Automated Geometric Correction of OpenStreetMap Buildings via Context- and Boundary-Aware Segmentation

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

[1] Geospatial Team, InnoPAM, Seoul, Republic of Korea – (ihlee, cjw)@innopam.com

[2] Department of Geoinformatics, University of Seoul, Seoul, Republic of Korea – (dlchd22, anbn, soar21, iplee)@uos.ac.kr

Keywords: OpenStreetMap, Boundary refinement, Building segmentation, SAM 2, SegFormer, True orthoimage

Abstract

OpenStreetMap (OSM) is a representative open geospatial platform that provides free access to major spatial objects, including buildings worldwide, constructed through crowdsourcing-based manual digitization. However, subjective differences among contributors and the absence of unified quality control standards have led to the accumulation of positional offsets and boundary shape errors in building polygons. To address this issue, studies using deep learning-based semantic segmentation for OSM quality improvement have been conducted. Nevertheless, Transformer-based segmentation models exhibit an under-segmentation tendency that merges adjacent buildings into a single object, along with limitations in precise boundary delineation. To overcome these challenges, this study proposes a two-stage framework that integrates SegFormer, which excels in global context recognition, with SAM 2, which is capable of precise boundary segmentation. In the first stage, SegFormer semantically segments building regions from a true orthoimage, and in the second stage, SAM 2 infers object-level precise boundaries using the bounding boxes of OSM polygons as box prompts. The two results are combined into a prior probability map, enabling uncertain boundary regions to be re-evaluated in an unsupervised manner. In experiments conducted over the Suseo-dong area in Gangnam-gu, Seoul, the proposed method achieved a BIoU of 70.40%, an improvement of 23.85 percentage points over OSM building data, with consistent performance gains across all evaluation metrics. This framework offers scalability applicable to any region worldwide without additional label construction, provided that high-resolution true orthoimagery and OSM data are available.

1. Introduction

OpenStreetMap (OSM) is a representative open geographic information platform that constructs major spatial objects worldwide—including buildings, roads, and land use—as vector data through crowdsourcing-based manual digitization and provides them free of charge (OpenStreetMap contributors, 2022). Currently, millions of volunteer contributors participate in manual digitization by interpreting imagery, and building data alone comprises over 500 million polygons globally (Biljecki et al., 2023). Given this openness and broad spatial coverage, OSM data serve as a key spatial resource across diverse fields, including disaster response and humanitarian mapping, urban spatial analysis, and training data construction for AI models.

However, the manual digitization approach based on voluntary participation by numerous contributors involves structural limitations: regional disparities in contributor engagement and the absence of unified quality control standards. Boundary shape inconsistencies arising from subjective differences among contributors accumulate throughout the data, frequently resulting in reduced positional accuracy and shape precision of building polygons (Brovelli et al., 2016; Moradi et al., 2023). Furthermore, when OSM data containing such errors are used as training labels for AI models, shape errors can be directly propagated into the learning process, undermining the reliability of performance evaluation.

To address these issues, studies using deep learning-based approaches for OSM quality improvement have been actively conducted. Zhuo et al. (2018) proposed a method that extracts building boundaries through deep learning-based semantic segmentation of UAV imagery and uses them as constraints to automatically correct positional and shape errors in OSM building polygons. Positional errors of OSM building boundaries, which previously reached several meters, were reduced to approximately 10 cm after correction, and robustness was validated across multiple datasets covering various building types. Meanwhile, Zhang et al. (2020) proposed an automated framework that applies deep learning not only to shape error correction but also to the detection of omission and commission errors. The method uses a pretrained FCN as a feature extractor, treats OSM as noisy labels to train a Random Forest model, and refines the results through superpixel-based graph cuts. However, these studies rely on CNN-based architectures, which have structural limitations in capturing global context and provide insufficient boundary delineation precision in complex urban environments.

Transformer-based architectures have overcome the limitations of CNN-based models in global context understanding and improved building boundary detection performance (Xie et al., 2021). However, in complex urban environments, they exhibit an under-segmentation tendency that merges adjacent buildings into a single object, and the model structure's focus on global feature extraction also limits precise boundary representation. To address this, methods applying boundary-specific loss functions or incorporating additional boundary information into the learning process have been proposed, but they not only require the construction of precise boundary label data but also introduce a task conflict problem in which specialization in boundary learning degrades global context recognition capability. Consequently, it is structurally difficult to simultaneously satisfy the two requirements of global context recognition and precise boundary segmentation within a single model. To overcome this limitation, a complementary approach in which different models share the responsibilities of global context recognition and boundary refinement is required. SAM 2 (Ravi et al., 2025) is a general-purpose segmentation model capable of inferring object-level precise boundaries using only box prompts without being tied to a specific class, and its ability to independently perform boundary refinement without additional training makes it well suited for such a complementary combination.

