A Local Rank-Based Calibration and Graph-Cut Refinement Framework for Building Change Detection

Chong Lee [1], Inhyeok Lee [2], Jangwoo Cheon [2], Bui Ngoc An [1], Juhee Lee [1], Impyeong Lee [1]

[1] Dept. of Geoinformatics, University of Seoul, 02504 Seoul, Republic of Korea – (dlchd22, anbn, soar21, iplee)@uos.ac.kr
[2] InnoPAM, Geospatial Team, 04366 Seoul, Republic of Korea – (ihlee, cjw)@innopam.com

Keywords: Building segmentation, Building change detection, Local Rank-Based Prior Calibration, Graph-Cut Refinement, Aerial imagery, Post-processing

Abstract

Accurate building change detection depends on how well building boundaries are delineated, as distortions and merging errors hinder reliable correspondence. In dense urban areas, deep learning models frequently merge adjacent buildings—especially within narrow gaps—producing structural inconsistencies that lead to change detection errors. We propose a post-processing method integrating Local Rank-Based Prior Calibration, which reinterprets Softmax probabilities as percentile-based local ranks, with Graph-Cut refinement for structural correction. The refined mask is matched with historical building data to classify four change types. Experiments using aerial imagery from Seoul show that the method reduces structural errors, lowering under-segmentation from 51.64% to 22.02% and improving Intersection over Union (IoU) from 0.748 to 0.759. In change detection, it increases the mean F1-score from 0.522 to 0.608 and improves all classes, including new construction, whose F1-score rises from 0.269 to 0.707. Ablation studies confirm that calibration and graph-based refinement both contribute to the improvements. These results show that stabilizing segmentation outputs enhances the reliability of building-level change detection in dense urban environments.

1. Introduction

The physical environment of cities is rapidly reshaping due to technological advancements and social changes (Wu et al., 2022). These transformations are particularly evident at the building level among various urban spatial elements (AURUM, 2022). National agencies, public institutions, and private enterprises periodically construct building datasets to reflect these changes (NGII, 2025). However, these datasets often suffer from long update intervals, reducing temporal currency due to their episodic acquisition cycles (Hartmann et al., 2016). Consequently, change detection approaches are increasingly adopted to mitigate these limitations by deriving building footprints from recent aerial imagery and comparing them with historical data.

Change detection determines whether a building remains unchanged, has newly appeared, or has been demolished by analyzing correspondence between building footprints across two temporal instances. The reliability of this correspondence depends on how precisely segmentation results at each time point delineate true building boundaries and structural forms. However, pixel-based metrics often fail to sufficiently capture structural errors. Myint et al. (2011) noted that metrics such as Intersection over Union (IoU) may not adequately reflect discrepancies when building merging occurs, potentially leading to an overestimation of performance.

Existing change detection studies have not sufficiently addressed the need to calibrate the quality of building segmentation used as input. Furthermore, deep learning–based building segmentation often produces merging errors between adjacent structures, which in turn propagate as additional inaccuracies during change detection. Post-processing techniques based on morphology, superpixels, and watershed algorithms have been utilized for boundary refinement (Neto and Dantas, 2024; Achanta et al., 2012; Wu and Li, 2022). Nevertheless, resolving merging errors that occur within narrow inter-building gaps remains fundamentally challenging. Model improvements incorporating boundary-aware designs and multi-scale receptive fields have strengthened boundary recognition (Wang et al., 2025; Ma et al., 2023). However, these models remain limited in reliably separating adjacent objects in dense urban areas. Techniques based on CRF and Graph Cut have improved accuracy near boundaries (Boykov and Jolly, 2001; Chen et al., 2018). However, merging issues in dense regions have continued due to the smoothing effect. Indistinct boundaries between adjacent buildings or residual object merging (Chen et al., 2023) still occur. These issues degrade the reliability of change detection.

To address these limitations, this study applies a rank-based post-processing method. This method reinterprets Softmax probabilities as percentile ranks within the local distribution (Lee et al., 2025). This approach evaluates pixels within narrow gaps based on their local rank within a window rather than their absolute values, thereby reducing incorrect foreground classification. By inputting the calibrated Softmax map into the Graph Cut algorithm, object separation is maintained even during the global optimization process. Consequently, this method mitigates merging errors and enhances object-level boundary clarity. These improvements ultimately increase the accuracy of correspondence formation and the stability of results in change detection.

This study aims to demonstrate how the proposed post-processing method improves the accuracy and reliability of building change detection compared with existing approaches.

