CCC 2025 **Proceedings of the Creative Construction Conference** (2025)
Edited by: Miroslaw J. Skibniewski & Miklós Hajdu & Žiga Turk
https://doi.org/10.22260/CCC2025/0052

BASIC STUDY OF STEEL STRUCTURE WORK AREA RECOGNITION MODEL FOR FIREPROOFING SPRAY ROBOT USING RGB-D IMAGES

Sangmin Lee, Sebeen Yoon, Mingyun Kang, Kangmin Bae, Minseung Cha,
Taehoon Kim
Seoul National University of Science and Technology, Seoul, Republic of Korea

Abstract
Most of the fireproofing spray work on steel structures in construction sites is still carried out manually, resulting concerns about safety and low productivity. Although robotic systems have been introduced to address these issues, existing approaches often lack autonomous work area recognition and require operator intervention. To overcome these limitations, this study proposes a model for identifying steel structures using RGB-D images and generating improved work area masks for fireproofing spray robot. The proposed model applies image segmentation model to detect steel beams and columns, followed by a two-step mask improvement process. Depth-based filtering removes outlier regions based on percentile thresholds, while HSV-based filtering eliminates color inconsistent background pixels. The proposed model improved the mean IoU from 87.67% with raw segmentation model to 90.99% with cleaner object boundaries. Proposed model can accurately recognize the work area and provide this information to the robot, enabling it to perform tasks autonomously.

Keywords: fireproofing robot, semantic segmentation, mask improvement, work area recognition

© 2025 The Authors. Published by the International Association for Automation and Robotics in Construction (IAARC) and Diamond Congress Ltd.

Peer-review under responsibility of the scientific committee of the Creative Construction Conference 2025.

1. Introduction

Fireproofing spray on steel structures is essential procedure to guarantee fire resistance in construction. Conventional method of fireproofing spray is conducted by human using spraying gun, lifting on an equipment like scissor lift. This method has several potential problems such as inconsistent coverage, low productivity, and fall hazards at height[1].

Adopting robotics is one of the solutions to solve these problems in fireproofing spray work. Several research have been conducted [1-3], however, most of research has focused on systems that rely on pre-defined inputs such as pre-mapped work areas rather than incorporating real-time perception and work-area recognition[1].

Image-based segmentation technology has been widely used to automatically recognize an area of interest. In particular, the approach using RGB images is being adopted in various studies due to its advantage in terms of cost and equipment installation. However, since RGB images do not contain depth information, there are limitations of difficulty for estimating the location within the real space or distinguishing the distance between objects. Furthermore, the recognition accuracy tends to decrease under various lighting changes or complex background conditions[4]. These limitations are more prominent in environments where the structure is complex and the object of work is not clearly distinguished, such as construction sites.

To overcome these issues, this study proposes an advanced work area recognition model for robots that utilize RGB-D images to detect steel structure. And by leveraging depth and color information, the model can effectively detect work areas for fireproofing spray. Therefore, the refined masks are utilized

Corresponding author email address : kimth@seoultech.ac.kr

to generate accurate 3D models of the target structures, which serve as a foundation for downstream applications such as autonomous spray-path planning and post-spray quality inspection.

2. Literature Review

2.1. Robotics in Construction sites

Robotic technologies have been actively studied for various construction tasks due to the safety risks faced on workers and the need for improved productivity. For example, Wu et al.[2] proposed a method using a dual-flow aggregation network to perform curtain wall frame segmentation, aiming to assist robot systems in improving safety during high-elevation curtain wall installation. Similarly, Wang et al.[3] introduced an integrated mortar spraying and plastering robot that automates repetitive wall construction tasks, thereby enhancing work efficiency and reducing labor intensity.

These researches highlight the growing feasibility and value of robotic systems in construction, suggesting that similar approaches could be effectively applied to fireproofing spray work, which involves repetitive high-risk tasks and demands uniform application quality.

