Instance Segmentation of Exterior Insulation Finishing System using Synthetic Datasets

Mingyun Kang[1] Sebeen Yoon[1] Juho Han[1] Sanghyeon Na[1] and Taehoon Kim[1]

[1]Architecture Engineering, Seoul National University of Science and Technology, Republic of Korea

kmg4312@seoultech.ac.kr, paul6452@seoultech.ac.kr, hanjooho0421@seoultech.ac.kr
nsh143@seoultech.ac.kr, kimth@seoultech.ac.kr

Abstract –
The quality inspection of adhesive of Exterior Insulation Finishing System (EIFS) is important because poor adhesive can lead to detachment of the insulation. Computer vision-based inspection stands out as a notable alternative. Recently, imaged-based deep learning model are widely used for the automated monitoring and inspection in construction field. To train the model, the relevant large datasets are essential. However, collecting datasets in the construction site is hazardous because of inherent risk of accidents. Also, synthetic datasets method which is one of alternatives to solve this problem are focused on fixed and regular shaped objects. To address these challenges, this study analyses the validity of synthetic datasets in terms of segmentation of adhesive in EIFS, which has irregular shape. For instance segmentation, the datasets were divided into two groups: (1) real datasets, composed of 100 actual photos, (2) mixed datasets, which combined 50 randomly sampled images from both synthetic datasets and real datasets. The mAP@50 of instance segmentation for real datasets and mixed datasets is 87% and 99%, respectively. This study prove that synthetic datasets can effectively train segmentation models, enabling the recognition of irregularly shaped objects and enhancing overall performance.

Keywords – Synthetic datasets, Image-based deep learning, Instance segmentation, Exterior insulation finishing system

1 Introduction

Exterior Insulation and Finishing System (EIFS) is a method of covering the entire exterior of building. There are primarily two methods for applying EIFS: the dry and wet processes. The wet process, the most common method, involves attaching insulating material to the structure, which is coated with an adhesive such as mortar [1,2]. Poor adhesive performance can lead to detachment of the insulating material. However, inspecting the application quality with limited supervisory personnel at construction sites is labor-intensive and virtually impractical in terms of time and cost. To address this challenge, automating the inspection of adhesive using deep learning-based computer vision can significantly enhance productivity by minimizing manpower and costs [2,3].

The relevant large datasets are essential for the deep learning. Existing open-source image datasets such as Common Objects in Context (COCO) [4] and the Stanford 2D-3D-Semantics Dataset [5] are available but, there are no datasets related to construction site, particularly, EIFS. Additionally, acquiring image datasets in construction site is difficult due to the inherent risk of accidents at construction sites. Moreover, the images obtained are often disordered and cluttered [6]. As an alternative, using web crawling techniques to acquire images allows the rapid collection of a large number of photos without visiting construction sites. However, a significant issue arises with the time-consuming process of labeling these images to transform them into a dataset suitable for training in computer vision [6,7].

Recently, a novel approach has been proposed to overcome the challenges previously mentioned: synthetic image modeling, which involves creating datasets for specific fields. By synthesizing in artificial modeled virtual environment, desired scenes can be created. The synthesized scenes are already data-encoded, thus reducing efforts for labeling and annotation [3,7,8]. However, existing studies have focused on fixed and regular shaped objects, thus applying instance segmentation with synthetic datasets of irregular shaped object is not explored well.

The purpose of this study is to conduct foundational research on the effectiveness of training an image segmentation model for segmenting dab and insulation in EIFS utilizing synthetic data, comparing the performance between real datasets and mixed datasets combining synthetic datasets.

2 Related Work

2.1 Computer vision in construction.

With the advancement of datasets and the corresponding development of computer vision, applications in the construction industry have also evolved. AI, along with more sophisticated object detection and segmentation techniques, not only reduce tedious, human-error-prone tasks but also lead to improved research in terms of labor intensity and time consumption.

Computer vision facilitates rapid and accurate material quantity estimation, thereby aids progress monitoring. Li et al. (2021) [9] proposed a real-time, cost-effective rebar counting scheme using the YOLOv3 detector. Wang et al. (2021) [10] utilized surveillance video to track information of precast walls such as numbers and position, transmitting the data in the Building Information Model (BIM) to achieve automatic progress using timestamp methods.

