification from the YOLOx and SOLOv2 models. On the one hand, if the precision of the model is too low, the user will have to remove a large number of selected MEP components (false positive). Conversely, if the recall rate is too low, many MEP components will not be detected and will have to be identified manually. This best possible balance for each model is identified based on the PR curve with an overlap threshold of 50%.

Additionally, the detection performance is analyzed in details, regarding the characteristics of the images and radiators, which are or tend to be detected, or not to be detected.

Table 1: Distribution of the number of images for training, validation and testing presented in the following results section

Case	Number of images		
	Training	Validation	Testing
YOLOx: image number variation	50	5	124
	100	10	124
	200	20	124
	400	40	124
	800	80	124
YOLOx: best case	396	44	124
SOLOv2: image number variation	42	11	156
	84	22	156
	169	43	156
	339	85	156
	678	170	156
SOLOv2: best case	400	40	124

Results

Object detection with YOLOx

To analyze the performance of the YOLOx model depending on the number of training images, the PR-curves of several variants are shown in Figure 5. It can be observed that the performance increases steadily when increasing from 50 to 400 images. The training result with 800 images show that the training process failed to make the most of the larger database.

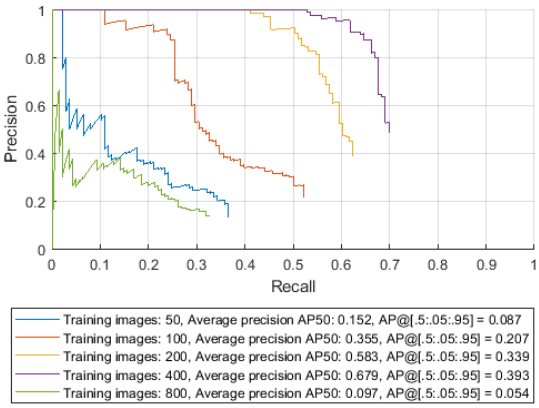

Figure 5: Influence of the number of training images on object detection with YOLOx: Precision-Recall curves

Investigating further with 400 images, the best model that could be achieved is presented in Figure 6. The PR curve is plot for several overlap thresholds. A significant decrease in the detection quality is observed already between 50% and 60% overlap threshold, indicating a relative mismatch between the ground truth and the detected bounding boxes. Nevertheless, close to 100% precision is ensured up to 75% recall with an overlap threshold of 50%. This indicates the radiators are mostly detected correctly, even if the bounding box size sometimes vary from the ground truth. The average precision for 50% overlap threshold AP50 reaches 80%, while the AP@[0.5:0.05:0.95] is 45%.

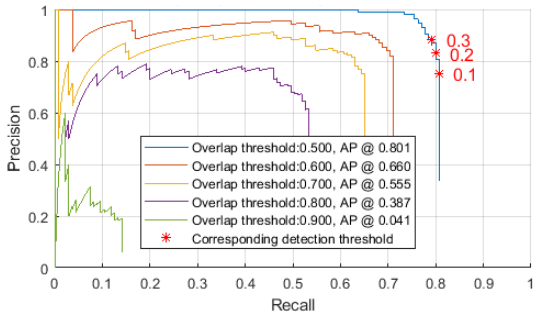

Figure 6: Best model for object detection with YOLOx: Precision-Recall curve for various overlap thresholds. AP50: 0.801; AP@[0.5:0.05:0.95]=0.449

Detection thresholds of 0.1 to 0.3 are shown in Figure 6. The best trade-off seems to be for 0.3: choosing a lower value (moving to right on the curve) decreases the precision, without increasing significantly the recall, while a higher value (moving to the left on the curve) would decrease significantly the recall. However, as illustrated in Figure 8 –which shows a selection of images, introduced later in this section–, the example of detection in image 30, evaluated as wrong, is actually relevant. This detection is considered wrong because the overlap ratio is below 50%. This underlines again the quality of the detection, despite a relative mismatch of the bounding boxes. With a confidence just over 0.1, this case encourages to lower the detection threshold. The precision for this detection threshold of 0.1 is 76% which means less than one fourth of the detected radiators are false positive. This threshold ensures the highest possible recall of 81%: indicating four fifth of all radiators are correctly detected.

