Domain-Adaptive Object Detection of Electrical Facilities for Enhanced Semantic Indoor Models

Lukas Arzoumanidis[1], Weilian Li[1,2], Youness Dehbi[1]

[1] Computational Methods Lab, HafenCity University, Hamburg, Germany - {firstname.lastname}@hcu-hamburg.de
[2] Faculty of Geosciences and Engineering, Southwest Jiaotong University, Chengdu, China

Keywords: domain-adaptive learning, electrical utilities, as-built BIM, semantic enrichment, indoor models, augmented reality

Abstract

Detecting visible electrical utilities is a prerequisite for developing advanced reasoning strategies to reconstruct hidden in-wall networks. This paper investigates the detection of visible power-related utilities using a domain-adaptive deep learning-based vision pipeline based on the YOLOv11-L, object detection model. Four publicly available datasets containing power sockets, power strips, and light switches were curated, relabeled, and merged into a unified training dataset of 3,459 images. The resulting model achieved a mean average precision (mAP) of 0.74 for power sockets and strips and 0.98 for light switches, demonstrating strong detection performance. Real-time evaluation on a low-cost smartphone via the Ultralytics HUB App indicates reliable detection in small-scale real-world environments and detected utilities could be integrated automatically into semantic indoor models using a marker-less referencing approach.The work further highlights broader applications, including Augmented Reality-based visualization to reduce cognitive load for project managers and inspectors or construction workers and electricians, and its potential use as input for existing and future reasoning methods for hidden-utility reconstruction. The prepared dataset, trained model and source code is available at: `https://github.com/hcu-cml/indoor-electrical-facility-detection`.

1. Introduction

The adoption of Building Information Modeling (BIM) has gained global significance, marking a decisive shift in the digital transformation of the construction sector (Murguia et al., 2023; Zhao et al., 2018). BIM has emerged as a key enabling technology for the full life cycle management of buildings, from initial design to eventual demolition (Costin et al., 2018). Broadly speaking, two principal categories of BIM can be distinguished. *As-planned BIM* refers to the model developed prior to construction, representing the intended design and execution plan. In contrast, *as-built BIM* describes the actual state of the built environment once construction is completed (Mengiste et al., 2024).

In Europe, the building stock is largely composed of older structures that pre-date modern documentation and digital modeling practices. Over 65% of residential and service buildings in Germany were constructed before 1980 (Gevorgian et al., 2021), and more than 40% of European residential buildings date to before the 1960s, with a substantial portion built prior to 1990 (Dovjak et al., 2019). Most of these buildings lack comprehensive documentation, and when records do exist, they are often available only in analog form and require semantic interpretation and digitization before they can be accessed in a digital format. This lack of structured information presents major challenges for maintenance, retrofitting, and smart-city integration. To address this, approaches for the reconstruction from dense observations such as laser scanning and Simultaneous Localization and Mapping (SLAM) are increasingly used to create as-built BIM models for existing buildings. These approaches enable the capture of both spatial geometry and semantic information about building elements. However, hidden utilities, e.g., electrical conduits or water-pipes, remain difficult to map with these methods, despite their importance within a complete BIM context (Volk et al., 2014).

Hidden utilities such as electrical conduits and water pipes cannot be captured by laser scanning or SLAM-based methods, even though they are essential components of a complete BIM representation (Volk et al., 2014). To address this gap, Dehbi et al. (2022) introduced an incremental, constraint-based reasoning approach for estimating hidden electrical networks. Their method can reduce verification effort by up to 80% compared to traditional procedures. The reasoner first generates candidate utility layouts and, after a limited number of point-based on-site measurements, refines these into a final network estimate. However, the method requires tedious manual pre-identification of both visible and hidden utilities, which is tedious and labor-intensive. When these utilities are obscured, e.g., by furniture, historical documents such as floorplans or technical drawings must be reviewed, and the relevant elements extracted to supply the required input. In order to eliminate the need for tedious manual pre-identification of visible utilities, this work aims to present a first approach toward an end-to-end pipeline for the semantic enrichment of indoor models by automatically detecting indoor electrical utilities using domain-adaptive deep learning, and by outlining a systematic approach for their subsequent integration into semantic indoor models or BIM.

The reasoning framework of Dehbi et al. (2022) includes a virtual-reality (VR) interface designed to support intuitive inspection and immersive 3D visualization. VR has proven effective for validating as-built states of concealed electrical networks (Knechtel et al., 2024), but its mobility and real-world integration remain constrained by hardware and tracking limitations (Zheng et al., 2023). Unlike VR, which generates a fully synthetic environment, Augmented Reality (AR) superimposes digital information onto the physical scene (Dudhee and Vukovic, 2020). Inspectors and electricians can view estimated cable paths directly overlaid on real built walls and building elements, enabling intuitive, context-aware maintenance and repair in existing buildings.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-2-2026-759-2026

Another application area of the proposed approach lies in construction progress management. Conventional progress management practices rely heavily on the manual organization and interpretation of construction drawings and site progress reports, which is costly, time-consuming, and labor-intensive (Wei et al., 2022; Pal and Hsieh, 2021). This challenge is particularly pronounced in indoor construction environments, where managing Mechanical, Electrical, and Plumbing (MEP) systems, especially electrical utilities, is inherently complex. Furthermore, progress assessment often lacks an integrated visualization platform, making it difficult for both project managers and on-site workers to intuitively evaluate construction status. As a result, timely decision-making is hindered, and construction workers or electricians have limited support for aligning installed utilities with as-planned designs. Integrating progress-related information, such as the installation status of electrical utilities, into BIM environments can enable faster identification of schedule deviations and allow workers to more intuitively detect and correct missing or incomplete MEP components on construction sites.

