Exploring Explainable AI for Symbol Detection in Construction Drawings

Jan Michalak[1] and Benedikt Faltin[1]
[1]Chair of Computing in Engineering, Ruhr-Universität Bochum, Universitätsstraße 150, 44801 Bochum, Germany
E-mail(s): jan.michalak@rub.de, benedikt.faltin@rub.de

Abstract: Having information in a digitally accessible format can significantly improve the efficiency of processes by enabling automation. In the construction industry, however, much building-related information is only available in analog form, mainly paper-based drawings, which hinders the introduction of digital processes. Artificial intelligence offers a practical approach for converting these analog construction drawings into digital information. Neural networks, such as the YOLO object detection model, can efficiently identify and extract crucial information from the drawings on a large scale. However, these models operate as "black boxes," making it challenging to understand the basis of their decisions. To address this problem, this study applies techniques from explainable artificial intelligence to gain insight into the inner workings of a neural network. In particular, YOLOv8 is trained to detect symbols in pixel-based drawings, and the Eigen-CAM visualization method is utilized to shed light on the network's internal decision-making processes. With the acquired knowledge about the inner decisions of the trained network during symbol detection, the training data set is refined to enhance the model's overall performance, leading to an improvement in the detection of mAP_{50} 10 points.

Keywords: XAI, Machine Learning, Symbol Detection, YOLO, Eigen-CAM

Erschienen in Tagungsband 35. Forum Bauinformatik 2024, Hamburg, Deutschland, DOI: 10.15480/882.13499

1 Introduction

Digital transformation is a crucial objective for today's society. Transforming analog processes into digital ones has the potential to make processes more efficient by unlocking automation. This task is particularly crucial for the construction industry. Since many buildings were erected before computers existed, thousands of documents containing critical aspects about the structures are still archived in analog form. However, extracting this information and transforming it into a digital building model can enhance the efficiency of maintenance management [1]. To make the information contained digitally accessible is a massive task.

A solution to this challenge may lie in artificial intelligence (AI). In the last decade, the capabilities of AI have advanced rapidly to a human degree and, in some cases, even surpassed it [2]. AI is already

being successfully deployed in various fields, such as medicine, autonomous driving, engineering, and transportation. As a result, using AI to convert information from analog documents in the construction industry is an area of research attracting much attention. However, many state-of-the-art AI models operate as "black boxes," and how decisions are made remains unknown. This lack of transparency leads to distrust among users and, therefore, to less acceptance from society to use these new tools - especially in sensitive work fields with a big influence on human life (e.g., medicine, finance, and construction) [3]. Explainable AI (XAI) can be used to gain insight into the model, making the "black box" more transparent and trustworthy while enabling the model to be improved by refinement of the training data set. Flaws in the training data could be directly observed and fixed, as demonstrated in work from the Fraunhofer Heinrich-Hertz-Institut [4]. This study employs XAI to gain insights into an object detection model trained to identify symbols in engineering drawings. By examining the model's internal mechanisms, the training data set can be refined to address weaknesses in the trained model. After retraining with the improved dataset, changes in detection accuracy are compared and evaluated.

2 Related Literature

The amount of research on symbol detection in technical drawings using object detection models is limited [5]. In the work by Zhao et al. [6] the object detection method YOLOv1 was used to efficiently and accurately detect structural components in scanned structural drawings. The same authors later published a further improved method in which they used R-CNN to get a more precise extraction of information for reconstructing digital building models [7].

Lu et al. [8] propose a different approach based on images and CAD drawings. This semi-automatic method reconstructs a digital model from images and CAD drawings. First, a circle Hough transform detects symbols in floorplan drawings, which is then combined with textual information extracted using OCR. In addition, a neuro-fuzzy network detects the building component class in images and identifies the material of the component surface. By integrating all this information, the building geometry is reconstructed and translated into an IFC-based model.

While some approaches utilize symbol detection methods for technical drawings, to the best of the author's knowledge, no research on applying explainable AI methods in this area exists. Consequently, the current state of XAI methods has been examined.

