Designing a Human-in-the-Loop System for Object Detection in Floor Plans

Johannes Jakubik*, **Patrick Hemmer***, **Michael Vössing, Benedikt Blumenstiel, Andrea Bartos, Kamilla Mohr**

Karlsruhe Institute of Technology
{johannes.jakubik, patrick.hemmer, michael.voessing, benedikt.blumenstiel}@kit.edu,
{andrea.bartos, kamilla.mohr}@alumni.kit.edu

Abstract

In recent years, companies in the Architecture, Engineering, and Construction (AEC) industry have started exploring how artificial intelligence (AI) can reduce time-consuming and repetitive tasks. One use case that can benefit from the adoption of AI is the determination of quantities in floor plans. This information is required for several planning and construction steps. Currently, the task requires companies to invest a significant amount of manual effort. Either digital floor plans are not available for existing buildings, or the formats cannot be processed due to lack of standardization. In this paper, we therefore propose a human-in-the-loop approach for the detection and classification of symbols in floor plans. The developed system calculates a measure of uncertainty for each detected symbol which is used to acquire the knowledge of human experts for those symbols that are difficult to classify. We evaluate our approach with a real-world dataset provided by an industry partner and find that the selective acquisition of human expert knowledge enhances the model's performance by up to 12.9%—resulting in an overall prediction accuracy of 92.1% on average. We further design a pipeline for the generation of synthetic training data that allows the systems to be adapted to new construction projects with minimal manual effort. Overall, our work supports professionals in the AEC industry on their journey to the data-driven generation of business value.

Introduction

Over the last few years, many companies in the Architecture, Engineering, and Construction (AEC) industry have started exploring how artificial intelligence (AI) can be used to improve work processes. However, so far, only a few medium- and large enterprises have utilized AI, and even fewer have managed to turn their data into actual business value (Blanco et al. 2018). At the moment, one of the most promising applications of AI is its use in augmenting the capabilities of employees by simplifying time-intensive and repetitive tasks. The idea is to employ AI to allow their employees to focus on other value-added activities—often those requiring human creativity (Dellermann et al. 2019).

The preliminary material take-off of construction projects, which is required to ensure efficient project execution, has high potential to benefit from the adoption of AI. For example, AI systems can be used to support humans in determining the bill of materials (BOM) required in all construction projects—a process also referred to as quantity determination. The goal of quantity determination is to calculate the actual and target quantities of products required for a project in accordance with the structure of the BOM (i.e., the relevant components). This task has widespread implications for several phases of the construction process. First, even before a contract is signed, a BOM is drawn up to document the required work, which affects the planning and scheduling of companies in the offering process. Second, the work preparation of the construction company—often conducted separately for different trades—requires the ordering of the correct quantities, which might have changed in the meantime. Third, for the billing, the actually executed quantities have to be determined—which once again might deviate from the target quantities.

Even though digital representations of buildings are increasingly being utilized today, they are far from common practice as the majority of the existing building stock is not represented in a digital format. Available data formats are often not compatible with one another due to the industry's high levels of technical fragmentation. Thus, companies are required to invest manual effort into counting relevant symbols in floor plans for the use case of quantity determination. This time-consuming task prevents domain experts from applying their scarce resources to other aspects of the construction process. To support domain experts with the determination of quantities for relevant symbols in the scope of BOM, we propose a human-in-the-loop system that allows companies to utilize semi-automatic floor plan analysis. The system relies on three components.

First, we train a model to detect symbols in floor plans based on Faster R-CNN architecture (Ren et al. 2017). The model's pyramid network including the classifier and bounding box regression heads are extended to a model ensemble to infer uncertainty estimates associated with the output of the classification for each symbol detected in the floor plan (Lakshminarayanan, Pritzel, and Blundell 2017). This enables the model to quantify which symbols are difficult to classify to seek for human assistance.

Second, symbols that are determined to have a high level of classification uncertainty are deferred to a human expert

*These authors contributed equally.

for revision to minimize the number of incorrectly detected or classified symbols. We present the detection results to domain experts in a so-called *Gallery View* sorted by descending uncertainty scores or grouped by predicted classes. The idea is to allow human experts to seamlessly correct misclassified symbols based on the classes present in the floor plans' legends which are automatically extracted when uploading the floor plans of a new construction project. Grouping the detected symbols by classes accelerates the identification of incorrectly classified symbols. In this context, we conduct a simulation study to analyze the effectiveness of several measures for uncertainty quantification that determine which instances are deferred to human experts. We find that nearly all considered uncertainty measures outperform random acquisition of human expert knowledge. In general, the acquisition of human expert knowledge results in an improved performance of up to 12.9% corresponding to an overall system accuracy of 92.1%.