Research has also been conducted specifically on fireproofing spray. Ikeda et al.[1] developed a fire resistive covering spraying robot that operates within pre-mapped work areas and follows pre-recorded spray areas. However, reliance on offline data prevents adaption to positioning errors or changing site conditions, and the lack of real time work area recognition means the system cannot distinguish worked areas from unworked areas- leading to material waste and quality problems.

2.2. RGB-D Camera in Construction sites

Construction sites are characterized by complex structures and dynamic environments, necessitating precise and real-time recognition of work areas. Traditional RGB-based recognition methods effectively extract color and shape information but lack depth perception, limiting their accuracy in three-dimensional space. To address this limitation, RGB-D cameras, which provide both color and depth information, have been increasingly utilized in construction automation research[5].

Recent studies have applied deep learning models for semantic segmentation of RGB-D images. For instance, Czerniawski et al.[6] developed a method to segment RGB-D images into 13 distinct classes of structural and non-structural building components using deep neural networks, thereby enabling detailed environmental understanding for construction progress monitoring. Meanwhile, Zheng et al.[7] proposed a semantic segmentation framework that leverages RGB-D data to enhance visual scene understanding in indoor environments. However, these methods rely on controlled datasets or synthetic data, making them less practical for deployment in dynamic and unstructured construction environments where acquiring high quality RGB-D data is often challenging. Furthermore, existing studies have primarily focused on object presence detection, with insufficient attention given to the recognition of work areas critical for robotic tasks. These limitations highlight the need for lightweight and scalable segmentation frameworks suited to construction sites. Therefore, this study proposes an RGB-based segmentation approach using widely accessible image data, complemented by depth-based post-process to improve accuracy, specifically targeting the recognition of fireproofing spray work areas.

3. Methodology

3.1. Framework

This study proposes an RGB-D image–based pipeline to automatically recognize work areas for fireproofing spray on steel structures. As shown in Fig. 1, a trained YOLOv11-seg model first detects beams and columns from RGB images, generating initial segmentation masks. However, due to environmental complexity, these masks often contain noise or incomplete boundaries. To address this, the pipeline applies a two-step refinement process that leverages depth and color information: depth-based filtering removes outliers based on geometric consistency, and HSV-based filtering eliminates color-inconsistent background pixels.