SHapley Additive exPlanation (SHAP) [9] is a method rooted in game-theory where each feature relevant to a decision is modeled as a player in a coalition game. This provides a measure of importance known as the Shapley value. Interpretations derived from SHAP can be both local and global and are consistent with each other. However, feature dependencies pose a challenge in SHAP since it assumes feature independence.

Class activation maps (CAMs) can only be applicable with Convolutional Neural Networks (CNNs) [10]. In CAMs the per-class weighted linear sum of visual patterns is used to localize features in a picture. However, CAMs have a drawback in that they cannot be applied to pretrained networks, and to networks, which do not consist of a CNN-based architecture.

Therefore, Selvaraju et al [11] introduce a gradient-weighted schema to the CAM approach (Grad-CAM) to extend its applicability to pretrained neural network architecture. However, it produces noisy visualizations and does not perform well with multiple instances of the same object in an image.

To overcome this issue, Chattopadhay et al. [12] smooth the visualizations by pixel-wise weighting the gradients (Grad-CAM++). Additionally, Grad-CAM++ can also work with multiple instances of the same feature in an image. The drawback of Grad-CAM and Grad-CAM++, is that both methods implement backpropagation and therefore assume that classifiers decide correctly.
In the work of Muhammad and Yeasin [13] this issue is addressed with their method Eigen-CAM. Eigen-CAM computes the principle components of the learned features of the network from its respective convolutional layer as shown in Section 3.2.

3 Methodology

In this study, the object detection model YOLOv8 [14] is combined with Eigen-CAM to investigate the network's inner decision-making when applied to the task of symbol detection. YOLOv8 was chosen because of its high accuracy among the available models and because it is well documented. The overall procedure is illustrated in Figure 1.

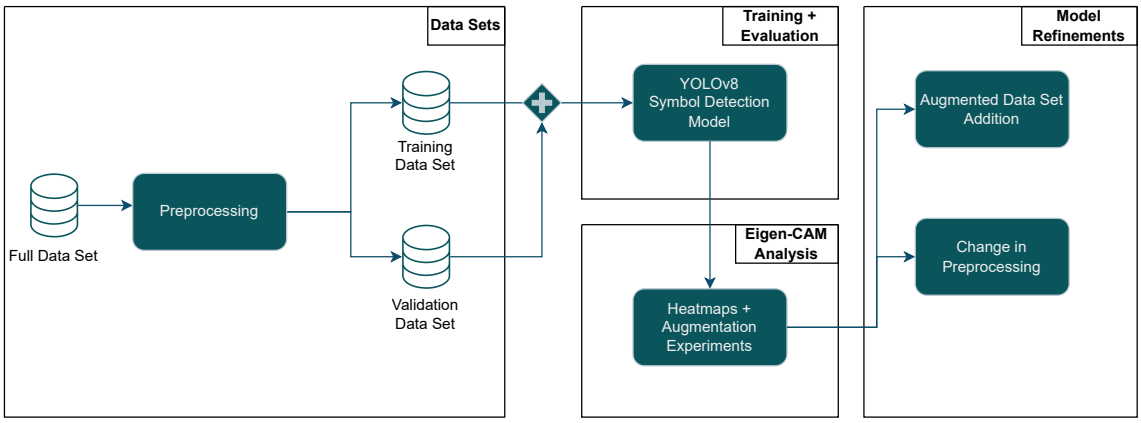

Figure 1: Overview of the structure of the study.

First, a data set consisting of 53 bridge construction drawings is collected and manually annotated as described in Section 3.1. This data is used for training YOLOv8. By utilizing Eigen-CAM, as detailed in Section 3.2, the performance of the trained YOLOv8 model is examined in Section 4. Based on these findings, the training data is extended to address the identified weaknesses of the model. The model is then retrained, and its performance is compared to the original results.