Third, we develop an approach that avoids extensive data labeling. This is particularly important as the system's adoption to new symbols or different construction projects likely requires different symbols to be detected. We reduce the reliance on manual labeling by training the model with synthetically generated floor plan data. This works by asking the user to highlight the required symbols on the floor plans' legend which are automatically extracted. We further extract the corresponding name of each symbol (i. e. , its label) by applying optical character recognition on the legend. The symbols are then positioned in various angles with different brightness, color, and contrast on background images from a set of reference backgrounds generated by domain experts. Thereby, our approach generates labeled training data without explicitly requiring labeling effort.

To summarize, our contribution is threefold. We propose a human-in-the-loop system that leverages human-AI collaboration in the scope of quantity determination for BOM in the AEC industry. We demonstrate with a technical experiment how *selectively* acquiring domain experts' knowledge can considerably improve the overall system performance while reducing the required manual effort. As existing work analyzing floor plans focuses mainly on residential construction, to the best of our knowledge, we are the first to extend AI for quantity determination to more complex large-scale construction plans. Lastly, we address the low level of standardization in the AEC industry by utilizing synthetically generated training data to allow the system to be conveniently adapted to different construction projects.

Related Work

In the following, we elaborate on relevant work regarding the utilization of AI in the context of floor plans and the application of human-in-the-loop systems in general.

Computer Vision in Floor Plans

Computer vision was recently leveraged in several studies to detect and classify symbols in floor plans. In this context, research developed models to detect several classes of furniture symbols, e.g., doors and tables in living units of residential buildings (e.g., Goyal et al. 2019, Rezvanifar, Cote, and Albu 2020, Ziran and Marinai 2018). Another stream of literature combines the detection of furniture symbols in floor plans with image captioning to generate textual descriptions of detected symbols in corresponding rooms of living units (Goyal, Chattopadhyay, and Bhatnagar 2021). A third field in prior literature developed models to segment architecture and furniture symbols in floor plans (e.g., Dong et al. 2021, Zhu et al. 2020). In this context, Fan et al. (2021) propose a single approach to both detect and segment furniture symbols in residential buildings. Compared to prior literature, we specifically design our model to detect professional and more complex symbols in floor plans based on real-world data from large industrial buildings. Furthermore, we account for real-world applicability by developing a human-in-the-loop system to acquire human expert knowledge in situations in which our model is uncertain about a specific prediction.

Applications of Human-in-the-Loop Systems

In machine learning research, human-in-the-loop systems have emerged as a viable means to acquire human knowledge in situations when the model indicates an increased uncertainty for the prediction of a specific instance (e.g., Amershi et al. 2014, Grønsund and Aanestad 2020). Therefore, human-in-the-loop systems constitute a critical component of deployed machine learning applications in domains where highly accurate model predictions are essential, e.g., in medicine (e.g., Budd, Robinson, and Kainz 2021, Holzinger 2016). In such complex settings, human expert knowledge is usually cost-intensive. Hence, instances that require human expertise should be selected carefully to limit the overall costs (e.g., Hemmer, Kühl, and Schöffer 2022, Jakubik et al. 2022). Due to the inherent complexity of the construction industry and the significance of highly accurate predictions, human-in-the-loop systems were previously leveraged in this domain (e.g., Karim et al. 2021). In this study, we extend the usage of human-in-the-loop systems to the detection of symbols in real-world floor plans of large-scale construction projects.

Approach

In the following, we outline our approach, which consists of a pipeline for synthetic data generation, the object detection model architecture, and the human-in-the-loop system. The latter is built on top of our model to acquire human expert knowledge. Figure 1 displays an overview of the entire process pipeline.

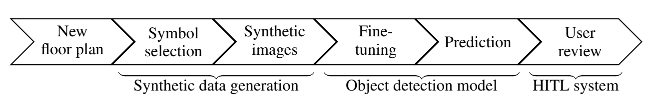


Figure 1: The pipeline of the deployed application. Our approach includes fine-tuning to detect symbols in new construction projects based on synthetic data, the object detection model, and the human-in-the-loop system (HITL).

