

Article

Improving Land Cover Classification Accuracy in Satellite Imagery Using Artistic Styles

Taeyeon Won ¹  and Yang Dam Eo ^{2,*} ¹ Department of Technology Fusion Engineering, Konkuk University, Seoul 05029, Republic of Korea; teadone@konkuk.ac.kr² Department of Civil and Environmental Engineering, Konkuk University, Seoul 05029, Republic of Korea

* Correspondence: eoandrew@konkuk.ac.kr; Tel.: +82-2-450-3078

Featured Application

Improving land cover classification using satellite imagery and artistic imagery.

Abstract

This study addresses ambiguity in spatial extent and boundaries in satellite image classification to improve the accuracy of fine-grained object-based Level 2 land cover classification. Unlike conventional data augmentation, we propose a novel Style-Adaptive U-Net that incorporates the visual characteristics of landscape paintings into classification learning. Specifically, we developed a lightweight CNN-based Art Encoder coupled with an Enhanced Style Feature Fusion (ESFF) module to inject artistic features into the network's feature representation. Based on visual features extracted from works by Egon Schiele, Van Gogh, Claude Monet, and Elyse Dodge, the model utilizes painting styles with distinct boundaries or strong textures to explicitly enhance the boundary recognition capability of objects. Experimental results demonstrate the efficiency and superiority of the proposed model. It achieves a peak Dice score of 0.7631, outperforming the baseline U-Net's 0.6512, and maintains a manageable processing load with only a 19% increase in parameters. Our comparative analysis shows a distinct representational mechanism by demonstrating that styles with explicit structural features (Schiele, Dodge) improve boundary discrimination, whereas styles emphasizing blurred transitions (Monet) yield limited functional gain. This validates our premise that the network actively utilizes artistic features as functional structural guidance rather than mere aesthetic enhancements, offering an efficient paradigm for resolving geographic ambiguity.

Keywords: U-Net; art-style transfer; style injection; CNN art encoder; boundary-aware; segmentation; generalization



Academic Editors: Andrea Prati,
Weitao Chen, Ailong Ma and
Guohua Wu

Received: 22 January 2026

Revised: 23 February 2026

Accepted: 2 March 2026

Published: 4 March 2026

Copyright: © 2026 by the authors.
Licensee MDPI, Basel, Switzerland.
This article is an open access article
distributed under the terms and
conditions of the [Creative Commons
Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Object-based remote sensing image classification faces increasing limitations as spatial resolution increases, particularly when relying on traditional segmentation and edge-based techniques. Segmentation techniques frequently exhibit two types of errors, namely over-segmentation and under-segmentation, and the determination of a balanced scale according to image characteristics has a substantial impact on classification [1]. In Object-Based Image Analysis (OBIA), final feature extraction and classification strongly depend on the quality of image segmentation, and the key challenge in image segmentation lies in selecting optimal

parameters and algorithms that can generate image objects corresponding to meaningful geographic objects [2].

Recent studies attempted various approaches to automate this process. Object-characteristic-based adaptive scale determination methods [3,4] estimate optimal scales by utilizing the average size of objects or Moran's Index within images. While these approaches reduce empirical parameter settings, they struggle to satisfy all objects with a single scale when image heterogeneity is high. Quality metric-based optimization methods [5] automatically explore parameters using classification quality evaluation metrics such as NSQI and AFI, but quality metrics can be biased toward specific classes or image characteristics, limiting their generalizability across diverse terrain conditions.

Bayesian optimization combined approaches [6] jointly optimize CNN-based features and MRS parameters to enhance boundary recognition performance. Despite their effectiveness, Bayesian search is computationally expensive, and in high-resolution images, training time becomes prohibitively long, limiting real-time applicability. Multi-Criteria Decision Making (MCDM) and fuzzy logic-based algorithm selection methods [7] propose selecting the most suitable classification algorithm according to image characteristics. Although these methods increase objectivity in algorithm selection, they suffer from complex fuzzy rule design and limited adaptability to new image types.

Recently, approaches combining deep learning with hyperparameter optimization have attracted increasing attention [6]. These methods automatically adjust parameters during model training using CNN or Transformer-based architecture. However, deep-learning models require massive training data and significant computing resources, and they remain sensitive to dataset quality and class imbalance. Despite these research efforts, several challenges persist. First, parameters optimized for specific regions often exhibit poor performance in other regions, limiting generalization across diverse terrain and sensor characteristics. Second, computational efficiency constraints hinder real-time processing and large-scale application in high-resolution images. Third, boundary ambiguity has not been fully resolved, with errors particularly prominent at urban–rural boundaries with many mixed pixels.

For edge-based classification techniques, accurately extracting building boundary information during training remains challenging, which leads to ambiguity in boundary regions [8]. Holistic Attention Edge Detection (HAE-RNet) [9] explicitly learns boundary information by introducing an Edge Extraction block into a CNN-based segmentation network. While this approach is effective in correcting boundary locations, boundaries of small objects in complex urban environments often remain incomplete.

The Probability-Guided Edge Enhancement Network (PEEN) [10] proposes a structure that iteratively refines boundary prediction based on probability functions. This method demonstrates advantages in reducing boundary uncertainty, but the iterative operations have high computational costs and limit real-time applicability. The Transformer-based Global and Edge Enhanced Transformer (GE-Swin) [11] designs a Dual Decoder to jointly learn global context and boundary information. This framework improves the performance of boundary recognition and small object classification, but the large number of model parameters increases memory requirements and makes it difficult to design a lightweight model.

EMNet [12] enhances boundary recognition of small objects in UAV imagery by adding Edge Feature Branch to Deeplabv3+ and applying Multi-Level Upsampling. However, the inclusion of the Edge Branch increases network complexity and limits generalization across diverse sensor data. Next, Hi-ResNet [13] improves boundary refinement by introducing a Class-Agnostic Edge-Aware Loss. This method improves performance through Attention

modules and multi-scale feature fusion, but the problem of boundary confusion between similar classes in complex backgrounds persists.

Furthermore, deep learning requires a large and comprehensive training dataset to develop generalizable models, but generating well-sampled and reliable training data remains challenging, requiring significant human resources to minimize the impact of noisy data [14,15].

A major limitation of existing approaches is that object-based classification in high-resolution satellite imagery suffers from limited accuracy due to boundary ambiguity and mixed pixel problems. Previous approaches attempt improvements through parameter optimization or boundary enhancement algorithms, but boundary recognition remains incomplete in complex terrain and diverse objects environments. To address this limitation, we utilize style injection-based deep learning. This approach recognizes that artists' paintings exhibit distinct patterns in boundary representation, texture emphasis, and color contrast, which are visual features that CNN and Attention-based networks are capable of learning. These styles are not merely visual transformations but contain structural signals such as strong contrast, nonlinear textures, and boundary emphasis, which can be utilized to enhance boundary discrimination in high-resolution satellite image classification where mixed pixels and boundary ambiguity frequently occur.

Therefore, by combining a style encoder with a U-Net architecture, we simultaneously extract and integrate boundary and texture information at multiple scales, improving object classification accuracy and generalization performance. Experiments quantitatively validated the effectiveness of this approach.