To analyze more in details the behavior of the best obtained YOLOx object detector, Figure 7 shows the distribution of the radiators of the test dataset, in terms of their size relative to the image (in x-axis) and in terms of image size (in y-axis).

Figure 8 complements the diagram by providing examples of images, for which problems occur. The radiators of the selected images in Figure 8 are identified in Figure 7 for improved readability. In Figure 7, the correctly detected radiators are shown in color dots, using a color scale to indicate the detection confidence. The great majority of the radiators are correctly detected with a very high confidence, close to 1 (yellow color). The missed radiators, which are not identified, are shown as red crosses. Interestingly, most missed radiators belong to large images (as image 119). Although the YOLOx algorithm accepts any input image size, in large images with side length over 2500 pixels, radiators covering more than 20% of the image are rarely detected. Other missed radiators are mostly small, covering less than 15% of the image. False positive detections are mostly small objects, below 20% of the image side length. Yet, most small radiators are correctly detected, as image 119 illustrates. For image 30, already discussed, it can be noted that the "img30" labels (missed radiator and false detection) in Figure 7 are close to each other. Image 5 gives another example of wrong detection (false positive). Images 5 and 53 reveal that the model achieves correct detections despite the radiator being partially hidden.

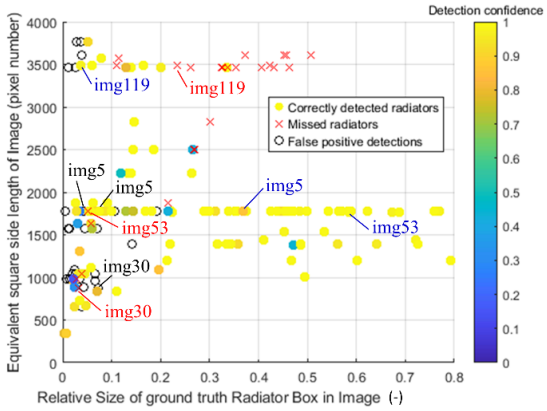

Figure 7: Ground truth characteristics of detected radiator masks (dots with the color scale indicating the detection confidence), as well as missed radiators and false positives, with reference to the radiator selection shown in Figure 8

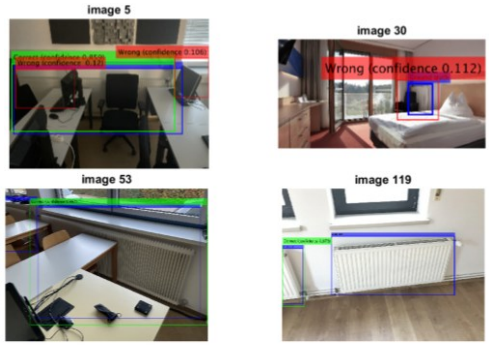

Figure 8: Detection of radiators with YOLOx in a selection of images, showing ground truth bounding boxes (in blue), correct detections (in green) and wrong detections (in red)

This analysis shows how the YOLOx model for radiator detection in images performs and gives insights of the model behavior, which will be evaluated relatively to the intended applications in the discussion section.

Instance segmentation with SOLOv2

The performance of the SOLOv2 model is analyzed in a similar way. Figure 9 plots the PR curves for various number of training images, showing quality improvement when increasing gradually from 42 images to 339 images. Only marginal improvement of the detection quality is observed between 339 and 678 training images: the average precision AP50 is respectively 0.746 and 0.749.

The best obtained SOLOv2 model used 400 images for training. The corresponding PR curves for various overlap thresholds are presented in Figure 10. The PR curves indicate that up to an overlap threshold of 80% the detection quality remains very high. This indicates a precise mask delimitation of the identified radiators. See for example in Figure 11, introduced later in this section, the identified masks in the images 4 and 29. This precise segmentation is confirmed by the AP@[0.5:0.05:0.95] which, at 64.5%, is significantly higher than the one of the YOLOx model (45%). For comparison, with 78.3%, the AP50 value of this SOLOv2 model for instance segmentation is almost as good as the one obtained with YOLOx for object detection (80%).