This study builds upon the context- and boundary-aware two-stage segmentation framework proposed by Lee et al. (2025) for verifying building depiction errors in national digital maps, and extends it to the problem of automatically improving the boundary quality of globally available OSM building data. National digital maps are subject to legal positional accuracy standards that guarantee an upper bound on positional error, whereas OSM building polygons are created based on contributors' subjective judgment, resulting in high variance and an uncertain upper bound of positional error. This fundamental difference necessitates re-exploration of key design parameters such as clipping margins and box prompt margins, as well as reconsideration of strategies for suppressing non-building object segmentation. Specifically, SegFormer (Xie et al., 2021), which excels in context-aware segmentation, is combined with SAM 2, which is specialized in boundary-aware segmentation, and OSM polygons are used as box prompts to structurally suppress under-segmentation. The goal of this study is to develop a scalable automated framework that uses OSM data itself as a segmentation guide without additional label construction, enabling automatic correction of positional and shape errors in OSM building polygons worldwide and uniform improvement of boundary quality.

2. Materials and Methods

2.1 Study Area and Data

2.1.1 True Orthoimage. This study covers the Suseo-dong area in Gangnam-gu, Seoul (approximately 2.86 km²) and uses a 2022 true orthoimage (GSD 0.12 m/px) provided by the Seoul Metropolitan Government. A true orthoimage is generated through orthorectification based on a digital surface model (DSM), eliminating the residual relief displacement of high-rise buildings that occurs in conventional orthoimages so that all objects correspond to their actual positions on the map.

2.1.2 OSM Building Data. OSM building polygons were obtained from the South Korea dataset distributed via Geofabrik (OpenStreetMap contributors, 2022), targeting the building layer (fclass: building). In imagery with a GSD of 0.12 m/px, small buildings have fewer pixels constituting their boundaries, making the boundary between buildings and background indistinct and potentially reducing the reliability of boundary discrimination. Considering this, building objects with an area less than 50 m² were excluded from the analysis. At a GSD of 0.12 m/px, 50 m² corresponds to a square with a side length of approximately 59 pixels. The final number of OSM building objects within the study area is 125.

2.1.3 Training Datasets for Building Segmentation. An integrated strategy was adopted for training data construction to simultaneously ensure model robustness and generalization performance. Benchmark datasets covering diverse regions and resolutions were used to secure robustness, and AI-Hub datasets reflecting building characteristics of the domestic environment were additionally integrated to supplement generalization performance for domestic data. The benchmark datasets include WHU (Ji et al., 2019), Inria (Maggiori et al., 2017), Alabama (Cao, 2022), Potsdam (ISPRS, 2015), and Vaihingen (ISPRS, 2014), while the domestic datasets include the national park change detection and Jeju Island land cover datasets from AI-Hub (National Information Society Agency, 2022). The collected data were divided into 512×512 pixel patches and standardized in PNG format, and patches containing building objects with an area of fewer than 10 pixels were excluded to prevent degradation of training efficiency. The final integrated dataset consists of 86,226 training patches and 12,171 validation patches (Table 1).

2.2 Overview of the Proposed Framework

The proposed framework builds upon the two-stage pipeline introduced by Lee et al. (2025), with the input data shifted from national digital maps to OSM building polygons and key experimental parameters such as clipping margins readjusted to account for the positional error characteristics of OSM data. The overall structure consists of two stages: (1) SegFormer-based context-aware segmentation and (2) SAM 2-based boundary-aware segmentation (Figure 1). This section describes the core components along with the modifications made for the application to OSM data. In the first stage, SegFormer semantically segments building regions from the entire input image to provide global contextual information. In the second stage, SAM 2 receives the bounding boxes of individual OSM polygons as box prompts and infers object-level precise boundaries. The two results are integrated into a prior probability map and fed back into the Mask decoder of SAM 2, enabling uncertain boundary regions to be re-evaluated in an unsupervised manner.