In the early stages of computer vision application in construction safety management, it primarily focused on determining the use of personal protective equipment such as workers' hard hats. Fang et al. (2018) [11] conducted research on non-hardhat-use detection using 100,000 photos from far-field surveillance at construction sites. Additionally, research on classifying cracked or damaged buildings has been actively pursued. Silva et al. (2018) [12] proposed a model that uses deep learning image classification to differentiate between concrete photos with and without cracks. Gao et al. (2018) [13] conducted structural damage recognition through spalling condition checks and evaluation of damage levels.

In the construction field of quality inspection, Xia et al. (2024) [14] conducted research on First Floor Elevation estimation using the YOLOv5 model and mobile LiDAR point clouds. Chen et al. (2021) [15] investigated quality of building façade using photos obtained from unmanned aerial vehicles.

2.2 Synthetic datasets related to construction industry

While computer vision has evolved alongside the advancement of datasets, there has been a shortage of datasets related to the construction field [6,7]. Acquiring real image data poses risks due to the nature of construction sites, also labeling acquired images manually requires significant time and labor.

To address this challenge, lots of research has been conducted to train computer vision models using synthetic datasets created using various methods related to construction sites. Soltani et al. (2016) [3] proposed a method to separately synthesize excavators and backgrounds. They found that, compared to traditional annotation methods, auto-annotation reduced the required time. Neuhausen et al. (2020) [8] aimed to create synthetic datasets using Blender to track worker productivity and safety management. Hong et al. (2021) [16] utilized the BIM model to generate synthetic datasets for infrastructure scenes. While these studies provided direction and validated the utility of synthetic datasets for computer vision, they focused on large objects. Lee et al. (2023) [17] conducted research of safety monitoring through synthetic datasets focusing on small-sized personal protective equipment detection using the Unity game engine.

Particularly, Wang et al. (2023) [18] studied synthetic datasets for rebar instance segmentation. They compared the performance using a Mask R-CNN model with various combinations of real datasets and synthetic dataset, concluding that experimental groups composed with both real and synthetic datasets appropriately are more robust for instance segmentation than using only real datasets.

Although these studies sufficiently validated the utility of synthetic datasets in the construction industry for computer vision, they targeted fixed and regular shapes (e.g., excavators, rebar, etc.). Therefore, in this study, we aim to create synthetic datasets for EIFS instance segmentation, targeting irregularly shaped objects that may vary in appearance over time, and validate their utility accordingly.

3 Methodology

This study analyses the effectiveness of training an image segmentation model using mixed datasets combined with synthetic data, compared to real datasets. Instance segmentation model was applied to segment dab and insulation in an image, thereby the classes were set as insulation and dab. Figure 1 illustrates the shape of EIFS.