Several detection thresholds (0.25, 0.3 and 0.35) are shown on the default PR curve for an overlap threshold of 50% in Figure 10. A detection threshold of 0.25 provides almost the highest recall possible (81%), at the cost of precision (only 70%). A detection threshold of 0.35 reduces the recall to 75%, but the precision increases to 87%. To avoid too many false positive detections, the best balance is chosen for a detection threshold of 0.35.

Figure 11 shows the distribution of the radiators of the test database, in terms of their pixel mask size relative to the image (in x-axis) and in terms of their pixel mask size relative to the bounding box (in y-axis). The pixel mask size is expressed in equivalent side length $\sqrt{A_{radiator}}$, as defined in the methodology section. Figure 12 complements the diagram of Figure 11, by providing examples of images with detection problems. It can be seen that radiator masks that are larger than 5% of the image side length are mostly detected with a confidence higher than 60% (light blue to yellow dots). Images 4 and 29, in Figure 12, illustrate the wide range of radiator sizes detected, from large to small. Smaller radiator masks tend to be detected with a lower confidence (blue dots). Most missed radiators (red crosses) are in the area below 10% of the image size or below 50% of radiator mask size relative to the radiator bounding box. This corresponds to significantly hidden radiators, like in image 67, or very small radiators like the one on the left edge of image 4. The only missed radiator of a consequent size, in the test dataset, is shown in figure 99. This shape and type of radiator is probably too little represented in the training database. Overall, this SOLOv2 model mostly misses small radiators relatively to the image size. Wrong detections occasionally occur as illustrated by image 29 and 67.

The analysis in this section shows the SOLOv2 model performance for radiator detection in images. It provides information on the behavior of the model, which are useful to assess the further steps for the intended application of the model.

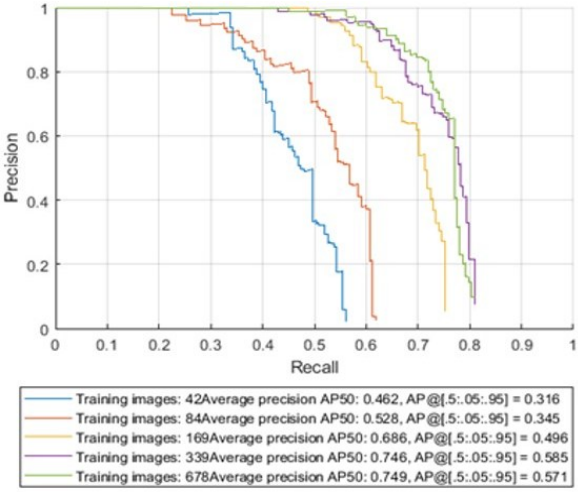

Figure 9: Influence of the number of training images on instance segmentation with SOLOv2: Precision-Recall curves

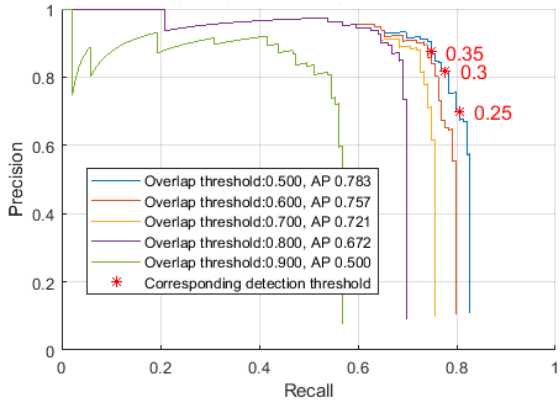

Figure 10: Best model for instance segmentation with SOLOv2: Precision-Recall curve for various overlap thresholds. AP50: 0.783; AP@[0.5:0.05:0.95]=0.645