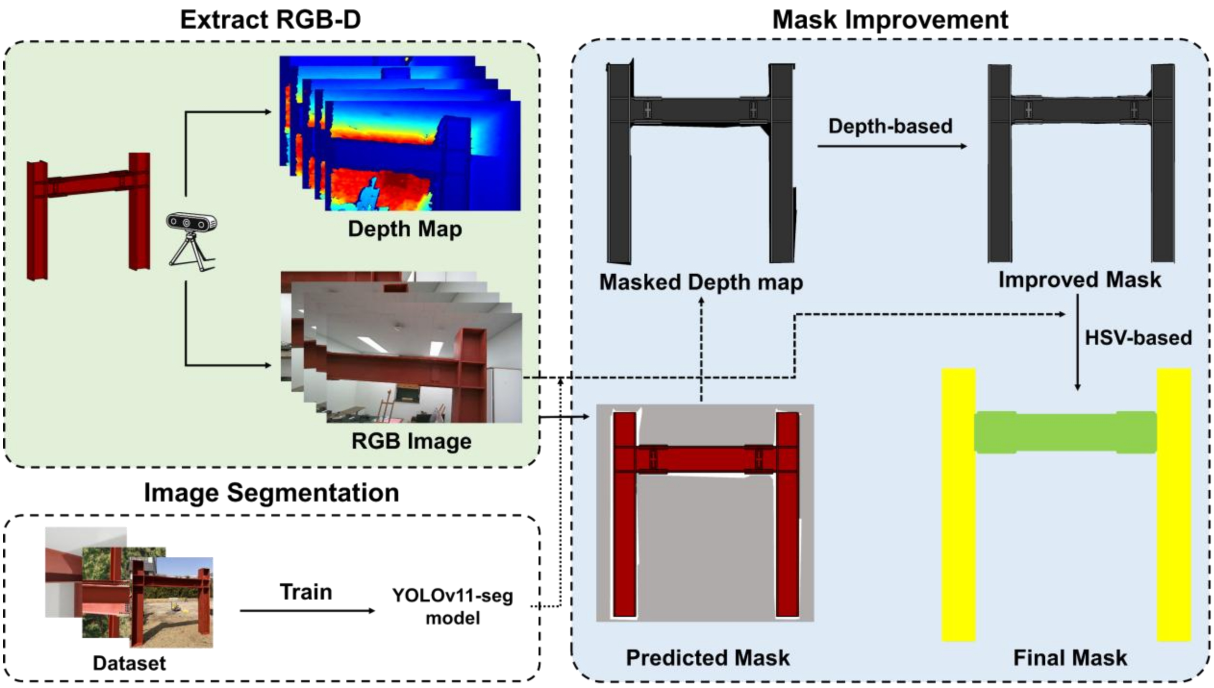
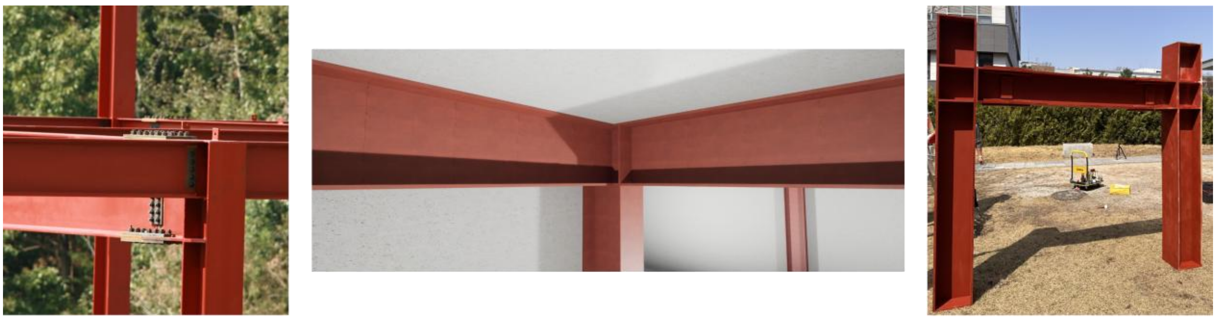


Fig. 1. Framework

3.2. Image Segmentation

The model is trained to identify two target classes: beam and column. As shown in Fig. 2, the YOLOv11-seg model was adopted and trained using a dataset composed of 250 images, including (a) 87 images of steel structures collected from web, (b) 100 synthetic images generated using Unreal Engine, and (c) 63 photographs of a mock-up. The mock-up was made of the same size as the actual steel structure and was manufactured by spraying rust preventive spray of the same color of the steel structure used in the actual construction site. These diverse data sources were combined to enhance the model's generalization capability across both real and artificial construction environments.



(a) Image via web crawling **(b) Synthetic data** **(c) Mock-up**

Fig. 2. Image datasets for training segmentation model; (a) Image via web crawling; (b) Synthetic data; (c) Mock-up

3.3. Extract RGB-D

For this study, an Intel RealSense D455 RGB-D camera was used to capture video of a steel structure mock-up as shown in Fig. 3. The camera was positioned approximately 1.2 meters away from the mock-up, with the web surface of the steel beam perpendicular to the camera. The RGB-D video was sampled at 20 frames per second (fps), generating a dataset of 47 pairs of RGB and depth images.