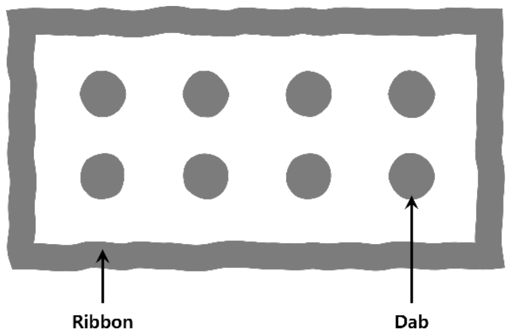

Figure 1. Ribbon and dab method

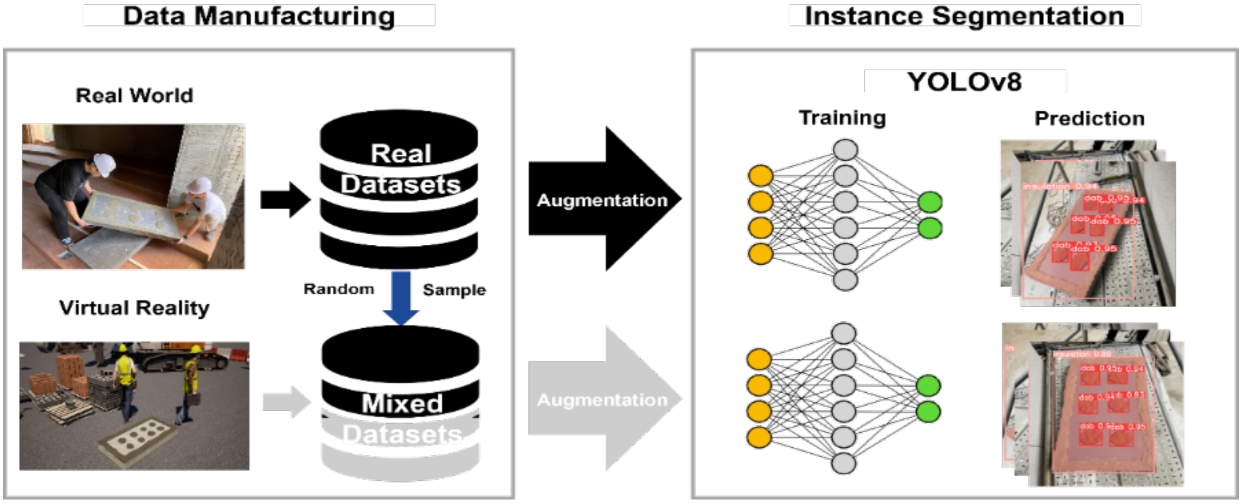

Figure 2. Framework of Study

As illustrated in Figure 2, the framework can be divided into two main phases. First, during the dataset manufacturing stage, photographs are collected from both real EIFS construction sites and virtual construction sites where EIFS is ongoing. Then, mixed datasets are made with synthetic datasets and randomly sampled real datasets. Both real datasets and mixed datasets are augmented before instance segmentation stage to prevent overfitting. Second, in the instance segmentation stage, the YOLOv8 model is used to train the datasets. The trained weights are then utilized to compare area segmentation in actual EIFS construction sites.

3.1 Data Manufacturing

To test the efficacy of synthetic datasets, the experimental datasets were divided into two control groups: (1) 100 numbers of realistic images from the construction site, and (2) a mixed dataset of 50 realistic and 50 synthetic images. For the real datasets, we visited three different EIFS construction sites, to acquire images for train model and testing. Labeling and annotation was carried out using Roboflow [19], and augmentation was performed to prevent overfitting.

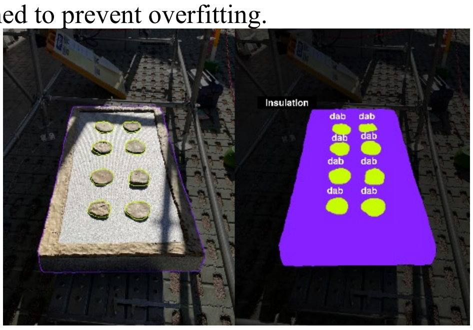

Figure 3. Images and annotation using Roboflow

Synthetic dataset environment was created using Unreal Engine 4, and the dataset synthesis model was developed utilizing the algorithm of NVIDIA's Deep Learning Dataset Synthesizer (NDDS) [20,21].

As illustrated in left side of Figure 4 virtual construction site was created using Unreal Engine 4. For the variation of datasets such as shade adding, point of view, worker's position were conducted. The right side of Figure 4 shows the annotation of the image. Every single color of image represents the class which is divided respectively and automatically connected the components of Unreal Engine 4.

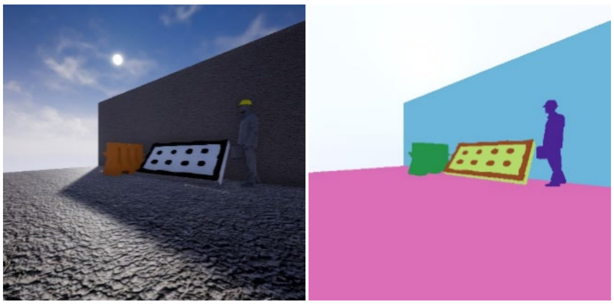

Figure 4. Virtual construction site of Unreal Engine 4(left) annotation of the image(right)

3.2 Instance Segmentation

The You Only Look Once (YOLO) model is one of the most renowned models for one-step object detection capabilities, offering rapid detection speed and high accuracy. The latest version of the YOLO series is YOLOv8 [22], which replaces the C3-module with C2f-module for robust gradient flow, adopts a discrete head structure, these modifications greatly improve the detection accuracy.