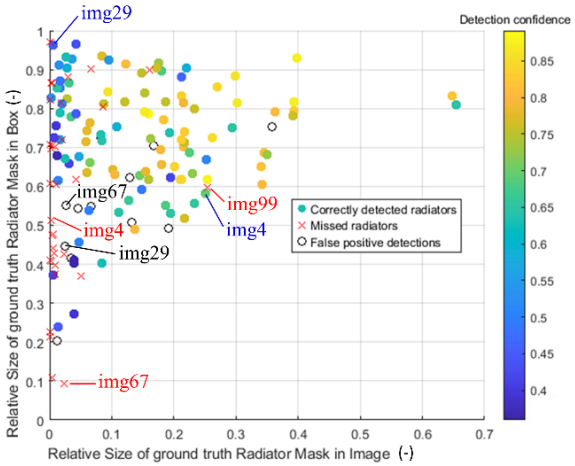

Figure 11: Ground truth characteristics of detected radiator masks (dots with the color scale indicating the detection confidence), as well as missed radiators and false positives, with reference to the radiator selection shown in Figure 12

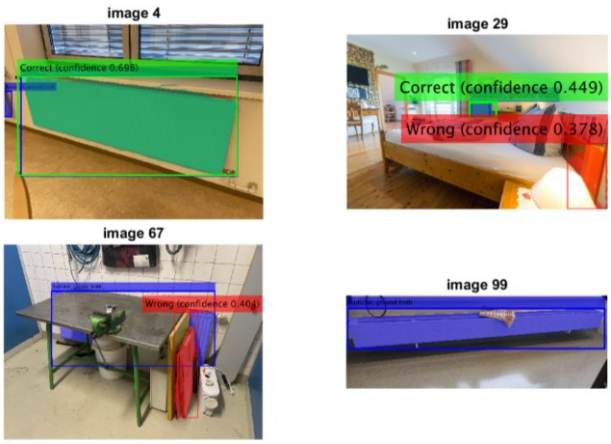

Figure 12: Segmentation of radiators with SOLOv2 in a selection of images, showing ground truth masks (in blue), correct detections (in green) and wrong detections (in red)

Discussion

The results analysis shows that the YOLOx model for radiator detection in images performs well with 80% average precision AP50. The SOLOv2 model for instance segmentation of radiators achieves 78% average precision AP50. These results are similar to the overall accuracy of 77.7% obtained by Wang et al. (2022) on various MEP components, as mentioned in the background analysis.

The constituted database has a great variety of images and object sizes and types. The high quality of the ground truth labelling ensured successful training on the data. It was demonstrated that 400 images are sufficient to achieve 80% of AP50 with YOLOx and 78% with SOLOv2. The difficulties encountered with higher number of images could be due to the diversity of the database. Nevertheless, YOLOx revealed a weakness for large images of more than 2500 pixels of average side length and radiators covering more than 20% of the image side length. By setting an upper limit to the image size, even higher object detection accuracy can be expected. Concerning SOLOv2 model, the analysis revealed that special cases, of radiators that are significantly hidden or of very small size relative to the image are often missed. These specific cases might be too little represented in the database to enable correct learning of such cases. On the positive side, this issue is not critical considering the intended scan-to-BIM application. Indeed, the geometry of hidden radiators can anyway not be scanned correctly by a laser scanner. And very small radiators in the image become larger when the camera moves closer to each object, removing the problem of low detection rate of small objects by SOLOv2. In any case, further investigation would be needed to harness the potential of more images –potentially more than 1000 in this perspective. This would enable to learn more difficult situations, where radiators are too hidden or of less common types.

For the scan-to-BIM application, it is crucial that well visible radiators are detected with precision. According to the high mean average precision AP@[0.5:0.05:0.95] of 0.645 achieved by the SOLOv2 model, the instance masks identified are very accurate, very close to the ground truth: 67% of the radiators are identified with an overlap ratio higher than 0.8 which is a very close match.