Source	Dataset	Patch size	Format	Training	Validation
Benchmark	WHU	512×512	*.png	4,317	737
	Inria	512×512	*.png	9,737	1,942
	Alabama	512×512	*.png	26,492	3,784
	Potsdam	512×512	*.png	1,881	268
	Vaihingen	512×512	*.png	230	32
Domestic	AI-Hub (National Park)	512×512	*.png	30,469	3,829
	AI-Hub (Jeju Island)	512×512	*.png	13,100	1,579
Total	-	-	-	86,226	12,171

Table 1. Datasets for training building segmentation models

https://doi.org/10.5194/isprs-archives-XLIX-B4-2026-151-2026 | © Author(s) 2026. CC BY 4.0 License.

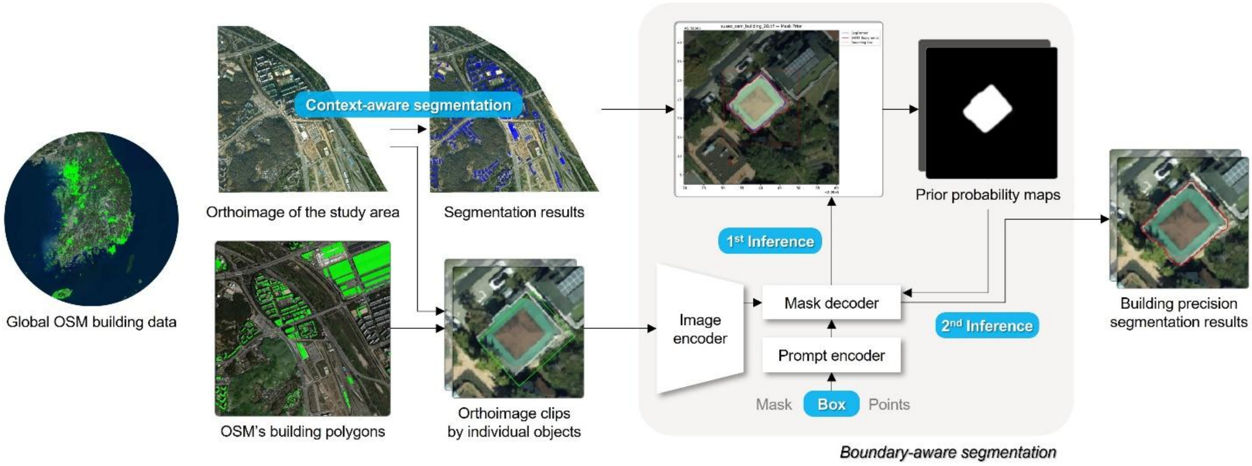

Figure 1. Overview of the proposed two-stage framework for automated geometric correction of OSM building polygons.

SegFormer was selected as the base model for the first stage. SegFormer achieves a large effective receptive field (ERF) through its hierarchical Transformer encoder, providing a structural advantage in global context recognition and making it well suited for detecting building regions across wide areas in complex urban environments. In addition, in the performance comparison experiment reported by Lee et al. (2025) using the same integrated dataset, SegFormer achieved the best performance across all metrics compared to major models such as DeepLabv3+, HRNet, and U-Net, and was therefore selected as the base model (Table 2).

SAM 2 was adopted as the model for the second stage. SAM 2 is a general-purpose segmentation model pretrained on large-scale segmentation data, capable of inferring object-level precise boundaries using only box prompts. It also features a modular architecture that allows the segmentation results of SegFormer to be directly input into the Mask decoder in the form of a prior probability map, making it structurally suited for complementary combination with SegFormer. In this study, the pretrained weights were used without additional fine-tuning, thereby clearly separating the learning-based context recognition of SegFormer from the zero-shot boundary refinement capability of SAM 2 and independently evaluating each model's contribution. The checkpoint used was sam2.1_hiera_base_plus. This checkpoint is relatively lightweight among the four checkpoints released by SAM 2 in terms of model size (80.8M parameters) and processing speed (64.1 FPS), while achieving a segmentation accuracy of J&F 78.2 on the SAV (Segment Anything Video) test bed, and was thus judged to offer the best balance between accuracy and efficiency. Here, J denotes region similarity and F denotes boundary accuracy.

Model	Recall (%)	Precision (%)	F1-Score (%)
DeepLabv3+	83.50	85.80	84.63
HRNet	86.30	82.10	84.15
PoolFormer	86.70	81.40	83.97
SegFormer	89.86	90.44	90.15
Swin	89.20	82.50	85.72
Twins	88.60	82.50	85.44
U-Net	79.60	83.60	81.55
UPerNet	83.50	82.50	83.00
ViT	86.80	84.10	85.43

Table 2. Performance comparison of building segmentation models (adapted from Lee et al., 2025).

