Automated Inspection of Obstructed Fire Extinguishers Using Amodal Instance Segmentation

Jan H. Heinbach[1] and Angelina Aziz[1]
[1]Chair of Computing in Engineering, Ruhr Universität Bochum, Universitätsstrasse 150, 44801 Bochum, Germany
E-mail(s): Jan.Heinbach@ruhr-uni-bochum.de, Angelina.Aziz@ruhr-uni-bochum.de

Abstract: Fire safety inspections are essential to ensure the safety of occupants during a fire outbreak. These inspections involve checking various items, such as fire safety equipment (FSE), to ensure they are unobstructed and fully functional. Recent research has highlighted the potential of machine learning and computer vision in automating and enhancing fire safety inspection processes using image analysis. However, identifying obstacles or potential obstructions in front of fire extinguishers effectively remains a challenge. The research focuses on reviewing documented images to identify instances where extinguishers are obstructed, either partially or fully. This study proposes a novel approach to address this challenge by combining modal and amodal instance segmentation models to evaluate the level of obstruction. One conducts classical (modal) instance segmentation of visible fire extinguisher parts, while the other performs amodal segmentation. Additionally, the annotation and dataset creation for amodal segmentation tasks is addressed. The study generates obstacles on images from both open-source and self-created datasets containing (unblocked) fire extinguishers. This approach requires only one modal annotation iteration to generate modal and amodal annotation data. Results demonstrate the effectiveness of the proposed approach in detecting covered or partially blocked extinguishers. Future research aims to refine amodal mask results and extend the approach to other FSE components, further enhancing fire safety inspection processes.

Keywords: Computer Vision, Instance Segmentation, Amodal Segmentation, Fire Safety Equipment

Erschienen in Tagungsband 35. Forum Bauinformatik 2024, Hamburg, Deutschland, DOI: 10.15480/882.13512

1 Introduction

Preventive fire protection is an essential component of construction and building management. Its main goal is to prevent fires and, in the event of a fire, to minimize the spread of fire and smoke to protect people and property. This begins in the planning phase, where the selection of appropriate materials and designs reduces the risk of fire. In building management and maintenance, preventive fire protection includes regular maintenance and inspections of fire safety equipment (FSE) — such as fire extinguishers — as well as training for residents and staff. Due to its importance, preventive fire protection and especially the fire safety inspections, which include inspection and maintenance of FSE, are mandatory in every building [1].

Fire safety inspections are essential to ensure the functionality and the visibility of FSE in case of a fire and therefore to ensure the security of occupants in the building. However, especially the inspection is a time consuming task. During a typical on-site fire safety inspection, a fire safety manager walks through the building while processing a checklist (e.g. [2]). Most of the inspection tasks are done by physical or visual review manually conducted by the fire safety manager. The inspections results are written down by hand or manually maintained with software support. In addition to the time consumption, this approach is prone to errors.

A crucial aspect of the inspection is to check whether FSE is blocked. As the fire safety manager often documents his inspection with images of the inspected FSE, the use of Artificial Intelligence (AI) to perform inspections based on these images is the most obvious approach. Computer vision methods can be used to inspect FSE on these images and detect blocked ones. Previous work from [3], can inspect simple FSE using visual object detection to detect obstacles in front of FSE. However, this approach has a limited amount of obstacles that could be detected, as each obstacle require training data to be detected as an individual class. As it is unrealistic to consider every possible obstacle as a single class, further work is required to enable the inspection of more complex objects.

In this paper, we want to improve the inspection and detection rate of fire extinguishers and be able to estimate blocked FSE behind any obstacle using trained (amodal) segmentation models, even if only a small part of the asset is visible.