The best balance has been determined to be a detection threshold of 0.1 for YOLOx, ensuring recall levels of 81%, while keeping the precision reasonably high at 76%. For SOLOv2, the best balance is obtained with 0.35 detection threshold, leading to 75% recall and a high precision of 87%. These values ensure maximal detection of the existing radiators (recall), while minimizing the false positives, which decrease precision. In the scan-to-BIM application, a false positive can be discarded by the user, while a missed radiator must be defined manually. From this point of view, the highest recall should be preferred. Nevertheless, too low precision means a significant number of false positive to discard, which is not desired either. Through the detailed analysis of the test database, it could be identified that most missed radiators are either small or significantly hidden. As discussed, these cases are not critical for the scan-to-BIM workflow. For object counting, the same effect can be expected regarding small missed radiators. This means the optimal balance can be shifted to a lower recall, minimizing the number of false positive. This should enhance the user experience, avoiding unnecessary discarding.

Conclusion and future work

This paper presented the performance evaluation of specially trained deep neural network for object detection and instance segmentation of radiators, as an example of MEP components. The YOLOx model for object detection achieves 80% of Average Precision AP50, while the SOLOv2 model for instance segmentation reaches 78% of AP50. From the Precision-Recall curves, the optimal detection threshold could be set to 0.1 for object detection and 0.35 for instance segmentation. The detailed analysis of the models behavior with the test database revealed their strengths and weaknesses. The YOLOx model handles small objects well, but should not be used with to large input images (more than 2500 pixel of side length). The SOLOv2 model struggles with significantly hidden radiators as well as very small radiators (less than 5% of mask size relative to image side length). These issues are not critical for the intended applications of object detection for quantity take-offs and instance segmentation for scan-to-BIM processes.

The quality of the constituted database enabled successful training with only 400 images. The procedure can therefore be used for implementing transfer learning to other MEP components.

The findings of this paper are crucial for combining LiDAR technologies with computer vision methods to provide an efficient workflow for accurate, semi-automated documentation of as-built information during the construction process. Again, this presented work is the first step to develop a semi-automatic workflow for object recognition of MEP components with deep neural networks in combination with laser scanning and photogrammetry within the research project *LiDAR4HVAC*. The continued developments should lead to a plugin for BIM that will allow the seamless transfer

of laser scan data (images and point clouds) into a BIM model for the investigated MEP components.

Acknowledgments

The authors would like to thank the companies involved in the *LiDAR4HVAC* project for providing images of radiators and for their support. This project (FFG Grant No. FO999895483) was funded by the program "COIN Netzwerke 14. Ausschreibung", which is managed by the Austrian Research Promotion Agency (FFG) in cooperation with the Austrian Federal Ministry of Labour and Economy (BMAW).

References

Abioye, S. O., Oyedele, L. O., Akanbi, L., Ajayi, A., Delgado, J. M. D., Bilal, M., Akinade, O. O. & Ahmed, A. (2021) Artificial intelligence in the construction industry: A review of present status, opportunities and future challenges. Journal of Building Engineering, 44, 103299.

Awed, J. & Fini, A. A. F. (2024) Governmentality in construction claim management: Role of smart data initiatives. Project Leadership and Society, 5, 100158.

Bosché, F., Turkan, Y., Haas, C. T., Chiamone, T., Vassena, G. & Ciribini, A. (2013) Tracking MEP installation works. In: 30th International Symposium on Automation and Robotics in Construction - ISARC 2013. Montréal, Canada, IAARC Publications, pp. 229-239.

Casini, M. (2022) Construction 4.0 advanced technology, tools and materials for the digital transformation of the construction industry. Duxfoird, Woodhead Publishing.

Disney, O., Roupé, M., Johansson, M., Ris, J. & Höglin, P. (2023) Total BIM on the construction site: a dynamic single source of information. Journal of Information Technology in Construction, 28, pp. 519-538.