2.3 Context-Aware Segmentation

SegFormer consists of a hierarchical Transformer encoder, the Mix Vision Transformer (MiT), and a lightweight All-MLP decoder. Through hierarchical token merging, it achieves a large multi-scale ERF, providing a structural advantage in effectively capturing global context information across wide areas. Training was conducted on the integrated dataset using the MMsegmentation framework (OpenMMLab, 2023), with the AdamW optimizer and Cross-Entropy Loss. A learning rate of 5.5×10^{-6}, which showed the most stable performance through comparative experiments, was adopted as the optimal value. The trained SegFormer semantically segments building regions based on global context, and its results are used as input for generating the prior probability map in the second stage.

However, due to the model structure's focus on global feature extraction, under-segmentation occurs at boundaries between adjacent buildings, where multiple buildings are merged and segmented as a single object. To address this, in the second stage, the box prompts of SAM 2 are used to restrict the search space to individual building units, thereby structurally suppressing under-segmentation and performing object-level precise boundary segmentation.

2.4 Boundary-Aware Segmentation

The second stage proceeds in the following order: initial inference using the bounding boxes of OSM polygons as prompts, generation of a prior probability map that integrates the segmentation results of SegFormer and SAM 2 as probability values, and secondary re-inference by feeding this map back into the Mask decoder of SAM 2. In the initial inference step, the true orthoimage was clipped for each individual OSM building polygon to construct input tiles (Figure 2). To ensure sufficient surrounding context while minimizing unnecessary areas, an evaluation of the effect of the clipping margin was conducted, and segmentation performance was found to saturate at 96 pixels (approximately 11.52 m) at a GSD of 0.12 m/px, which was adopted as the optimal value. In Lee et al. (2025), performance saturated at 80 pixels (approximately 9.60 m) when national digital maps were used. In contrast, OSM building polygons are not subject to legal positional accuracy standards and, as reported by Brovelli et al. (2016), exhibit average positional errors of approximately 0.8 m in both the X and Y directions, with no guaranteed upper bound on error (Moradi et al., 2023).

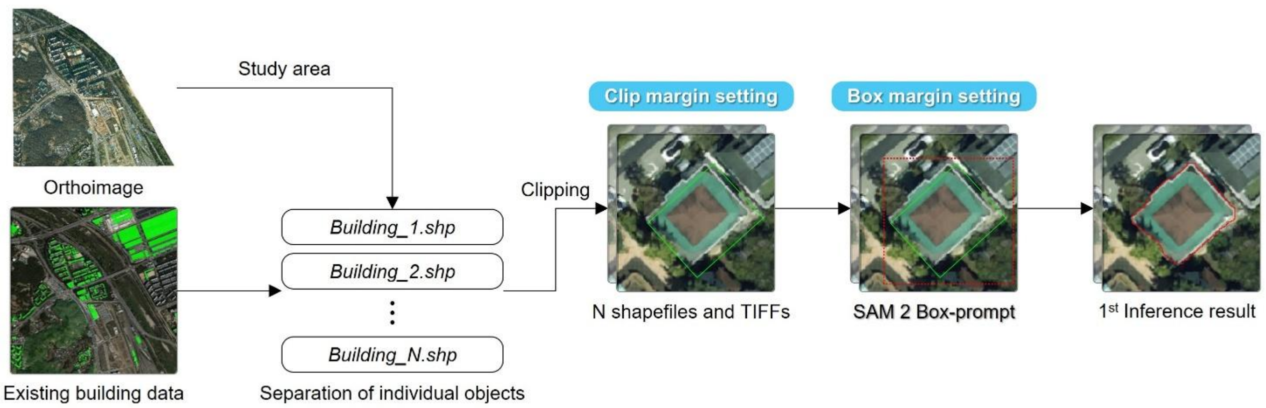

Figure 2. SAM 2 input configuration: clipping of the true orthoimage by individual OSM building polygons with margin.

Given the high variance and uncertain upper bound of positional errors inherent in OSM data, a wider margin is required to prevent actual buildings from falling outside the clipping area.