Synthetic Data Generation

The availability of large amounts of labeled data is crucial for the success of deep learning models in computer vision However, as the generation of manually labeled data is cost-intensive and the use of real-world data can lead to privacy issues, the use of synthetically generated data is prevalent in recent literature (e.g., Hinterstoisser et al. 2018, Nikolenko 2021). For object detection, a composition-based approach is increasingly gaining traction, in which cropped foreground symbols are positioned on different backgrounds (e.g., Dwibedi, Misra, and Hebert 2017). We follow this approach and extend it to the construction domain by extracting symbols and their class label based on optical character recognition methods (Smith 2007) from the legend of the floor plans (see Figure 1). The extracted symbols are then positioned on a set of empty reference background images generated by domain experts. We further employ the following additional data augmentation techniques on the training data aiming at increasing the robustness of the synthetically generated data as proposed in recent literature (Dwibedi, Misra, and Hebert 2017). Symbols are rotated between 0 and 359 degrees with an angle drawn from a discrete uniform distribution $\phi \sim \mathcal{U}(0, 359)$ before being positioned on the background. Moreover, blurring based on three different filters is applied to the symbols. The probability for the choice of a specific filter is given by $\psi \sim \mathcal{U}(0, 2)$ drawn from a discrete uniform distribution before being positioned on the background. Lastly, modifications in terms of brightness, color, contrast, and sharpness are performed on the composed plans with an intensity drawn from a continuous uniform distribution $\rho \sim \mathcal{U}(0.5, 1.5)$. With our synthetic data generation pipeline, we overcome the problem of insufficient available floor plan training data. Our approach allows us to create an arbitrary number of synthetic plans for model training while bypassing costly manual labeling. Moreover, it can be easily transferred to new floor plans with different symbols.

Object Detection Model

Following recent literature in the field of object detection in floor plans (e.g., Goyal et al. 2019, Ziran and Marinai 2018), we employ the Faster-RCNN architecture as the basis for the symbol detection pipeline (Ren et al. 2017). Additionally, we leverage the concept of deep ensembles to infer uncertainty estimates for each identified symbol. This technique has been demonstrated to generate not only high quality but also calibrated predictive uncertainty estimates (Lakshminarayanan, Pritzel, and Blundell 2017) that are used to decide which symbols are deferred to a human expert for subsequent revision in the human-in-the-loop system. Besides the possibility to generate high quality prediction uncertainty estimates, prior research has demonstrated the ability of ensembles to positively contribute to improved model performance (e.g., Kuncheva and Whitaker 2003). Additionally, we benchmark the performance of the described object detection model against a standard Faster-RCNN object detector that does not leverage uncertainty estimation (Softmax) and the approach of inferring uncertainty estimates through Monte-Carlo Dropout (Gal and Ghahramani 2016). In detail, we incorporate an ensemble of the feature pyramid network including both the classifier and bounding box regression heads that consists of five separate models respectively. For each ensemble, the model predictions (i.e., confidence scores of the classifier and bounding box coordinates of the regression) are calculated separately. The predictions are then averaged over all ensembles, while the spread of the predictions implicitly indicates the uncertainty.

Human-in-the-Loop System

Our human-in-the-loop system is tailored to iteratively present instances with uncertain prediction outcome together with their predicted classes to a human expert. Therefore, our approach does not only acquire human expert knowledge but also interacts with the human expert by suggesting a predicted class.

In case, that the predicted class does not match the ground truth class, the human expert corrects the model prediction by discriminating the ground truth from the predicted class. This has shown to be an effective approach in recent literature (e.g., Liu et al. 2013, Wang et al. 2016). Central to the human-in-the-loop system is (1) the meaningful selection of instances and (2) the sequence in which they are queried to a human expert to acquire knowledge. We address both challenges with so-called instance selection mechanisms. Each mechanism includes a specific acquisition function $a(\cdot)$ to measure the model uncertainty. Based on the model uncertainty, the mechanism selects the instance with the maximum uncertainty and passes this instance to the human expert. Specifically, we select an instance $x^\star$ to be labeled next by the human expert by maximizing the acquisition function $x^\star = \arg\max_x a(x)$. In our experiments, we study the effects of a range of acquisition functions on the overall model performance. We introduce the utilized acquisition functions in Table 1, where $\mathbb{P}(y = c|x)$ denotes a likelihood model over the set of classes $c \in \mathcal{C}$. Note that we resort to the ground truth labels for the symbols passed to the human expert during the human-in-the-loop system evaluation. This is in line with real-world quantity determination where a very high labeling quality is required to precisely estimate costs. Furthermore, only detected symbols can be presented to the human, whereas undetected symbols remain incorrectly classified as background in our simulations.

