Journal of Digital Media & Culture Technology Vol.5, No.2, December (2025), pp.173-184
ISSN 2983-0664(p) 2983-0850(e) *http://dx.doi.org/10.29056/jdmct.2025.12.08*

YOLOv8 Detection of Traditional Korean Architectural Components : Synthetic vs. Real-World Imagery

Jung Eun Yu[1], JungYoon Kim[2*]

Abstract

Media facade-based projection mapping requires precise alignment between architectural structures and projected imagery, yet outdoor environments often demand repeated manual calibration due to external factors. To address this limitation, reliable recognition of architectural structural components must be established in advance. As a foundational step toward automated projection mapping, this study quantitatively evaluates object detection performance for structural components in traditional Korean architectural images. A YOLOv8 model was trained separately on Unreal Engine-based synthetic images and real-world images, and detection performance was compared across six architectural components. The results indicate that the synthetic data-based model performs adequately for components with clear global shapes but shows limitations for those relying on fine patterns or surface textures, whereas the real-world image-based model demonstrates improved overall performance by capturing diverse visual cues. By revealing component-specific detection characteristics and data-dependent performance gaps, this comparative analysis provides practical guidance for selecting training data strategies and designing structure-aware alignment modules in automated projection mapping pipelines. These findings provide a technical basis for incorporating structural recognition into automated projection mapping systems.

Keyword : Computer Vision, Projection Mapping, YOLOv8, Synthetic Data, Architectural Object Detection

1. Introduction

A media facade refers to a technology that projects digital imagery onto building exteriors,

1 Department of Game and Interactive Media, Gachon University, Seongnam, Korea [Graduate Student]
e-mail: je20102810@gachon.ac.kr

2 Department of Game Media, College of IT Convergence, Gachon University, Seongnam, Korea [Professor]
e-mail: kjyoon@gachon.ac.kr (Corresponding author)

* This research was supported by Culture, Sports and Tourism R&D Program through the Korea Creative Content Agency grant funded by the Ministry of Culture, Sports and Tourism in 2023 (Project Name: Cultural Technology Specialist Training and Project for Metaverse Game, Project Number: RS-2023-00227648, Contribution Rate: 100%)

* A preliminary version of this paper was presented at the 2025 Winter Conference of the Next-Generation Convergence Information Service Society (NCISS), held at Jeju National University, Jeju, Republic of Korea, December 14-16, 2025.

Received(October 29, 2025), Review Result(1st: November 18, 2025), Accepted(December 12, 2025), Published(December 31, 2025)

transforming static urban surfaces into dynamic public screens and expanding urban space into a large-scale visual medium [1]. Owing to these characteristics, media facades have been increasingly adopted in advertising, public art, and performance production. In South Korea, they have steadily expanded around major landmarks such as Gwanghwamun, the Sejong Center for the Performing Arts, and Dongdaemun Design Plaza (DDP) [2]. In addition, media facades have been actively introduced to traditional architectural sites, including Gyeongbokgung Palace, Seokjojeon Hall of Deoksugung Palace, and Changgyeonggung Palace. This trend reinforces their cultural and artistic significance by enabling contemporary reinterpretations of cultural heritage and extending visitor experiences [3].

Projection mapping, as the core technology underlying media facades, involves the precise alignment of digital content with the boundaries of physical structures such as buildings [4]. However, projection mapping in outdoor environments is affected by various external factors, including projector displacement, illumination conditions, and weather variations. These factors reduce projection accuracy and impose structural limitations that require repeated manual calibration. To address these issues, approaches such as camera-based feedback, marker-based alignment, and projection error correction algorithms have been proposed. Nevertheless, fully automating projection mapping under complex environmental variations remains challenging [5][6].

