Stepwise Optimization and Ensemble Pipeline for Building Change Detection in High Resolution Satellite Imagery Using Mamba-Based Model

DongHyuk Jin [1], Junhwa Chi[2]

[1] Department of Data Engineering, Pukyong National University, Busan, Republic of Korea – easternjin@pukyong.ac.kr
[2] Division of Data Information Sciences, Pukyong National University, Busan, Republic of Korea – jchi@pknu.ac.kr

Keywords: Change Detection, Satellite Imagery, Mamba Architecture, Remote Sensing, Ensemble Strategy, Optimization

Abstract

We propose a systematic stepwise optimization pipeline for building change detection in dense urban environments using high-resolution CAS500-1 satellite imagery. To support robust model development, we constructed a dataset comprising 3,816 bi-temporal patch pairs across 28 urban regions. The framework employs a Mamba-based architecture as the baseline, leveraging its efficient global context modeling capability for binary change detection. The pipeline integrates three sequential optimization stages to enhance detection accuracy and stability. First, we evaluated normalization techniques tailored for 12-bit radiometric resolution, comparing percentile-based scaling, gamma correction, and log transformations. Second, we implemented an augmentation strategy that extends standard geometric transformations with optical and temporal methods to improve generalization in structurally complex urban settings. Third, we explored various ensemble configurations, including confidence-weighted and hierarchical aggregation to mitigate individual model scale limitations. Performance was validated through multi-faceted evaluation metrics covering pixel-level, contour-based, and object-based metrics. Experimental results demonstrate that gamma-based normalization, comprehensive augmentation, and hierarchical ensemble consistently outperform baseline configurations across multiple evaluation metrics. The final optimized pipeline achieved an F1-Score of 0.8070, making a significant improvement over the 0.7629 baseline. This work provides an extensible framework for operational satellite-based change detection and establishes a practical foundation for future ensemble-based architectures.

1. Introduction

Change detection is one of the key tasks in remote sensing, providing essential information for urban expansion analysis and environmental monitoring (Viana et al., 2019). While traditional methods laid the groundwork, recent studies have increasingly adopted deep learning architectures to handle the complexity of high-resolution satellite imagery (Khelifi and Mignotte, 2020). Notable developments include the Siamese CNN-based SNU-Net (Fang et al., 2022) and the Transformer-based ChangeFormer (Bandara and Patel, 2022), both of which have improved the extraction of multi-scale temporal features. Recently, Mamba-based models such as ChangeMamba (Chen et al., 2024a) have demonstrated superior performance on benchmark datasets by utilizing state space models to capture long-range dependencies with high computational efficiency.

Among these applications, building change detection (BCD) is particularly critical for systematic urban monitoring and planning (Chen et al., 2024b). However, several inherent challenges hinder the reliability of automated BCD systems. High-resolution satellite imagery often contains significant noise arising from low-illumination conditions in dense urban corridors, which can be mistakenly identified as structural changes. Furthermore, the extreme class imbalance—characterized by a low ratio of changed pixels compared to unchanged areas—complicates model training (Li et al., 2025). These environmental factors, combined with the high cost of acquiring accurately labeled multi-temporal datasets (Lu et al., 2008), necessitate more comprehensive approaches that go beyond simple architectural improvements, particularly those that address data quality and model robustness in complex urban environments.

The limitations of previous models in complex urban scenes highlight the need for a more comprehensive pipeline that integrates data-centric optimization with complementary strategies. Data-centric optimization focuses on improving the quality of input imagery to ensure that the model learns from more representative features (Sambasivan et al., 2021). This is particularly important for satellite imagery, where radiometric inconsistencies and misregistration between different timestamps can lead to false positives (Zhu et al., 2017). By implementing optimization techniques specifically designed for the spectral characteristics of satellite sensors, it is possible to mitigate the effects of varying solar angles and atmospheric conditions before the feature extraction stage.

The need for such optimization becomes more critical when aiming to maximize the operational utility of the recently launched CAS500-1 satellite, especially in high-density urban environments like those found in South Korea. These areas feature complex spatial arrangements with minimal spacing between buildings, which often result in severe occlusion and intricate shadow patterns (Liasis and Stavrou, 2016). To effectively leverage the high-precision land observation data provided by the national asset, a specialized approach is required to effectively handle its unique characteristics.

Furthermore, an ensemble strategy is adopted as a robust approach to refine and strengthen the individual model’s results (Ganaie et al., 2022). Single-model configurations, even within advanced architectures, may exhibit localized biases that lead to inconsistent boundary predictions. To explore the potential of result enhancement, this study integrates an ensemble of multiple ChangeMamba variants, specifically across Tiny, Small, and Base scales. By implementing and comparing various ensemble techniques, the pipeline aims to aggregate the complementary strengths of these different model scales, resulting in a more stable and generalized prediction. Through this approach, the proposed stepwise optimization and multi-scale ensemble framework is designed to provide a robust methodological foundation for implementing task-specific processing and optimal modeling strategies, ultimately achieving high-accuracy change detection across the complex urban satellite imagery.