SAM 2 supports three prompt types: mask, point, and box. Point prompts are sensitive to the placement of positive and negative points, potentially leading to unstable segmentation results, while mask prompts require prior mask information. In contrast, box prompts can explicitly restrict the search space, resulting in stable segmentation. Accordingly, the bounding boxes of individual OSM polygons were used as prompts to structurally restrict the search space to individual building units and suppress under-segmentation. A margin of 8 pixels (approximately 0.96 m) was added to the bounding box of each OSM polygon to accommodate the potential positional errors of OSM. Since SAM 2 is not a building-specific model, objects other than buildings may also be segmented within the bounding box. Therefore, the mask with the largest area among the segmentation results within the bounding box was regarded as a building and adopted as the initial inference result.

A prior probability map was generated to integrate the segmentation results of SegFormer and the initial inference results of SAM 2 (Figure 3). SegFormer is specialized in building classification and excels in global context recognition but may produce under-segmentation, while SAM 2 excels in boundary refinement but may also segment non-building objects. After aligning the two results to the same coordinate system and resolution, pixel-level set operations were performed: intersection regions where both results agree were assigned a value of 1.00, and complement regions where both agree on background were set to 0.00. Symmetric difference regions where the two results disagree were regarded as uncertainty regions where neither model is confident, and a low probability value of 0.003 was assigned. This low value is designed to prevent the Mask decoder of SAM 2 from being biased by strong prior signals and to encourage re-evaluation of these regions based on visual features in the image. In particular, given that the two models have different segmentation characteristics, the approach of assigning a weak signal to disagreement regions contributes to probabilistically integrating the complementary judgments of the building-trained SegFormer and the general-purpose boundary segmentation of SAM 2.

The generated prior probability map was binarized and then smoothed with a Gaussian blur (Equation 1) to alleviate spatial discontinuities near boundaries:

$$G(x,y) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \tag{1}$$

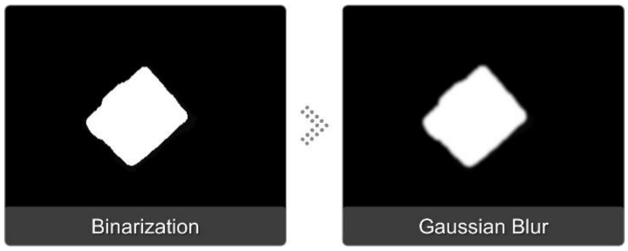

Figure 3. Method for generating the prior probability map from SegFormer and SAM 2 initial inference results.

where x and y are the horizontal and vertical distances from the kernel center, and σ is the standard deviation of the Gaussian distribution that determines the extent of the blur effect. A standard deviation of σ=1.5 was applied; this value was chosen to confine the smoothing effect to pixels near boundaries without affecting the high-confidence regions inside the mask. This map was then fed back into the Mask decoder of SAM 2 to perform secondary re-inference. Through this process, the model re-evaluates uncertain boundary regions in an unsupervised manner and infers precise boundaries based on visual features in the image.

2.5 Post-Processing

Since the segmentation results of SAM 2 are output as pixel-level raster masks, post-processing was performed to convert them into a form comparable to OSM polygons and to minimize unnecessary noise and topological complexity. First, morphological operations were applied to the raster masks to remove fine gaps inside the masks and small-scale noise along the boundaries, after which the masks were converted to vector polygons. Specifically, opening and closing operations were performed sequentially, each applied once using a 5×5 elliptical kernel: the opening operation removes small protrusions and noise along the boundaries, while the closing operation fills fine gaps inside buildings to ensure the continuity and completeness of the masks.

When raster-based segmentation results are converted to vectors, the pixel grid structure produces staircase artifacts that generate numerous unnecessary vertices. The Douglas-Peucker algorithm (Ramer, 1972) was therefore applied to simplify the boundaries. The tolerance was set to the GSD of the input image (0.12 m/px), limiting the boundary shape error to within one pixel. This prevents increased errors caused by unnecessary vertices during object-level shape comparison with OSM polygons and minimizes topological complexity.

https://doi.org/10.5194/isprs-archives-XLIX-B4-2026-151-2026

2.6 Evaluation Metrics

Since this study aims to improve the boundary quality of OSM building polygons, pixel-level quantitative evaluation was conducted to assess both overall segmentation accuracy and boundary precision. The ground truth (GT) data were taken from the re-digitized label data constructed by Lee et al. (2025). The study area contains 125 OSM building polygons; however, since the GT labels subdivide boundaries between adjacent buildings into individual objects, 135 objects were used for evaluation. Based on this, the segmentation results of the OSM building data, standalone SegFormer, and the proposed method were compared with the GT to evaluate pixel-level performance.