2 Related Background

2.1 Digitalization of FSE inspections

Due to the huge potential of automation, there is a lot of research going on to support during FSE inspections [4]–[8], as well as to increase the level of automation in the context of preventive fire protection in general [9], [10]. Due to the fact that the use of visual object recognition enables immediate deployment in existing and new buildings, computer vision methods are a widely used approach. Oh et al. [8] create an approach to use object recognition on augmented reality glasses to support the fire prevention officer during the fire safety inspection or even allow general public to contribute to the fire safety inspection. Similar to that approach, Bayer and Aziz [4] as well as Corneli et al. [6] enable and evaluate real-time detection of different FSE inside buildings, while Kumar et al. [7] detect FSE at construction sites. More enhanced approaches, as from Alayed et al. [5] or previous work from us [3], use the detection of FSE to perform actual inspections. Alayed et al. [5] are inspecting fire extinguishers using object detection to classify if a certain inspection criteria is met. In [3], object detection is used to detect FSE and obstacles to identify obstacles that block FSE.

2.2 Amodal Object Detection

Recent work in computer vision aims to develop a new computer vision technique besides classification, detection with bounding boxes and classical instance segmentation: Amodal segmentation [11], [12].

Classification, as the simplest task, aims to classify a whole image to predefined classes. Object detection with bounding boxes, aims to identify and locate objects in images using bounding boxes. Among others, R-CNN [13], the You Only Look Once (YOLO) [14] architecture series with YOLOv8 [15] as a current state of the art version, are capable of both tasks. Classical instance segmentation, hereafter referred to as modal segmentation, aims to not only locate the object but identifies each pixel belonging to the object. Exemplarily, Mask R-CNN [16] and YOLOv8 [15] are modal segmentation approaches. The difference from modal to amodal segmentation is visualized in Figure 1. While modal segmentation aims to identify the visible part of an instance (visible on the right), an amodal segmentation aims to identify both the visible and the occluded parts of an instance (visible on the left). [11], [17]

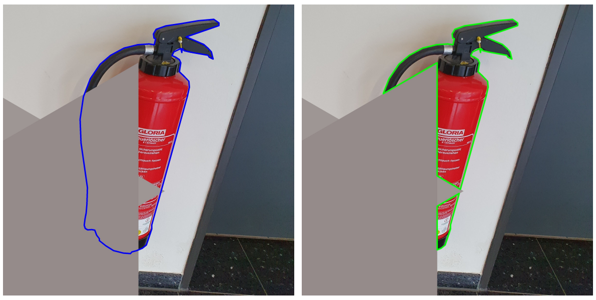


Figure 1: Amodal segmentation annotation (left, blue) and (well known) modal segmentation annotation (right, green). In the modal segmentation the mask is split up in two parts.

While this might be an solvable task to a human [18], the computational amodal segmentation is still subject of current research [12], [19], [20]. A lot of different approaches are studied in this context. The method from Li and Malik [11] uses R-CNN [13] modal segmentation results as a basis to perform a bounding box expansion. *SharpMask* [21] is an Convolutional Network [22], specially for amodal segmentation. In contrast to that, some approaches just use state of the art modal segmentation solutions with new training data [12].

3 Methodology

Since amodal segmentation is still subject of current research and — to the best of our knowledge — there is no state-of-the-art amodal segmentation solution available, we use a straight forward approach and create amodal training data to train YOLOv8 [15] modal segmentation models as amodal ones.

3.1 Amodal Annotation

The creation of amodal segmentation data is challenging, as the generation of annotations is complicated. In contrast to modal segmentation, where the modal segmentation mask can be definitely determined by humans based on images, parts of the amodal mask are invisible (see Figure 1). Therefore, these human annotations are inaccurate and might not be the ground truth [11].

For this reason, we collect images with completely visible fire extinguishers, create modal annotations and generate artificial obstacles. Subsequently, the previously modal annotation can be used as amodal annotation, as the fire extinguisher was completely visible before. The new modal annotation can be calculated by calculating the difference between the fire extinguisher and the artificially generated obstacles. No manual amodal annotation is required and the amodal annotation definitely represents the ground truth.

Figure 2: Generated artificial obstacles. *Type 1* (monochrome simple shapes) on the left and *Type 2* (rectangles, filled with patterns) on the right.