Stable automation of projection mapping requires a prior capability to reliably recognize the structural components of target buildings. Architectural structures directly influence the separation of projection regions and the alignment process. Recognition at the level of structural components therefore constitutes foundational information for an automated correction pipeline. Modern architecture presents difficulties in this regard because of its high degree of formal freedom and design diversity, which complicates the construction of generalized recognition models for structural patterns [7][8]. In contrast, traditional Korean public architecture exhibits characteristic regularity, with structural units such as columns, windows, and eaves arranged in repetitive and rule-based configurations. This property provides a comparatively favorable environment for evaluating structure-based recognition techniques.

To address this issue, this study conducts automatic detection of major structural components in traditional Korean architectural images and quantitatively analyzes the results by comparing synthetic and real-world images. Object detection models were trained using 3D rendering-based synthetic images and real-world images captured from real-world heritage sites, and detection performance differences between the two data types were evaluated. By examining the potential and limitations of recognizing traditional architectural structures, this study presents a starting point for structure-aware research toward future automation of projection mapping systems.

2. Theoretical Background

2.1 Projection Mapping and Structural Alignment

Projection mapping is a technique that aligns digital content with structures in three-dimensional space and projects imagery along the contours of a target object [9]. Previous studies on projection mapping have primarily proposed calibration methods that align projector and camera coordinate systems based on geometric structural information, such as the position, shape, and surface boundaries of the projection target. These methods perform projection coordinate transformations derived from this alignment. By treating structural information as a prerequisite for projection alignment and optical calibration, such approaches aim to minimize distortion on non-planar or complex surfaces and to ensure stable projection quality [10].

Projection mapping in outdoor and architectural environments requires precise geometric alignment for non-planar and complex structural surfaces. Maintaining stable projection quality is difficult because of various environmental factors, including illumination changes, shadow formation, defocus, and projection latency. Prior studies have sought to mitigate these issues through geometric alignment, photometric compensation, defocus correction, and shadow removal techniques. Nevertheless, complete automation and real-time deployment remain limited because of environmental uncertainty and hardware constraints in outdoor settings [11]. Accordingly, recent research has explored self-calibrating projection mapping frameworks that employ structured light and camera feedback to automatically estimate geometric parameters and to reduce reliance on static configurations [6][12].

2.2 Computer Vision-Based Architectural Structure Recognition

Research on computer vision-based architectural structure recognition has progressed toward automatically separating and identifying facade elements or structural components from images. Recent studies have applied deep learning-based segmentation techniques to recognize repetitive and regular architectural elements such as windows, doors, and walls. In particular, it has been reported that incorporating prior knowledge of the morphological regularity and proportional relationships of architectural components can contribute to improved recognition performance. These approaches focus on extracting structural information from architectural images through object detection and segmentation, and they have been extended to a range of architectural and urban research fields, including urban environment analysis, building condition assessment, and spatial reconstruction [7][13].

However, architectural form varies substantially depending on building type and style. In freely composed and irregular facade structures, the lack of traditional normalized patterns introduces additional difficulties in defining and recognizing boundaries between components. For example, the facades of irregular modern buildings have been reported to increase the complexity of applying conventional segmentation techniques [8].