Experimental Setup

In the following, we describe our experimental setup. We introduce the dataset, the evaluation metrics, and provide details on our implementation.

Data

Our experiments are based on a real-world industry dataset that was gathered by an industry partner and labeled by nine domain experts over the span of three weeks. The dataset consists of 44 two-dimensional floor plans with symbols from a total of 39 different classes. The domain experts annotated a total of 5,907 symbols in the floor plans. In general, the classes are construction project-specific depending on the corresponding legend included in floor plans. We

Selection Mechanism	Description
Upper Bound and Baseline	
- Oracle	Select the instance that accelerates overall model accuracy most strongly
- Random	Select a random instance
Uncertainty measures	
- Margin	Select the instance with the smallest margin between first and second prediction: $a(x) = -(\mathbb{P}(y = c_1(x)\|x) - \mathbb{P}(y = c_2(x)\|x))$ with c_i being the class with i-th highest confidence of x
- Confidence	Select the instance with the lowest confidence: $a(x) = -max_c \mathbb{P}(y = c\|x)$
- Entropy	Select the instance with the highest entropy: $a(x) = \mathbb{H}[y\|x] = -\sum_c \mathbb{P}(y = c\|x)\, log(\mathbb{P}(y = c\|x))$
- BALD	Select the instance with Bayesian Active Learning by Disagreement (BALD) (Houlsby et al. 2011): $a(x) = \mathbb{H}[y\|x] - E_{p(\omega)}[\mathbb{H}[y\|x, \omega]]$ with ω denoting a model in the ensemble
- Mean Std.	Select the instance with the highest standard deviation averaged over all classes: $a(x) = \frac{1}{\|C\|} \sum_c \sqrt{Var_\omega[\mathbb{P}(y = c\|x, \omega)]}$ with ω denoting a model in the ensemble
- Var. Ratio	Select the instance with the highest ratio of ensemble predictions not being the mode class (Gal, Islam, and Ghahramani 2017): $a(x) = 1 - max_y \mathbb{P}(y\|x)$

Table 1: Instance selection mechanisms for the human-in-the-loop system evaluated in this study.

present a small excerpt of one of the floor plans in Figure 2. We use these floor plans as the test set. By leveraging our synthetic data generation pipeline, we generate 20,000 synthetic images, which we use for training and validation.

Metrics

We evaluate our system in two stages, with the object detection model in stage 1 and the human-in-the-loop system in stage 2. Following common practice, our object detection model is evaluated based on the mean Average Precision (mAP). We set an intersection-over-union (IoU) threshold of 0.5 for the Average Precision (AP) computation as in the PASCAL VOC Challenge (Everingham et al. 2015). The mAP metric then refers to the averaged AP calculated for each class, that is quantifying the area under the precision-recall curve. For details on the calculations of the mAP, we refer to Everingham and Winn (2011). In the second stage, we evaluate the human-in-the-loop system. Note that our human-in-the-loop system is tailored to the classification of

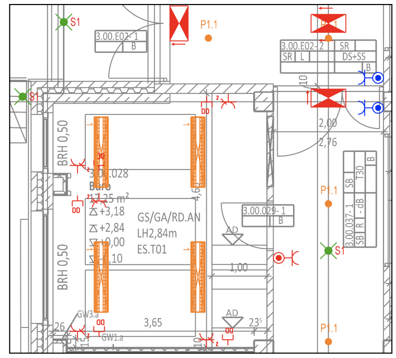


Figure 2: Small excerpt from a real-world floor plan.

detected regions of interest. As the detection of symbols itself is not corrected in our system, the recall will not change significantly, which implies that mAP would not adequately represent the performance of the human-in-the-loop system. Therefore, we measure the performance of the human-in-the-loop system with classification accuracy. This metric refers to the percentage of correct predictions given the number of detected and undetected symbols.