Mobile AR applications allow users to freely navigate a building while accessing digital utility layouts, supporting immediate comparison between predicted and observed conditions, as illustrated in Figure 1. Hence, a comparison between the as-planned BIM model and the as-is site condition enables early detection of schedule deviations. AR systems can additionally provide real-time measurements, and lower cognitive load by presenting information directly within the spatial context. As a result, the second aim of this work is to present an initial step toward AR-based visualization of both visible and concealed electrical utilities in existing and under-construction buildings.

The remainder of the paper is structured as follows. Section 2 reviews recent advances in AR-based utility visualization and reasoning techniques for predicting hidden in-wall networks. Section 3 outlines the preparation and normalization of the training dataset, and Section 4 details our domain-adaptive deep learning methodology for utility detection, including its integration and implementation within the Ultralytics HUB App. Section 5 presents the performance evaluation of the proposed approach and discusses its potential incorporation into mobile AR frameworks and advanced reasoning systems for hidden-utility estimation. Section 6 outlines a possible integration of the proposed approach into semantic indoor models, discussing general aspects of camera pose estimation, coordinate transformations, and mapping strategies, and indicating how detected utilities could be localized within the room geometry. Finally, Section 7 concludes the paper and highlights directions for future research.

2. Related work

In the domain of as-built BIM, two principal paradigms can be distinguished: (1) reconstruction from dense observations using scan-to-BIM strategies, and (2) reasoning-based methods that operate on sparse observations supported by strong prior knowledge. Within the first paradigm, (Dehbi et al., 2021) introduced an optimal scan-planning framework that enforces network connectivity to guide the acquisition of three-dimensional indoor models. Recent advances in Scan-to-BIM research have focused on improving the accuracy and completeness of automatically reconstructed indoor models. Jarząbek-Rychard and Maas (2023) highlight the importance of uncertainty assessment in these workflows and propose a geodetic stochastic

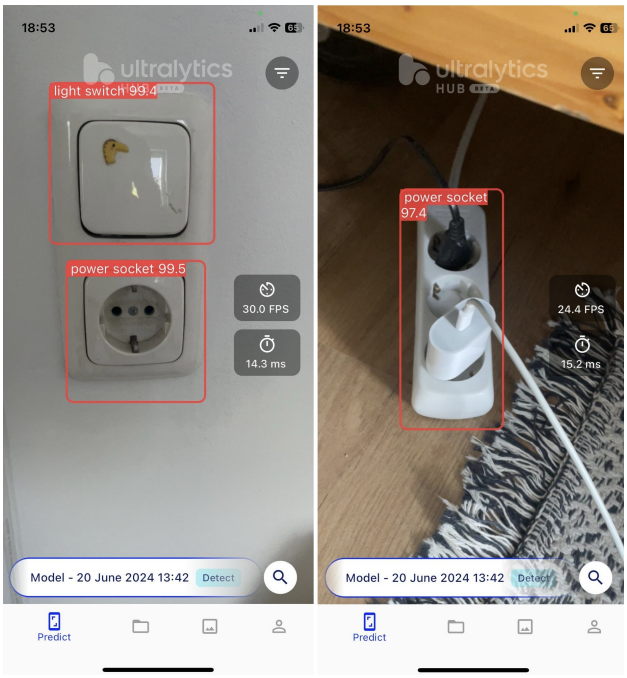


Figure 1. Mobile AR application that allows users to freely navigate a room while detecting visible electrical utilities.

modeling approach that incorporates statistical uncertainty directly into the reconstruction process, enabling systematic evaluation of dimensional accuracy with built-in self-verification. Complementing this, Mahmoud et al. (2025) introduce a deep learning-based framework capable of reconstructing both parametric BIM elements and non-parametric semantic indoor models from point clouds, including complex furniture that traditional methods often neglect. Their integration of deep learning and model-driven reconstruction addresses challenges such as incomplete data and clustering errors, producing geometrically faithful and semantically rich 3D representations.

In contrast, only a limited number of approaches have addressed the reconstruction of indoor models from sparse observations in combination with reasoning processes that do not depend on direct surveying. To enhance the semantic richness of indoor models by incorporating a-priori hidden utilities and infrastructure, such as electrical lines, additional reasoning-based strategies have been developed. In this context, (Dehbi et al., 2022, 2017) proposed an incremental, constraint-driven reasoning approach for estimating as-built electrical line routing in buildings, supported by informed sampling of electrical line hypotheses. To support user interaction with this reasoning framework, (Knechtel et al., 2024) presented an immersive virtual-reality environment for visualising and validating the inferred electrical line paths, as illustrated in Figure 5. However, their method assumes known locations of consumers, such as power sockets and light switches, a dependency that motivates the approach advanced in this work.

2.1 Computer vision for construction management

Shamsollahi et al. (2024) present a deep learning-based approach for automated progress monitoring in dense indoor construction environments. Their method leverages the YOLACT++ instance segmentation model with deformable convolutions to detect Mechanical, Electrical, and Plumbing (MEP) components from digital images. To address the visual

complexity of indoor scenes, the authors employ transfer learning, extensive data augmentation, and synthetically generated images that simulate realistic conditions such as occlusion, clutter, and variable lighting. They further investigate different combinations of real and synthetic training data to improve model robustness. The approach aims to reduce manual and error-prone inspections by enabling more reliable automated monitoring of indoor construction progress.