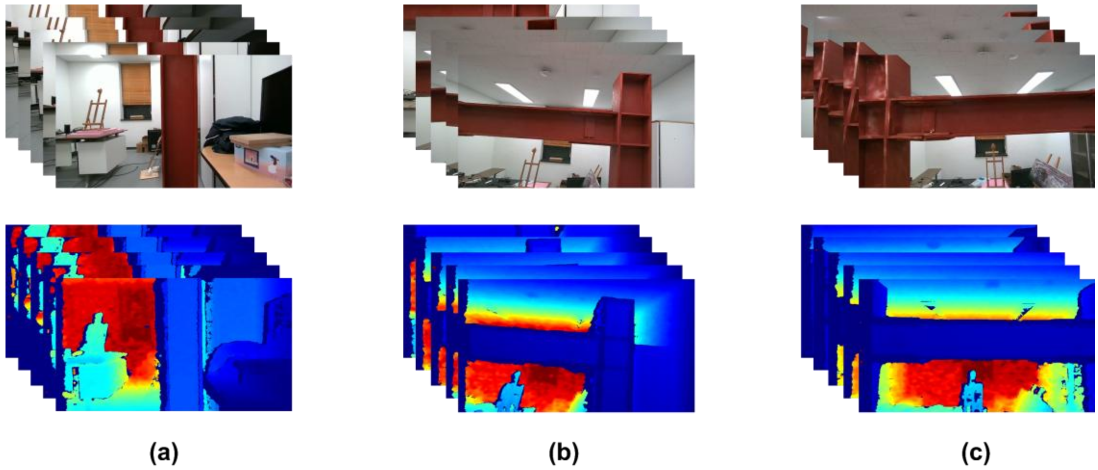


Fig. 3. RGB and Depth image extracted from RealSense (a) column; (b) beam and column; (c) beam and columns

3.4. Mask Improvement

To enhance the quality and reliability of the predicted masks, two complementary refinement strategies are introduced: depth-based filtering and HSV-based filtering.

3.4.1. Depth-based mask improvement

Depth refinement leverages the assumption that individual structural elements exhibit relatively consistent depth distributions. After the YOLO-predicted mask is expanded via dilation, the valid depth pixels within this region are collected. The 1st and 99th percentiles of these depth values are used to define lower and upper thresholds. Any mask pixel associated with a depth value outside this range is removed, except in cases where the depth value is zero, which is retained to prevent unnecessary erosion of the mask. This percentile-based filtering effectively eliminates outlier regions, such as background elements or overlapping objects, while preserving the structural consistency of the target area.

3.4.2. HSV-based mask improvement

To further eliminate background pixels based on appearance, the RGB image is converted into HSV color space. Within the YOLO mask region, the mean hue is computed, and all pixels within ±15 degrees of this value are retained. In addition, saturation and value thresholds are imposed to ensure only vivid and structurally relevant areas are kept.

4. Results and Discussions

The evaluation of the segmentation performance was conducted using the Intersection over Union (IoU) metric, computed for each target class on a per-image basis. The final mean IoU (mIoU) was then calculated by averaging the class-wise IoU scores across all test images which is consistent with the general approach discussed by Wang et al[8].

The results show that the proposed model achieved significantly higher accuracy than the original YOLOv11-based segmentation. Specifically, the IoU improved from 0.8537 to 0.8807 for the beam class (+2.7%p) and from 0.8999 to 0.9339 for the column class (+3.4%p). The mean IoU also increased from 0.8767 to 0.9099, demonstrating consistent gains across both classes. These improvements were particularly evident in terms of background elimination and boundary refinement, as shown in Table 1. Despite the added improvement steps, the total processing time per image pair remained efficient, only 13 seconds for 47 images, supporting the practical applicability to real-time robotic workflows.

In addition, Fig .4 provides a visual comparison between ground truth (GT) mask, the raw YOLO mask, and the final mask generated by our model. Visually, the improved masks better delineate the object boundaries and effectively remove irrelevant background elements.

Table 1. Quantitative evaluation (Bold values indicate better performance)

Model	Class	IoU	mIoU
YOLOv11-seg	Beam	0.8537	0.8767
	Column	0.8999	
Ours	Beam	**0.8807**	**0.9099**
	Column	**0.9339**	