2. Related Work: Style Transfer

Style transfer refers to the technique of transferring visual characteristics, such as style, texture, and color, from one image to the structural content of another. This technique is applied in a variety of fields, from artistic expression to data augmentation.

Initially, it was primarily implemented using CNN-based VGG networks, which combined statistical representations of style images derived from Gram Matrices and preserved the structural content of the source image [16]. As the field evolved, learning-based transformations using Generative Adversarial Networks (GANs) were introduced, enabling more consistent and flexible style representations that extend beyond single-image transformations.

GAN-based methods offer a distinct advantage in that they can capture the unique structural patterns and contrast characteristics associated with individual painting styles, rather than simply transferring texture or color distributions. This approach has been utilized to increase dataset diversity for edge recognition beyond simple visual variation and improve model performance [17].

Recent research efforts have focused on integrating transformer architectures with attention mechanisms, complementing existing CNN and GAN-based style transfer approaches. Unlike methods that rely solely on statistical transformations of feature maps, these approaches leverage the spatial relationships and semantic context inherent in the data, enabling more relevant style translation [18].

The introduction of the Vision Transformer (ViT) architecture marked a significant advance by enabling image processing at a patch-based granularity that captures long-range dependencies. This capability facilitates the simultaneous extraction and application of global style patterns, such as composition and contrast distribution, and local textures, including brush strokes and boundary representations. By addressing the local receptive field limitations inherent in CNN-based methods, ViT-based approaches improve style consistency and boundary stability [19].

Fine attention mechanisms, including self-attention and cross-attention, allow for precise pixel-level tuning of the mapping relationship between data and style. These attention-based architectures assign different weights to style information across specific regions, such as sky, water, and building boundaries, allowing the model to focus on semantically important areas while avoiding unnecessary distortion. In this context, attention serves as a key component in achieving semantic integration of data and style, rather than mere visual fusion [20].

The integration of Transformer and Attention modules demonstrates that models can simultaneously enhance global style understanding and local detail representation, achieving performance improvements over conventional CNN and GAN-based approaches. These improvements lead to meaningful gains in applications that require precise boundary recognition, including classification models and representation learning.

3. Experiment

3.1. Study Area and Satellite Data

We employed 4-channel WorldView-3 satellite imagery, including Red, Green, Blue, and Near-Infrared bands, together with corresponding medium-level land cover classification maps as experimental data. The satellite imagery was acquired on 27 October 2020, and the study area is defined by northwest coordinates (nwLat = 37.58390400, nwLong = 127.03631700) and southeast coordinates (seLat = 37.51983500, seLong = 127.11545400).

The corresponding land cover maps were obtained as shapefile vector data for the same region in 2020 through the Environmental Spatial Information Service of the Ministry of Climate, Energy and Environment (<https://aid.mcee.go.kr/map/map.do>, accessed on 1 August 2025). After confirming that the attribute table of the vector data contained Level 1, Level 2, and Level 3 codes, the coordinate system was converted to WGS 1984 to ensure consistency with the satellite imagery. The shape (.shp) data were then rasterized to the same spatial resolution as the satellite imagery, ensuring spatial alignment between the two datasets. The resulting satellite imagery and land cover maps are illustrated in Figures 1 and 2, respectively.



Figure 1. WorldView-3 satellite images used in the experiment.



Figure 2. Level 2 land cover map used in the experiment.

To construct a dataset with uniform image sizes, both the satellite imagery and land cover maps were divided into 512 px × 512 px patches. A total of 775 image patches were generated, of which 582 images were used for training and 193 images were used for testing, with the random seed fixed at 42 for consistency. This sampling was conducted uniformly across the entire dataset.

Figure 3 shows the mask distribution in the experimental area using a pie chart. Figure 4 shows the colors and legends visualized in the pie chart and Level 2 masks. These colors are listed in the official data of the Environmental Spatial Information Service of the Ministry of Climate, Energy, and Environment. Based on Level 2 codes, the dominant classes include transportation areas (15), inland water (71), residential areas (11), artificial grasslands (42), and commercial and business areas (13). These five classes account for 75.6% of the entire dataset.

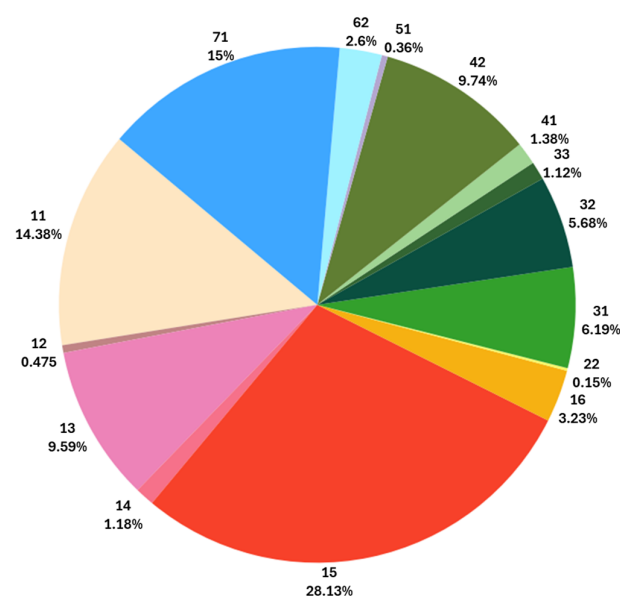


Figure 3. Distribution chart of Level 2 pixels used in the experiment. (Level 2 code, Pixel Ratio) Referring to Figure 4 for item names by Level 2 code.

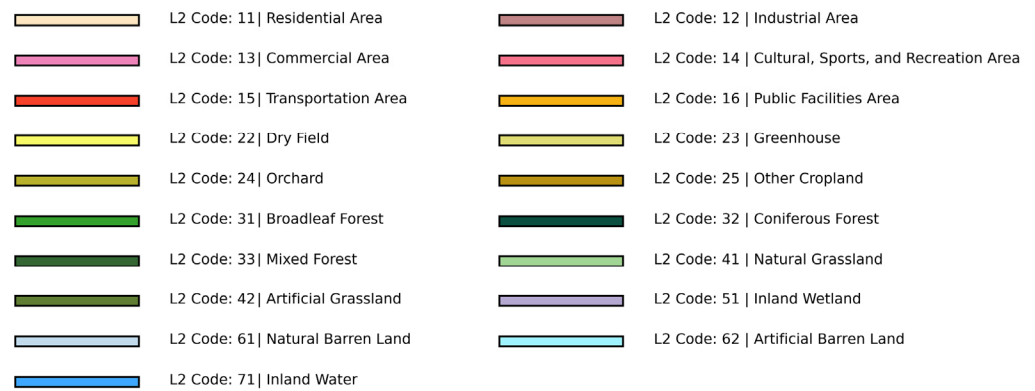


Figure 4. Colors and names according to Level 2 code.

3.2. Selected Artists and Their Visual Characteristics

For satellite image transformation through style injection, we built a training dataset by selecting four artists with diverse and unique painting styles. These selection criteria were expected to yield more discriminatory results, focusing on distinguishing between artists whose works exhibited blurred boundaries and softly diffused textures, and those whose works exhibited strong textures with distinct edges.