2. Methodology

2.1 Overview

The proposed pipeline accepts the Softmax output from a segmentation model as input. A post-processing module is then

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-4-2026-95-2026 | © Author(s) 2026. CC BY 4.0 License.

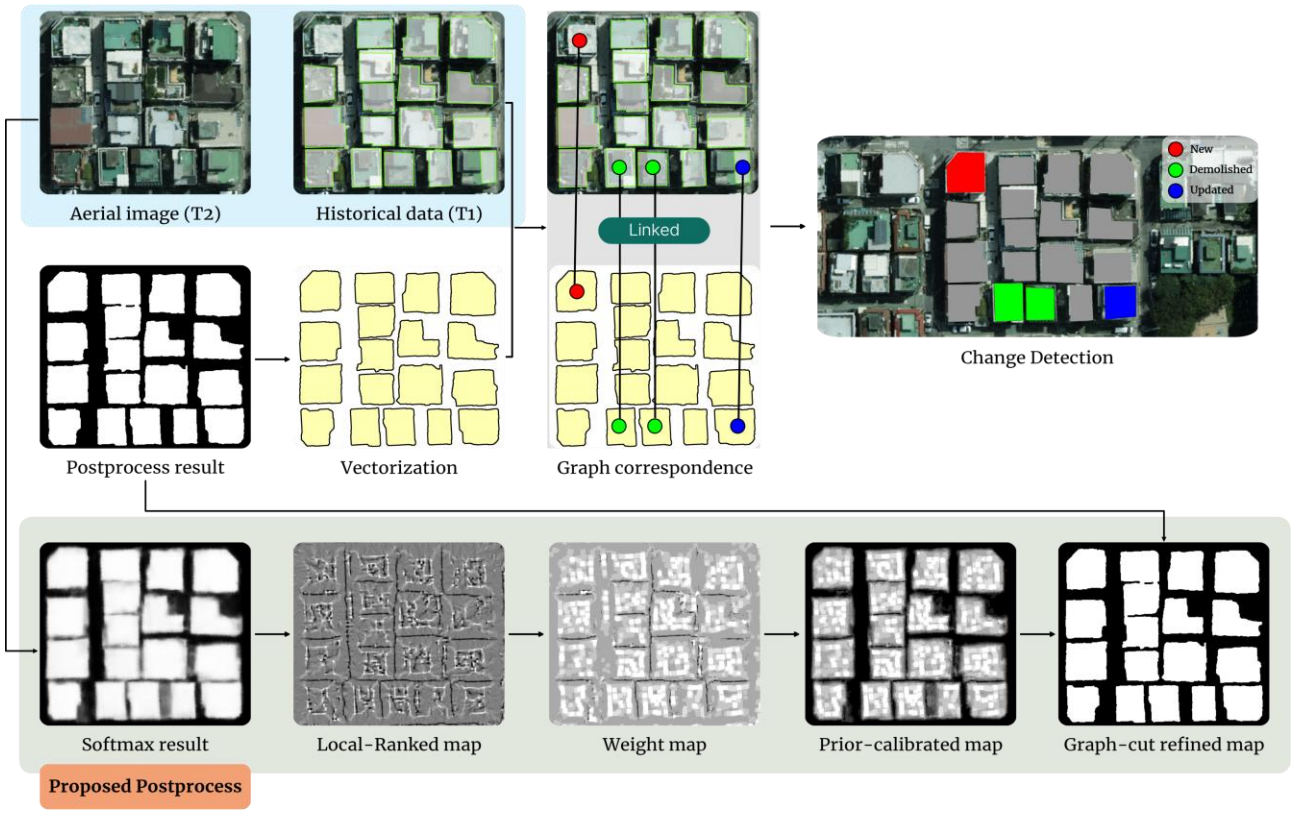

Figure 1. Pipeline overview illustrating the impact of the proposed post-processing module

applied, and the resulting building mask is matched with historical building data to determine change types. The post-processing stage, detailed in Figure 1, comprises local rank-based probability calibration and Graph-Cut-based binarization. This approach effectively reduces merging errors between adjacent buildings and enhances boundary quality. The generated building mask for the current epoch is matched with historical data and classified into categories: new, demolished, updated, or unchanged. The overall architecture is decoupled from the model training phase. It is designed to be modularly integrated with various segmentation models.