We generated two different types of obstacles, exemplarily seen in Figure 2. The first type are different simple geometric shapes in the average pixel colour value of the image (*Type 1*). The second obstacle type are rectangles, filled with different patterns (*Type 2*).

3.2 Data

To start the amodal annotation mentioned before, data from various sources is collected. More precisely, out of 888 different images, 388 are real world images, mostly self-made in a mixture of buildings in Germany (primarily university related buildings and fire stations) and from the FireNet dataset from University College in London [23]. The remaining 500 images are synthetically created.

Based on these original images (without obstacles), three different amodal datasets are generated. 876 images with *Type 1* obstacles and 862 images with *Type 2* obstacles are generated. The resulting datasets are listed in Table 1. According to Subsection 3.1, each dataset contains two labels, one with modal annotation and one with amodal annotations.

Table 1: Overview about created amodal datasets.

Dataset name	Contained obstacles	Total Image amount
Shape Obstacles	*Type1*, original	1764
Pattern Obstacles	*Type2*, original	1750
Shape and Pattern Obstacles	*Type1*, *Type2*, original	2626

3.3 YOLOv8 Training

Each of the created datasets from Table 1 is used to train two different YOLOv8 [15] segmentation models. For this purpose, the datasets are separated in 70% training (*train*), 20% validation (*val*) and

10% test (*test*) data. In a first iteration, the modal (classic) annotations are used to train YOLOv8 [15] models for modal instance segmentation. The second training iteration uses the amodal annotations in the same YOLOv8 training setup, aiming to create an YOLOv8 amodal segmentation model. All trainings are performed on 6 NVIDIA A100-SXM4-40GB GPUs with the same setup. The used YOLOv8 parameters [1] are IMGSZ set to 640, BATCH set to 72 and PATIENCE set to 50. All other parameters are default values. The PATIENCE parameter allows to terminate training if no progress is made in 50 epochs. Table 2 list the effective training epochs of the trained YOLOv8 models and the short names of the models.

Table 2: Effective training epochs of the different YOLOv8 models.

Model	**Training Epochs**
Shape Amodal (SA)	352
Shape Modal (SM)	414
Pattern Amodal (PA)	416
Pattern Modal (PM)	282
Shape Pattern Amodal (SPA)	345
Shape Pattern Modal (SPM)	492

4 Evaluation

Subsequently, the trained models are evaluated. As evaluation metric the mean Intersection over Union (mIoU), which is the average IoU over all images, is used. If the mask of a fire extinguisher was divided by an obstacle, the two resulting modal segmentations are considered together as one when calculating the IoU. In addition to the evaluation, the *occlusion rate* is calculated. The *occlusion rate* indicates how much of an extinguisher (the amodal segmentation) is hidden (missing in the modal segmentation). We suggest a threshold of 0.05 for considering the fire extinguisher as blocked, but further work is required here. The detailed evaluation results and exemplarily detection results can be found in Table 3. The modal models (SM, PM, SPM) show good results. Reviewing the amodal models (SA, PA, SPA), the PA and SPA model perform better than the SA model (only *Type 1* obstacles), clearly visible when reviewing Example 2. In total, it definitely points out, that the amodal segmentation of fire extinguishers performs surprisingly good.

5 Conclusion

The paper presents a promising approach to improve the automation of FSE inspections, using amodal segmentation. The straight forward approach to train YOLOv8 models — which originally perform modal segmentations — with amodal training data works amazingly well. The amodal annotation strategy ensures that annotations present the ground truth. Future work aims to integrate the models in an inspection tool (as proposed in [3]) to further support the fire safety manager by detecting blocked FSE regardless of the blocking object. It could be extended to multiple different FSE or other inspection tasks, possibly in combination with mobile robots as in [24].

[1] All parameters can be found here: https://docs.ultralytics.com/modes/train

Furthermore, the amodal segmentation approach, as well as the amodal annotation strategy, can be used to detect and predict hidden parts of objects in multiple different areas, not only in building an construction industry, as amodal segmentation is a current research topic.