Building upon this framework, this study proposes a stepwise optimization pipeline that integrates normalization, augmentation, and multi-scale ensemble strategies based on ChangeMamba, specifically tailored for high-resolution CAS500-1 imagery to improve robustness and detection accuracy in dense urban environments.

2. Data and methods

2.1 Dataset

The BCD dataset used in this study is based on high-resolution imagery acquired from CAS500-1, Korea's national land observation satellite launched in 2021. CAS500-1 is equipped with the AEISS-C sensor, providing a spatial resolution of 0.5 m for panchromatic and 2 m for multispectral bands. In this research, Level-2G pan-sharpened products with a spatial resolution of 0.5 m, provided through the National Geography Information Platform (National Geographic Information Institute, 2023), were utilized to ensure high-fidelity urban monitoring.

Specifically, this study leverages the CAS500-1 BCD dataset, previously constructed to support deep learning-based change detection (Jin et al., 2025). The dataset was constructed following the pipeline illustrated in Figure 1.

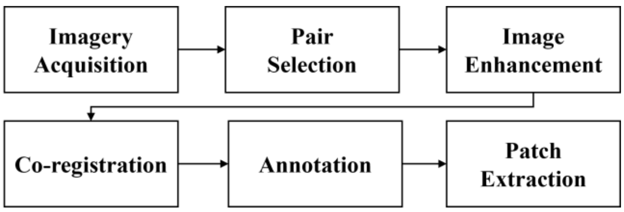

Figure 1. Construction pipeline of the CAS500-1 BCD dataset.

The finalized dataset provides a diverse representation of complex urban structural variations. As shown in Figure 2, the dataset includes representative urban regions such as Seoul, Busan, and Daejeon, along with paired pre-event (T1), post-event (T2), and ground truth (GT) samples. To ensure objective model evaluation, the dataset consists of 3,816 patches (256×256 pixels), which were partitioned into training, validation, and test sets at a ratio of 70%, 15%, and 15%, respectively. By adopting this pre-established and validated dataset, this study focuses on optimizing the change detection pipeline, enabling a systematic evaluation of data-centric strategies and ensemble techniques for high-resolution satellite assets.

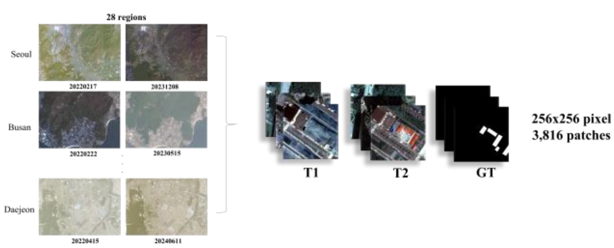

Figure 2. Overview of the CAS500-1 BCD dataset.

2.2 Model

In this study, ChangeMamba was adopted as the core architecture, extending the VMamba (Liu et al., 2024) for BCD. As illustrated in Figure 3, the model extracts and fuses bi-temporal features through a multi-stage structure. The backbone utilizes Visual State Space (VSS) blocks to capture global contextual information from each input image by modeling spatio-temporal tokens. To integrate these features, Spatio-Temporal State Space blocks and Fusion blocks are employed across multiple stages. This design enables efficient global feature learning with linear computational complexity, facilitating robust detection of structural changes.

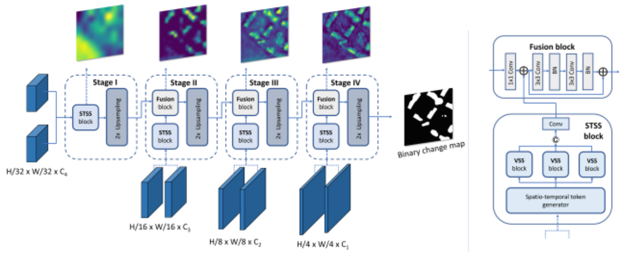

Figure 3. Overview of the ChangeMamba architecture for BCD (Modified from Chen et al., 2024a).

The architecture is configured into three variants with different model complexities: Tiny, Small, and Base. The differences among these variants in terms of feature map channel dimensions, the number of VSS blocks, and total parameter counts are summarized in Table 1.

Model Scale	Channels	VSS Block	Params(M)
Tiny	768	10	32
Small	768	21	52
Base	1024	21	92

Table 1. Architectural configurations of ChangeMamba across different model scales (Tiny, Small, and Base).

These multi-scale variants provide complementary representations, enabling more effective ensemble learning in the later stage. The Base model was selected as the primary single model for the initial optimization stages and baseline experiments, as it provides a strong performance baseline across various change detection benchmarks. To further enhance the stability and precision of the final predictions, all three variants were integrated during the ensemble stage.

2.3 Stepwise Optimization

The proposed stepwise optimization pipeline is designed to systematically improve detection performance and stability for CAS500-1 satellite imagery using a Mamba-based architecture. As illustrated in Figure 4, the pipeline follows a structured sequence starting from the BCD dataset and progressing through three primary stages. This stepwise design allows each component to be independently optimized while progressively improving the overall detection performance.