3.1 Data set

A data set of 53 drawings is collected to train and validate the object detection model. These drawings were provided by various industry partners. The data set identified four distinct significant symbol classes: axis symbols, section symbols, gradient symbols, and elevation symbols, as shown in Figure 2. These symbols were manually labeled, resulting in a total of 4,787 labeled instances. Given the relatively small size of these symbols compared to the overall drawings, the images were divided into smaller patches as proposed by Faltin et al. [15]. This method aligns with the findings presented in [16], which demonstrated that tiling large images enhances the detection rate of small objects. Finally, the annotations are exported using the Datumaro library[1] in the YOLO format, making them

[1]https://github.com/openvinotoolkit/datumaro (Accessed: 06 July 2024)

compatible with the YOLOv8 network. The data set is further divided into training and validation data sets. The training data set comprised 14,000 images containing 4,432 instances, while the evaluation set included 1,080 images with 345 instances.

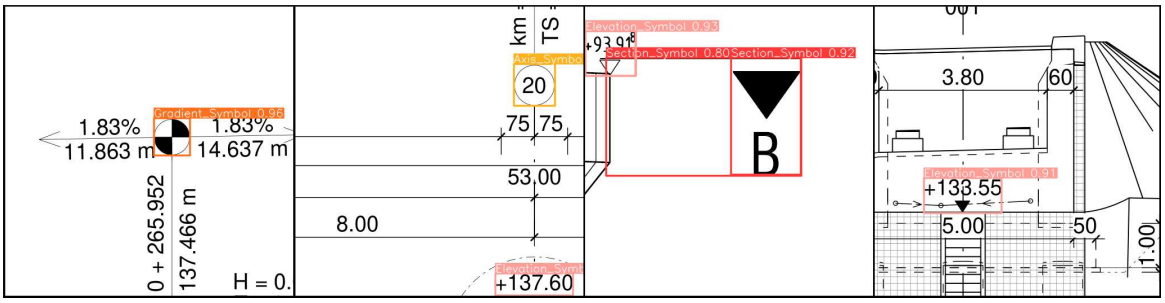

Figure 2: Exemplary representation of typical symbols contained in construction drawings. From left to right: gradient symbol, axis symbol, section symbol, and elevation symbol.

3.2 Eigen-CAM

The XAI method Eigen-CAM is utilized to gain insight into the patterns the YOLOv8 network is learning for symbol detection. The central idea of Eigen-CAM is to identify the principal components of the individual layers in the CNN, as illustrated in Figure 3, and visualize them based on an input image. For the kth-layer the output O_k for an input image I is determined by the combined weight matrix W_k, as described by the following equation

$$O_k = W_k^T I \tag{1}$$

By singular value decomposition (SVD), the principal components of O_k can be calculated with

$$O_{L=k} = U \Sigma V^T \tag{2}$$

Here, matrix U is an orthogonal matrix, with its columns representing the left singular vectors. The diagonal matrix Σ contains the singular values along its main diagonal. Matrix V is also orthogonal, with its columns forming the right singular vectors.
The class activation map, L is given by the projection of O_k onto the first eigenvector, as expressed by

$$L = O_k V_1 \tag{3}$$

Here, V_1 represents the first eigenvector in the matrix V. The visualization is achieved through a normalized heatmap highlighting areas in the image representing saliency features learned by the object detection model.

Applying this technique to the symbol detection network, it is possible to determine which pixels/features have the greatest impact on the predictions. The training data can be improved by analyzing errors in the predictions to mitigate these effects.

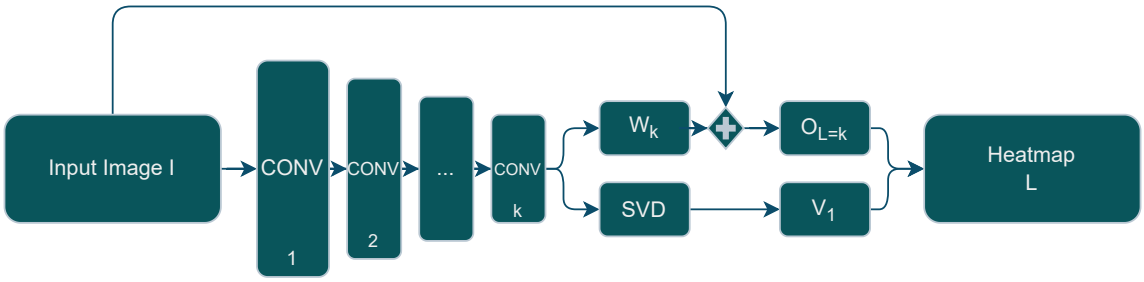

Figure 3: The input image I is processed by multiple convolutional layers $CONV_i$. Eigen-CAM visualizes the pixels of highest importance for the prediction by calculating the eigenvalues of the target layer k using SVD and using the projection of O_k onto V_1.