Recall, Precision, F1-Score, IoU, and BIoU were used as evaluation metrics. Recall is defined as the proportion of actual building pixels correctly detected, Precision as the proportion of predicted building pixels that are actual buildings, F1-Score as the harmonic mean of the two metrics, and IoU as the intersection of the predicted and GT masks divided by their union.

BIoU quantifies boundary positional precision by generating buffers of radius r around the predicted boundary (∂P) and the GT boundary (∂G) and computing the IoU of the two buffer regions (Equation 2). This metric is well suited for directly evaluating the improvement in building boundary quality, which is the objective of this study (Cheng et al., 2021).

$$\mathrm{BIoU} = \frac{|B_r(\partial G) \cap B_r(\partial P)|}{|B_r(\partial G) \cup B_r(\partial P)|} \tag{2}$$

where ∂G is the boundary of the GT mask, ∂P is the boundary of the predicted mask, B_r is an operator that generates a buffer of radius r, and |·| denotes the area of the corresponding region. Cheng et al. (2021) recommend setting the buffer radius to 0.5–2% of the image diagonal length; in this study, the buffer radius was set to 0.75 m, corresponding to approximately 1% of the diagonal length of a 512×512 pixel tile.

3. Results

3.1 Quantitative Results

To quantitatively verify the effect of improving OSM building boundary quality, which is the core objective of this study, the pixel-level segmentation performance of the OSM building data, standalone SegFormer, and the proposed method (SegFormer + SAM 2) were compared (Table 3). The OSM building data showed reasonable performance in terms of overall region detection, with a Recall of 89.95%, Precision of 84.64%, F1-Score of 87.22%, and IoU of 77.33%. However, the BIoU, which represents boundary precision, was only 46.55%. This quantitatively demonstrates that shape errors originating from subjective differences among contributors and manual digitization without unified standards have accumulated throughout the data, supporting the need for boundary quality improvement of OSM building data.

The standalone SegFormer showed improvement in overall region detection performance compared to the OSM building data, with a Recall of 92.18%, Precision of 87.15%, F1-Score of 89.60%, and IoU of 84.76%. However, the BIoU remained at 68.22%. This indicates that while global context-based semantic segmentation is effective for building region detection, its structural focus on global feature extraction limits the precision of boundary representation between adjacent buildings.

Metric	OSM	SegFormer	Ours
Recall (%)	89.95	92.18	91.63
Precision (%)	84.64	87.15	92.21
F1-Score (%)	87.22	89.60	91.92
IoU (%)	77.33	84.76	85.36
BIoU (%)	46.55	68.22	70.40

Table 3. Pixel-level segmentation performance: OSM vs. SegFormer vs. proposed method

In contrast, the proposed method achieved a Recall of 91.63%, Precision of 92.21%, F1-Score of 91.92%, and IoU of 85.36%, showing consistent performance improvement across all metrics. In particular, the BIoU reached 70.40%, an improvement of 23.85 percentage points over the OSM building data and 2.18 percentage points over the standalone SegFormer. Recall decreased slightly by 0.55 percentage points compared to the standalone SegFormer (92.18%), which is attributable to the box prompts of SAM 2 explicitly restricting the search space to individual building units, causing some portions of adjacent building regions to be treated as background. However, this was compensated by the 5.06 percentage point improvement in Precision and the significant improvement in BIoU, a result consistent with the objective of boundary precision improvement in this study.

These results quantitatively demonstrate that the two-stage pipeline combining SegFormer's global context-based semantic segmentation with SAM 2's object-level boundary refinement effectively corrects positional and shape errors in OSM data.

3.2 Qualitative Results

To visually examine how the performance improvement identified in the quantitative results is reflected in the actual segmentation results, the segmentation results of the OSM building data, standalone SegFormer, and the proposed method were compared through three representative cases (Figure 4). The three cases demonstrate the improvement achieved by the proposed method along three axes: OSM positional offset correction, building shape reproduction, and under-segmentation suppression.