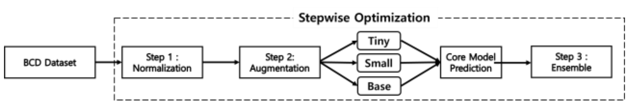

Figure 4. Stepwise optimization pipeline.

The optimization pipeline begins with normalization to align the radiometric characteristics of the satellite imagery. This is followed by data augmentation to improve model generalization across diverse urban environments. In the next phase, multi-scale variants (Tiny, Small, and Base) generate independent predictions. Finally, these predictions are integrated through an ensemble strategy to produce a robust and accurate change map.

2.3.1 Normalization

The optimization process begins with normalization strategies tailored for the high radiometric resolution of 12-bit CAS500-1 imagery. Since raw 12-bit satellite data contains a much broader dynamic range than standard 8-bit images, appropriate radiometric processing is essential to ensure that the model effectively learns structural features without being affected by sensor noise or extreme pixel values (Chavez, 1996). To identify the optimal preprocessing strategy for dense urban scenes, we adopte percentile-based normalization as a baseline and further evaluate gamma correction and log transformation (Al-Safar, 2021) to compare their respective impacts on detection performance. The correction formulas for each method are provided in Equations (1) - (4).

$$I_{percentile}(x,y) = \frac{clip(I(x,y) - P_2, 0, P_{98} - P_2)}{P_{98} - P_2}, \quad (1)$$

$$I_{\gamma}(x,y) = (I_{percentile}(x,y))^{\frac{1}{\gamma}}, \quad (2)$$

$$I_{log}(x,y) = log(1 + I(x,y)), \quad (3)$$

$$I_{log-transformed} = I_{norm}\left(I_{log}(x,y)\right) \quad (4)$$

where x, y = image coordinates
$I(x,y)$ = input pixel value
P_2, P_{98} = 2nd and 98th percentile pixel value
$I_{percentile}(x,y)$ = percentile normalized output
γ = gamma coefficient
$I_{\gamma}(x,y)$= gamma-corrected output
$I_{log}(x,y)$ = log-transformed value
$I_{log-transformed}$ = log-transformed normalized output

As illustrated in Figure 5, each normalization technique modifies the radiometric range with a distinct objective. Percentile-based normalization clips extreme values and utilizes the intensity range between 2% and 98% to reduce the impact of outliers and sensor noise. Building upon this, gamma correction redistributes pixel intensity levels, enabling a more effective representation of non-linear variations. In contrast, log transformation rescales the intensity distribution by emphasizing lower-range values, thereby stabilizing the overall dynamic range during training. These differences allow us to evaluate how each method affects feature representation under varying radiometric conditions.

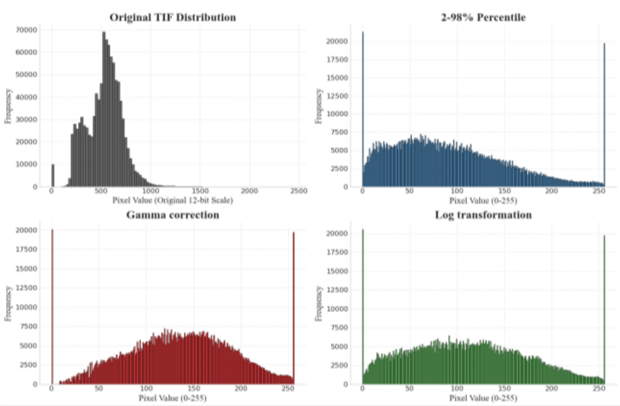

Figure 5. Histogram comparison of different normalization techniques (percentile-based, gamma correction, and log transformation).

2.3.2 Augmentation

The second stage of the stepwise optimization introduces a comprehensive data augmentation strategy to enhance model robustness under diverse environmental conditions in satellite imagery (Hao et al., 2023). In this study, we establish a geometric augmentation baseline, including random cropping, horizontal and vertical flipping, and rotation (Yu et al., 2017). Building upon this baseline, we introduce and evaluate additional optical and temporal augmentation techniques to assess their contributions to the model generalization capability.

Optical augmentation is designed to simulate radiometric inconsistencies and sensor-induced artifacts commonly observed in high-resolution satellite imagery. These transformations are applied in real-time during training as images are fed into the model. As illustrated in Figure 6, three types of filters—brightness and contrast adjustment, Gaussian noise injection, and blur or sharpening—are applied. Each filter is independently applied with a probability of 20%, allowing the model to encounter diverse image conditions and reducing overfitting to specific distortions. Brightness and contrast adjustments simulate variations in solar illumination and atmospheric haze, while Gaussian noise improves robustness against sensor-induced artifacts. Additionally, blur and sharpening filters account for variations in image focus and environmental conditions that may obscure structural edges.

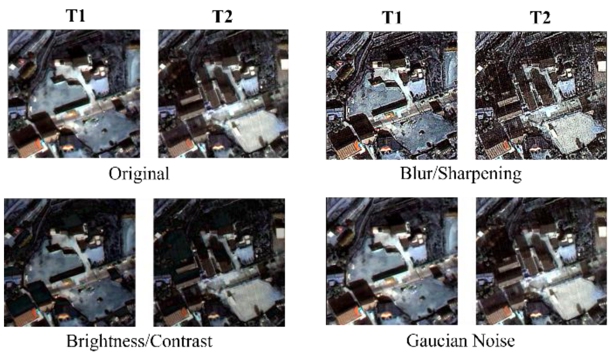

Figure 6. Example of optical augmentation techniques.