Implementation Details

We use the 20,000 synthetically generated images and assign 14,000 to the training set and 6,000 to the validation set. The images have a resolution of 1,024×1,024 pixels. We evaluate the model performance on the real-world floor plans. As their resolution is larger than 1,024×1,024 pixels, we cut each plan into individual overlapping images, each with a resolution of 1,024×1,024 pixels. As a result, we obtain 3,995 floor plan crops for the evaluation (i. e. , test data). The model is trained for 500 epochs with Stochastic Gradient Descent as optimizer while performing early stopping on the validation set. We train the model with a learning rate of 0.001 and use a ResNet-50 as the model backbone. For further details on the Faster-RCNN architecture, we refer to Ren et al. (2017).

Experimental Results

In this section, we first present the performance of our object detection model on the real-world industry dataset. Second, we show how the acquired human expert knowledge enhances the overall model performance.

Evaluation of Object Detection Model

Overall, the object detection model achieves high performance in both the detection of symbols and the subsequent classification of detected symbols. This is indicated by a mAP score of 82.7% for a given IoU threshold of 0.5. The corresponding interpolated precision-recall curve in Figure 3 represents the trade-off between the precision and recall for different levels of model confidence (i. e. , confidence thresholds). Moreover, we compare our deep ensemble-based model with a standard Faster R-CNN (Softmax) and one that infers uncertainty estimates through

Monte-Carlo Dropout (Gal and Ghahramani 2016) (see Figure 3). In line with prior literature on classification (Lakshminarayanan, Pritzel, and Blundell 2017), deep ensembles yield the best performance. Due to space constraints, we refrain from reporting results on Monte-Carlo Dropout and Softmax uncertainty calculations in the following sections.

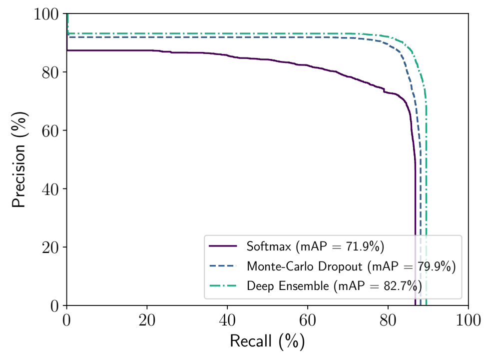


Figure 3: Interpolated precision-recall curve of the object detection models over the set of classes (IoU threshold: 0.5).

Evaluation of Human-in-the-Loop System

In the following, we present the results of our model when acquiring human expert knowledge with the human-in-the-loop system. For this, our model calculates uncertainty scores for all detected and subsequently classified symbols. The model then presents the detected symbols given a descending uncertainty score to the human expert. By correcting the symbols in descending order proposed by our system, the accuracy finally increases by 12.9% to 92.1% as displayed in Table 2. The accuracy difference to 100% can be attributed to the symbols remaining undetected by the model. In this context, budget refers to the deferred share of detected symbols per floor plan. For a labeling budget of 100%, the resulting performance is independent of the instance selection mechanism. Furthermore, the evaluation in Table 2 demonstrates that the increase is primarily driven by the correction of background areas incorrectly detected as symbols which has a larger impact on precision compared to recall. On average, the model detects 14.7 (i. e., 10.7% of the detected objects) background areas as symbols per floor plan compared to 3.4 incorrectly classified symbols (2.5%) and 119.7 correct predictions (86.9%).

Our results further suggest that all evaluated instance selection mechanisms support the overall model performance. The detailed performances of the mechanisms are presented in Figure 4 and Table 3, where the accuracy is evaluated over the percentage of detected symbols per floor plan that are deferred to the human expert. Our simulation indicates that all evaluated selections mechanisms, except for Variation Ratio (Var. Ratio), outperform random selection (e. g., 2.5% to 3.0% higher accuracy at 50% budget for all mechanisms except Variation Ratio). For a limited labeling budget, these mechanisms attain similar performances. With an increasing budget, BALD and Mean Standard Deviation (Mean Std.) achieve a slightly better performance than the remaining mechanisms. However, these differences are subject to uncertainty. Given a higher labeling budget above 75%, Variation Ratio has a similar performance as the other methods.