2.3 Synthetic Data in Vision-Based Recognition

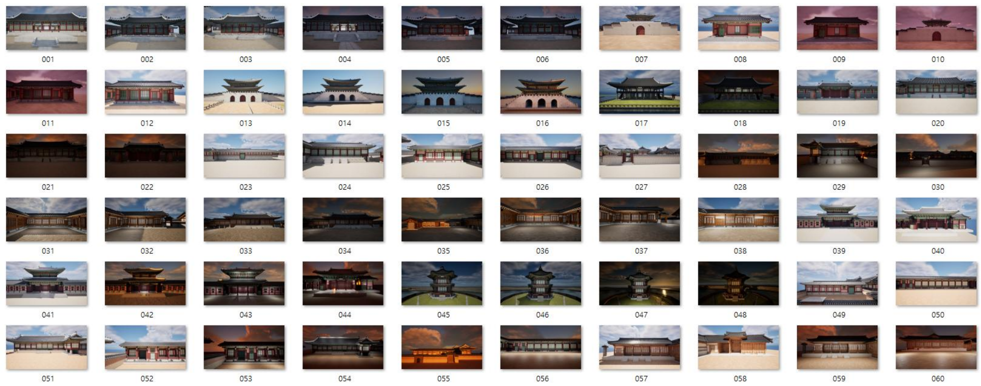
Synthetic data generated using 3D models can reduce the cost and constraints associated with real-world data acquisition. Such data also enable precise control over viewpoint, illumination, and layout conditions. For these reasons, their use has expanded across computer vision research. Synthetic images allow automatic labeling and have been proposed as an alternative training data source to address difficulties in obtaining repeated data under specific conditions [14]. Recently, virtual environments based on game engines such as Unreal Engine have been adopted as research platforms for synthetic data generation and simulation. These environments have been reported to support the training and validation of computer vision algorithms [15].

However, synthetic data have limitations in fully reproducing visual variations observed in real environments, including illumination changes, noise, and material properties. As a result, discrepancies in data distributions can arise when compared with real-world data. This difference is commonly described as the domain gap. Numerous studies have consistently reported that models trained solely on synthetic data may exhibit degraded performance when applied to real-world data [16].

3. Methodology

This study was conducted to construct an object detection model for automatically detecting structural components of traditional architecture and to compare performance differences between Unreal Engine-based synthetic images and real-world image-based datasets. The synthetic dataset was constructed as shown in [Fig. 1] using high-resolution 3D digital heritage models provided by the Korea Heritage Service [17].

These 3D digital models accurately represent the forms of actual cultural heritage assets. They were imported into the Unreal Engine 5.4.4 environment, where frontal images were rendered from identical camera viewpoints under two conditions, daytime and evening. A total of 60 synthetic images were generated using 10 architectural models, including Gangnyeongjeon, Geonchunmun, Gwanghwamun, Gyeonghoeru, Gyotaejeon, Hamhwadang, Heungbokjeon, Heungryemun, Hyangwonjeong, Jagyeongjeon, and Unhyeongung. The rendered PNG files, created using the High Resolution Screenshot function in Unreal Engine, were converted to JPEG format with a quality setting of 90% using Python to ensure consistency in post-processing.



[Fig. 1] Synthetic Images Generated from Traditional Architectural 3D Models Provided by the Korea Heritage Service

The real-world dataset consisted of 58 images selected from publicly available resources on the National Heritage Portal, focusing on images in which the frontal structure of buildings could be clearly identified [18]. To minimize bias caused by differences in composition and perspective when compared with the synthetic images, the selected images were constrained to views as close to frontal as possible.

Both datasets were annotated according to an identical definition of structural components. The labeled categories included Roof, Pyeonaek, Column, Gate, Door, and Changho. All bounding boxes were manually annotated by the researcher using the LabelImg tool.

For object detection, YOLOv8 was employed, which is based on the latest anchor-free architecture in the YOLO family. The synthetic image dataset and the real-world image dataset were trained independently. Identical training conditions were applied to ensure a fair comparison. The main parameters included 100 epochs, a batch size of 4, and an input image size of 640. Model performance was evaluated using 30 real-world test images that were not included in training. Evaluation metrics included True Positive (TP), False Positive (FP), and False Negative (FN), as well as Precision, defined as the proportion of correctly detected objects among predicted objects, and Recall, defined as the proportion of correctly detected objects among existing ground-truth objects [19].

The performance comparison between the two models was conducted by directly contrasting the metrics obtained under the same evaluation procedure. This comparison was intended to provide a quantitative basis for examining performance differences between the synthetic dataset-based model and the real-world dataset-based model.