Doukari, O., Kassem, M. & Greenwood, D. (2023) Building Information Modeling. In: Disrupting Buildings - Digitalisation and the Transformation of Deep Renovation, Cham, Switzerland, Springer Nature, pp. 39-52.

Ge, Z., Liu, S., Wang, F., Li, Z., & Sun, J. (2021) Yolox: Exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430.

Hart, L., Knoblach, S., & Möser, M. (2023) Automated pipeline reconstruction using deep learning & instance segmentation. ISPRS Open Journal of Photogrammetry and Remote Sensing, 9, 100043.

Jiang, P., Ergu, D., Liu, F., Cai, Y. & Ma, B. (2022) A review of yolo algorithm developments, Procedia Computer Science, 199, pp. 1066-1073.

Kufuor, J., Mohanty, D. D., Valero, E. & Bosché, F. (2021) Automatic MEP component detection with deep learning. In: 25th International Conference on Pattern Recognition – ICPR 2020, Milan, Italy, Springer International Publishing, pp. 373-388.

Mirzaei, K., Arashpour, M., Asadi, E., Masoumi, H., Bai, Y. & Behnood, A. (2022) 3D point cloud data processing with machine learning for construction and infrastructure applications: A comprehensive review. Advanced Engineering Informatics, 51, 101501.

Monsberger, M. (2019) CHALLENGES AND RISKS RELATING TO MEP ENGINEERING IN LARGE BUILDING CONSTRUCTION PROJECTS. In: 10th International Structural Engineering and Construction Conference - ISEC 2019. Chicago, USA, ISEC Press.

Noichl, F., Braun, A. & Borrmann, A. (2021) "BIM-to-Scan" for Scan-to-BIM: Generating Realistic Synthetic Ground Truth Point Clouds based on Industrial 3D Models. In: Proceedings of the 2021 European Conference on Computing in Construction – EC^3, pp. 164-172.

Perez-Perez, Y., Golparvar-Fard, M., & El-Rayes, K. (2021) Scan2BIM-NET: Deep learning method for segmentation of point clouds for scan-to-BIM. Journal of Construction Engineering and Management, 147(9), 04021107.

Rosebrock, A. (2017) Deep Learning for Computer Vision with Python – Starter Bundle, PyImageSearch.

Tofan, B. A., Llamas, J. M., Alvarez, S. & Gutiérrez, M. C. (2020) Methods for surveying and diagnostics of HVAC systems in the existing buildings, Deliverable report 1.2.

Wang, B., Chen, Z., Li, M., Wang, Q., Yin, C. & Cheng, J. C. P. (2024) Omni-Scan2BIM: A ready-to-use Scan2BIM approach based on vision foundation models for MEP scenes. Automation in Construction, 162, 105384.

Wang, B., Wang, Q., Cheng, J. C., Song, C., & Yin, C. (2022) Vision-assisted BIM reconstruction from 3D LiDAR point clouds for MEP scenes. Automation in Construction, 133, 103997.

Wang, B., Yin, C., Luo, H., Cheng, J. C. & Wang, Q. (2021) Fully automated generation of parametric BIM for MEP scenes based on terrestrial laser scanning data. Automation in Construction, 125, 103615.

Wang, X., Zhang, R., Kong, T., Li, L., & Shen, C. (2020) Solov2: Dynamic and fast instance segmentation. Advances in Neural information processing systems, 33, pp. 17721-17732.

Wei, W., Lu, Y., Lin, Y., Bai, R., Zhang, Y., Wang, H., & Li, P. (2023) Augmenting progress monitoring in soil-foundation construction utilizing SOLOv2-based instance segmentation and visual BIM representation. Automation in Construction, 155, 105048.

Yarmohammadi, S. & Ashuri, B. (2015) Exploring the approaches in the implementation of BIM-based MEP coordination in the USA. Journal of Information Technology in Construction (ITcon), 20, pp. 347-363.

Yin, C., Wang, B., Gan, V. J., Wang, M., & Cheng, J. C. (2021) Automated semantic segmentation of industrial point clouds using ResPointNet++. Automation in Construction, 130, 103874.