2.2 Local rank-based prior calibration and graph cut

The proposed post-processing technique takes the Softmax probability map from a segmentation model as input and integrates Local Rank-Based Prior Calibration with global optimization via Graph-Cut. This process is performed to mitigate the merging issues between adjacent buildings that frequently occur in simple threshold-based binarization. It involves transforming the probability s(p) of each pixel p into a local rank within a local w x w window. This definition is presented in Eq. 1. The percentile is calculated as the proportion of all pixels q within the window that satisfy the condition $s(q)$ is less than or equal to $s(p)$, where p denotes the target pixel, q represents neighbouring pixels, and s(.) signifies the Softmax probability. A non-linear weighting function is applied to the calculated percentile to derive the calibration coefficient w(p). The decision to skip calibration is determined based on the threshold δ, as defined in Eq. 2. Consequently, the Softmax probability $s(p)$ is combined with the calibration coefficient $w(p)$to generate the calibrated probability $s_{cal}(p)$, as formulated in Eq. 3. In this framework, w, α, and δ denote the window size, the percentile-based decay strength, and the calibration skip threshold, respectively; these parameters control the scope of local statistical reflection, the degree of suppression near boundaries, and the stability of internal regions.

$$\text{percentile}(p) = \frac{|\{\, q \in \text{window}(p) \mid s(q) \le s(p) \,\}|}{w^2}, \quad (1)$$

$$w_p = \begin{cases} s(p), & \text{if } |\bar{s}_{10\%} - s(p)| < \delta \\ (\text{percentile}(p))^{\alpha}, & \text{otherwise} \end{cases}, \quad (2)$$

$$s_{cal}(p) = w_p \cdot s(p), \quad (3)$$

The calibrated probability map is subsequently utilized as input for Graph-Cut optimization, which separates the foreground from the background by minimizing an energy function composed of a data term and a boundary term. The structure of this function is presented in Eq. 4. The smoothing intensity of the boundary term is controlled by λ; a larger value simplifies the boundaries, while a smaller value preserves detailed structures. Ultimately, the binarization process, which integrates percentile-based calibration with Graph-Cut optimization, enables more precise segmentation of boundaries between building objects.

$$E(L) = \sum_{p} R(p, l_p) + \sum_{(p,q)\in N} V(p,q), \quad (4)$$

2.3 Change detection

The change detection procedure in this study uses building segmentation results from current aerial imagery as input. Vector building data from the past time point is used as the reference. Polygons from the two time points are matched through spatial overlap analysis. Correspondence between objects is then

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-4-2026-95-2026

established according to their degree of overlap. Depending on the combination of object counts across the two epochs, these correspondence relationships are categorized into six types: 0:1, 1:0, 1:1, 1:N, N:1, and N:N. The 0:1 type corresponds to buildings that newly appear in the current time point and is classified as new construction. Conversely, the 1:0 type represents buildings that existed previously but are absent in the current data and is classified as demolition.

For 1:1 correspondence, if the overlap ratio exceeds a predefined threshold, the object is considered identical and classified as no change. If the ratio falls below the threshold, it is treated as an update, reflecting morphological changes or reconstruction. The 1:N, N:1, and N:N types represent structural alterations such as splitting, merging, or complex transitions and are all grouped into the update category. Ultimately, all objects are classified into four final categories: new construction, demolition, update, and no change. To ensure a fair comparison of the post-processing module's performance, this object-based change detection procedure was applied uniformly across all experiments.

3. Experiments

3.1 Study Area and Dataset

The study area for this research is Mangubon-dong in Jungnang-gu, Seoul. The region contains a diverse mixture of building ages and land-use types and is undergoing active urban transformation. These characteristics provide an optimal environment for validating the performance of the proposed building change analysis framework. The data for the current epoch consists of aerial orthophotos acquired in 2022, from which relief displacement and parallax errors have been eliminated through rigorous orthorectification. Because building footprints in orthophotos are geometrically aligned with the ground surface, this imagery improves boundary extraction accuracy during segmentation. The spatial location of the study area and the urban structure derived from the orthophoto are illustrated in Figure 2. The reference data for the past epoch is the 2020 digital building map provided by the National Geographic Information Institute (NGII). This dataset constitutes the national standard building information verified through aerial photo interpretation and quality inspection. In this study, the dataset was used as the historical baseline for comparison with the building footprints of the current epoch.