Temporal augmentation addresses the bi-temporal nature of change detection through an image exchange strategy. During training, the chronological order of T1 and T2 images is randomly swapped with a 30% probability. This approach prevents the model from overfitting to a specific temporal direction, ensuring that changes are recognized as structural discrepancies regardless of sequence.

2.3.3 Ensemble

The final stage of the stepwise optimization pipeline integrates predictions from three multi-scale ChangeMamba variants (Tiny, Small, and Base) to produce a robust and accurate change mask. Although these models share the same architecture, their differing parameter configurations lead to complementary error distributions. To exploit this diversity, we compare four ensemble architectures, each distinguished by its decision-making strategy.

As a baseline ensemble strategy, averaging methods are first evaluated to provide a simple yet effective integration of multi-scale predictions. Average Ensemble serves as a fundamental fusion strategy (Dietterich, 2000) to reduce prediction variance by aggregating outputs from the Tiny, Small, and Base models. As illustrated in the parallel framework of Figure 7, this strategy integrates multi-scale variants using either uniform or confidence-based weighting to generate a unified change mask. In the Simple Average approach, each model is assigned equal weight, and the ensemble probability is computed as the arithmetic mean of the softmax outputs. To further refine this, the Weighted Average Ensemble assigns dynamic weights by emphasizing high-confidence predictions using squared probabilities. For both averaging methods, a fixed threshold of 0.5 is applied to the ensemble probability to obtain the final binary change map. The formulations for both simple and weighted averaging methods are provided in Equations (5) - (6).

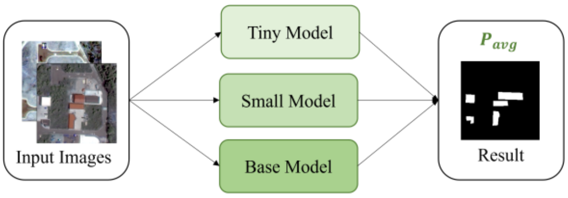

Figure 7. Schematic of the Average Ensemble strategy.

$$P_{sim}(x,y) = \sum_{i=1}^{3} \frac{1}{3} \cdot P_i(x,y), \tag{5}$$

$$P_{wei}(x,y) = \frac{\sum_{i=1}^{3} P_i(x,y)^2}{\sum_{i-1}^{3} P_i(x,y)}, \tag{6}$$

where x, y = image coordinates
i = index of core model scale
$P_i(x,y)$ = output probabilities map by model
$P_{sim}(x,y)$ = simple average probability map
$P_{wei}(x,y)$ = weighted average probability map

Hierarchical Ensemble implements a multi-pass decision strategy (Viola and Jones, 2000) designed to improve detection accuracy through a sequential model hierarchy. As illustrated in Figure 8, the input data is evaluated in a predefined order, either in a descending sequence from the Base model to the Tiny model or in an ascending sequence from the Tiny model to the Base model. At each stage, confidence thresholds are used to assign labels to high-certainty pixels, while uncertain pixels are passed to the next model in the hierarchy for further evaluation. Specifically, pixels with probabilities exceeding an upper threshold (0.80) for change or falling below a lower threshold (0.05) for non-change are assigned final labels and excluded from subsequent processing. Remaining pixels that are not finalized in earlier stages are classified by the final model using a standard threshold of 0.5 to determine the final change map. This hierarchical strategy effectively reduces both false positives and false negatives by progressively refining predictions across models with different capacities.

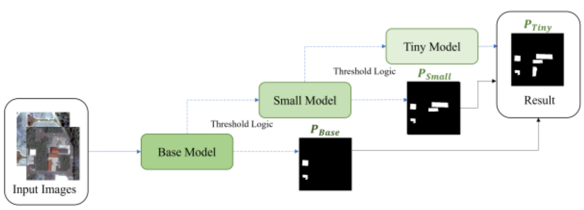

Figure 8. Schematic of the Hierarchical Ensemble strategy.

Stacking Ensemble adopts a meta-learning framework (Wolpert, 1992) to identify the optimal combination of predictions from multi-scale variants by treating their outputs as input features for a secondary model. As illustrated in Figure 9, the prediction maps generated by the three core models are used as input features for training the meta-model. These probability maps from the Tiny, Small, and Base models are flattened and stacked pixel-wise to form a high-dimensional feature space, which is paired with corresponding ground truth labels to train and evaluate the meta-model. During testing, the prediction maps from each core model are fed into the trained meta-model as input vectors. The output of the meta-model is used as the final change detection map. To identify the most effective architecture for this fusion task, three base algorithms—Random Forest (Breiman, 2001), Logistic Regression (Hosmer et al., 2013), and LightGBM (Ke et al., 2017) —were trained and compared. This stacking approach enables the model to learn optimal combinations of multi-scale predictions, capturing complex relationships that cannot be achieved through simple averaging.