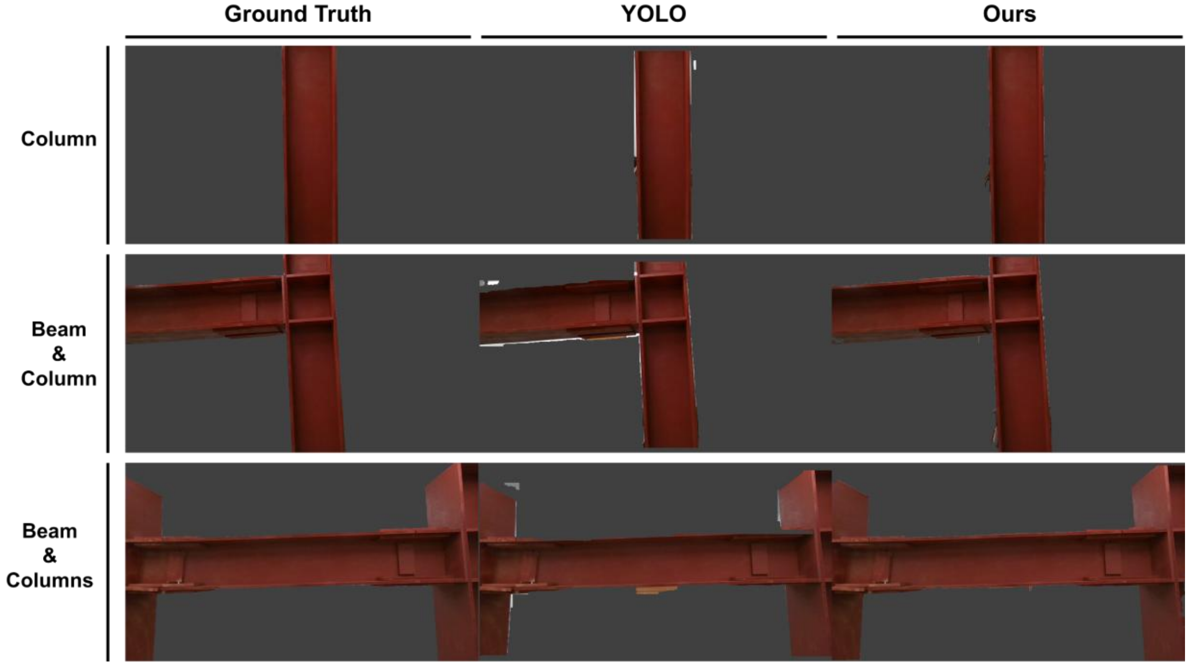


Fig. 4. Qualitative evaluation

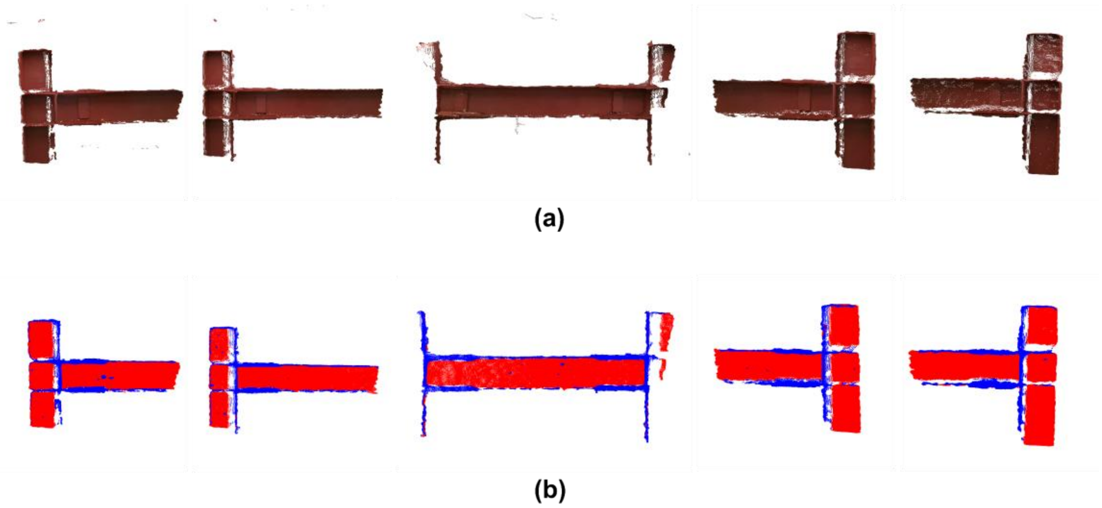


Fig. 5. (a) 3D models of detected work areas; (b) classified as web(red) and flange(blue)