In this study, four painters were selected to inject style, each exhibiting a unique visual style, as evidenced by their unique brushstrokes, color contrast patterns, and boundary expressions. These artists are Claude Monet, Van Gogh, Elyse Dodge, and Egon Schiele, and their representative works can be seen in Figure 5.



Figure 5. Works of artists used in the study: (a) Claude Monet; (b) Van Gogh; (c) Elyse Dodge; (d) Egon Schiele.

Claude Monet, a leading Impressionist painter, is recognized for a smooth painting style. The predominant characteristics observed in his works include ambiguous boundaries, gradual color transitions, and a compositional structure that emphasizes overall visual continuity [21]. Injecting these characteristics is expected to facilitate analysis of how segmentation models respond to ambiguous boundaries.

In contrast, Van Gogh's style is characterized by bold brushwork and dynamic textures. His paintings typically exhibit distinct color contrasts and expressive lines, accompanied by irregular and nonlinear texture patterns [22]. By injecting these characteristics, the classification model is guided to learn adaptability to nonlinear and irregular spatial patterns commonly observed in satellite imagery.

Elyse Dodge represents a contemporary artistic style that emphasizes pronounced contrast and clear structural separation between forms [23]. This style is incorporated into training with the expectation that it encourages segmentation models to distinguish object boundaries more clearly, thereby enhancing boundary recognition capability.

Finally, Schiele is well known for strong contours and form-centered compositions, with dominant characteristics that include rough lines and extreme color contrasts [24].

Similar to the case of Dodge, these characteristics are expected to support form-based analysis and improve edge inference capability.

In essence, while Monet and Van Gogh emphasize visual impressions and sensory perception, Schiele and Dodge emphasize structural interpretation of forms and boundaries. This contrast in characteristics provides suitable experimental conditions for analyzing the impact of style injection on land cover classification problems. In this study, we examined how transforming satellite images using the visual characteristics of each artist, and subsequently injecting the features at different stages of feature extraction, influences land cover class inference.

3.3. Proposed Model Architecture and Experimental Setup

3.3.1. Style-Adaptive U-Net Architecture

Unlike conventional approaches that generate new images through style injection, this study focuses on directly embedding artistic style information as conditional input into the representation space of the U-Net model, as shown in Figure 6. Inspired by the style encoding and injection framework of StyleID, we designed a multi-scale feature injection architecture specialized for land cover classification, as illustrated in Figure 7a.

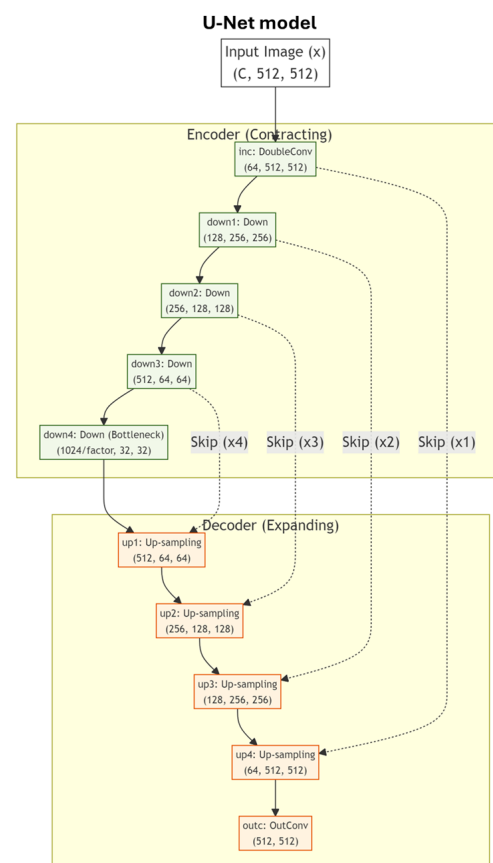


Figure 6. Structure of the U-Net model.

The standard U-Net was selected as the base architecture due to its balanced complexity and proven performance in various segmentation benchmarks. To effectively integrate artistic characteristics into this backbone, we designed a novel Art Encoder based on the feature extraction concepts of StyleID [25]. As shown in Figure 7a, the Art Encoder possesses a hierarchical structure that extracts diverse visual features across multiple scales through sequential convolutional stages and pooling operations.

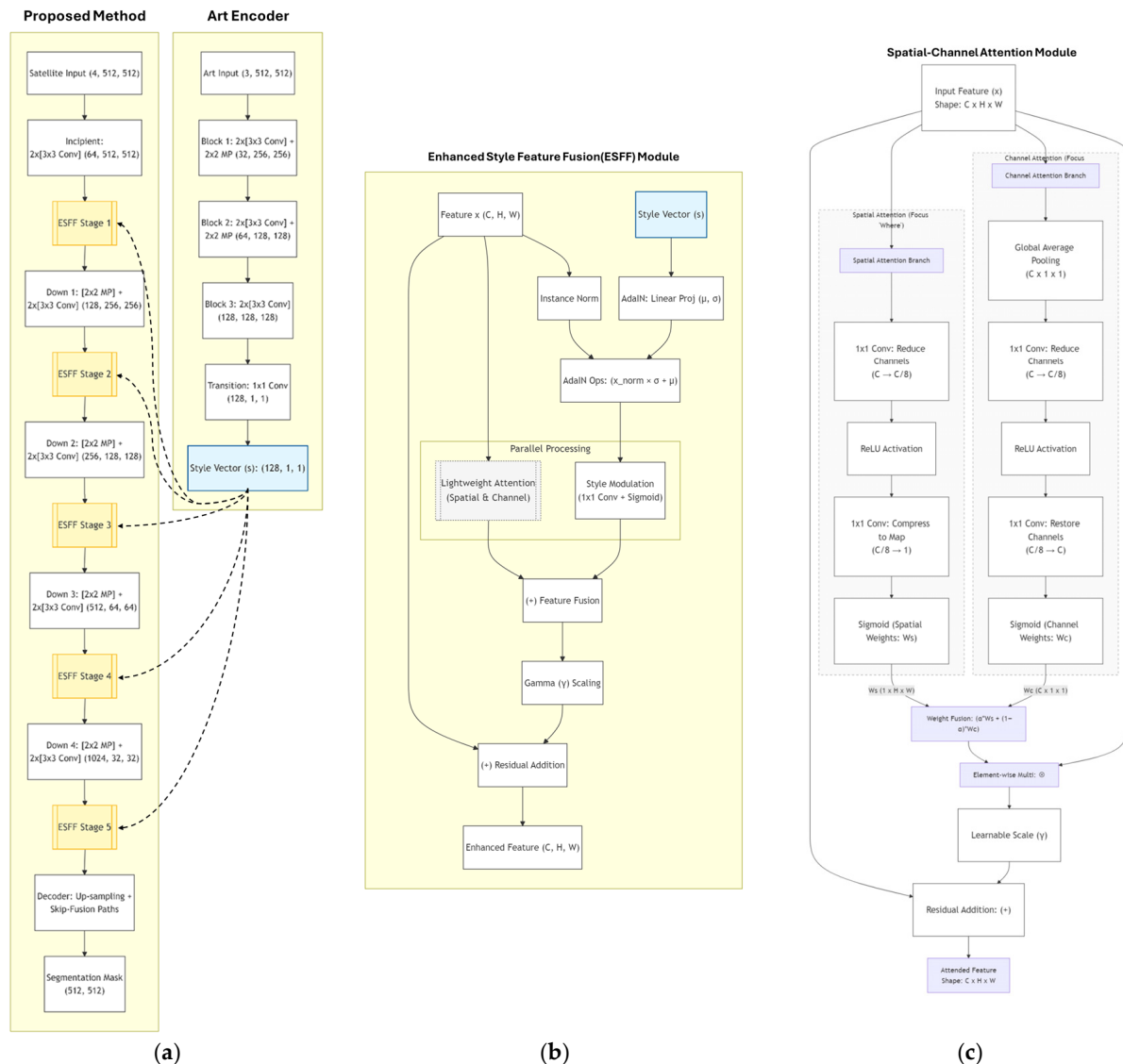


Figure 7. Structure of the models used in the experiment: (a) Structure of Proposed model; (b) Structure of ESFF module; (c) Structure of Spatial-Channel Attention module.