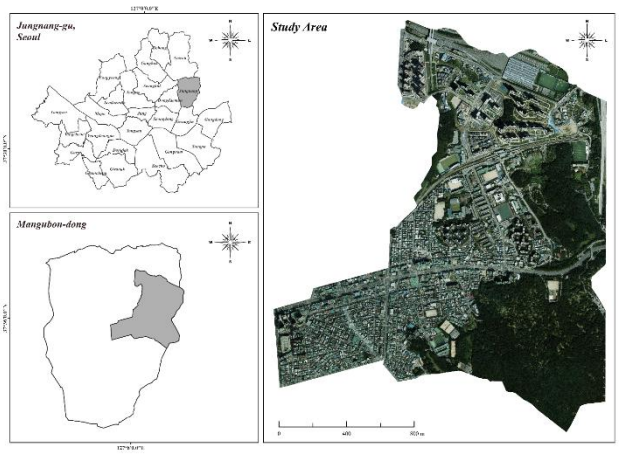

Figure 2. Mangubon-dong, Jungnang-gu, Seoul.

3.2 Ground Truth

To validate the change detection results, the digital building map provided by the National Geographic Information Institute (NGII) was utilized as the ground truth. Building polygons from 2020 and 2022 were registered to a unified coordinate system, and errors such as omissions and duplications were corrected through rigorous manual inspection. The registered datasets were used to establish spatial correspondence between polygons. Based on this correspondence, object-level change types were classified into new construction, demolition, update, and no change. Consequently, the compiled change type dataset served as the reference baseline for the quantitative evaluation of the change detection performance.

Model	Dataset	
	Alabama[a)]	Inria[b)]
DeepLabV3+	79.92	74.09
HRNet	74.15	74.47
SegFormer [c)]	76.1	77.84
Swin	75.7	77.85
Twin	74.71	76.14
Unet	72.13	72.79
ViT	60.22	65.8

Table 1. Comparison of building segmentation performance across different models on the Alabama and Inria datasets. Data sources: a) Microsoft (2018), b) Maggiori et al. (2017). c) Xie et al. (2021)

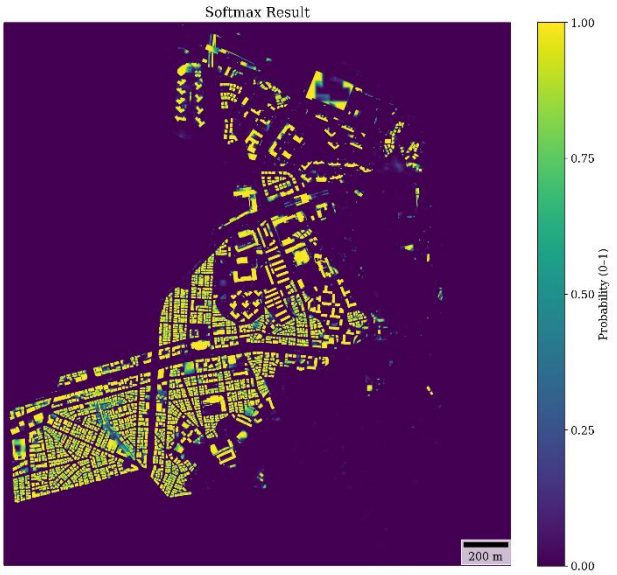

Figure 3. Spatial distribution of building-class confidence predicted by the SegFormer model.

3.3 Model and Data Preparation

In this study, the Transformer-based model SegFormer was employed for building segmentation. SegFormer has been reported to demonstrate stable performance across diverse urban scenes, a consistency that is further corroborated in Table 1. Pre-training was conducted using the Alabama and Inria datasets, which were partitioned into approximately 3,000 tiles of 512 x 512 pixels. Subsequently, fine-tuning was performed utilizing domestic UAV imagery. During the inference phase, input images were segmented into 512 x 512 tiles, with a 50% overlap

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-4-2026-95-2026 | © Author(s) 2026. CC BY 4.0 License.

applied to mitigate boundary uncertainty. The tile-level Softmax outputs were integrated into a continuous probability map covering the entire study area. Figure 3 presents the visualization results of this probability-based confidence distribution. The finally generated Softmax probability map serves as the input for the post-processing stage.

3.4 Evaluation Metrics

In this study, separate evaluation frameworks were applied to reflect the different objectives of the post-processing and change detection stages. For the post-processing stage, both pixel-level and object-level evaluations were utilized to verify segmentation accuracy and structural errors. Pixel-level evaluation employs Precision (Eq. 5), Recall (Eq. 6), F1-score (Eq. 7), and Intersection over Union (IoU) (Eq. 8). In this evaluation, True Positives (TP), False Positives (FP), and False Negatives (FN) are defined based on the agreement between the prediction and the ground truth.