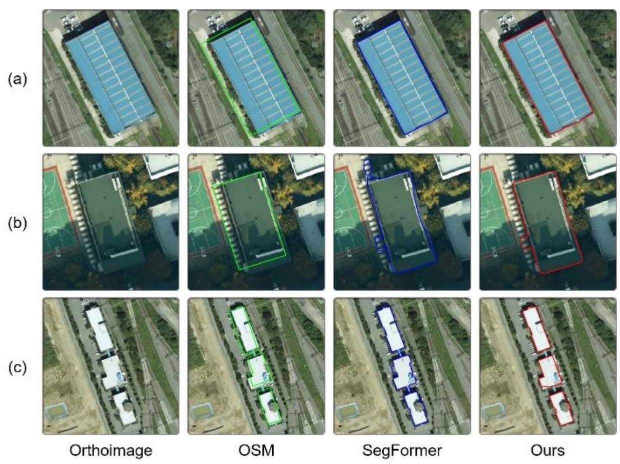

Figure 4. Qualitative comparison of segmentation results: Orthoimage, OSM, SegFormer, and proposed method (Ours). (a) OSM positional offset correction, (b) building shape refinement, (c) under-segmentation suppression.

The first case (Fig. 4a) involves a building with a large positional offset in the OSM polygon and is intended to verify whether the proposed method is robust to OSM positional errors. The OSM polygon is clearly displaced from the actual building location, resulting in a pronounced spatial mismatch with the building boundary. The standalone SegFormer generally captured the building region based on global context, but its alignment precision with the building boundary was limited. In contrast, the proposed method ensured that the actual building was included within the search space through the clipping margin and bounding box margin designed to accommodate OSM positional errors. Based on this, it accurately inferred the actual building boundary using visual features in the image and effectively corrected the positional and shape errors of OSM.

The second case (Fig. 4b) examines shape reproduction precision for an isolated building. The OSM polygon represents the overall outline of the building in a simplified form, failing to capture detailed shape features such as protrusions and steps. The standalone SegFormer exhibited shape distortion, with irregular boundary noise appearing along parts of the building outline and fine structural features being omitted. In contrast, the proposed method combined the global context information of SegFormer with the boundary refinement capability of SAM 2 through the prior probability map, reproducing the detailed contours of the building more precisely and increasing the similarity to the actual shape compared to OSM and SegFormer.

The third case (Fig. 4c) evaluates the under-segmentation suppression effect for a facility where multiple buildings are arranged consecutively. The OSM polygons delineate boundaries at the individual building level; however, the boundaries between buildings and detailed shapes are not sufficiently captured, resulting in low shape precision. The standalone SegFormer, due to its structural focus on wide-area feature extraction, produced under-segmentation that merged adjacent buildings into a single connected mass, making one-to-one correspondence with individual buildings difficult. In contrast, the proposed method used the bounding boxes of individual OSM polygons as box prompts to structurally restrict the search space to building-level units, thereby suppressing the under-segmentation observed in SegFormer, accurately separating each building as an independent object, and faithfully reproducing the boundaries and detailed shapes between buildings.

In summary, the proposed method simultaneously achieves robustness against OSM positional errors, precise shape reproduction, and object separation, visually supporting the quantitative metrics in Table 3, particularly the improvement in BIoU.

4. Conclusions

This study applied and extended the two-stage segmentation framework of Lee et al. (2025) to OSM data in order to automatically improve the boundary quality of crowdsourcing-based OSM building data. In the first stage, SegFormer semantically segments building regions based on global context, and in the second stage, SAM 2 infers object-level precise boundaries using the bounding boxes of OSM polygons as box prompts. The prior probability map generated by combining the results of the two stages probabilistically integrates the two opposing requirements of global context recognition and precise boundary segmentation, and guides SAM 2 to re-evaluate disagreement regions in an unsupervised manner. The experimental results show that the proposed method achieved a Recall of 91.63%, Precision of 92.21%, F1-Score of 91.92%, and IoU of 85.36%, with consistent performance improvement across all metrics. In particular, the BIoU, a boundary precision metric, reached 70.40%, an improvement of 23.85 percentage points over the OSM building data. This quantitatively demonstrates that the two-stage pipeline effectively corrects boundary shape errors in OSM that have accumulated from subjective differences among contributors and manual digitization processes. Furthermore, the proposed framework can use OSM data itself as a segmentation guide without additional label construction, provided that high-resolution true orthoimagery is available for regions where OSM building polygons exist. This demonstrates that the framework is a scalable automated solution capable of automatically improving building boundary quality in any region worldwide.

The experiments were conducted in a single area of Suseo-dong, Gangnam-gu, Seoul (2.86 km²), and generalization performance for regions with different urban morphologies and OSM contribution levels requires further verification. In addition, the proposed method focuses on shape error correction; omission errors (buildings present in imagery but unregistered in OSM) and commission errors (buildings registered in OSM but actually demolished), which frequently occur in OSM, are beyond the detection scope of the current framework. Moreover, building boundary occlusion by surrounding objects such as trees is a factor that degrades boundary precision, and a separate mechanism is required to handle this issue.