4. Results

This study compares and analyzes the object detection performance of two YOLOv8 models trained separately on a synthetic image dataset and a real-world image dataset. Both models were evaluated using the same set of 30 test images. Precision, Recall, and F1-score were calculated based on class-wise True Positive (TP), False Positive (FP), and False Negative (FN) values.

4.1 Performance of the Synthetic Image-Based Model

The performance results of the YOLOv8 model trained on synthetic images are presented in [Table 1]. The overall performance metrics were Precision 0.6517, Recall 0.3500, and F1-score 0.456. Precision was relatively maintained, whereas Recall showed a substantial decline. This result indicates that the model made comparatively accurate predictions for detected objects but failed to detect a large number of existing objects.

[Table 1] Detection Performance of the YOLOv8 Model Trained on Synthetic Images

Class	#GT	#Detected	TP	FP	FN	Precision	Recall
Roof	44	54	39	15	5	0.7222	0.8864
Pyeonaek	18	0	0	0	18	-	0
Column	163	72	65	7	98	0.9027	0.3988
Gate	7	4	1	3	6	0.25	0.1429
Door	27	45	9	36	18	0.2	0.3333
Changho	66	0	0	0	66	-	0
Total	325	175	114	61	211	0.6517	0.35

In the class-wise analysis, Roof achieved relatively high performance, with a Precision of 0.7222, a Recall of 0.8864, and an F1-score of 0.796. This outcome can be attributed to the large area and clearly defined geometric boundaries of roof structures.

Conversely, Gate and Door showed low Precision values of 0.25 and 0.20, respectively. Their Recall values also remained limited at 0.1429 and 0.3333. In the case of Door, both open and closed states were present, and this variation produced substantial changes in visual patterns despite representing the same structural component. As a result, detection performance decreased for components that rely heavily on internal pattern information. Gate also exhibited limited performance. Although its arched shape is a distinctive geometric feature, the visual context within bounding boxes varied considerably depending on whether the structure was open or closed.

Column recorded a high Precision of 0.9027, but Recall decreased to 0.3988 because of a large number of false negatives. This result suggests that repetitive arrangements and perspective variations caused some columns to visually merge with background structures.

Pyeonaek and Changho did not produce any correct detections, resulting in a Recall of 0. In particular, Changho, which consists of thin and repetitive line patterns, showed markedly low detection performance because object boundaries were not clearly distinguishable.

4.2 Performance of the Real-World Image-Based Model

The performance results of the YOLOv8 model trained on real-world images are summarized in [Table 2]. The overall Precision was 0.8033, Recall was 0.5971, and F1-score was 0.6874. Compared with the synthetic image-based model, both Recall and F1-score showed overall improvement. This outcome indicates that training with real-world visual diversity contributed to expanding the detection range of the model.

[Table 2] Detection Performance of the YOLOv8 Model Trained on Real-World Images

Class	#GT	#Detected	TP	FP	FN	Precision	Recall
Roof	44	54	38	16	6	0.7037	0.8636
Pyeonaek	18	10	9	1	9	0.9000	0.5000
Column	163	95	83	12	80	0.8737	0.508
Gate	7	7	7	0	0	1.0	1.0
Door	27	13	12	1	15	0.9231	0.4444
Changho	66	61	45	16	21	0.7377	0.682
Total	325	240	194	46	131	0.8033	0.5971

Roof maintained similar performance to the synthetic-based model, with a Precision of 0.7037, a Recall of 0.8636, and an F1-score of 0.7769. This result suggests that the global shape and clear structural boundaries of roofs remain stable features in real-world images. Gate achieved Precision and Recall values of 1.0, indicating that all ground-truth objects were correctly detected. This finding confirms that the arched structural characteristic was consistently captured under real-world conditions.

Changho also demonstrated balanced detection performance, with a Precision of 0.7377, a Recall of 0.682, and an F1-score of 0.709.