3.3.2. Technical Mechanisms: AdaIN and Attention

To prioritize model lightweight design and training efficiency, we opted for a CNN-based Art Encoder rather than a Transformer-based architecture. The stylistic features generated by this encoder are injected into each stage of the U-Net model through an injection module that combines Adaptive Instance Normalization (AdaIN) and Attention mechanisms, as shown in Figure 7b,c:

- **Adaptive Instance Normalization (AdaIN):** This mechanism allows the model to learn the statistical characteristics specifically the mean and variance of style images while preserving the original structural form of the satellite imagery [26]. It enables the network to capture distinctive style characteristics without introducing unwanted color distortions;
- **Spatial and Channel-specific Attention:** To mitigate potential information loss during feature synthesis, we integrated attention modules that selectively emphasize spatial regions and channels with high semantic significance. This directs the model's focus toward regions with clear boundaries or complex textures, enhancing boundary identification while maintaining computational efficiency.

3.3.3. End-to-End Style Integration

By utilizing the Enhanced Style Feature Fusion (ESFF) module in Figure 7b, the stylistic feature maps are combined with the U-Net's convolutional and up-convolutional modules, ensuring continuous interaction between content and style information throughout the entire network. By propagating these features via skip connections, the model learns to utilize stylistic information internally rather than relying on simple external mapping. This end-to-end synergy allows the model to leverage both local and global artistic cues to effectively resolve spatial uncertainties and boundary ambiguities in land cover classification.

3.3.4. Experimental Setup and Implementation

In this study, we applied the basic U-Net model and four style injection intensities (soft, balanced, strong, and ultra-strong) to four artists, for a total of 17 experimental configurations. All learning parameters, except for those required by the proposed methodology, were set identically across all experiments to ensure a fair comparison. The specific learning parameter settings for the basic U-Net model and each style injection intensity are summarized in Table 1. To prioritize computational efficiency, the proposed Style-Adaptive U-Net was designed with a streamlined parameter count of 36,976,328, representing a moderate 19% increase compared to the baseline U-Net (31,039,379 parameters). The model size expanded from 118.4 MB to 141.1 MB, which remains manageable for high-resolution satellite image processing tasks.

Table 1. Parameters used in training.

Common Parameters		Individual Parameters					
Epoch	300	Experiment name	U-Net	soft	balanced	strong	ultra-strong
Batch size	4	Art style	N/A		Gogh, Monet, Dodge, Schiele		
Learning rate	1×10^{-5}	Style gamma	N/A	1.0	1.2	1.5	2.0
Scale	1.0	Style temperature	N/A	1.0	0.6	0.4	0.3
Number of classes	19	Adaptive scale	N/A	0.8	1.0	1.2	1.5
Input channels	4	Detail weight initialization	N/A	Uniform	Ones	Random	Random

In this framework, style gamma functions as a coefficient that controls the influence range of style signals during model training, thereby determining the intensity at which style features are reflected. Style temperature is a parameter that adjusts the sharpness of style feature distributions, while adaptive scale controls the magnitude of style feature influence, with higher values increasing the impact of style characteristics. Detail weight initialization refers to the initial weight settings for detailed texture elements to which styles are applied. Training and inference were performed for all 17 experimental configurations, and model performance was evaluated using Accuracy, Dice score, and Buffer-based Accuracy.

4. Results

We conducted a total of 17 experiments using the base U-Net model and our style injection-based U-Net model. Each experiment was configured by applying four style injection strengths, namely soft, balanced, strong, and ultra-strong, to four artist styles (Monet, Van Gogh, Schiele, and Dodge). Performance evaluation was conducted using Dice score, Accuracy, and Buffer-based Accuracy metrics with tolerance ranges of 1 px, 2 px, and 3 px. Dice score and Accuracy are employed as standard metrics for evaluating land cover classification performance, while Buffer-based Accuracy is additionally applied to assess boundary recognition performance by introducing a tolerance range around object boundaries. All experimental results are summarized in Table 2.

Table 2. Result of 17 experiments.

Experiment Name	Dice Score	Accuracy	Buffer-Based Accuracy		
			Buffer_1	Buffer_2	Buffer_3
Baseline U-Net	0.6512	0.7137	0.7335	0.7509	0.7659
Gogh balanced	0.7038	0.7079	0.7265	0.7426	0.7565
Gogh soft	0.7211	0.7444	0.7645	0.7819	0.7966
Gogh strong	0.7468	0.7478	0.7682	0.7860	0.8014
Gogh ultra-strong	0.7494	0.7391	0.7597	0.7779	0.7936
Monet balanced	0.7319	0.7162	0.7365	0.7548	0.7710
Monet soft	0.7328	0.7504	0.7710	0.7891	0.8046
Monet strong	0.7245	0.7473	0.7677	0.7855	0.8009
Monet ultra-strong	0.7410	0.7494	0.7697	0.7875	0.8028
Dodge balanced	0.754	0.7289	0.7492	0.7675	0.7835
Dodge soft	0.7484	0.7427	0.7634	0.7818	0.7978
Dodge strong	0.7422	0.7494	0.7700	0.7881	0.8038
Dodge ultra-strong	0.6946	0.7379	0.7581	0.7757	0.7908
Schiele balanced	0.7440	0.7515	0.7723	0.7908	0.8070
Schiele soft	0.7441	0.7508	0.7716	0.7899	0.8059
Schiele strong	0.7416	0.7489	0.7692	0.7867	0.8015
Schiele ultra-strong	0.7631	0.7428	0.7631	0.7807	0.7958

As shown in Table 2, most style injection models achieved higher Dice scores and Accuracy compared to the base U-Net model. Among all experimental configurations, the Schiele style with ultra-strong settings attained the highest Dice score of 0.7631, whereas the Schiele style with balanced settings yielded the best performance in both Accuracy and Buffer-based Accuracy metrics.

With respect to Buffer-based Accuracy, style injection models demonstrated consistently higher performance than the base U-Net across all tolerance ranges (1 px, 2 px, and 3 px), and the performance values increased as the tolerance range expanded.

When comparing results by artist style, experiments applying the Monet style exhibited relatively lower Dice scores and Accuracy compared to other styles. In contrast, experiments applying Van Gogh, Schiele, and Dodge styles showed higher Dice scores than the base U-Net across most style injection strength settings.