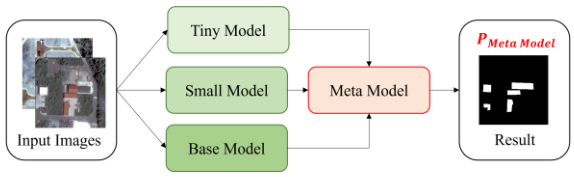

Figure 9. Schematic of the Stacking Ensemble strategy.

Dynamic Ensemble employs an input-aware selection mechanism (Jacobs et al., 1991) to assign the most suitable model among the three core variants for each input scene. As illustrated in Figure 10, the process begins with a selector module that evaluates the radiometric characteristics of the input image pair. To enable this adaptive selection, the validation stage plays a key role, where statistical features extracted from T1 and T2 images—such as mean, standard deviation, and absolute pixel differences—are used as input features for training the selector model. These features are paired with target labels indicating the optimal Expert Model, defined as the variant (Tiny, Small, and Base) that achieves the best F1-score for each validation patch. During testing, features extracted from each image pair are fed into the trained selector to predict the most reliable expert model. The prediction from the selected expert model is then used as final change detection map. To ensure consistency across the ensemble framework, the same three base algorithms used in the stacking ensemble were also evaluated as selector models. This dynamic selection strategy allows the framework to adaptively choose the most suitable model for each scene, improving robustness under varying radiometric and structural conditions.

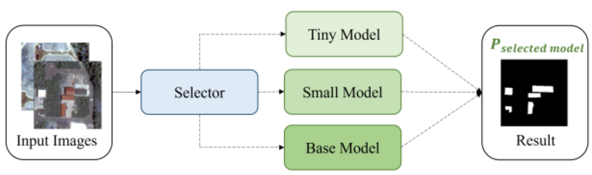

Figure 10. Schematic of the Dynamic Ensemble strategy.

2.4 Evaluation Metrics

To objectively evaluate the practical detection capability of the proposed model, multi-faceted evaluation metrics are introduced, extending beyond conventional pixel-based accuracy (Hussain et al., 2013). Standard pixel-based metrics—including Precision, Recall, F1-Score, Intersection over union (IoU), and Kappa coefficient (Kappa) —are used to provide a comprehensive statistical overview of model performance. However, BCD in high-resolution satellite imagery often suffers from a severe class imbalance, where unchanged regions vastly outnumber changed pixels. In such scenarios, overall accuracy can be misleading, as

https://doi.org/10.5194/isprs-archives-XLIX-B3-2026-601-2026 |

it may fail to reflect the model's ability to detect the minority class of interest. Therefore, this study emphasizes change pixel-based metrics to better evaluate the model's sensitivity to structural variations under imbalanced conditions.

In addition to pixel-level analysis, object-based (Blaschke, 2010) and contour-based (Maninis et al., 2018) metrics are employed to assess the geometric and structural integrity of detected building structures. Object-based analysis focuses on identifying individual building entities, enabling the removal of insignificant noise and evaluating the model's applicability in real-world urban monitoring scenarios. Contour-based analysis measures the precision of building boundaries, evaluating how effectively the model preserves fine structural details and edges. This comprehensive evaluation framework is essential for validating robust detection performance in high-precision mapping applications. By combining pixel-level, object-level, and contour-based metrics, the evaluation provides a more comprehensive assessment of both detection accuracy and structural consistency. Example of these multi-faceted metrics are provided in Figure 11.

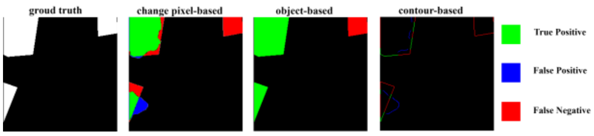

Figure 11. Examples of multi-faceted evaluation metrics.

3. Results

3.1 Analysis of Normalization Optimization

To determine the most effective input representation for 12-bit CAS500-1 imagery, three normalization techniques were compared. The quantitative results, categorized into overall pixel-based and change pixel-based metrics, are summarized in Table 2.

(a)	Prec	Rec	F1	IoU	Kappa
Percentile	0.8026	0.7269	0.7629	0.6167	0.7400
Gamma	0.7981	0.8007	0.7994	0.6659	0.7789
Log	0.7809	0.7733	0.7770	0.6354	0.7544
(b)	Prec	Rec	F1	IoU	Kappa
Percentile	0.6410	0.5856	0.5718	0.4833	0.5486
Gamma	0.6691	0.6450	0.6226	0.5345	0.6007
Log	0.6488	0.6438	0.6111	0.5179	0.5865

Table 2. Evaluation metrics of normalization optimization: (a) overall pixel-based, (b) change pixel-based.

As shown in Table 2, gamma correction consistently achieved the best performance across all evaluated metrics. In the overall pixel-based evaluation, it achieved the best F1-score (0.7994) and IoU (0.6659). In the change pixel-based evaluation, which more directly reflects the model's ability to detect structural changes, gamma correction achieved higher F1-score (0.6226) and IoU (0.5345) compared to the percentile and log methods.