Furthermore, in the domain of construction progress management, the detection and temporal mapping of construction processes and their subsequent integration with BIM have been actively investigated. One approach employs Mask R-CNN to automatically assess wall construction progress across an entire floor, with the resulting progress information directly integrated into a BIM environment (Wei et al., 2022). Another approach presents a computer vision-based method for automatically detecting interior partition components and inferring their construction state from 2D digital images. Specifically, the method identifies studs, insulation, electrical outlets, and multiple drywall states, including installed, plastered, and painted conditions (Hamledari et al., 2017).

2.2 AR for underground and in-wall MEP utilities

A growing body of research investigates the use of augmented and mixed reality technologies to improve the visualization and management of underground utilities. Fenais et al. (2019) developed a mobile AR-GIS system that supports real-time mapping and cloud-based data sharing, addressing long-standing communication gaps between utility owners and construction personnel. Complementing this, Muthalif et al. (2024) introduced six interactive mixed-reality visualization methods and evaluated their effectiveness in terms of scene complexity, parallax, occlusion handling, and depth perception, demonstrating that different techniques offer distinct advantages for subsurface interpretation. Similarly, Tarek and Marzouk (2022) presented a handheld AR application that integrates BIM inspection data and cloud-based workflows to provide more robust performance for infrastructure operation and maintenance tasks. Providing broader context, Muthalif et al. (2022) offered a comprehensive review and taxonomy of AR visualization methods for subsurface utilities, identifying persistent challenges, such as poor depth cues, parallax effects, limited positional accuracy, and incomplete utility location data, and outlining key research gaps. Collectively, these studies demonstrate the increasing capacity of AR and MR technologies to improve the accuracy, usability, and overall effectiveness of utility visualization, and they highlight the substantial potential of these tools for representing as-built wall utilities and broader utility networks.

3. Dataset preparation

This section describes the preparation of the normalized dataset used to train our object detection model. Since the automatic detection and mapping of electrical utilities must perform reliably across a wide range of power socket, power plug, and light-switch designs, we compiled a large, heterogeneous, and normalized dataset for training and validation. The dataset integrates images of electrical utilities from multiple providers and diverse sources, ensuring broad variability in appearance and strengthening the model’s generalization capability.

In this work, we merged different datasets from Roboflow (Roboflow, Inc., 2025) and converted them to the YOLO annotation format, enabling the use of the Ultralytics ecosystem of pretrained YOLO models and facilitating rapid deployment of custom-trained models on mobile devices.

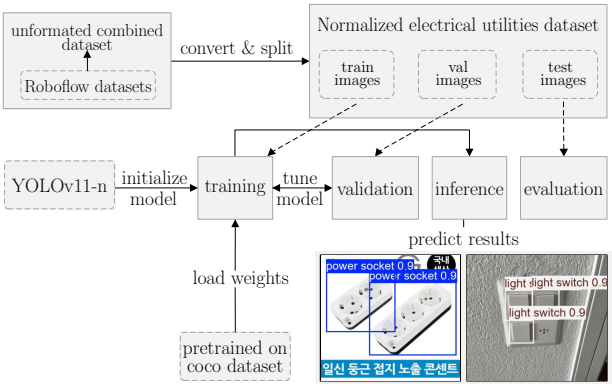


Figure 2. Overview of our approach for domain-adaptive electrical utilities detection.

Table 1. Structure of the prepared dataset.

datasets and classes	num. of images
power sockets (Roboflow, 2024)	1,590
power sockets (Roboflow, 2025)	1,003
light switches (Engelbracht, 2024)	715
power strips (Roboflow, 2022)	151

As shown in Table 1, the combined dataset form Roboflow comprises 3,459 images distributed across the three classes. The majority of samples (2,593) correspond to power sockets, while the smallest subset (151) belongs to the power strip class. All images in the prepared dataset exhibit varying resolutions, lighting conditions, and orientations. The prepared dataset was partitioned into 80% for training, 10% for testing, and 10% for validation and is publicly available by following this link: `https://zenodo.org/uploads/18835199`.

4. Methodology

Using the training subset, we train the YOLOv11 model to simultaneously detect three distinct electrical utilities, as illustrated in Figure 2. For our experiments, we adopted the Ultralytics[1] implementation of YOLOv11,. We selected the *Large* variant, which offers a greater number of learnable parameters than the *Regular* and *Small* versions and was pretrained on the COCO dataset prior to fine-tuning. The model was chosen for its strong generalization capability and widespread adoption, such as for the detection of tiny rooftop objects (Arzoumanidis et al., 2025). Compared with earlier versions, YOLOv11, incorporates improved spatial attention mechanisms, enabling the network to better focus on salient image regions and thereby achieve higher detection accuracy, particularly for small or partially occluded objects (Khanam and Hussain, 2024).

During training, model performance is assessed using two primary loss components: bounding box loss (box loss) and classification loss (cls loss). Classification loss measures the discrepancy between the predicted and true class labels, while bounding box loss quantifies the agreement between predicted

[1] `https://docs.ultralytics.com/models/yolo11/`

and ground-truth bounding boxes by accounting for factors such as aspect ratio and the spatial offset between box centers. The latter is typically computed using an IoU-based loss function.