Lastly, the results indicate different capabilities of mechanisms to identify misclassified background objects or incorrectly classified symbols, as the latter has a particular influence on the recall. Variation Ratio is the only selection mechanism that outperforms random selection on recall at 50% Budget by a large margin. Thus, in contrast to the remaining mechanisms, Variation Ratio enhances the identification of incorrectly classified symbols compared to misclassified background objects.

Metric	0% Budget	100% Budget
Accuracy	81.6	92.1
Precision	88.1	100.0
Recall	90.3	92.1

Table 2: Averaged results (in %) from 44 floor plans without and with full acquisition of human expert knowledge.

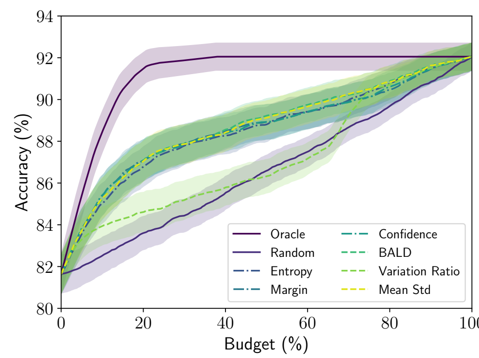


Figure 4: Experimental results for the human-in-the-loop system. Budget refers to the proportion of detected symbols that are deferred to the human expert. The accuracy is averaged over 44 floor plans and visualized with the variance.

Implications for Practitioners

Our work includes several implications for the development and application of real-world object detection models and human-in-the-loop systems. First, our evaluation suggests that, apart from one exception, all consulted acquisitions of human expert knowledge support the model in classifying detected symbols. Second, for varying levels of labeling budget, we observe differences in the performance of the instance selection mechanisms with different effects on precision and recall. Therefore, our work informs practitioners about the meaningful choice of the instance selection mechanism. Third, our work demonstrates that object detection can successfully be leveraged in complex construction settings of large-scale projects. Furthermore, synthetic training data can be a suitable means to reduce intensive labeling

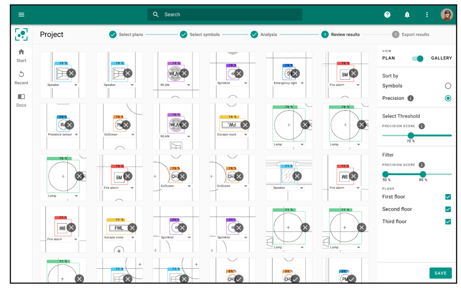
Selection Mechanism	Accuracy at 50% Budget	Precision at 50% Budget	Recall at 50% Budget
UPPER BOUND AND BASELINE			
- Oracle	92.1	100.0	92.1
- Random	86.6	94.1	91.0
UNCERTAINTY MEASURES			
- Margin	88.9	96.4	90.9
- Confidence	88.9	96.4	90.9
- Entropy	88.8	96.3	90.9
- BALD	89.2	96.7	91.1
- Mean Std.	89.1	96.6	91.0
- Var. Ratio	86.3	93.4	91.5

Table 3: Averaged results (in %) from 44 floor plans with 50% budget for the acquisition of human expert knowledge.

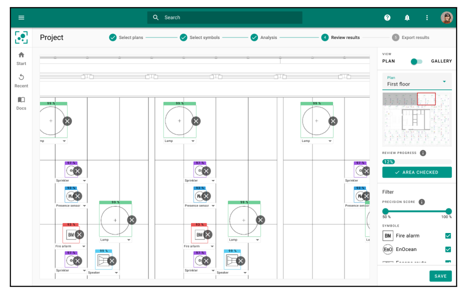
efforts while achieving remarkable performance results on real-world floor plans. Nevertheless, as with any other research, our work is not free of limitations as it builds upon a specific object detection model (i. e. , Faster R-CNN) which requires fine-tuning for the detection of additional symbols. Therefore, practitioners should take into consideration that the model needs to be customized to specific symbols in real-world settings. Thus, we particularly incorporate the generation of synthetic training data in our approach, such that the model adoption can be executed automatically.