The consistent performance across both overall and change-specific metrics suggests that gamma correction effectively preserves structural features while mitigating radiometric noise in 12-bit satellite imagery. Figure 12 illustrates these differences, where gamma-corrected images exhibit enhanced boundary clarity and contrast in dense urban shadow regions, supporting more reliable feature extraction. This indicates that gamma correction is particularly effective for handling the wide dynamic range and shadow variations inherent in high-resolution urban satellite imagery.

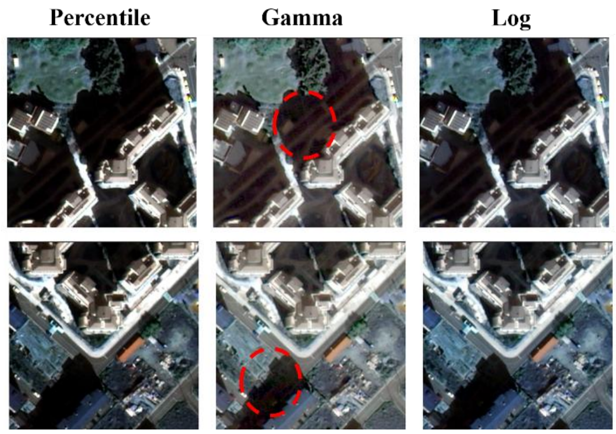

Figure 12. Visual comparison of building structures across different normalization techniques.

3.2 Impact of Data Augmentation Strategies

The second stage of optimization evaluates the effect of various data augmentation strategies on model generalization capability. Geometric augmentation serves as the baseline, while optical and temporal augmentations are introduced to simulate complex imaging conditions. The quantitative results for both overall pixel-based and change pixel-based metrics are presented in Table 3.

(a)	Prec	Rec	F1	IoU	Kappa
Geo	0.7981	0.8007	0.7994	0.6659	0.7789
Geo+Opt	0.8059	0.7988	0.8023	0.6699	0.7823
Geo+Tem	0.8082	0.7930	0.8006	0.6675	0.7804
All	0.7855	0.8169	0.8009	0.6679	0.7801
(b)	Prec	Rec	F1	IoU	Kappa
Geo	0.6691	0.6450	0.6226	0.5345	0.6007
Geo+Opt	0.6657	0.6567	0.6262	0.5347	0.6036
Geo+Tem	0.6621	0.6395	0.6198	0.5319	0.5981
All	0.6663	0.6769	0.6381	0.5487	0.6151

Table 3. Evaluation metrics of augmentation optimization: (a) overall pixel-based, (b) change pixel-based.

While individual strategies improved specific metrics, the comprehensive augmentation strategy achieved the best performance in terms of F1-score (0.6381) and IoU (0.5487) for change pixels. These results suggest that incorporating optical noise and temporal order swapping improves the model's ability to handle radiometric variations and diverse change patterns in multi-temporal satellite imagery. The increase in recall indicates improved sensitivity to structural changes in high-density urban environments. Therefore, the comprehensive augmentation strategy is adopted as the final configuration to ensure robust change detection performance. This improvement is particularly important in dense urban environments, where radiometric inconsistencies and structural complexity often lead to missed detections.

3.3 Evaluation of Ensemble Techniques

The final stage of the optimization pipeline evaluates various ensemble strategies to integrate predictions from the Tiny, Small, and Base variants of the ChangeMamba architecture. For each ensemble category, the best-performing configuration was

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLIX-B3-2026-601-2026 | © Author(s) 2026. CC BY 4.0 License.

selected for cross-comparison: Confidence Weighted average for average-based, descending for hierarchical-based, LGBM for stacking-based, and Random Forest for dynamic-based ensembles. The performance comparison across these strategies is shown in Figure 13.

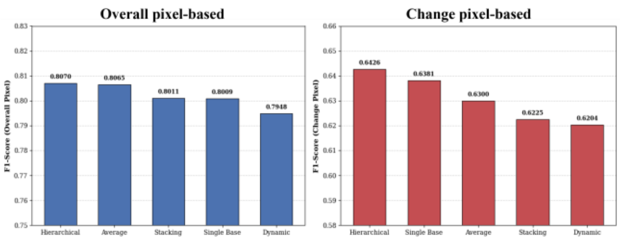

Figure 13. Performance comparison of different ensemble configurations in terms of F1-Score.

To further validate detection performance, we conducted a multi-faceted evaluation of the three top-performing models: the single Base model, the descending Hierarchical and the Weighted Average ensemble. The quantitative results across pixel-based, contour-based, and object-based metrics are summarized in Table 4.

(a)	Base single	Hierarchical	Average
F1	0.8009	0.8070	0.8065
IoU	0.6679	0.6764	0.6758
Kappa	0.7801	0.7788	0.7870
(b)	Base single	Hierarchical	Average
F1	0.6381	0.6426	0.6300
Prec	0.5487	0.5537	0.5442
Rec	0.6151	0.6195	0.6081
(c)	Base single	Hierarchical	Average
F1	0.5739	0.5774	0.5707
Prec	0.6077	0.6085	0.6142
Recall	0.5784	0.5868	0.5666
(d)	Base single	Hierarchical	Average
GT		983	
TP	790	789	766
FP	215	203	172

Table 4. Evaluation metrics of ensemble optimization: (a) overall pixel-based, (b) change pixel-based, (c) contour-based, (d) object-based.