For post-training evaluation, we use Precision, Recall, and mean Average Precision (mAP) as performance metrics. Computing mAP requires first determining the Intersection over Union (IoU) between predicted and ground-truth bounding boxes, along with the corresponding Precision and Recall values. IoU quantifies the ratio of the overlap area to the union area of the two boxes. Precision is defined as the proportion of true positive detections among all predictions, whereas Recall represents the proportion of true positives relative to all ground-truth instances. An IoU threshold is applied to ensure that only predictions with sufficient spatial overlap are counted as true positives. Precision-Recall (PR) curves are generated for each class by ranking predictions according to their confidence scores and computing Precision and Recall across varying thresholds. Class-wise Average Precision (AP) values are then aggregated to obtain the overall mAP. In this work, we additionally report mAP at an IoU threshold of 0.50 (mAP_{50}), corresponding to the average precision evaluated at IoU = 0.5.

The Ultralytics HUB App[2] provides a mobile platform for deploying and testing custom-trained YOLO models on consumer-grade devices. It supports real-time object detection, enabling users to stream camera input directly through the model for immediate inference results. The app allows seamless integration of domain-adaptive or fine-tuned YOLO models, offering high responsiveness even on low-cost smartphones due to its optimized inference engine. In addition to real-time visualization, the HUB App facilitates model management, performance monitoring, and rapid experimentation in real-world environments, making it well-suited for field testing of computer vision workflows such as indoor utility detection and semantic enrichment of indoor models.

Transfer learning has become a central strategy in deep learning, allowing models trained on large, generic datasets to be adapted efficiently to new tasks with limited annotated data (Pan and Yang, 2010; Yosinski et al., 2014). Instead of training a model from scratch, knowledge encoded in pre-trained feature representations is reused and fine-tuned for a more specific downstream application. This substantially reduces training time, improves generalization, and significantly lowers the data requirements for specialized tasks.

Domain adaptation represents a more targeted form of transfer learning, addressing scenarios in which the source domain, used for training, and the target domain, used for deployment, exhibit distributional differences (Wang and Deng, 2018; Ganin et al., 2016). These discrepancies, commonly referred to as *domain shift*, often degrade model performance when applying vision models trained on curated datasets to real-world conditions, such as indoor environments with varied lighting, occlusions, and sensor characteristics. Domain-adaptive methods mitigate this effect by aligning feature distributions across domains, either through adversarial learning, feature-space regularization and normalization, or style-transfer techniques. This allows the model to retain generalizable high-level representations while adapting to the specific visual characteristics of the target environment.

In the context of indoor utility detection, domain adaptation is particularly relevant thanks to publicly available datasets differ substantially from real indoor scenes captured on-site. Through adaptive fine-tuning, specialized models can maintain robust performance despite variations in textures, materials, cable designs, and sensor noise, making domain adaptation an essential component for reliable real-time detection and semantic enrichment of indoor models.

[2] https://docs.ultralytics.com/hub/app/

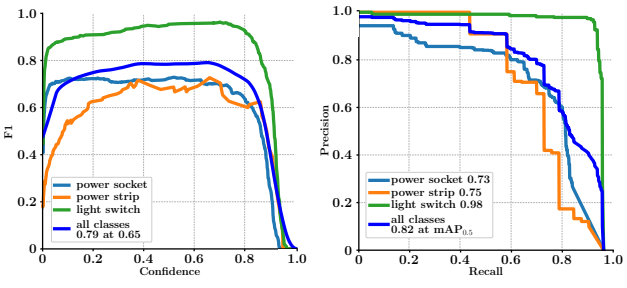
5. Experimental results

To detect, classify, and map the different classes of roof material, we employed a pretrained deep learning-based object detection model, specifically YOLOv11-L, which contains 25.3 million parameters and requires 86.9 billion floating-point operations (FLOPs), as provided by Ultralytics. The model was pretrained on the COCO dataset[3], which comprises 80 object categories. Our implementation is based on PyTorch and utilizes CUDA acceleration. All experiments were performed using PyTorch version 2.2.0 with CUDA 12.8 on an NVIDIA RTX 2000 Ada GPU equipped with 8 GB of VRAM. Default hyperparameter settings were used, and the model was trained for 100 epochs.

The experimental results indicate notable performance differences across the training, validation, and testing phases for all three target classes. Throughout training, both the training and validation loss functions of both models exhibit a consistent downward trend. As shown in Figure 4a, the bounding box loss and classification loss decrease within a similar value range. This behavior is expected, as both losses correspond to related regression tasks: the bounding box loss evaluates the accuracy of spatial localization, while the classification loss measures the correctness of predicted class labels. A closer comparison of the training and validation losses reveals that, after approximately 25 epochs, the classification loss begins to decline more rapidly than the bounding box loss. This suggests that the model learns to correctly classify objects earlier than it fully converges on precise spatial localization. When evaluating the mAP, mAP_{50}, Precision, and Recall during validation, all metrics demonstrate an increasing trend across epochs, as illustrated in Figure 4b. The consistent decrease in both training and validation losses, coupled with the increase in validation metrics, indicates effective model learning and generalization.

To assess the performance of our model trained on the normalized dataset, we conducted evaluations using a corresponding normalized test dataset. We computed both the F1-confidence curve and the Precision-Recall curve, which together characterize the model's detection quality, as shown in Figure 3. The

[3] https://cocodataset.org



(a) F1-confidence curve (b) Precision-Recall curve

Figure 3. Per-class F1 score-confidence curves and precision-recall curves for the predicted electrical utilities on our normalized test dataset.

Table 2. Per-class and overall bounding box Precision, Recal, mAP and mAP_{50} after inference on the test dataset.