YOLOv8-seg is an instance segmentation model derived from YOLOv8. YOLOv8-seg consists of five models: 8n-seg, 8s-seg, 8m-seg, l-seg, and x-seg. Starting with the lightest model, 8n-seg, the amount of computation increases gradually towards 8x-seg. Consequently, mAP and processing speed also increase. The selected model was YOLOv8x-seg in this study, which is the highest accuracy model in the aspect of mAP.

4 Experimental Study

4.1 Evaluation Metrics

In this section, we aim to discuss the results of tests conducted using YOLOv8 to evaluate the validity of instance segmentation in both real datasets and mixed datasets. The data were split into training, validation, and test datasets at a ratio of 80:10:10. The parameters for the training model were set as follows: 200 epochs, a learning rate of 0.01, and a batch size of 16.

The performance metric for the segmentation algorithm was mean Average Precision at 50 (map@50). Average Precision (AP) is the area under the precision-recall curve for a specific class, and the mAP is the average of these AP values across all classes. The map@50 specifically refers to the mAP calculated with an Intersection over Union (IoU) threshold set at 0.5, meaning that predictions must overlap at least 50% with the ground truth to be considered correct. The equation of AP, mAP are demonstrated below.

$$\mathrm{AP} = \lim_{n\to\infty} \sum_{k=1}^{n} (R_k - R_{k-1})P_k = \int_0^1 p(r)\,dr \tag{1}$$

$$\mathrm{mAP} = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{2}$$

The Intersection over Union (IoU) is calculated as the ratio of the area of overlap between the Ground Truth and the Detection to the sum of their areas (Equation(3))

$$\mathrm{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \tag{3}$$

$$\text{precision} = \frac{TP}{TP + FP} \tag{4}$$

$$\text{recall} = \frac{TP}{TP + FN} \tag{5}$$

True positive (TP) refers to the count of instances correctly predicted as true by the model, false positive (FP) denotes instances wrongly predicted as true while they are false, false negative (FN) represents instances incorrectly predicted as false that are actually true. Precision (Equation(4)) is the proportion of correct predictions among the results predicted by the model, while Recall (Equation(5)) is the proportion of actual true instances that the model correctly predicts as true. Both Precision and Recall are adjusted based on the confidence level setting used as the model's decision criterion. Raising the decision criterion increases Precision, whereas lowering it enhances Recall.

4.2 Test of Real Datasets

The result of Real datasets at the best confidence level, the values of precision, recall and map@50 are 0.89, 0.96, and 0.87 respectively. The highest performance at the 198th epoch are depicted in the Confusion Matrix shown in the following Figure 5. According to the confluence matrix, the model trained by real datasets predicted the dab well at 1.0, while the prediction of insulation was relatively low at 0.71.

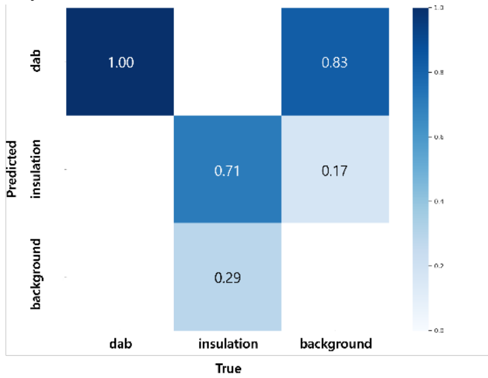

Figure 5. Confusion matrix of real datasets

4.3 Test of Mixed Datasets

The results of the 100 numbers of mixed datasets which is composed with 50 numbers of synthetic datasets and 50 randomly sampled real datasets. The Precision, Recall, and map@50 at the best confidence level are 0.99, 0.94, and 0.95, respectively. The confusion matrix in the subsequent Figure 6 illustrates the performance outcomes for the mixed datasets dataset using the model that achieved its peak performance at the 175th epoch. According to the confusion matrix, it can be seen that model trained by synthetic datasets predicted well as 1.0 for both dab and insulation.