As shown in Table 4, the Hierarchical Ensemble consistently achieved strong performance across most evaluation metrics. It achieved the highest F1-scores (0.8070 and 0.6426) in overall pixel-based and change pixel-based evaluations, respectively. Its advantage is particularly evident in the contour-based evaluation, achieving an F1-score of 0.5774 and recall of 0.5868, indicating improved boundary preservation.

In the object-based evaluation, the Hierarchical Ensemble reduced the number of False Positives (from 215 to 203) compared to the single Base model. This reduction in false alarms, along with high precision in change-pixel detection, indicates that the hierarchical strategy effectively filters noise while maintaining sensitivity to structural changes in dense urban environments. These improvements are visually supported in Figure 14, where the Hierarchical Ensemble reduces both False Positives (blue) and False Negatives (red) compared to the baseline. This result suggests that the sequential decision process effectively leverages the complementary strengths of multi-scale models, allowing high-confidence predictions to be fixed early while refining uncertain regions in later stages.

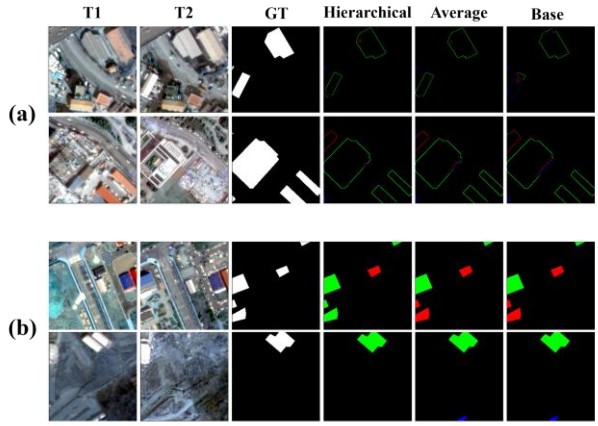

Figure 14. Qualitative comparison of change detection results: (a) contour-based, (b) object-based evaluations.

3.4. Overall Performance Analysis

The comprehensive evaluation of the final optimized pipeline demonstrates a clear performance improvement over the baseline configuration. By integrating key techniques at each stage—gamma correction for radiometric normalization, comprehensive augmentation for improved generalization, and the Hierarchical Ensemble for final decision-making—the proposed framework achieved a final F1-score of 0.8070. This corresponds to a 5.7% improvement over the baseline F1-score of 0.7629, which relied on standard normalization and a single-scale architecture. These results demonstrate the effectiveness of the proposed stepwise optimization framework in improving both detection accuracy and robustness in complex urban environments.

4. Conclusion

This study presents a stepwise optimization and ensemble pipeline for BCD in dense urban environments. By integrating CAS500-1 satellite imagery with a state-of-the-art Mamba-based architecture, the framework demonstrates strong potential for systematic land monitoring. Experimental results indicate that task-specific normalization and comprehensive augmentation play a key role in ensuring radiometric robustness. Building upon these components, multiple ensemble strategies were evaluated using pixel-, contour-, and object-based metrics to identify the most effective fusion approach. Finally, the Hierarchical Ensemble was identified as the most effective approach, leveraging the complementary strengths of multi-scale variants to achieve balanced and high-fidelity results. These results demonstrate the effectiveness of the proposed stepwise optimization pipeline for high-precision satellite-based change detection.

Despite these advancements, certain limitations remain to be addressed in future work. First, although advanced ensemble techniques such as Stacking and Dynamic Ensemble were introduced, their performance gains were limited in this study. This is primarily attributed to insufficient fine-tuning of their meta-learners and selection logic, suggesting that these complex strategies have not yet reached their full potential. To overcome this, future research should focus on a more rigorous optimization of hyperparameters and training protocols specifically tailored for these ensemble strategies. Second, the current approach relies on a single-architecture-based ensemble using only different scales of ChangeMamba, which may possess inherent structural biases. To move beyond the limitations of a single backbone, exploring multi-architecture ensemble strategies that integrate fundamentally different models, such as Diffusion models or Vision Transformers, is a promising direction.

Nevertheless, the proposed stepwise optimization workflows and the multi-faceted evaluation framework provide valuable insights and a methodological foundation for future extensions. The effective handling of high-resolution spectral characteristics and stable performance across various urban scenarios suggest that this work can serve as a practical baseline for future satellite missions and ensemble-based applications.

Acknowledgements

This work was supported by the Electronics and Telecommunications Research Institute (ETRI) grant (202502180001), and by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2023-00242528).

References

Al-Safar, A.A.Y., 2021. Enhancing spatial characteristics of satellite images. Int. Conf. Contemp. Inf. Technol. Math. (ICCITM), Mosul, Iraq, 191–196.