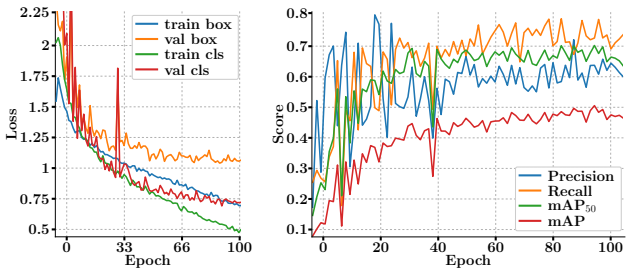
class	Precision	Recall	mAP_{50}	mAP
power socket	0.8	0.62	0.73	0.47
power strip	0.89	0.61	0.74	0.59
light switch	0.95	0.96	0.98	0.76
overall	0.88	0.73	0.82	0.61

results indicate that the highest overall mean F1 score across all classes is 0.79, reached at a confidence threshold of 0.65. This represents a strong outcome given the heterogeneity of the dataset, as discussed in Section 3. An examination of individual classes shows that power sockets and power strips follow similar F1 score trends across confidence thresholds, whereas light switches stand out as a positive outlier. The F1-confidence curve further demonstrates that the peak F1 score for light switches occurs at a comparatively higher confidence threshold, as illustrated in Figure 3a. This suggests that the model achieves more accurate and more confident predictions for light switches than for the average of all classes, despite the fact that this class contains only about half as many training instances as the power socket class.

To further substantiate the previous findings, Figure 3b illustrates the Precision-Recall curve. In agreement with the F1-confidence results, the Precision-Recall curve demonstrates distinctly higher Precision and Recall values for the light switch class, whereas the other two classes exhibit comparatively lower performance. As previously noted, the power socket and power strip classes remain close to the overall class average. The highest mAP_{50} value, computed as the area under the Precision-Recall curve at an IoU threshold of 0.50, is 0.82, indicating strong overall performance across confidence thresholds when all classes are considered together.

As shown in Table 2, the mAP values for the power socket and power strip classes are nearly identical, which may indicate a weak separability between these classes in the training images and their corresponding annotations. This can be explained by the fact that power strips typically consist of one or more power sockets combined with a power cable. Consequently, if the model fails to detect the power cable in a given instance, it may default to predicting a power socket rather than a power strip, leading to ambiguous classifications. This behavior is illustrated in Figure 6. However, the light switch class, despite having only half as many training instances, exhibits markedly higher performance.

As illustrated in Figure 1, our model, deployed in the Ultralytics Hub App, successfully detects all electrical utilities across all classes in each scene during the camera traversal through different rooms, consistently producing high-confidence predictions.



(a) losses (b) metrics

Figure 4. Training and validation loss, validation mAP, Precision and Recall.

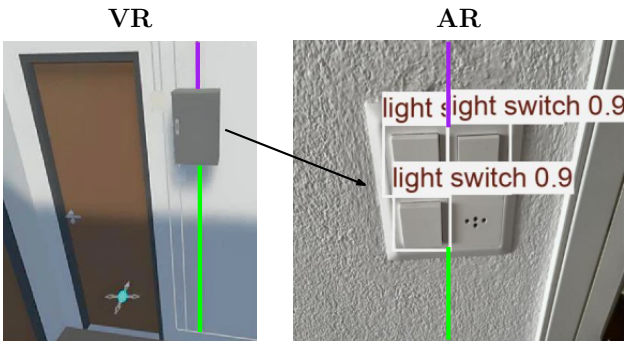


Figure 5. Illustration of the VR-based visualization from Knechtel et al. (2024) (left) and a corresponding AR-based visualization enabled by our approach (right).

Given that power strips are underrepresented in the Roboflow dataset, it is plausible that the training process suffers from class imbalance with respect to the power strip category. To mitigate the effects of class imbalance for underrepresented objects such as power strips, class-specific data augmentation can be applied. Although the dataset exhibits substantial heterogeneity in terms of surrounding environments (i.e., background clutter) and lighting conditions under which the images were captured, these variations are unlikely to severely degrade model performance. However, to address potential inaccuracies arising from low-light conditions, the integration of LiDAR data acquired from mobile devices, such as Apple iPhone or iPad models, could be beneficial. For example, depth information obtained from close-range scans could be used to better discriminate electrical utilities from non-relevant objects. Collectively, these factors may significantly improve the robustness of the proposed approach.

6. Integrating detections into semantic indoor models

To explore the integration of detected electrical utilities into semantic indoor models using low-cost mobile sensors, mobile Apple devices are considered as a representative platform due to their integrated LiDAR sensors, which may offer advantages for electrical utility detection, particularly critical under low-light conditions such as those encountered at construction sites. In this context, ARKit, first released in 2017[4], provides core augmented reality capabilities related to tracking and sensing. Specifically, ARKit handles the continuous estimation of the device’s camera pose (Six Degrees of Freedom (6DoF)) and establishes the world coordinate reference frame required for spatial tracking.