References

[1] Bundesministerium für Arbeit und Soziales (BMAS), *Technische Regeln für Arbeitsstätten: Maßnahmen gegen Brände (ASR A2.2)*, German, GMBl, 2018.

[2] Bundesministerium des Innern, für Bau und Heimat (BMI), *Brandschutzleitfaden für Gebäude des Bundes*, Arbeitskreis Brandschutzleitfaden, editor, Jun. 2019.

[3] J.-H. Heinbach and A. Aziz, "Visual partial inspection of fire safety equipment using machine learning", en, in *34th Forum Bauinformatik / 34. Forum Bauinformatik*. Bochum: Ruhr-Universität Bochum, 2023, pp. 324–331. DOI: 10.13154/294-10096. [Online]. Available: https://hss-opus.ub.ruhr-uni-bochum.de/opus4/10096.

[4] H. Bayer and A. Aziz, "Object detection of fire safety equipment in images and videos using yolov5 neural network", en, in *Proceedings of 33. Forum Bauinformatik*, München, 2022.

[5] A. Alayed, R. Alidrisi, E. Feras, S. Aboukozzana, and A. Alomayri, "Real-time inspection of fire safety equipment using computer vision and deep learning", en, *Engineering, Technology & Applied Science Research*, vol. 14, no. 22, pp. 13 290–13 298, Apr. 2024. DOI: 10.48084/etasr.6753.

[6] A. Corneli, B. Naticchia, M. Vaccarini, F. Bosché, and A. Carbonari, "Training of yolo neural network for the detection of fire emergency assets", in *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, vol. 37, IAARC Publications, 2020, pp. 836–843.

[7] S. Kumar, H. Gupta, D. Yadav, I. A. Ansari, and O. P. Verma, "Yolov4 algorithm for the real-time detection of fire and personal protective equipments at construction sites", *Multimedia Tools and Applications*, vol. 81, no. 16, pp. 22 163–22 183, 2022.

[8] D.-S. Oh, J. H. Jeon, J.-K. Kim, K. B. Lee, and S. G. Hong, "Inspection system of firefighting facilities based on yolov4 using augmented reality glass", in *2022 13th International Conference on Information and Communication Technology Convergence (ICTC)*, IEEE, 2022, pp. 1604–1608.

[9] F. M. Talaat and H. ZainEldin, "An improved fire detection approach based on yolo-v8 for smart cities", en, *Neural Computing and Applications*, vol. 35, no. 28, pp. 20 939–20 954, Oct. 2023. DOI: 10.1007/s00521-023-08809-1.

[10] Y.-J. Chen, Y.-S. Lai, and Y.-H. Lin, "Bim-based augmented reality inspection and maintenance of fire safety equipment", *Automation in Construction*, vol. 110, p. 103 041, Feb. 2020. DOI: 10.1016/j.autcon.2019.103041.

[11] K. Li and J. Malik, "Amodal instance segmentation", en, in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, editors, Cham: Springer International Publishing, 2016, pp. 677–693. DOI: 10.1007/978-3-319-46475-6_42.

[12] J. Ao, Q. Ke, and K. A. Ehinger, “Image amodal completion: A survey”, *Computer Vision and Image Understanding*, vol. 229, p. 103 661, Mar. 2023. DOI: 10.1016/j.cviu.2023.103661.

[13] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation”, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.

[14] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection”, 2016, pp. 779–788. [Online]. Available: https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Redmon_You_Only_Look_CVPR_2016_paper.html.

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

[16] K. He, G. Gkioxari, P. Dollar, and R. Girshick, “Mask r-cnn”, 2017, pp. 2961–2969. [Online]. Available: https://openaccess.thecvf.com/content_iccv_2017/html/He_Mask_R-CNN_ICCV_2017_paper.html.

[17] Y. Zhu, Y. Tian, D. Metaxas, and P. Dollar, “Semantic amodal segmentation”, en, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI: IEEE, Jul. 2017, pp. 3001–3009. DOI: 10.1109/CVPR.2017.320. [Online]. Available: http://ieeexplore.ieee.org/document/8099803/.