Future work will verify generalization potential through multi-region validation covering diverse urban morphologies and OSM contribution levels, and will advance the framework into a more comprehensive automated OSM quality management system by incorporating omission and commission error detection capabilities.

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).

References

Biljecki, F., Chow, Y. S., Lee, K., 2023. Quality of crowdsourced geospatial building information: A global assessment of OpenStreetMap attributes. *Build. Environ.*, 237, 110295. doi.org/10.1016/j.buildenv.2023.110295.

Brovelli, M. A., Minghini, M., Molinari, M. E., Zamboni, G., 2016. Positional accuracy assessment of the OpenStreetMap buildings layer through automatic homologous pairs detection: The method and a case study. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLI-B2, 615-620. doi.org/10.5194/isprsarchives-XLI-B2-615-2016.

Cao, D., 2022. Alabama Buildings Segmentation. Available online: kaggle.com/datasets/meowmeowplus/alabama-buildings-segmentation (30 September 2025).

Cheng, B., Girshick, R., Dollár, P., Berg, A. C., Kirillov, A., 2021. Boundary IoU: Improving object-centric image segmentation evaluation. In: *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, TN, USA, June 20-25, pp. 15329-15337. doi.org/10.1109/cvpr46437.2021.01508.

ISPRS, 2014. 2D Semantic Labeling - Vaihingen data. Available online: isprs.org/resources/datasets/benchmarks/UrbanSemLab/2d-sem-label-vaihingen.aspx (30 September 2025).

ISPRS, 2015. 2D Semantic Labeling Contest - Potsdam. Available online: isprs.org/resources/datasets/benchmarks/UrbanSemLab/2d-sem-label-potsdam.aspx (30 September 2025).

Ji, S., Wei, S., Lu, M., 2019. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. *IEEE Trans. Geosci. Remote Sens.*, 57(1), 574-586. doi.org/10.1109/tgrs.2018.2858817.

Lee, I., Lee, C., Cheon, J., An, B. N., Lee, D., Lee, I., 2025. Automatic verification of building depiction errors in national digital maps via context- and boundary-aware segmentation. *Korean Journal of Remote Sensing*, 41(5), 749-767. doi.org/10.7780/kjrs.2025.41.5.5.

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*, Fort Worth, TX, USA, July 23-28, pp. 3226-3229. doi.org/10.1109/igarss.2017.8127684.

Moradi, M., Roche, S., Mostafavi, M. A., 2023. Evaluating OSM building footprint data quality in Québec province, Canada from 2018 to 2023: A comparative study. *Geomatics*, 3(4), 541-562. doi.org/10.3390/geomatics3040030.

National Information Society Agency, 2022. AI-Hub. Available online: aihub.or.kr (30 September 2025).

OpenMMLab, 2023. OpenMMLab Semantic Segmentation Toolbox and Benchmark. Available online: github.com/open-mmlab/mmsegmentation (30 September 2025).

OpenStreetMap contributors, 2022. South Korea (Version 2022-01-01) [Data set]. Geofabrik. Available online: download.geofabrik.de/asia/south-korea.html (30 September 2025).

Ramer, U., 1972. An iterative procedure for the polygonal approximation of plane curves. *Comput. Graph. Image Process.*, 1(3), 244-256. doi.org/10.1016/s0146-664x(72)80017-0.

Ravi, N., Gabeur, V., Hu, Y.-T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., Mintun, E., Pan, J., Alwala, K. V., Carion, N., Wu, C.-Y., Girshick, R., Dollár, P., Feichtenhofer, C., 2025. SAM 2: Segment anything in images and videos. In: *Proceedings of the 13th International Conference on Learning Representations (ICLR 2025)*.

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., Luo, P., 2021. SegFormer: Simple and efficient design for semantic segmentation with transformers. In: *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, pp. 12077–12090.

Zhang, Y., Zhu, Y., Li, H., Chen, S., Peng, J., Zhao, L., 2020. Automatic change detection between outdated building maps and new VHR images based on pre-trained fully convolutional feature maps. *Sensors*, 20(19), 5538. doi.org/10.3390/s20195538.

Zhuo, X., Fraundorfer, F., Kurz, F., Reinartz, P., 2018. Optimization of OpenStreetMap building footprints based on semantic information of oblique UAV images. *Remote Sens.*, 10(4), 624. doi.org/10.3390/rs10040624.