Building upon the foundation of ARKit, RealityKit (introduced in 2019[5]) provides advanced functionalities for 3D rendering and scene management. In our approach, these capabilities can be leveraged to visualize indoor models derived from as-planned or as-built BIM data, including the precise geometry of Mechanical, Electrical, and Plumbing (MEP) utilities and architectural openings. Operating at a higher level of abstraction, the RoomPlan API (2022[6]) facilitates parametric room scanning by

[4] https://developer.apple.com/augmented-reality/arkit/
[5] https://developer.apple.com/documentation/realitykit
[6] https://developer.apple.com/augmented-reality/roomplan/

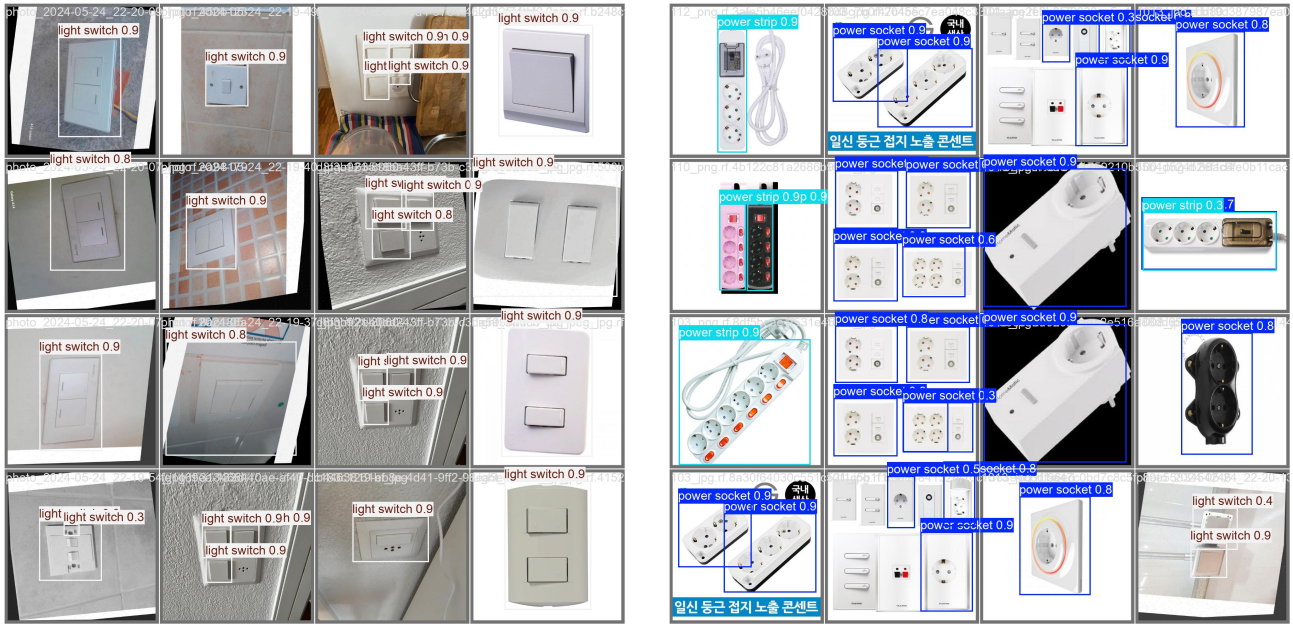


Figure 6. A diverse set of predicted electrical utilities identified in our normalized test dataset.

synthesizing device sensor data. The scanning process initiates at the device's start position, establishing the world origin $(0, 0, 0)$ for the AR session. By fusing LiDAR measurements with optical data, the framework reconstructs the environment's spatial geometry and classifies features into semantic categories, such as walls, doors, and windows.

To integrate our detection pipeline into Apple's AR environment, we employ the Ultralytics implementation of YOLOv11 for CoreML[7]. This enables the direct deployment of the trained YOLOv11 model within Apple's ARKit framework, thereby eliminating the need to rely on the Ultralytics HUB application at runtime. The integration is straightforward, as the CoreML-based implementation supports a plug-and-play deployment workflow and offers clear advantages over previous approaches, such as the method proposed by Le et al. (2021).

To integrate semantic information, such as the location and type of electrical utilities, derived from computer vision inference, with a semantic indoor model or the architectural specifications of a BIM, the spatial mapping from the two-dimensional image plane of the capturing device to the fixed three-dimensional coordinate system of the indoor model or BIM.

For simplification, we consider the integration into a BIM in the following. The initial phase involves translating the two-dimensional output of the object detection model into three-dimensional spatial coordinates. The detection model operates on 2D pixel buffers, returning a bounding box vector $B = (w, h)$ within the image space. To resolve the depth ambiguity inherent in monocular vision, a raycasting operation is performed. Utilizing the device's pinhole camera model and intrinsic matrix K, a ray $\vec{r}$ is projected from the bounding box centroid $(\frac{w}{2}, \frac{h}{2})$ into the scene. The intersection of this ray with the geometry G, generated by the device's LiDAR sensor using the RoomPlan API, yields the spatial coordinate P_{AR}

$$P_{AR} = O_{cam} + d \cdot K^{-1} \begin{bmatrix} \frac{w}{2} \\ \frac{h}{2} \\ 1 \end{bmatrix}, \tag{1}$$

where O_{cam} is the camera origin in the local session space, and d represents the depth value obtained from the LiDAR depth map at the intersection point.

A fundamental challenge in this integration is the discrepancy between the stochastic initialization of the AR session's local coordinate system (S_{AR}) and the absolute coordinate system defined within the BIM (S_{BIM}). The AR session typically initializes its world origin $(0, 0, 0)$ at the device's start position, necessitating a transformation to map detected electric utilities to the coordinate system of the BIM.

To accurately map the detected utilities, a rigid body transformation matrix T must be computed such that

$$P_{BIM} = T \times P_{AR}, \tag{2}$$

where P_{BIM} represents the target coordinates within the building model. The matrix $T \in \mathbb{R}^{4\times4}$ encapsulates both the rotational matrix (R) and the translation vector ($\mathbf{t}$) required to align the two reference frames using homogeneous coordinates

$$T = \begin{bmatrix} R & \mathbf{t} \\ 0 & 1 \end{bmatrix}. \tag{3}$$

To empirically derive the transformation matrix T, two primary methodologies are proposed.