Table 1: Performance comparison after Eigen-CAM Analysis.

	mAP_{50}	Precision	Recall
Pre Eigen-CAM Analysis	82.93	91.40	86.42
Synthetic Data Set Addition	85.21	85.77	85.21
Change in Preprocessing (Overlap)	92.28	92.40	90.58

4 Results and Discussion

The state-of-the-art object detection model YOLOv8 is trained using a NVIDIA A100 SXM4 40 GB graphics card. A batch size of 128 was utilized for training, with a maximum of 100 epochs. The backpropagation process employed the AdamW optimizer with a learning rate of 0.001 and a momentum of 0.9 for the first 10,000 iterations. For more iterations the SGD optimizer with a learning rate of 0.01 and a momentum of 0.9 is used. The trained network achieves a high mAP_{50} as shown in Table 1 for symbol detection in the construction drawings.

To gain further insight into the detection model, the Eigen-CAM method is used to visualize the areas the model considers for predicting the different symbols, as shown in Figure 4.

The class activation map for the axis symbols, gradient symbols and elevation symbols shows that the model's attentive area is larger than the symbol itself. The results indicate that the model predicts symbols by focusing on their surrounding context. For instance, the text around the gradient symbol typically describes the gradient parameters. The model, therefore, searches for parameter texts in the region of the potential gradient symbol and makes a positive prediction if this context information is present.
The interpretation for the section symbol is not as straight forward. Counterintuitively, the detection

Table 2: Performance comparison after Eigen-CAM Analysis of elevation symbol detections. True positive and false positive values as instances and normalized values.

	TP (inst.)	TP (norm.)	FP (inst.)	FP (norm.)
Pre Eigen-CAM Analysis	273	0.87	42	0.93
Synthetic Data Set Addition	282	0.90	56	0.98
Change in Preprocessing (Overlap)	326	0.93	7	0.87

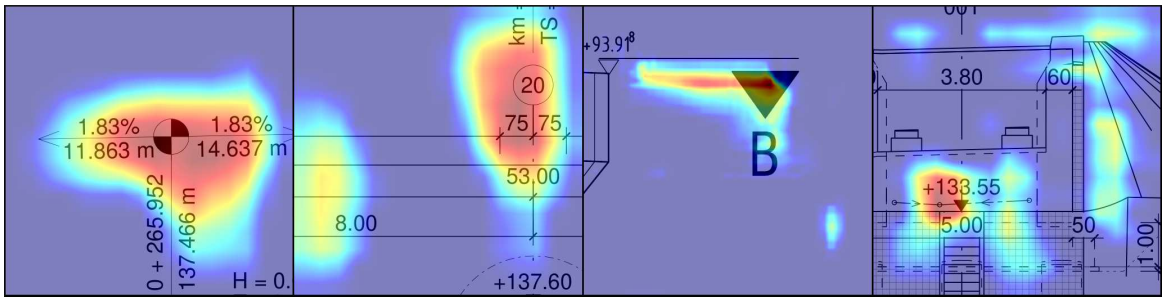

Figure 4: Exemplary illustration of the Eigen-CAM heatmaps for the layer 21 of the detection seen in 2. The red-colored parts mark highly significant parts, while the blue-colored parts are insignificant for the prediction.

head for small objects shows a fitting heatmap for the section symbol, even though its size is relatively large compared to other symbols. This can be explained by the distinctive geometrical features of the section marker, which are less dependent on the context of the image.