$$Precision = \frac{TP}{TP + FP}, \tag{5}$$

$$Recall = \frac{TP}{TP + FN}, \tag{6}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}, \tag{7}$$

$$IoU = \frac{TP}{TP + FP + FN}, \tag{8}$$

Precision and Recall assess prediction accuracy and the extent of missed detections, respectively. The F1-score evaluates the balanced performance of these two indicators, while IoU measures the degree of spatial overlap with the reference data. Object-level evaluation utilizes Average Recall (AR), Average Precision (AP), Under-Segmentation Rate (USR), and Over-Segmentation Rate (OSR) to assess the separation accuracy and structural errors of building objects. In the change detection stage, only object-level evaluation is applied to calculate Precision, Recall, and F1-score for each change type. Here, TP, FP, and FN are defined based on the agreement between the predicted class and the reference class. Recall evaluates the detection performance for changed objects, while Precision assesses whether unchanged objects are incorrectly classified as changed.

4. Result and Discussion

4.1 Post-processing Validation

The proposed post-processing technique was evaluated both quantitatively and qualitatively against threshold-based binarization, a baseline model, and graph-based post-processing. As detailed in the analysis based on Table 2, the proposed method demonstrated the lowest structural error rates, achieving a USR of 22.02% and an OSR of 0.30%. This performance effectively suppressed merging and over-segmentation in areas adjacent to boundaries. Unlike threshold-based approaches, which showed large performance variation depending on parameter settings, the proposed technique mitigated boundary-probability instability through local rank-based calibration.

This mechanism ensured stable object separation even within narrow gaps between buildings. Furthermore, AR and AP metrics indicated stable object correspondence performance across diverse conditions. Improvements in boundary quality were also evident in pixel-level metrics, with an IoU of 0.7591 and an F1-score of 0.8631. Collectively, the proposed technique delivered consistent enhancements in reducing structural errors and improving boundary separation performance.

Figure 4 presents comparative results regarding the distributions of TP, FP, and FN for each post-processing technique. In the Baseline method, FP bridges narrow gaps between two buildings, resulting in merging errors.

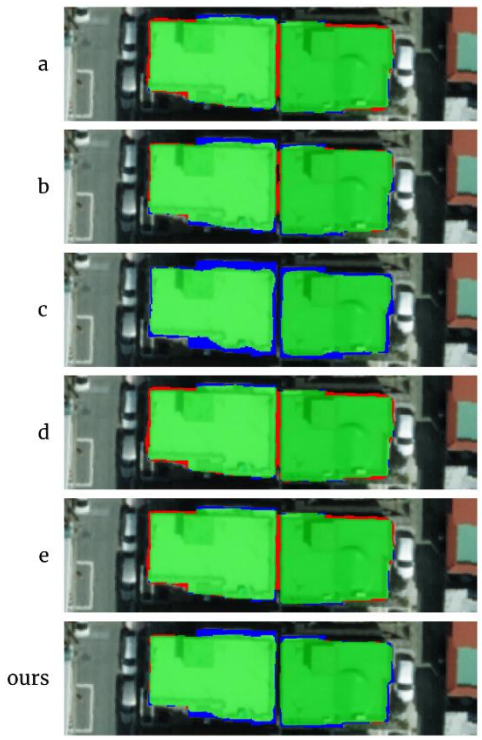

Figure 4. Qualitative comparison of post-processing methods. Green represents true positives, red represents false positives, and blue represents false negatives. (a) Baseline, (b) thr0.7, (c) thr0.9, (d) Otsu, (e) Grid-CRF, and (f) Proposed method.

method	Objects	US	USR	OS	OSR	AR	AP	IoU	Rec	Pre	F1
thr@0.7	2688	755	28.09	12	0.45	0.4811	0.5849	0.7545	0.8512	0.8691	0.86
thr@0.9	2688	160	5.95	33	1.23	0.5274	0.6159	0.7018	0.7433	0.9263	0.8248
otsu	2688	1486	55.28	11	0.41	0.2882	0.4563	0.7436	0.8963	0.8136	0.853
grid-crf	2688	1425	53.01	9	0.33	0.3069	0.4731	0.7465	0.8902	0.8222	0.8548
Baseline	2688	1388	51.64	11	0.41	0.3142	0.4721	0.7475	0.8907	0.8231	0.8555
ours	2688	592	22.02	8	0.3	0.517	0.6006	0.7591	0.8467	0.8801	0.8631

Table 2. Quantitative comparison of structural and detection metrics for all post-processing methods (US: under-segmented objects, OS: over-segmented objects)

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-4-2026-95-2026 | © Author(s) 2026. CC BY 4.0 License.