In contrast, Column maintained a high Precision of 0.8737, but Recall was limited to 0.508 because 80 false negatives were observed. This result can be attributed to illumination changes, shadow effects, partial occlusion, and visual interference from background structures in real environments, which obscured the boundaries of repetitive column structures. Door also showed a low Recall of 0.4444 despite a high Precision of 0.9231. This tendency appears to stem from visual variations caused by opening and closing states, as well as differences in material and color across individual doors. Pyeonaek exhibited some improvement compared with the synthetic dataset, yet Recall remained at 0.5 even in the real-world model. This result suggests that Pyeonaek remains a challenging class because of variability in typography, materials, proportions, and additional factors such as lighting conditions and surface aging in real environments.

4.3 Comparison between Synthetic Image-Based and Real-World Image-Based Models

A comparison between the synthetic image-based model and the real-world image-based model shows that the two models exhibited similar trends in terms of Precision, whereas clear differences were observed in Recall and F1-score. The synthetic image-based model achieved a Precision of 0.6517, a Recall of 0.3500, and an F1-score of 0.456. In contrast, the real-world image-based model recorded a Precision of 0.8033, a Recall of 0.5971, and an F1-score of 0.6874, indicating an overall improvement in detection performance.

In particular, the real-world model demonstrated reduced false negatives across multiple classes, resulting in higher Recall values. This outcome suggests that training with data reflecting real-world visual characteristics contributes to expanding the range of detectable objects.

5. Discussion

This study confirms that differences in visual characteristics between synthetic images captured from 3D models in Unreal Engine and real-world images have a meaningful impact on the object detection performance of YOLOv8. Across the overall metrics, the real-world image-based model achieved higher Recall and F1-score than the synthetic image-based model. This improvement can be interpreted as the result of incorporating diverse visual cues present in real environments, including surface textures, illumination variations, and signs of material aging. In contrast, the synthetic image-based model exhibited cases in which detection did not occur for certain classes or false negatives increased substantially. These results indicate limitations for structural components that cannot be sufficiently represented by shape information alone.

At the class level, Roof showed relatively high performance in both models. Because Roof occupies a large area and presents a clear global shape, it tended to be learned and detected in a stable manner even under conditions with limited texture or fine surface cues. In contrast, Gate experienced a pronounced performance degradation in the synthetic image-based model despite its distinctive arched geometry. This is likely because the visual context inside the bounding box varies significantly depending on whether the gate is open or closed. Even when the same arched boundary is present, changes in the internal background, such as passageways, walls, or shadows, hinder consistent feature learning. The perfect detection of Gate in the real-world image-based model suggests that environmental visual cues commonly accompanying arched structures in real scenes, including lighting, shadows, and spatial depth, may function as effective contextual information for identifying the presence of this object.

Door also exhibited high detection difficulty in both models. Substantial variation in internal patterns caused by opening and closing states, combined with differences in surrounding decoration, materials, and colors, contributed to this tendency. In the real-world image-based model, high Precision accompanied by low Recall indicates that the model detected only a limited subset of highly confident cases, leaving a considerable number of false negatives. For Column, the repetitive arrangement led to frequent overlap between adjacent objects, while perspective changes altered column width and contrast. These factors increased missed detections.

Changho and Pyeonaek represent classes with particularly low detection performance in the synthetic image-based model. Changho consists of thin, repetitive line patterns. In Unreal Engine renderings, anti-aliasing and resolution or sampling characteristics can cause such line structures to fragment or merge at the pixel level, resulting in unstable pattern cues. In real-world images, however, naturally occurring factors such as shadows, dust, wear, and uneven illumination may instead emphasize boundaries, leading to improved performance. Pyeonaek shows substantial variation in typography, proportions, and materials. In Unreal Engine captures, fine surface details such as engraved or embossed characters, wood grain, and paint diffusion tend to be simplified. As a result, edge and texture cues required by the model may not be sufficiently provided. This limitation likely contributed to frequent detection failures in the synthetic dataset and to the still limited Recall observed in the real-world dataset.