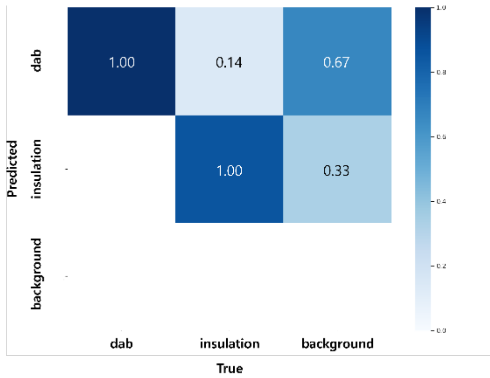

Figure 6. Confusion matrix of mixed datasets

4.4 Comparative Analysis

The difference in map@50 between real datasets and mixed datasets was more marginal than expected. However, the trends observed in YOLOv8's validation, as shown in Figure 7, indicated distinct learning patterns for real datasets and mixed datasets. While the number of datasets may not be sufficient, obtaining real datasets involved visiting construction sites to capture photographs, as mentioned in the introduction, which presents hazards and challenges in image labeling. In contrast, with synthetic datasets, just a few clicks can modify the environment, allowing for the manufacture of datasets with varied characteristics.

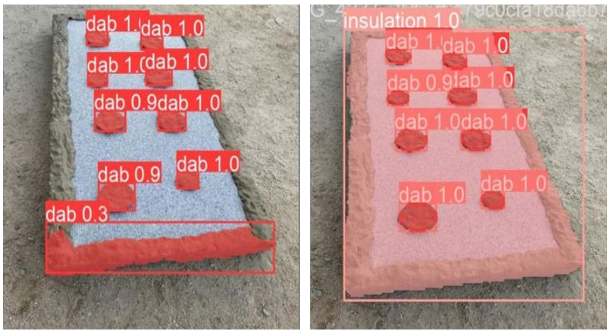

Figure 7. Prediction of real EIFS image with model trained by real datasets (left) and mixed datasets (right)

5 Conclusions

This study analyzed the effectiveness of training a deep learning model for segmenting irregular shaped dab and insulation in EIFS utilizing synthetic datasets, compared to real datasets. In this study, images of EIFS in construction sites were acquired from both real world and virtual environments to create datasets. To evaluate the segmentation performance of EIFS, the subjects were categorized into two classes, dab and insulation, and trained using the YOLOv8x-seg model. The map@50 results used as performance indicators for real datasets and mixed datasets were 0.87 and 0.95, respectively.

Utilizing synthetic datasets can reduce the risk of visiting construction sites with safety accidents. Also, generating synthetic datasets enable the creation of various compositions, environments, and scenarios with little effort. Furthermore, research suggested segmentation of irregular shaped object and potential model robustness as well. Thus, proposed method can be utilized in other applications such as concrete crack.

For further research, the effectiveness of applying synthetic datasets in the instance segmentation on various objects will be explored with a large size of datasets. Additionally, the comprehensive automatic supervision model will be studied as well as segmentation of ribbons with dab and insulation, to calculate the area of EIFS segments in pixels.

6 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 and National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 1711191456).

7 References

[1] J. H. Kim, S. M. Kim, and J. T. Kim, "Evaluation on Weather Resistance of Dry and Wet External Thermal Insulation Walls according to EOTA Test Standard," Journal of Korean Institute of Architectural Sustainable Environment and Building Systems, vol. 15, no. 4, pp. 405-418, 2021.

[2] S. Yoon, M. Kang, H. Jang, and T. Kim, "Computer Vision-based Automated Adhesive Quality Inspection Model of Exterior Insulation and Finishing System,". Journal of the Korea Institute of Building Construction, vol. 23, no. 2, pp. 165–173, Apr. 2023.

[3] M. M. Soltani, Z. Zhu, and A. Hammad, Automated Annotation for Visual Recognition of Construction Resources Using Synthetic Images, Automation in Construction, vol. 62, pp. 14-23, 2016.