For the thr@0.9 method, the elevated threshold leads to a substantial increase in FN, causing prominent detection losses across the entire extent of both buildings. The Otsu method shows increased FN near boundaries and the simultaneous appearance of FP in some areas, producing an unstable boundary representation. In the case of Grid-CRF, FP are still observed within the gaps between buildings, indicating that merging errors are not completely resolved. In contrast, the proposed method (Ours) demonstrates that FP rarely appears in the inter-building gaps, while FN are maintained at a minimal level along the boundary lines.

The proposed methods depend on the quality of the initial segmentation and require additional computation. Nevertheless, they consistently suppress misclassification near boundaries. This approach secures a balance between Precision and Recall while reducing under-segmentation at the object level. By clarifying the boundaries between buildings, the method mitigates structural errors inherent in existing techniques. Consequently, the method alleviates both the Precision–Recall imbalance and the boundary distortion observed in threshold-based approaches. As a result, it satisfies the reliability and consistency requirements for change detection preprocessing.

4.2 Impact on Change Detection

Change detection performance was evaluated by comparing segmentation results from the Baseline and existing post-processing techniques using the same input imagery with historical building data. As shown in Table 3, the proposed method achieved a Precision of 0.651, Recall of 0.774, and F1-score of 0.707 for new constructions. This reflects reductions in both false positives and false negatives compared with conventional methods. The Baseline showed degraded performance due to a tendency to extract existing and newly constructed buildings as a single object. The thr@0.9 method also produced more missed detections for new constructions because of the elevated threshold. In the case of Grid-CRF, excessive boundary smoothing failed to improve the detection accuracy of new constructions. For the demolition class, the proposed method secured a Precision of 0.322 and Recall of 0.583 by reducing false positives. For the update class, stable object shapes in the current imagery enabled the identification of partial structural changes, yielding a Recall of 0.912. Furthermore, for the no-change class, the boundaries of existing buildings were stably preserved, yielding an F1-score of 0.795.

As summarized in Table 4**,** the Full Model includes both Prior Calibration and Graph-based Postprocessing and represents the performance achieved when the entire change detection procedure is applied. Ablation results demonstrated that the Full Model attained the highest performance, recording an MR of 0.733, an MP of 0.5208, and an MF of 0.6078.

Method	Type	GT	Pred	TP	FN	FP	R	P	F	MR	MP	MF
Baseline										0.639	0.4395	0.522
	new	159	674	112	47	562	0.704	0.166	0.269			
	demolished	48	47	25	23	22	0.521	0.532	0.526			
	updated	195	2423	108	5	1529	0.956	0.066	0.123			
	unchanged	4894	1848	918	1529	6	0.375	0.994	0.545			
thr 0.7										0.7245	0.4335	0.5432
	new	159	634	124	35	510	0.78	0.196	0.313			
	demolished	48	63	28	20	35	0.583	0.444	0.505			
	updated	195	1834	106	7	970	0.938	0.099	0.178			
	unchanged	4894	2938	1461	986	8	0.597	0.995	0.746			
thr 0.9										0.7458	0.4048	0.5251
	new	159	475	118	41	357	0.742	0.248	0.372			
	demolished	48	132	32	16	100	0.667	0.242	0.356			
	updated	195	1707	97	16	612	0.858	0.137	0.236			
	unchanged	4894	3534	1752	695	15	0.716	0.992	0.832			
otsu										0.6305	0.445	0.5221
	new	159	711	111	48	600	0.698	0.156	0.255			
	demolished	48	44	25	23	19	0.521	0.568	0.543			
	updated	195	2513	109	4	1621	0.965	0.063	0.118			
	unchanged	4894	1668	828	1619	6	0.338	0.993	0.505			
grid-crf										0.6498	0.4665	0.5431
	new	159	370	111	48	259	0.698	0.3	0.42			
	demolished	48	55	28	20	27	0.583	0.509	0.544			
	updated	195	2319	107	6	1531	0.947	0.065	0.122			
	unchanged	4894	1830	908	1539	7	0.371	0.992	0.54			
ours										0.733	0.5208	0.6078
	new	159	189	123	36	66	0.774	0.651	0.707			
	demolished	48	87	28	20	59	0.583	0.322	0.415			
	updated	195	1413	103	10	787	0.912	0.116	0.205			
	unchanged	4894	3262	1622	825	9	0.663	0.994	0.795			

Table 3. Comparison of Change Detection Performance Using Segmentation Outputs Generated by Different Postprocessing Methods

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-4-2026-95-2026

When Prior Calibration was removed, MF decreased to 0.5415, confirming that rank-based probability calibration contributes to improving the F1-score in change detection. This improvement stems from enhanced stability in object correspondence, achieved by mitigating local distortions in Softmax probabilities within complex building layouts or narrow gaps. Conversely, removing Postprocessing caused declines in both MR and MP, reducing MF to 0.522. This indicates that structural correction through boundary refinement is essential for reliable correspondence. Ultimately, the Full Model, which integrates both components, was found to maintain the most stable balance between Recall and Precision, significantly enhancing the overall F1-score.