In summary, synthetic data enabled a certain level of detection performance for structural components dominated by clear global shape information, such as Roof. However, it showed limitations for components influenced by state changes, fine repetitive patterns such as Changho, or surface texture and textual information such as Pyeonaek. These findings suggest that synthetic data-based training can be effective for learning structural form, but additional consideration of diverse real-world visual conditions is necessary for practical deployment.

6. Conclusion

This study automatically detected six structural components in frontal images of traditional Korean architecture and compared the object detection performance of YOLOv8 models trained on Unreal Engine-based synthetic images and real-world images. Quantitative evaluation using the same test images confirmed that the visual characteristics of the training data directly influence detection performance.

The experimental results show that the synthetic image-based model achieved a certain level of performance for structural components with clear global shape information, such as Roof. However, detection performance was substantially limited for components affected by open-closed state variations, fine repetitive patterns such as Changho, or surface texture and textual information such as Pyeonaek. In contrast, the real-world image-based model learned diverse visual cues present in real-world environments environments, including illumination, shadows, and traces of aging. As a result, overall Recall and F1-score improved, and the presence of structural components was recognized more consistently.

These findings indicate that automatic recognition of structural components in traditional architecture depends not only on geometric form but also on visual context formed in real environments. Differences in detection characteristics across components further demonstrate that projection region segmentation and structure-based alignment required for projection mapping cannot be handled uniformly. Component-specific visual conditions must therefore be considered.

By quantitatively analyzing object detection performance for structural components in frontal images of traditional architecture, this study serves as a foundational validation step toward the automation of projection mapping systems. Through comparative analysis of detection feasibility and limitations across components, the study clarifies the visual requirements and technical constraints associated with projection region segmentation and structure-based alignment. The results can inform subsequent work on enhanced structure recognition using camera feedback, projection error correction, and automated projection mapping pipelines that incorporate structure-level content alignment.

Future research will address the identified limitations by expanding the range of target buildings and increasing sample size, while strengthening data composition to include environmental variations such as illumination, distance, viewing angle, and occlusion. In addition, learning strategies that combine synthetic and real-world data will be explored to improve robustness against domain differences. The study will further extend toward a projection mapping system that integrates automatic correction mechanisms informed by component-specific error patterns. Through these steps, future work will systematically evaluate the reliability of image-based structural recognition for projection region segmentation and alignment, as well as methods for mitigating recognition errors in real projection environments.

References

[1] S. Gehring, A. Wiethoff, “Interaction with Media Façades”, Informatik-Spektrum, vol. 37, no. 5, July 2014, pp. 474-482, doi: 10.1007/s00287-014-0818-0.

[2] J. S. Yu, J. H. Chung, “Study of the Image Design Components of Urban Landscape Media Facade”, The Korea Academia-Industrial cooperation Society, vol. 15, no. 11, November 2014, pp. 6478-6483, doi: 10.5762/KAIS.2014.15.11.6478.

[3] Y. J. Lee, “Palace Media Art Must Evolve to Become a Key Tourism Resource - Lee Chang-geun's Travel & Art”, edaily.co.kr, https://www.edaily.co.kr/News/Read?newsId=03493206638762640, (accessed September 28, 2025).

[4] K. B. Kim, “Analysis of 3D Graphics Utilization and Visual Composition in Projection Mapping”, Journal of Digital Art Engineering & Multimedia, vol. 12, no. 2, June 2025, pp. 149-160, doi: 10.29056/jdaem.2025.06.01.

[5] J. Deglint, A. Cameron, C. Scharfenberger, H. Sekkati, M. Lamm, A. Wong, D. A. Clausi, “Auto-calibration of a projector-camera stereo system for projection mapping”, Journal of the Society for Information Display, vol. 24, no. 8, August 2016, pp. 510-520, doi: 10.1002/jsid.464.