A deterministic method relies on a common reference anchor, e.g., QR code, existing in both the physical space and the digital model. By detecting the anchor in the AR session at position C_{AR} and identifying its corresponding coordinate C_{BIM} in the model, the translational offset Δ can be calculated.

[7] https://www.ultralytics.com/blog/bringing-ultralytics-yolo11-to-apple-devices-via-coreml

Assuming a pre-aligned azimuthal orientation (e.g., aligning the user's start orientation with the BIM North), the transformation simplifies to a vector addition of the offset

$$\Delta = C_{BIM} - C_{AR}. \tag{4}$$

Consequently, for every detected electric utility i, the final coordinate is derived as

$$\begin{bmatrix} x_{BIM} \\ y_{BIM} \\ z_{BIM} \end{bmatrix}_i = \begin{bmatrix} x_{AR} \\ y_{AR} \\ z_{AR} \end{bmatrix}_i + \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix}. \tag{5}$$

For a marker-less referencing approach, the parametric output of the RoomPlan framework could be utilized. The system captures an indoor model as simplified geometries of structural elements (walls, openings), as illustrated in Figure 7. This local 3D indoor model is subsequently registered against the structural subset of the BIM. This registration is achieved manually by the operator, who places a virtual anchor at a pre-determined reference point (e.g., the lower-left corner of the room) to align the device's start orientation with the BIM's coordinate system (e.g., North).

Figure 7. Exemplary office room captured in 3D using RoomPlan API. Windows, doors, floor and walls are detected and reconstructed in a local coordinate reference system.

7. Conclusion & Outlook

This paper presents an initial step toward the automatic semantic enrichment of indoor electrical utilities. A normalized training dataset was created by aggregating and harmonizing four publicly available datasets, each containing a single category of electrical fixtures. All annotations were cleaned, merged, and converted into the YOLO annotation format. The trained model achieved a mean average precision (mAP) of 0.74 for power sockets and power strips and 0.98 for light switches, demonstrating strong detection performance. The model was further deployed within the Ultralytics HUB App to enable real-time detection on low-cost mobile devices, where it produced promising results in small-scale practical testing. Besides, a marker-less referencing approach, for the automatic integration of predicted electrical utilities into a semantic indoor model has been presented, as illustrated in Figure 5. In addition, the detected electrical utilities can be visualized through augmented reality (AR) systems, as can be seen in Figure 7, reducing cognitive load for inspectors and electricians and facilitating intuitive, context-aware construction, maintenance and repair operations.

The enriched indoor model can also benefit advanced reasoning strategies, such as those used to infer hidden electrical utilities or reconstruct in-wall electrical networks (Dehbi et al., 2022), by directly providing reliable observations without the need for manual annotation. Another approach to hidden-utility reasoning involves adapting existing Common Sense Knowledge Graphs (CSKG) (Ilievski et al., 2021), specifically tailored for HVAC and electrical utility reasoning in indoor environments. Prior work on Spatial Commonsense Graphs for object localisation in 3D scenes has already demonstrated the feasibility of this direction and provides a strong foundation for future research (Giuliari et al., 2022).

References

Arzoumanidis, L., Li, W., Knechtel, J., Kosmayadi, Y., Dehbi, Y., 2025. Automatic Detection of Tiny Drainage Outlets and Ventilations on Flat Rooftops from Aerial Imagery. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, X-G-2025, 125–132.

Costin, A., Adibfar, A., Hu, H., Chen, S. S., 2018. Building Information Modeling (BIM) for Transportation Infrastructure – Literature Review, Applications, Challenges, and Recommendations. *Automation in construction*, 94, 257–281.

Dehbi, Y., Haunert, J.-H., Plümer, L., 2017. Stochastic and Geometric Reasoning for Indoor Building Models with Electric Installations – Bridging the Gap between GIS and BIM. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4, 33–39.

Dehbi, Y., Knechtel, J., Niedermann, B., Haunert, J.-H., 2022. Incremental Constraint-based Reasoning for Estimating As-built Electric Line Routing in Buildings. *Automation in Construction*, 143, 104571.

Dehbi, Y., Leonhardt, J., Oehrlein, J., Haunert, J.-H., 2021. Optimal Scan Planning with Enforced Network Connectivity for the Acquisition of Three-dimensional Indoor Models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 180, 103–116.

Dovjak, M., Slobodnik, J., Krainer, A., 2019. Deteriorated Indoor Environmental Quality as a Collateral Damage of Present Day Extensive Renovations. *Journal of Mechanical Engineering/Strojniški Vestnik*, 65(1).

Dudhee, V., Vukovic, V., 2020. Superimposing Building Information Models in Augmented Reality. *20th International Conference on Construction Applications of Virtual Reality: Enabling the Development and Implementation of Digital Twins*, Teesside University, 11–18.

Engelbracht, T., 2024. SpotLight: Light Switch Dataset. https://universe.roboflow.com/timengelbracht/spotlight-light-switch-dataset. Visited: 2026-02-03.

Fenais, A., Ariaratnam, S. T., Ayer, S. K., Smilovsky, N., 2019. Integrating Geographic Information Systems and Augmented Reality for Mapping Underground Utilities. *Infrastructures*, 4(4).

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V., 2016. Domain-adversarial Training of Neural Networks. *J. Mach. Learn. Res.*, 17(1), 2096–2030.