To enhance the detection model, the focus was shifted to the elevation symbol due to its high occurrence in the dataset. It had the lowest true positive detection rate and the highest false positive rate, as shown in Table 2. Often dimension symbols are incorrectly classified as an elevation symbol. This may be due to the tiling process in which a part of the elevation symbol is cut off. This leaves a geometric structure that could be mistaken for the dimension symbol. Due to the tiling process, the accuracy of predictions can decrease for symbols with nondistinctive geometric forms.

By analyzing the Eigen-CAM heatmaps, it is concluded that the elevation marker's false positive detections are mainly due to a larger field of attention, which also contains a significant amount of white space. Two different approaches are implemented to refine the model's field of attention: First, synthetic training data was created by copying tiles containing elevation symbols, then partially hiding parts of the images with white space and rearranging the elevation symbols into new positions. Second, the preprocessing of the drawings is modified, by including a overlap of 50 pt in the tiling of the construction plans. While the first approach shows no significant improvements, as illustrated in Table 1, the second approach significantly improves detection accuracy. The inclusion of a overlap lead to a increase of the accuracy by 10 points. Furthermore, this change is evident in the heatmaps, as shown in Figure 5. The model's field of attention is smaller for each symbol compared to the attention fields before the preprocessing change.

5 Conclusion and Outlook

The proposed study utilized the XAI method Eigen-CAM to analyze how the YOLOv8 network makes decisions when detecting symbols in bridge construction drawings. While YOLOv8 performed well in detecting symbols, Eigen-CAM provided insights into its decision-making process, particularly highlighting issues with uncertain or incorrect predictions. This information guided adjustments to the data set, improving the network's performance. Retraining on the expanded data set resulted in an approximately 11% improvement in mAP_{50}, showcasing the potential of XAI methods to enhance symbol detection accuracy.

Despite these promising findings, further research is needed. Normalization should be applied to

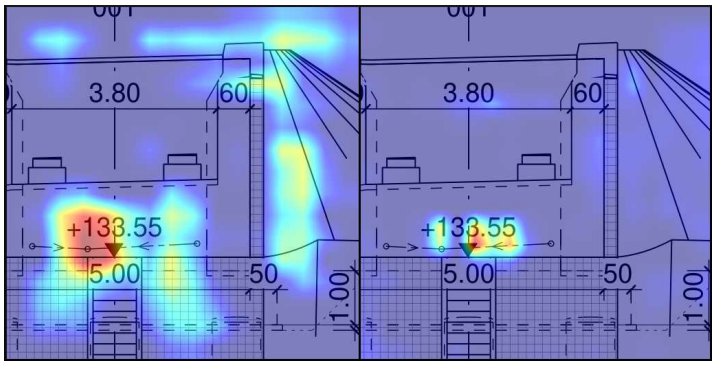

Figure 5: Comparison of Eigen-CAM heatmaps: Model without overlap (left) and with overlap (right) in tiling of the pictures in data preprocessing.

Eigen-CAM visualizations to ensure better heatmap comparability across different layers and images. Concatenating normalized Eigen-CAM activation maps globally may yield additional valuable insights. Moreover, exploring optimal tile overlap in relation to XAI methods such as Eigen-CAM could further enhance detection metrics, as demonstrated in this study's use of a tiled approach with overlap.

References

[1] R. Sacks, A. Kedar, A. Borrmann, *et al.*, "Seebridge as next generation bridge inspection: Overview, information delivery manual and model view definition", *Automation in Construction*, vol. 90, pp. 134–145, 2018. DOI: https://doi.org/10.1016/j.autcon.2018.02.033. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0926580517303977.

[2] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning", *Nature*, vol. 521, no. 7553, pp. 436–444, May 28, 2015. DOI: 10.1038/nature14539. [Online]. Available: https://www.nature.com/articles/nature14539 (visited on 11/18/2023).

[3] R. Confalonieri, L. Coba, B. Wagner, and T. R. Besold, "A historical perspective of explainable artificial intelligence", *WIREs Data Mining and Knowledge Discovery*, vol. 11, no. 1, e1391, Jan. 2021. DOI: 10.1002/widm.1391. [Online]. Available: https://wires.onlinelibrary.wiley.com/doi/10.1002/widm.1391 (visited on 11/17/2023).