Configuration	MR	MP	MF
Full Model	**0.733**	**0.5208**	**0.6078**
w/o Prior Calibration	0.668	0.4538	0.5415
w/o Postprocessing	0.639	0.4395	0.522

Table 4. Ablation results for Prior Calibration and Postprocessing

As illustrated in Figure 5, for the new case, the proposed technique clearly segmented buildings in the current epoch. Consequently, it reliably identified objects absent in the historical data as new constructions. In the demolished case, misclassification of the background was effectively suppressed. As a result, regions that existed in the past but were not detected in the current epoch were consistently classified as demolition For the updated case, segmentation reflected partial structural changes. Although correspondence with historical data was maintained, the altered morphologies were clearly revealed. With object matching preserved, variations in boundaries and structures were distinctly represented, enabling accurate classification of the update type. Finally, in the unchanged case, buildings in the current epoch appeared nearly identical to those in the past, resulting in the formation of stable object correspondences.

The change detection procedure in this study was designed with a simplified structure that directly establishes correspondence between historical building data and current segmentation results. This configuration aims to rigorously validate the impact of segmentation quality on change detection performance. A significant observation is that even within such a simple procedural framework, performance varies considerably depending on the sophistication of the post-processing. Existing change detection methods typically rely on overlap ratios or simple correspondence rules. As a result, their reliability is easily compromised when merging or boundary errors occur during segmentation. Prior Calibration and Graph-based Postprocessing enhanced the structural stability of the segmentation. As a result, the classification accuracy for all change types was significantly improved.

5. Conclusion

In this study, Prior Calibration and Graph-based Postprocessing were applied to mitigate segmentation errors. The resulting outputs were then used as inputs for change detection. The proposed technique demonstrated an improvement in overall change detection performance compared to existing methods, with increases of 0.094 in MR, 0.0813 in MP, and 0.0858 in MF. Performance enhancements were observed across all change types, as Prior Calibration corrected local probabilities and Graph-based Postprocessing refined boundaries and structures. These results imply that even with the advancement of change detection techniques, prioritizing stable and accurate building segmentation remains essential. This study confirmed that the reliability of change detection and the accuracy of object correspondence can be significantly elevated solely through effective post-processing. Future work will focus on further refining the criteria for discriminating change types based on object-level change patterns, such as variations in boundary, shape, and area.

Acknowledgements

This work is supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant R&D program of Digital Land Information Technology Development funded by the Ministry of Land, Infrastructure and Transport (MOLIT) (Grant RS-2022-00142501).

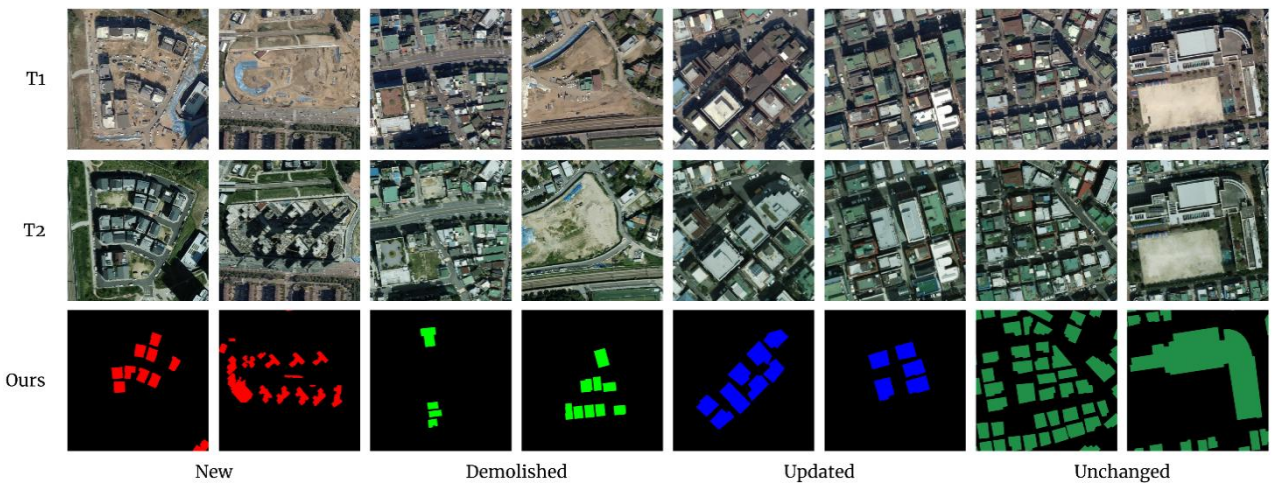

Figure 5. Qualitative examples of change detection results using the proposed segmentation and postprocessing framework across four change types (new, demolished, updated, unchanged).