Bandara, W.G.C., Patel, V.M., 2022. A transformer-based siamese network for change detection. Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS), 207–210.

Blaschke, T., 2010. Object based image analysis for remote sensing. ISPRS J. Photogramm. Remote Sens., 65(1), 2–16.

Breiman, L., 2001. Random forests. Mach. Learn., 45(1), 5–32.

Chavez, P.S., 1996. Image-based atmospheric corrections—revisited and improved. Photogramm. Eng. Remote Sens., 62(9), 1025–1035.

Chen, H., Song, J., Han, C., Xia, J., Yokoya, N., 2024a. ChangeMamba: Remote sensing change detection with spatiotemporal state space model. IEEE Trans. Geosci. Remote Sens., 62, 1–20.

Chen, Z., Yang, B., Zhu, R., Dong, Z., 2024b. City-scale solar PV potential estimation on 3D buildings using multi-source RS data: A case study in Wuhan, China. Appl. Energy, 359, 122720.

Dietterich, T.G., 2000. Ensemble methods in machine learning. Int. Workshop Multiple Classifier Systems, 1–15.

Fang, S., Li, K., Shao, J., Li, Z., 2022. SNUNet-CD: A densely connected siamese network for change detection of VHR images. IEEE Geosci. Remote Sens. Lett., 19, 1–5. doi: 10.1109/LGRS.2021.3056416.

Ganaie, M.A., Hu, M., Malik, A.K., Tanveer, M., Suganthan, P.N., 2022. Ensemble deep learning: A review. Eng. Appl. Artif. Intell., 115, 105151.

Hao, X., Liu, L., Yang, R., Yin, L., Zhang, L., Li, X., 2023. A review of data augmentation methods of remote sensing image target recognition. Remote Sens., 15(3), 827.

Hosmer, D.W., Lemeshow, S., Sturdivant, R.X., 2013. Applied logistic regression. Wiley.

Hussain, M., Chen, D., Cheng, A., Wei, H., Stanley, D., 2013. Change detection from remotely sensed images: From pixel-based to object-based approaches. ISPRS J. Photogramm. Remote Sens., 80, 91–106.

Jacobs, R.A., Jordan, M.I., Nowlan, S.J., Hinton, G.E., 1991. Adaptive mixtures of local experts. Neural Comput., 3(1), 79–87.

Jin, D., Park, J.S., Jang, Y., Lee, S., Chi, J., 2025. GeoAI dataset for building change detection from CAS500-1. GEO DATA, 7(3), 216–229.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., 2017. LightGBM: A highly efficient gradient boosting decision tree. Adv. Neural Inf. Process. Syst., 30.

Khelifi, L., Mignotte, M., 2020. Deep learning for change detection in remote sensing images: Comprehensive review and meta-analysis. IEEE Access, 8, 126385–126400.

Li, J., He, W., Li, Z., Guo, Y., Zhang, H., 2025. Overcoming the uncertainty challenges in detecting building changes from remote sensing images. ISPRS J. Photogramm. Remote Sens., 220, 1–17.

Liasis, G., Stavrou, S., 2016. Satellite images analysis for shadow detection and building height estimation. ISPRS J. Photogramm. Remote Sens., 119, 437–450.

Liu, Y., Tian, Y., Zhao, Y., Yu, H., Xie, L., Wang, Y., Ye, Q., Liu, Y., 2024. Vmamba: Visual state space model. arXiv preprint arXiv:2401.10166.

Lu, D., Batistella, M., Moran, E., 2008. Integration of Landsat TM and SPOT HRG images for vegetation change detection in the Brazilian Amazon. Photogramm. Eng. Remote Sens., 74, 421–430.

Maninis, K.-K., Pont-Tuset, J., Arbeláez, P., Van Gool, L., 2018. Convolutional oriented boundaries: From image segmentation to high-level tasks. IEEE Trans. Pattern Anal. Mach. Intell., 40(4), 819–833.

National Geographic Information Institute,2023. Geospatial Information Platform. National Geographic Information Institute. https://map.ngii.go.kr Accessed 1 March 2026.

Sambasivan, N., Kapania, S., Highfill, H., Akrong, D., Paritosh, P., Aroyo, L.M., 2021. Everyone wants to do the model work, not the data work: Data cascades in high-stakes AI. Proc. CHI Conf. Hum. Factors Comput. Syst., 1–15.

Viana, C.M., Oliveira, S., Oliveira, S.C., Rocha, J., 2019. Land use/land cover change detection and urban sprawl analysis. In: Spatial Modeling in GIS and R for Earth and Environmental Sciences, Elsevier, 621–651.

Viola, P., Jones, M., 2001. Rapid object detection using a boosted cascade of simple features. Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 1–8.

Wolpert, D.H., 1992. Stacked generalization. Neural Netw., 5(2), 241–259.

Yu, X., Wu, X., Luo, C., Ren, P., 2017. Deep learning in remote sensing scene classification: A data augmentation enhanced convolutional neural network framework. GISci. Remote Sens., 54(5), 741–758.

Zhu, X.X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: A comprehensive review and list of resources. IEEE Geosci. Remote Sens. Mag., 5(4), 8–36.