Furthermore, the refined masks were utilized to reconstruct clean 3D models of the target steel structures, as illustrated in Fig. 5(a). These models provide intuitive visualizations of the recognized work areas and serve as a geometric basis for higher-level robotic planning tasks. To visualize the recognized work areas in 3D, point clouds are constructed from the refined RGB-D masks. First, statistical outlier removal (SOR) is applied to reduce noise. Then, planar segmentation using RANdom SAmple Consensus (RANSAC) is performed to extract the web surface of the steel structure. Points belonging to the extracted plane are classified as the web (red), while the remaining points are considered as flanges (blue), as shown in Fig. 5(b).

5. Conclusion

This study proposed an RGB-D image–based segmentation framework to automatically recognize work areas for fireproofing spray on steel structures in construction sites. By combining a YOLOv11-based segmentation model with depth and HSV-based refinement strategies, the proposed method significantly improved the accuracy and reliability of the segmentation masks. The refined masks not only exhibited cleaner boundaries and reduced background noise but also supported 3D reconstruction and web–flange classification, providing a foundation for robotic spray-path planning and quality inspection.

Experimental results on RGB-D mock-up data demonstrated that the proposed method achieved notable improvements in class-wise and mean IoU scores from 87.67% to 90.99% compared to the original segmentation results. Despite the refinement steps, the method remained computationally efficient, confirming its suitability for real-time robotic applications.

The proposed model can accurately recognize the work area in real-time and provide this information to the robot, enabling it to perform spray task autonomously. However, this method was only tested in a controlled mock-up environment. For the further study, we would collect more data and test from actual construction sites.

Acknowledgements

This research was supported by a grant (RS-2022-00143493) from Digital-Based Building Construction and Safety Supervision Technology Research Program funded by Ministry of Land, Infrastructure and Transport of Korean Government.

References

[1]]Y. Ikeda, H. Segawa, and N. Yabuki, “Development and Application of a Fire Resistive Covering Spraying Robot to Building Construction Site,” Proceedings of the 37th International Symposium on Automation and Robotics in Construction (ISARC), Oct. 2020, doi: 10.22260/isarc2020/0203.

[2] D. Wu et al., “Curtain wall frame segmentation using a dual-flow aggregation network: Application to robot pose estimation,” Automation in Construction, vol. 168, p. 105816, Dec. 2024, doi: 10.1016/j.autcon.2024.105816.

[3] Y. Wang et al., “Mortar spraying and plastering integrated robot for wall construction,” Automation in Construction, vol. 165, p. 105533, Sep. 2024, doi: 10.1016/j.autcon.2024.105533.

[4] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 4, pp. 834–848, Apr. 2018, doi: 10.1109/tpami.2017.2699184.

[5] Z. Wang, Y. Zhang, K. M. Mosalam, Y. Gao, and S. Huang, “Deep semantic segmentation for visual understanding on construction sites,” Computer-Aided Civil and Infrastructure Engineering, vol. 37, no. 2, pp. 145–162, May 2021, doi: 10.1111/mice.12701.

[6] T. Czerniawski and F. Leite, “Automated segmentation of RGB-D images into a comprehensive set of building components using deep learning,” Advanced Engineering Informatics, vol. 45, p. 101131, Aug. 2020, doi: 10.1016/j.aei.2020.101131.

[7] Z. Zheng, L. Mentzer, B. Iskender, M. Price, C. Prendergast, and A. Cloitre, “A pipeline for segmenting and structuring RGB-D data for robotics applications,” arXiv preprint arXiv:2410.17988, Oct. 2024. [Online]. Available: https://doi.org/10.48550/arXiv.2410.17988

[8] Z. Wang et al., “Revisiting evaluation metrics for semantic segmentation: Optimization and evaluation of fine-grained intersection over union,” GitHub Repository, 2023. [Online]. Available: https://github.com/zifuwanggg/JDTLosses