Path to Deployment

We developed our approach for object detection in floor plans towards real-world application with a focus on the acquisition of human expert knowledge through interaction mechanisms. For the design of the mechanisms, we identified and incorporated specific industry needs by involving a large-scale construction partner (Drees & Sommer) in the development process right from the beginning of our work. The overall design of the interaction mechanisms aims at providing guidance during the process of correcting misclassified symbols to ensure that domain experts see merit in including our application in their daily business. Specifically, we developed two interaction mechanisms which are depicted in Figure 5. That is, (a) the *Gallery View* which aims at correcting misclassifications (see Figure 5a) and (b) the *Plan View* that allows to correct errors in the detection of our model (see Figure 5b). The Gallery View visualizes classifications with uncertain prediction outcomes. Here, we utilize our acquisition functions to measure the model confidence and sort the detections accordingly. Thus, the goal of the first mechanism is to identify and correct symbols that were incorrectly classified into another class. In contrast, in the Plan View, human experts are guided throughout the plan to draw bounding boxes for symbols that the model did not detect in the first place. Note that this interaction mechanism is necessary to achieve an overall classification accuracy of 100%, as the Gallery View only allows to correct symbols that were previously detected. Undetected symbols are not part of the Gallery View and, therefore, require the human expert to inspect the floor plan for undetected symbols. As a next step, we will deploy our application prototype on-site at the case company to evaluate both mechanisms with end users as part of the human-centered development of our application. More specifically, we aim at analyzing the overall impact of our application in real-world usage regarding time savings and costs. The ongoing evaluation of the interaction mechanisms in daily business is therefore subject to future research on human-AI collaboration.



(a) Gallery View



(b) Plan View

Figure 5: Interaction mechanisms for human-AI collaboration on the task of quantity determination.

Conclusion

In this paper, we propose a human-in-the-loop system for object detection in the context of quantity determination in the AEC industry. The model acquires human expert knowledge for detected symbols which are characterized by high prediction uncertainty resulting in considerable performance improvements in terms of accuracy.

We further account for the real-world applicability of our approach by proposing a pipeline for the generation of synthetic training data to reduce costs and effort for manual labeling. Thus, our work is tailored to supporting practitioners in the AEC industry to generate data-driven business value.

Acknowledgments

We express our gratitude to *digitales bauen part of DREES & SOMMER* for facilitating and supporting this research. The authors acknowledge the financial support by the Federal Ministry for Economic Affairs and Climate Action of Germany in the project Smart Design and Construction (project number 01MK20016F).

References

Amershi, S.; Cakmak, M.; Knox, W. B.; and Kulesza, T. 2014. Power to the People: The Role of Humans in Interactive Machine Learning. *AI Magazine*, 35(4): 105–120.

Blanco, J. L.; Fuchs, S.; Parsons, M.; and Ribeirinho, M. J. 2018. Artificial Intelligence: Construction Technology's next Frontier. *Building Economist*, 1: 7–13.

Budd, S.; Robinson, E. C.; and Kainz, B. 2021. A Survey on Active Learning and Human-in-the-Loop Deep Learning for Medical Image Analysis. *Medical Image Analysis*, 102062.

Dellermann, D.; Ebel, P.; Söllner, M.; and Leimeister, J. M. 2019. Hybrid Intelligence. *Business & Information Systems Engineering*, 61(5): 637–643.

Dong, S.; Wang, W.; Li, W.; and Zou, K. 2021. Vectorization of Floor Plans Based on EdgeGAN. *Information*, 12(5): 206.

Dwibedi, D.; Misra, I.; and Hebert, M. 2017. Cut, Paste and Learn: Surprisingly Easy Synthesis for Instance Detection. In *International Conference on Computer Vision*, 1310–1319.

Everingham, M.; Eslami, S. M. A.; Van Gool, L.; Williams, C. K. I.; and Zisserman, A. 2015. The Pascal Visual Object Classes Challenge: A Retrospective. *International Journal of Computer Vision*, 111(111): 98–136.

Everingham, M.; and Winn, J. 2011. The Pascal Visual Object Classes Challenge 2012 (VOC2012) Development Kit. *Pattern Analysis, Statistical Modelling and Computational Learning, Tech. Rep*, 8: 5.

Fan, Z.; Zhu, L.; Li, H.; Chen, X.; Zhu, S.; and Tan, P. 2021. FloorPlanCAD: A Large-Scale CAD Drawing Dataset for Panoptic Symbol Spotting. *arXiv preprint arXiv:2105.07147*.