Gevorgian, A., Pezzutto, S., Zambotti, S., Croce, S., Oberegger, U. F., Lollini, R., Kranzl, L., Muller, A., 2021. European Building Stock Analysis. *Eurac Research: Bolzano, Italy*.

Giuliari, F., Skenderi, G., Cristani, M., Wang, Y., Del Bue, A., 2022. Spatial Commonsense Graph for Object Localisation in Partial Scenes. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 19496–19505.

Hamledari, H., McCabe, B., Davari, S., 2017. Automated Computer Vision-based Detection of Components of Under-construction Indoor Partitions. *Automation in Construction*, 74, 78-94.

Ilievski, F., Szekely, P., Zhang, B., 2021. CSKG: The CommonSense Knowledge Graph. R. Verborgh, K. Hose, H. Paulheim, P.-A. Champin, M. Maleshkova, O. Corcho, P. Ristoski, M. Alam (eds), *The Semantic Web*.

Jarząbek-Rychard, M., Maas, H.-G., 2023. Modeling of 3D Geometry Uncertainty in Scan-to-BIM Automatic Indoor Reconstruction. *Automation in Construction*, 154, 105002.

Khanam, R., Hussain, M., 2024. YOLOv11: An Overview of the Key Architectural Enhancements.

Knechtel, J., Li, W., Orgeig, Y., Haunert, J.-H., Dehbi, Y., 2024. Immersive Virtual Reality to Verify the As-built State of Electric Line Networks in Buildings. T. H. Kolbe, A. Donaubauer, C. Beil (eds), *Recent Advances in 3D Geoinformation Science (Proceedings of the 18th 3D GeoInfo Conference 2023, Munich)*, Springer Nature Switzerland, Cham, 129–143.

Le, H., Nguyen, M., Yan, W. Q., Nguyen, H., 2021. Augmented Reality and Machine Learning Incorporation Using YOLOv3 and ARKit. *Applied Sciences*, 11(13).

Mahmoud, M., Zhao, Z., Chen, W., Adham, M., Li, Y., 2025. Automated Scan-to-BIM: A Deep Learning-based Framework for Indoor Environments with Complex Furniture Elements. *Journal of Building Engineering*, 106, 112596.

Mengiste, E., Garcia de Soto, B., Hartmann, T., 2024. Automated Integration of As-is Point Cloud Information with As-planned BIM for Interior Construction. *International Journal of Construction Management*, 24(2), 137–150.

Murguia, D., Vasquez, C., Demian, P., Soetanto, R., 2023. BIM Adoption Among Contractors: A Longitudinal Study in Peru. *Journal of Construction Engineering and Management*, 149(1), 04022140.

Muthalif, M. Z. A., Shojaei, D., Khoshelham, K., 2022. A review of augmented reality visualization methods for subsurface utilities. *Advanced Engineering Informatics*, 51, 101498.

Muthalif, M. Z. A., Shojaei, D., Khoshelham, K., 2024. Interactive Mixed Reality Methods for Visualization of Underground Utilities. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 92(6), 741-760.

Pal, A., Hsieh, S.-H., 2021. Deep-learning-based Visual Data Analytics for Smart Construction Management. *Automation in Construction*, 131, 103892.

Pan, S. J., Yang, Q., 2010. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345-1359.

Roboflow, 2022. Power strip dataset. https://universe.roboflow.com/test-gxlza/power-strip-cosfv. Visited: 2026-02-03.

Roboflow, 2024. Socket rocket dataset. https://universe.roboflow.com/test-gvutb/socket_rocket. Visited: 2026-02-03.

Roboflow, 2025. Power socket detection dataset. https://universe.roboflow.com/yolov5-power-socket-detection/power-socket-detection-rofcv. Visited: 2026-02-03.

Roboflow, Inc., 2025. Roboflow public datasets. https://public.roboflow.com/. Accessed: 2025-11-15.

Shamsollahi, D., Moselhi, O., Khorasani, K., 2024. Automated Detection and Segmentation of Mechanical, Electrical, and Plumbing Components in Indoor Environments by Using the YOLACT++ Architecture. *Journal of Construction Engineering and Management*, 150(8), 04024100.

Tarek, H., Marzouk, M., 2022. Integrated Augmented Reality and Cloud Computing Approach for Infrastructure Utilities Maintenance. *Journal of Pipeline Systems Engineering and Practice*, 13(1), 04021064.

Volk, R., Stengel, J., Schultmann, F., 2014. Building Information Modeling (BIM) for Existing Buildings-Literature Review and Future Needs. *Automation in construction*, 38, 109–127.

Wang, M., Deng, W., 2018. Deep Visual Domain Adaptation: A Survey. *Neurocomputing*, 312, 135–153.

Wei, W., Lu, Y., Zhong, T., Li, P., Liu, B., 2022. Integrated Vision-Based Automated Progress Monitoring of Indoor Construction Using Mask Region-Based Convolutional Neural Networks and BIM. *Automation in Construction*, 140, 104327.

Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How Transferable Are Features in Deep Neural Networks? *Advances in Neural Information Processing Systems (NeurIPS)*, 3320–3328.

Zhao, X., Pienaar, J., Gao, S., 2018. Critical Risks Associated with BIM Adoption: A Case of Singapore. *Proceedings of the 21st International Symposium on Advancement of Construction Management and Real Estate*, Springer, 585–596.

Zheng, X., Sun, S., Cao, Y., Li, J., Ding, D., Li, Z., 2023. Extended VR: Exploring the Integration of VR Experiences and Real-World Engagement. *Companion Publication of the 2023 ACM Designing Interactive Systems Conference*, 184–187.