Results based on style injection intensity showed that, for most configurations, Dice score and buffer-based accuracy tended to improve as injection intensity increased from soft to balanced, strong, and ultra-strong. However, some experiments achieved higher performance with balanced or strong settings than with ultra-strong.

5. Discussion

In this study, we compared and analyzed U-Net-based classification models with various artist styles and style injection intensities. As a result, we confirmed that most adaptive style injection models achieved higher Dice score and accuracies than the baseline U-Net model. In particular, the Schiele style with the ultra-strong setting recorded a Dice score of 0.7631, showing the best performance among 17 experimental configurations. The buffer-based accuracy index also showed an overall increasing trend at tolerance levels of 1 px, 2 px, and 3 px, and the Schiele style with the balanced setting showed the best performance. These results demonstrate that the boundary structure of the classification class remains stable even when a limited boundary tolerance is applied, suggesting that adaptive style injection contributes to improving boundary recognition and structural consistency. This suggests that the structural cues provided by Schiele's unique and distinct

style serve as more meaningful cues to the network than vaguely changing color-based features, enabling more robust feature representations in complex landscapes.

The performance boost observed in this study is not merely a result of adding visual variety, but a consequence of a deliberate architectural choice. By substituting the compute-intensive Transformer encoders typically used in style transfer with a custom-built, lightweight CNN Art Encoder, we pinpointed local patterns and structural signals—such as global mood and gritty textures—essential for land cover classification. This integration into the U-Net’s multi-stage architecture creates a seamless internal synergy that bridges the gap between artistic texture and geographic content through end-to-end learning. Crucially, while the addition of the Art Encoder resulted in a moderate 19% increase in parameter count, this complexity remains significantly lower than that of heavy Transformer-based models, yet provides substantial gains in boundary recognition and classification accuracy. This balance demonstrates that the proposed method serves as a functional and efficient enrichment of the model’s interpretative capacity rather than a cosmetic overlay.

Land cover classification results, especially, remained remarkably stable even under the ultra-strong style injection settings. Across all experimental configurations, the semantic consistency of the output masks was preserved, indicating that the incorporation of artistic style features does not compromise the inherent stability of the base classifier. Instead, style injection appears to enrich the model’s representation space without inducing class collapse or semantic distortion. This enhanced generalization capability is further supported by the buffer-based accuracy results. As the buffer tolerance increased from 1 pixel to 3 pixels, model performance consistently and steadily improved, suggesting that the proposed approach effectively mitigates minor boundary localization errors that commonly occur in high-resolution satellite imagery. These findings imply that the proposed method extends beyond the simple transfer of artistic elements and serves as a functional mechanism for addressing subtle boundary ambiguities.

In particular, spatial uncertainty frequently observed in urban–rural transition zones with ambiguous boundaries were more effectively resolved using the proposed model. By leveraging artistic styles as structured learning signals, the model improves geographic precision in challenging boundary regions. Consequently, the Schiele-style injection produced substantially sharper and more coherent boundaries compared to both the Monet-style injection and the baseline U-Net model, reinforcing the importance of boundary-emphasizing artistic characteristics for land cover classification tasks.

The differences in results when using works by different artists demonstrate the important relationship between visual features and classification accuracy. Monet’s fluid color transitions and blurred boundaries, while aesthetically pleasing, provided little functional benefit for edge recognition, resulting in lower Dice scores. In contrast, Schiele and Dodge’s strong contrasts and sharp outlines served as a kind of structural signal to the network. These distinct stylistic elements directly compensate for the mixed pixel problem, providing a stronger discriminative signal. As a result, the Schiele style injection results produced much sharper and more consistent boundaries compared to the Monet style injection and the baseline U-Net. The lack of functional gain in the blurred, changing styles further emphasizes that the network utilizes artistic cues as structural guidance rather than aesthetic enhancement. Consequently, the diffusion of edges in Monet’s style limits the encoder’s discriminatory power.

However, although there was a tendency for boundary perception to improve as stylistic cues were strengthened, this was not the case in some results. Optimal performance often emerged at balanced config or Strong config rather than the ultra-strong config extreme. This indicates a critical threshold where artistic textures transition from being a helpful discriminative cue to a source of spectral distortion. Therefore, finding these optimal

configs is important, as over-injection can distort key features of the original image. Thus, the effectiveness of the proposed method lies in its ability to balance stylistic emphasis with the preservation of the source data's geographic integrity. Our findings demonstrate that while stylistic features can improve boundary ambiguity, there is a quantifiable threshold above which stylistic noise begins to overwhelm geographic content, and that careful empirical tuning of injection parameters is required for practical applications.

Beyond the immediate accuracy gains, the core strengths of this framework are its robustness and generalizability. The model maintained stable land cover inference without class structure collapse, regardless of the artistic style or injection intensity applied. This confirms that artistic style injection serves as a functional enrichment of the model's interpretive capacity rather than a cosmetic overlay. This adaptability suggests that these artistic cues can be successfully applied to a variety of geographic regions and sensor types, opening up a new paradigm in which non-traditional visual data, such as artwork, serve as powerful learning cues for remote sensing applications.

Figures 8–13 below show a comparison table for experimental results. This table includes satellite images with histogram equalization for easy pixel separation, real-world data, U-Net model results, Schiele's super-strong model results, Schiele's balanced model results, and Monet's balanced model results. Furthermore, the visualized scenes were selected to contain many Level 2 classes or be distinct urban areas.