[4] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, et al., Microsoft COCO Common Objects in Context, in Computer Vision–ECCV 2014 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13, Springer International Publishing, 2014, pp. 740-755.

[5] I. Armeni, S. Sax, A. R. Zamir, and S. Savarese, Joint 2D-3D-Semantic Data for Indoor Scene Understanding, arXiv preprint arXiv1702.01105, 2017.

[6] H. Tajeen and Z. Zhu, Image dataset development for measuring construction equipment recognition performance, Automation in Construction, vol. 48, pp. 1-10, 2014.

[7] J. G. Lee, J. Hwang, S. Chi, and J. Seo, Synthetic Image Dataset Development for Vision-Based Construction Equipment Detection, Journal of Computing in Civil Engineering, vol. 36, no. 5, Art. no. 04022020, 2022.

[8] M. Neuhausen, P. Herbers, and M. König, Using Synthetic Data to Improve and Evaluate the Tracking Performance of Construction Workers on Site, Applied Sciences, vol. 10, no. 14, Art. no.

4948, 2020.
[9] Y. Li, Y. Lu, and J. Chen, "A deep learning approach for real-time rebar counting on the construction site based on YOLOv3 detector," Automation in Construction, vol. 124, p. 103602, 2021.
[10] Z. Wang, Q. Zhang, B. Yang, T. Wu, K. Lei, B. Zhang, and T. Fang, "Vision-based framework for automatic progress monitoring of precast walls by using surveillance videos during the construction phase," Journal of Computing in Civil Engineering, vol. 35, no. 1, pp. 04020056, 2021.
[11] Q. Fang, H. Li, X. Luo, L. Ding, H. Luo, T. M. Rose, and W. An, "Detecting non-hardhat-use by a deep learning method from far-field surveillance videos," Automation in Construction, vol. 85, pp. 1-9, 2018.
[12] W. R. L. da Silva and D. S. de Lucena, “Concrete Cracks Detection Based on Deep Learning Image Classification,” The 18th International Conference on Experimental Mechanics, Jun. 2018, doi: 10.3390/icem18-05387.
[13] Y. Gao and K. M. Mosalam, "Deep transfer learning for image-based structural damage recognition," Computer-Aided Civil and Infrastructure Engineering, vol. 33, no. 9, pp. 748-768, 2018.
[14] J. Xia and J. Gong, "Computer vision based first floor elevation estimation from mobile LiDAR data," Automation in Construction, vol. 159, p. 105258, 2024.
[15] K. Chen, G. Reichard, A. Akanmu, and X. Xu, "Geo-registering UAV-captured close-range images to GIS-based spatial model for building façade inspections," Automation in Construction, vol. 122, p. 103503, 2021.
[16] Y. Hong, S. Park, H. Kim, and H. Kim, "Synthetic Data Generation Using Building Information Models," Automation in Construction, vol. 130, Art. no. 103871, 2021.
[17] H. Lee, J. Jeon, D. Lee, C. Park, J. Kim, and D. Lee, "Game engine-driven synthetic data generation for computer vision-based safety monitoring of construction workers," Automation in Construction, vol. 155, p. 105060, 2023.
[18] H. Wang, Z. Ye, D. Wang, H. Jiang, and P. Liu, "Synthetic Datasets for Rebar Instance Segmentation Using Mask R-CNN," Buildings, vol. 13, no. 3, p. 585, 2023.
[19] Roboflow, https://roboflow.com, Accessed: 13/03/2024
[20] T. To, J. Tremblay, D. McKay, Y. Yamaguchi, K. Leung, A. Balanon, J. Cheng, W. Hodge, and S. Birchfield, "NDDS: NVIDIA Deep Learning Dataset Synthesizer," 2018. [Online]. Available: https://github.com/NVIDIA/Dataset_Synthesizer
[21] J. Tremblay, T. To, and S. Birchfield, “Falling Things: A Synthetic Dataset for 3D Object Detection and Pose Estimation,” 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Jun. 2018, doi: 10.1109/cvprw.2018.00275.
[22] YOLOv8, https://github.com/ultralytics/ultralytics, Accessed: 13/03/2024