Gal, Y.; and Ghahramani, Z. 2016. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. In *International Conference on Machine Learning*, 1050–1059.

Gal, Y.; Islam, R.; and Ghahramani, Z. 2017. Deep Bayesian Active Learning with Image Data. In *International Conference on Machine Learning*, 1183–1192.

Goyal, S.; Chattopadhyay, C.; and Bhatnagar, G. 2021. Knowledge-driven Description Synthesis for Floor Plan Interpretation. *International Journal on Document Analysis and Recognition*, 1–14.

Goyal, S.; Mistry, V.; Chattopadhyay, C.; and Bhatnagar, G. 2019. BRIDGE: Building Plan Repository for Image Description Generation, and Evaluation. In *International Conference on Document Analysis and Recognition*, 1071–1076.

Grønsund, T.; and Aanestad, M. 2020. Augmenting the Algorithm: Emerging Human-in-the-Loop Work Configurations. *The Journal of Strategic Information Systems*, 29(2).

Hemmer, P.; Kühl, N.; and Schöffer, J. 2022. Utilizing Active Machine Learning for Quality Assurance: A Case Study of Virtual Car Renderings in the Automotive Industry. In *Proceedings of the 55th Hawaii International Conference on System Sciences*.

Hinterstoisser, S.; Lepetit, V.; Wohlhart, P.; and Konolige, K. 2018. On Pre-Trained Image Features and Synthetic Images for Deep Learning. In *European Conference on Computer Vision*.

Holzinger, A. 2016. Interactive Machine Learning for Health Informatics: When do we Need the Human-in-the-Loop? *Brain Informatics*, 3(2): 119–131.

Houlsby, N.; Huszar, F.; Ghahramani, Z.; and Lengyel, M. 2011. Bayesian Active Learning for Classification and Preference Learning. *arXiv preprint arXiv:1112.5745*.

Jakubik, J.; Blumenstiel, B.; Vössing, M.; and Hemmer, P. 2022. Instance Selection Mechanisms for Human-in-the-Loop Systems in Few-Shot Learning. In *International Conference on Wirtschaftsinformatik*.

Karim, M. M.; Qin, R.; Chen, G.; and Yin, Z. 2021. A Semi-Supervised Self-Training Method to Develop Assistive Intelligence for Segmenting Multiclass Bridge Elements from Inspection Videos. *Structural Health Monitoring*.

Kuncheva, L. I.; and Whitaker, C. J. 2003. Measures of Diversity in Classifier Ensembles and their Relationship with the Ensemble Accuracy. *Machine Learning*, 51(2): 181–207.

Lakshminarayanan, B.; Pritzel, A.; and Blundell, C. 2017. Simple and Scalable Predictive Uncertainty Estimation Using Deep Ensembles. In *Neural Information Processing Systems*, 6405–6416.

Liu, C.; Loy, C. C.; Gong, S.; and Wang, G. 2013. POP: Person Re-identification Post-rank Optimisation. In *International Conference on Computer Vision*, 441–448.

Nikolenko, S. 2021. *Synthetic Data for Basic Computer Vision Problems*, 161–194. Springer.

Ren, S.; He, K.; Girshick, R.; and Sun, J. 2017. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *Transactions on Pattern Analysis and Machine Intelligence*, 39(6): 1137–1149.

Rezvanifar, A.; Cote, M.; and Albu, A. B. 2020. Symbol Spotting on Digital Architectural Floor Plans Using a Deep Learning-based Framework. In *Computer Vision and Pattern Recognition Workshops*, 568–569.

Smith, R. 2007. An Overview of the Tesseract OCR Engine. In *International Conference on Document Analysis and Recognition*, volume 2, 629–633.

Wang, H.; Gong, S.; Zhu, X.; and Xiang, T. 2016. Human-in-the-Loop Person Re-identification. In *European Conference on Computer Vision*, 405–422.

Zhu, R.; Shen, J.; Deng, X.; Walldén, M.; and Ino, F. 2020. Training Strategies for CNN-based Models to Parse Complex Floor Plans. In *International Conference on Software and Computer Applications*, 11–16.

Ziran, Z.; and Marinai, S. 2018. Object Detection in Floor Plan Images. In *IAPR Workshop on Artificial Neural Networks in Pattern Recognition*, 383–394.