[6] F. Li, H. Sekkati, J. Deglint, C. Scharfenberger, M. Lamm, D. Clausi, J. Zelek, A. Wong, “Simultaneous projector-camera self-calibration for three-dimensional reconstruction and projection mapping”, IEEE Transactions on Computational Imaging, vol. 3, no. 1, March 2017, pp. 74-83, doi: 10.1109/TCI.2017.2652844.

[7] G. Zhang, Y. Pan, L. Zhang, “Deep learning for detecting building façade elements from images considering prior knowledge”, Automation in Construction, vol. 133, January 2022, pp. 104016, doi: 10.1016/j.autcon.2021.104016.

[8] J. Wei, Y. Hu, S. Zhang, S. Liu, “Irregular Facades: A Dataset for Semantic Segmentation of the Free Facade of Modern Buildings”, Buildings, vol. 14, no. 9, August 2024, pp. 2602, doi: 10.3390/buildings14092602.

[9] D. J. Kim, B. E. Lee, “Development of a Mapping Generator Based on Extended Reality Content”, Journal of Digital Art Engineering & Multimedia, vol. 15, no. 11, December 2022, pp. 441-450, doi: 10.29056/jdaem.2022.12.10.

[10] A. Grundhöfer, D. Iwai, “Recent advances in projection mapping algorithms, hardware and applications”, Computer graphics forum, vol. 37, no. 2, May 2018, pp. 653-675, doi: 10.1111/cgf.13387.

[11] D. Iwai, “Projection mapping technologies: A review of current trends and future directions”, Proceedings of the Japan Academy, Series B, vol. 100, no. 3, March 2024, pp. 234-251, doi: 10.2183/pjab.100.012.

[12] M. Sugimoto, D. Iwai, K. Ishida, P. Punpongsanon, K. Sato, “Directionally decomposing structured light for projector calibration”, IEEE Transactions on Visualization and Computer Graphics, vol, 27, no. 11, August 2021, pp. 4161-4170, doi: 10.1109/TVCG.2021.3106511.

[13] M. B. Starzyńska-Grześ, R. Roussel, S. Jacoby, A. Asadipour, “Computer vision-based analysis of buildings and built environments: A systematic review of current approaches”, ACM Computing Surveys, vol. 55, no. 13s, July 2023, pp. 1-24, doi: 10.1145/3578552.

[14] H. Schieber, K. C. Demir, C. Kleinbeck, S. H. Yang, D. Roth, “Indoor Synthetic Data Generation: A Systematic Review”, Computer Vision and Image Understanding, vol. 240, March 2024, pp. 03907, doi: 10.1016/j.cviu.2023.103907.

[15] J. E. Yu, Y. H. Kim, “Unreal Engine Meets Computer Vision: Technological Evolution and Trends”, Journal of Digital Media & Culture Technology, vol. 4, no. 2, December 2024, pp. 163-176, doi: 10.29056/jdmct.2024.12.07.

[16] K. Man, J. Chahl, “A review of synthetic image data and its use in computer vision”, Journal of Imaging, vol. 8, no. 11, November 2022, pp. 1-33, doi: 10.3390/jimaging8110310.

[17] KOREA HERITAGE SERVICE, “KOREA HERITAGE SERVICE Fab”, fab.com, https://www.fab.com/sellers/KOREA%20HERITAGE%20SERVICE, (accessed September 28, 2025).

[18] Korea National Heritage Portal, “Korea National Heritage Portal”, www.heritage.go.kr, https://www.heritage.go.kr/main/?v=1765519483507, (accessed September 28, 2025).

[19] L. A. Nguyen, M. D. Tran and Y. S. Son, “Empirical Evaluation and Analysis of YOLO Models in Smart Transportation”, Empirical evaluation and analysis of yolo models in smart transportation, vol. 5, no. 4, November 2024, pp. 2518-2537, doi: 10.3390/ai5040122.