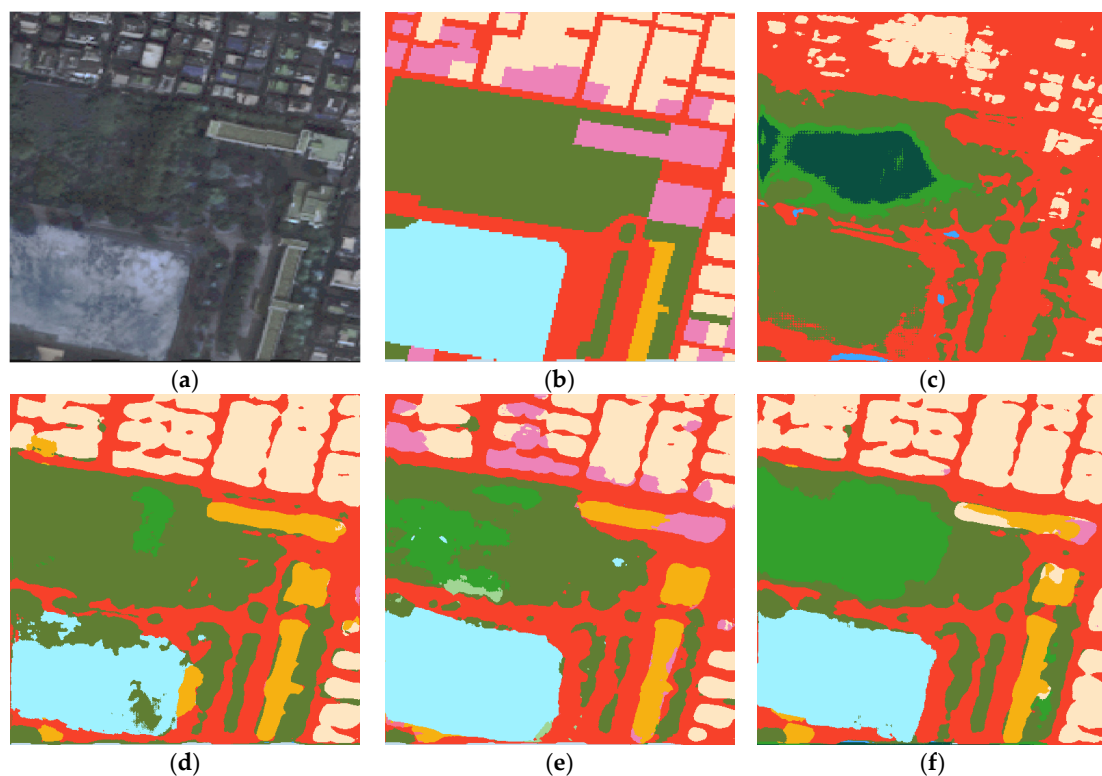


Figure 8. Grid images comparing results for scene 2: (a) satellite imagery; (b) ground truth; (c) U-Net; (d) Schiele ultra-strong; (e) Schiele balanced; (f) Monet balanced.

As can be seen in Figures 8, 11 and 13, for classes that accounted for a small part of the training data, there were frequent instances of misestimation as majority class pixels or of them being fragmented and estimated as different classes. Misestimation also frequently occurred for classes that occurred due to simple color differences. This problem occurred across all models, from the basic U-Net model to the proposed methodology, and it was assumed that this was due to an insufficient amount of training data for the relevant classes. However, as can be seen in Figure 10, the Schiele ultra-strong experiment produced clear

boundary results, and fewer fragments were observed for chunky pixels. In general, the results of the proposed methodology mostly showed better classification results than the results of the existing U-Net model.

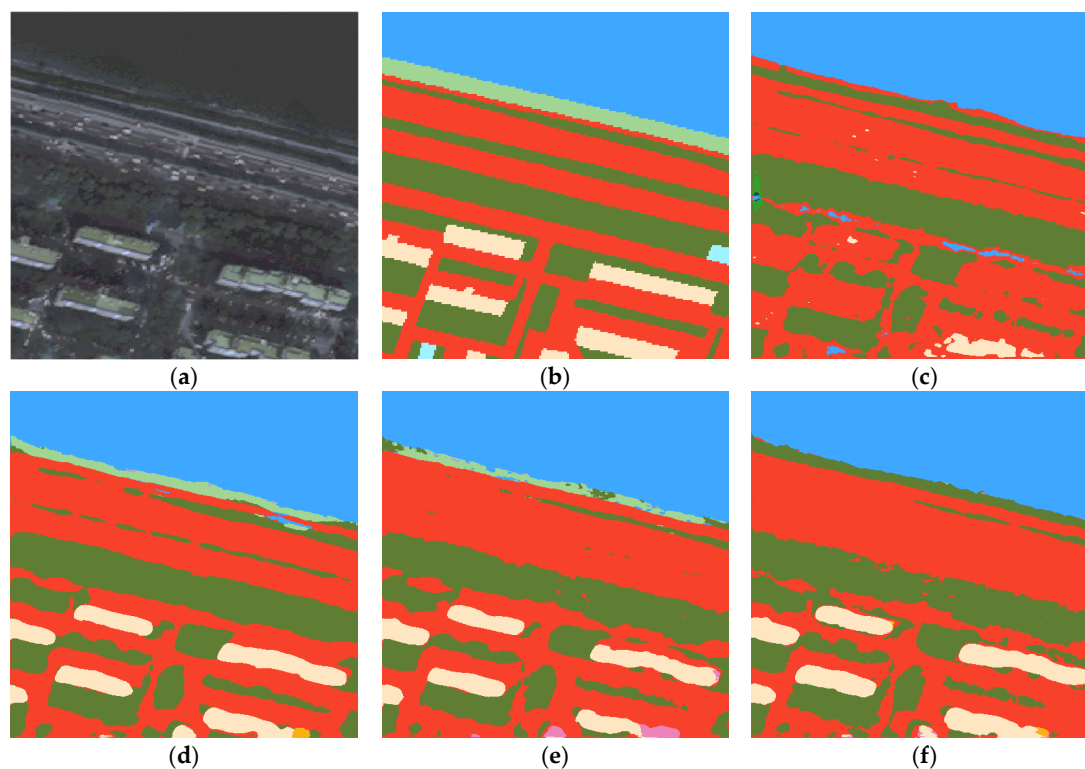


Figure 9. Grid images comparing results for scene 125: (a) satellite imagery; (b) ground truth; (c) U-Net; (d) Schiele ultra-strong; (e) Schiele balanced; (f) Monet balanced.

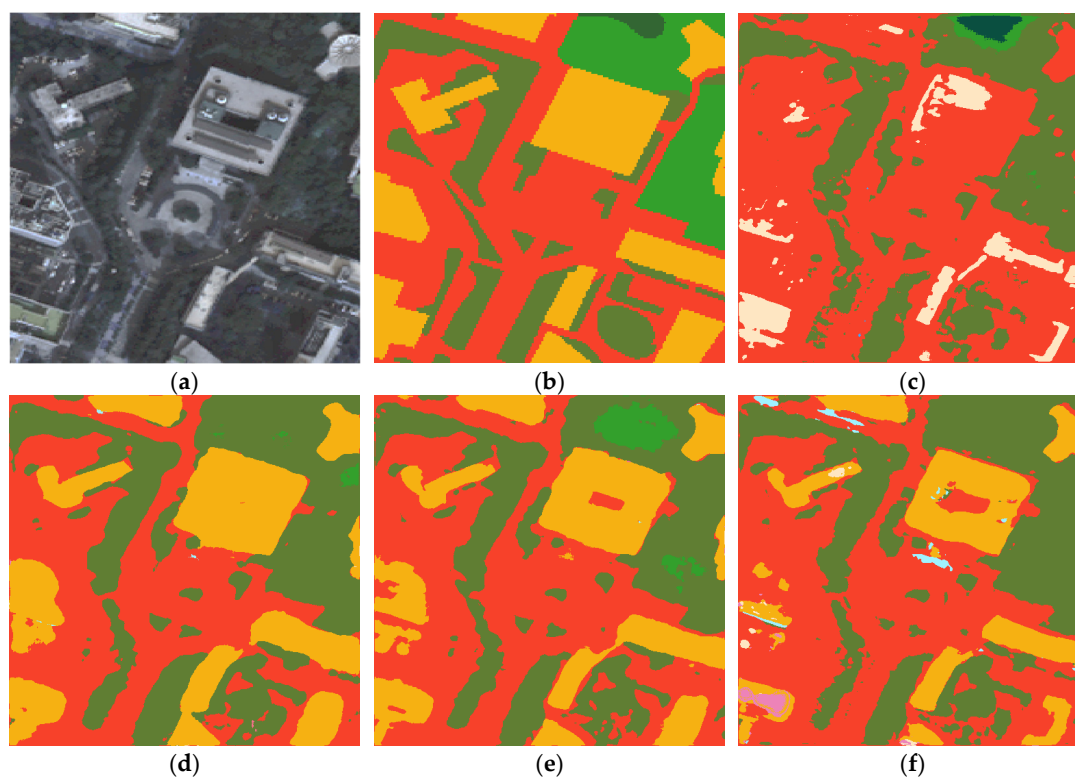


Figure 10. Grid images comparing results for scene 262: (a) satellite imagery; (b) ground truth; (c) U-Net; (d) Schiele ultra-strong; (e) Schiele balanced; (f) Monet balanced.