References

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S., 2012. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11), pp. 2274–2282. https://doi.org/10.1109/TPAMI.2012.120

AURUM, 2022. auri, url: https://www.aurum.re.kr/Research/PostView.aspx?mm=1&ss=1&pid=23487, accessed on 17/11/2025.

Boykov, Y. Y., Jolly, M.-P., 2001. Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. In *Proceedings of the 2001 Eighth IEEE International Conference on Computer Vision (ICCV)*, Vancouver, BC, Canada, July 7–14, pp. 105–112. https://doi.org/10.1109/ICCV.2001.937505

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A. L., 2018. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4), pp. 834–848. https://doi.org/10.1109/TPAMI.2017.2699184

Chen, S., Ogawa, Y., Zhao, C., Sekimoto, Y., 2023. Large-scale individual building extraction from open-source satellite imagery via super-resolution-based instance segmentation approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 195, pp. 129–152. https://doi.org/10.1016/j.isprsjprs.2022.11.006

Hartmann, A., Meinel, G., Hecht, R., Behnisch, M., 2016. A Workflow for Automatic Quantification of Structure and Dynamic of the German Building Stock Using Official Spatial Data. *ISPRS International Journal of Geo-Information*, 5(8), 142. https://doi.org/10.3390/ijgi5080142

Lee, C., Lee, I., Cheon, J., An, B. N., Lee, J., Lee, I., 2025. Building segmentation refinement via local rank-based calibration and graph cut. *Korean Journal of Remote Sensing*, 41(5), pp. 813–828. https://doi.org/10.7780/kjrs.2025.41.5.9

Ma, X., Ma, M., Hu, C., Song, Z., Zhao, Z., Feng, T., et al., 2023. LoG-CAN: Local-global class-aware network for semantic segmentation of remote sensing images. *arXiv preprint* arXiv:2303.07747. https://doi.org/10.48550/arXiv.2303.07747

Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P., 2017. Can semantic labeling methods generalize to any city? The Inria aerial image labeling benchmark. In *Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Fort Worth, TX, USA, July 23–28, pp. 3226–3229. https://doi.org/10.1109/IGARSS.2017.8127684

Microsoft, 2018. US building footprints, url: https://github.com/Microsoft/USBuildingFootprints, accessed on 29/09/2025.

Myint, S. W., Gober, P., Brazel, A., Grossman-Clarke, S., Weng, Q., 2011. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sensing of Environment*, 115(5), pp. 1145–1161. https://doi.org/10.1016/j.rse.2010.12.017

NGII, 2025. Map production, url: https://www.ngii.go.kr/kor/content.do?sq=207, accessed on 17/11/2025.

Neto, A., Dantas, D., 2024. Building damage segmentation after natural disasters in satellite imagery with mathematical morphology and convolutional neural networks. In *Proceedings of the 2024 26th International Conference on Enterprise Information Systems*, Angers, France, Apr. 28–30, pp. 828–836. https://doi.org/10.5220/0012706300003690

Wang, J., Chen, T., Zheng, L., Tie, J., Zhang, Y., Chen, P., et al., 2025. A multi-scale remote sensing semantic segmentation model with boundary enhancement based on UNetFormer. *Scientific Reports*, 15, 14737. https://doi.org/10.1038/s41598-025-99663-9

Wu, Y., Li, Q., 2022. The algorithm of watershed color image segmentation based on morphological gradient. *Sensors*, 22(21), 8202. https://doi.org/10.3390/s22218218202

Wu, Y., Wu, S., Qiu, X., Wang, S., Yao, S., Li, W., Xia, S., 2022. Integrated evaluation method of the health-related physical environment in urbanizing areas: A case study from a university campus in China. *Frontiers in Public Health*, 10, 801023. https://doi.org/10.3389/fpubh.2022.801023

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., Luo, P., 2021. SegFormer: Simple and efficient design for semantic segmentation with transformers. *arXiv preprint* arXiv:2105.15203. https://doi.org/10.48550/arXiv.2105.15203