[4] A. Rommel and W. Samek, *Cebit 2017: Analyse-software für neuronale netze – dem computer beim denken zuschauen*, 2017. [Online]. Available: https://www.hhi.fraunhofer.de/fileadmin/PDF/CC/PM/2017/PI_FraunhoferHHI_Dem_Computer_beim_Denken_zuschauen_CeBIT_d.pdf.

[5] C. F. Moreno-García, E. Elyan, and C. Jayne, "New trends on digitisation of complex engineering drawings", *Neural Computing and Applications*, vol. 31, no. 6, pp. 1695–1712, Jun. 2018. DOI: 10.1007/s00521-018-3583-1. [Online]. Available: http://dx.doi.org/10.1007/s00521-018-3583-1.

[6] Y. Zhao, X. Deng, and H. Lai, "A deep learning-based method to detect components from scanned structural drawings for reconstructing 3d models", *Applied Sciences*, vol. 10, no. 6, 2020. DOI: 10.3390/app10062066. [Online]. Available: https://www.mdpi.com/2076-3417/10/6/2066.

[7] Y. Zhao, X. Deng, and H. Lai, “Reconstructing bim from 2d structural drawings for existing buildings”, *Automation in Construction*, vol. 128, p. 103 750, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:236240776.

[8] Q. Lu, L. Chen, S. Li, and M. Pitt, “Semi-automatic geometric digital twinning for existing buildings based on images and cad drawings”, *Automation in Construction*, vol. 115, p. 103 183, 2020. DOI: https://doi.org/10.1016/j.autcon.2020.103183. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0926580519315079.

[9] S. M. Lundberg and S. Lee, “A unified approach to interpreting model predictions”, *CoRR*, vol. abs/1705.07874, 2017. arXiv: 1705.07874. [Online]. Available: http://arxiv.org/abs/1705.07874.

[10] B. Zhou, A. Khosla, À. Lapedriza, A. Oliva, and A. Torralba, “Learning deep features for discriminative localization”, *CoRR*, vol. abs/1512.04150, 2015. arXiv: 1512.04150. [Online]. Available: http://arxiv.org/abs/1512.04150.

[11] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-cam: Visual explanations from deep networks via gradient-based localization”, in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 618–626. DOI: 10.1109/ICCV.2017.74.

[12] A. Chattopadhay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, “Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks”, in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Lake Tahoe, NV: IEEE, Mar. 2018, pp. 839–847. DOI: 10.1109/WACV.2018.00097. [Online]. Available: https://ieeexplore.ieee.org/document/8354201/ (visited on 11/17/2023).

[13] M. B. Muhammad and M. Yeasin, “Eigen-CAM: Class activation map using principal components”, in *2020 International Joint Conference on Neural Networks (IJCNN)*, Glasgow, United Kingdom: IEEE, Jul. 2020, pp. 1–7. DOI: 10.1109/IJCNN48605.2020.9206626. [Online]. Available: https://ieeexplore.ieee.org/document/9206626/ (visited on 11/29/2023).

[14] G. Jocher, A. Chaurasia, and J. Qiu, *Ultralytics yolov8*, version 8.0.0, 2023. [Online]. Available: https://github.com/ultralytics/ultralytics.

[15] B. Faltin, P. Schönfelder, and M. König, “Inferring interconnections of construction drawings for bridges using deep learning-based methods”, in *ECPPM 2022 - eWork and eBusiness in Architecture, Engineering and Construction 2022*, 1st ed., London: CRC Press, Mar. 8, 2023, pp. 343–350. DOI: 10.1201/9781003354222-44. [Online]. Available: https://www.taylorfrancis.com/books/9781003354222/chapters/10.1201/9781003354222-44 (visited on 04/06/2024).

[16] F. O. Unel, B. O. Özkalayci, and C. Çigla, “The power of tiling for small object detection”, *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 582–591, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:198903617.