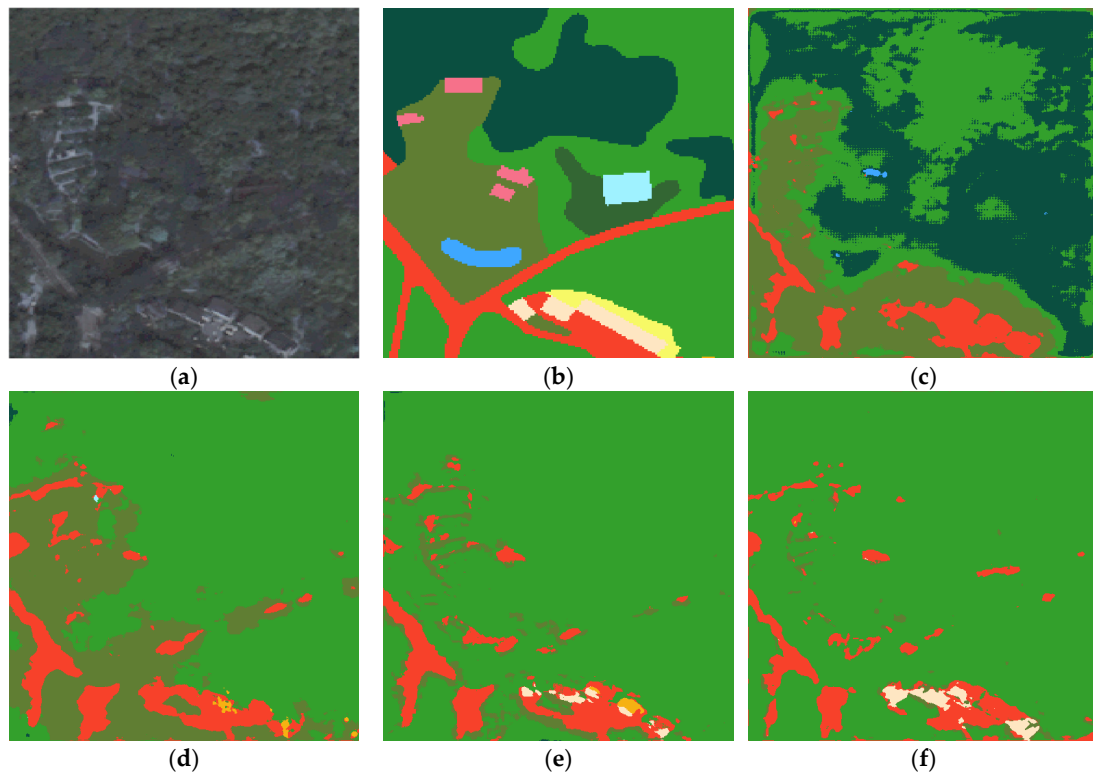


Figure 11. Grid images comparing results for scene 397: (a) satellite imagery; (b) ground truth; (c) U-Net; (d) Schiele ultra-strong; (e) Schiele balanced; (f) Monet balanced.

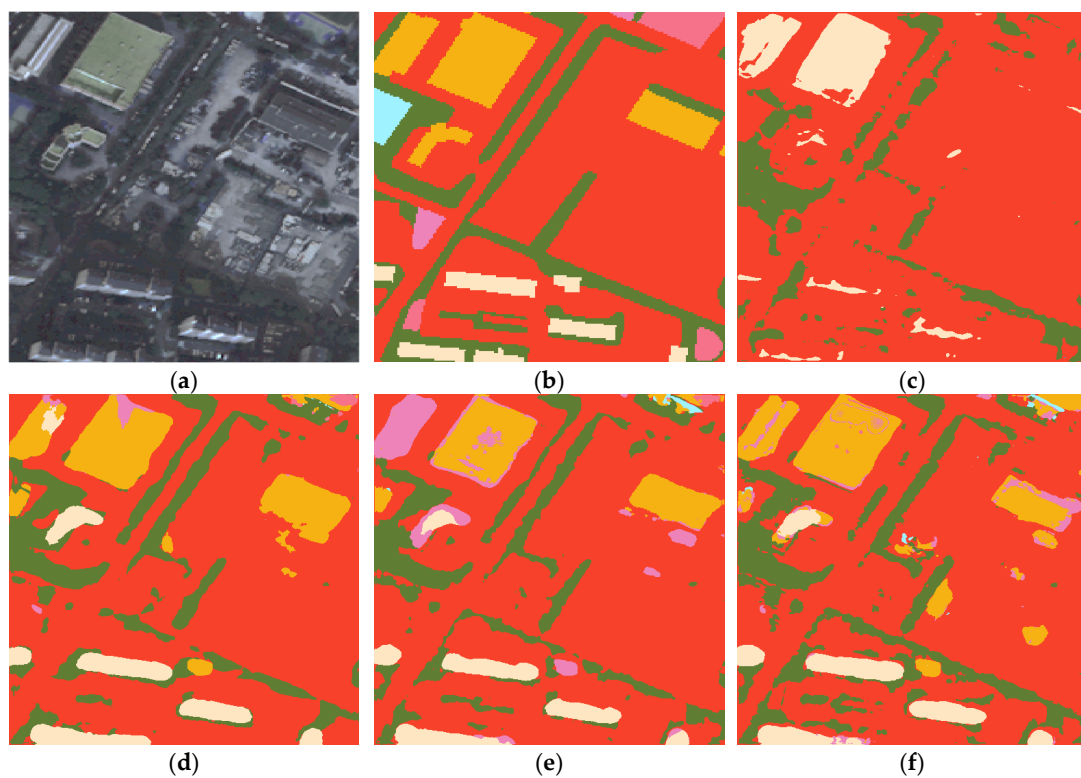


Figure 12. Grid images comparing results for scene 559: (a) satellite imagery; (b) ground truth; (c) U-Net; (d) Schiele ultra-strong; (e) Schiele balanced; (f) Monet balanced.