[18] G. Kanizsa, “Organization in vision: Essays on gestalt perception”, en, *(No Title)*, [Online]. Available: https://cir.nii.ac.jp/crid/1130000796894898688.

[19] P. O. Pinheiro, R. Collobert, and P. Dollar, “Learning to segment object candidates”, no. arXiv:1506.06204, Sep. 2015. DOI: 10.48550/arXiv.1506.06204. [Online]. Available: http://arxiv.org/abs/1506.06204.

[20] K. Xu, L. Zhang, and J. Shi, “Amodal completion via progressive mixed context diffusion”, en, no. arXiv:2312.15540, Dec. 2023. [Online]. Available: http://arxiv.org/abs/2312.15540.

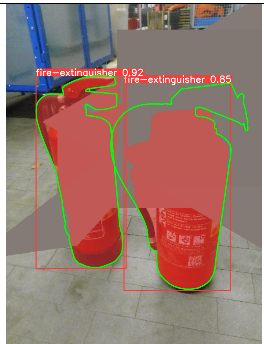
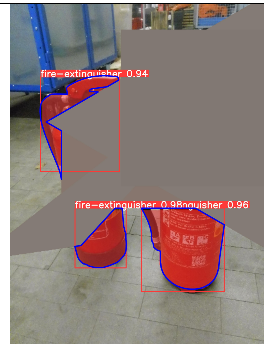
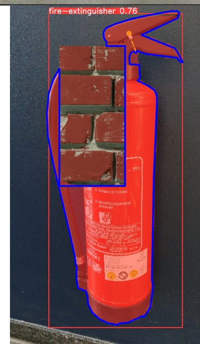
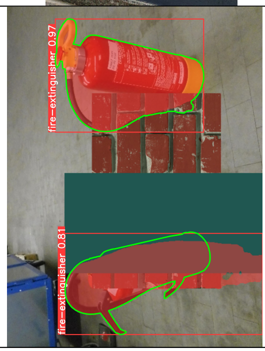
[21] P. O. Pinheiro, T.-Y. Lin, R. Collobert, and P. Dollàr, “Learning to refine object segments”, no. arXiv:1603.08695, Jul. 2016. DOI: 10.48550/arXiv.1603.08695. [Online]. Available: http://arxiv.org/abs/1603.08695.

[22] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition”, en, *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998. DOI: 10.1109/5.726791.

[23] J. Boehm, F. Panella, and V. Melatti, *Firenet*, Jul. 2019. DOI: 10.5522/04/9137798.v1. [Online]. Available: https://rdr.ucl.ac.uk/articles/dataset/FireNet/9137798.

[24] A. Aziz, P. Herbers, H. Bayer, and M. König, “Fully autonomous fire safety equipment inspection missions on a legged robot”, EN, pp. 804–812, Jan. 2024. DOI: 10.1061/9780784485224.097.

Table 3: Evaluation data of the trained YOLOv8 models (full names : Table 2). Left : Amodal Segmentation models, Right : modal segmentation models. The mIoU is the average IoU over *test*, *train*, *val* and full dataset. Example 1 out of *test* dataset, Example 2, 3 : real world images (not part of the dataset). Red : Segmentation masks, Green : amodal ground truth, Blue : modal ground truth. Rotations only for visualization.

Model	**mIoU**	**Example 1**	**Example 2**	**Example 3**	**Model**	**mIoU**	**Example 1**
SA	*train:* 0.86 *test:* 0.82 *val:* 0.85 *total:* 0.85				SM	*train:* 0.85 *test:* 0.81 *val:* 0.84 *total:* 0.85	
PA	*train:* 0.88 *test:* 0.85 *val:* 0.83 *total:* 0.86				PM	*train:* 0.87 *test:* 0.80 *val:* 0.80 *total:* 0.84	
SPA	*train:* 0.91 *test:* 0.88 *val:* 0.88 *total:* 0.89				SPM	*train:* 0.90 *test:* 0.88 *val:* 0.86 *total:* 0.88