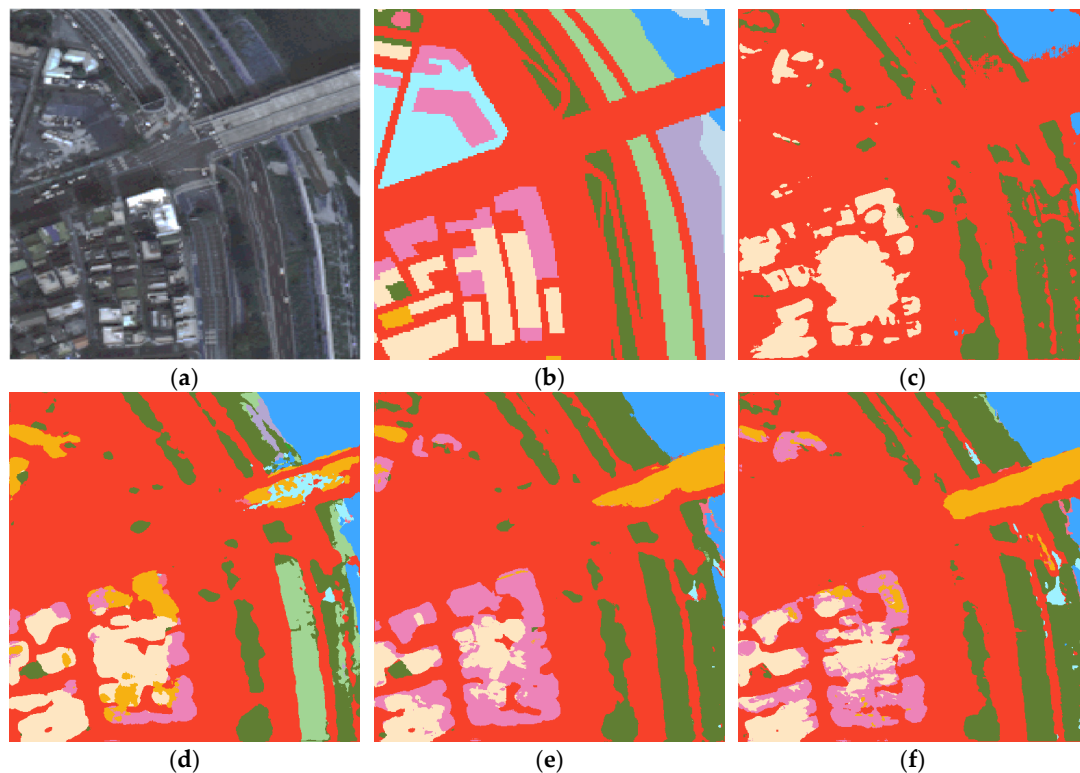


Figure 13. Grid images comparing results for scene 728: (a) satellite imagery; (b) ground truth; (c) U-Net; (d) Schiele ultra-strong; (e) Schiele balanced; (f) Monet balanced.

6. Conclusions

We proposed a method for injecting artistic style information into U-Net-based land cover classification models and evaluated its effectiveness through experiments with various artist styles and style injection strengths. To achieve this, we designed a CNN-based Art Encoder, applied a multi-stage style injection structure that injects style features at multiple stages of the U-Net encoder–decoder, and trained the entire model in an end-to-end manner.

Most style injection models achieved consistent performance improvements in Dice score and Accuracy compared to the baseline U-Net model. Especially when applying artist styles with distinct visual characteristics, namely Schiele, Van Gogh, and Dodge, boundary recognition performance improved overall, and Buffer-based Accuracy metrics also outperformed higher values than those of the baseline U-Net and Monet style injection models. These results demonstrate that the proposed model maintains structurally stable prediction results even when allowing pixel-level error ranges of 1 to 3 pixels around the boundary. In contrast, the Monet style yielded performance comparable to existing methods or improvements within the margin of error but showed relatively lower performance overall. This outcome suggests that expression characteristics emphasizing blurred boundaries and smooth color transitions have limited effectiveness for boundary recognition, which constrains classification performance.

To summarize, our results demonstrate that artistic style injection can effectively enhance the edge recognition capability of land cover classification models, going beyond simple visual transformation or data augmentation. By leveraging the structural and textural characteristics of the styles of artworks as functional learning cues, the proposed approach demonstrates the potential to improve generalization performance across diverse image conditions. However, because our experiments used data from a single region and sensor, generalization performance across heterogeneous sensors and different regions was not directly evaluated. Future research should utilize datasets from diverse regions and

sensor types and integrate style injection with transformer- and diffusion-based methods to verify more robust generalization performance.

Author Contributions: Conceptualization, Y.D.E.; methodology, Y.D.E. and T.W.; software, T.W.; validation, T.W.; formal analysis, T.W.; investigation, T.W.; resources, Y.D.E.; data curation, T.W.; writing—original draft preparation, T.W.; writing—review and editing, Y.D.E.; visualization, T.W.; supervision, Y.D.E.; project administration, Y.D.E.; funding acquisition, Y.D.E. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIT) (No. 2019R1A2C1085618).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: While writing this manuscript/research, the author used ChatGPT-5.2 for English proofreading and contextual editing. The authors reviewed and edited the results and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Liu, D.; Xia, F. Assessing object-based classification: Advantages and limitations. *Remote Sens. Lett.* **2010**, *1*, 187–194. [\[CrossRef\]](#)
2. Hossain, M.D.; Chen, D. Segmentation for Object-Based Image Analysis (OBIA): A review of algorithms and challenges from remote sensing perspective. *ISPRS J. Photogramm. Remote Sens.* **2019**, *150*, 115–134. [\[CrossRef\]](#)
3. Hao, S.; Cui, Y.; Wang, J. Segmentation Scale Effect Analysis in the Object-Oriented Method of High-Spatial-Resolution Image Classification. *Sensors* **2021**, *21*, 7935. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Ikokou, G.B.; Smit, J. A Technique for Optimal Selection of Segmentation Scale Parameters for Object-oriented Classification of Urban Scenes. *S. Afr. J. Geomat.* **2013**, *2*, 358–369.
5. Chen, Y.; Chen, Q.; Jing, C. Multi-resolution segmentation parameters optimization and evaluation for VHR remote sensing image based on meanNSQI and discrepancy measure. *J. Spat. Sci.* **2021**, *66*, 258–278. [\[CrossRef\]](#)
6. Azeez, O.S.; Shafri, H.Z.M.; Alias, A.H.; Haron, N.A. A Joint Bayesian Optimization for the Classification of Fine Spatial Resolution Remotely Sensed Imagery Using Object-Based Convolutional Neural Networks. *Land* **2022**, *11*, 1905. [\[CrossRef\]](#)
7. Janusonis, E.; Kazakeviciute-Januskeviciene, G.; Bausys, R. Selection of Optimal Segmentation Algorithm for Satellite Images by Intuitionistic Fuzzy PROMETHEE Method. *Appl. Sci.* **2024**, *14*, 644. [\[CrossRef\]](#)
8. Hosseinpour, H.; Samadzadegan, F.; Javan, F.D. A Novel Boundary Loss Function in Deep Convolutional Networks to Improve the Buildings Extraction from High-Resolution Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 4437–4454. [\[CrossRef\]](#)
9. Zhang, Y.; Yang, R.; Dai, Q.; Zhao, Y.; Xu, W.; Wang, J.; Wang, L. Boosting Semantic Segmentation of Remote Sensing Images by Introducing Edge Extraction Network and Spectral Indices. *Remote Sens.* **2023**, *15*, 5148. [\[CrossRef\]](#)
10. Yu, C.; Zuo, Y.; Zhang, Q.; Wang, Y. Probability-Guided Edge Enhancement Network for Remote Sensing Image Semantic Segmentation. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5519717. [\[CrossRef\]](#)
11. Wang, H.; Li, X.; Huo, L.; Hu, C. Global and edge enhanced transformer for semantic segmentation of remote sensing. *Appl. Intell.* **2024**, *54*, 5658–5673. [\[CrossRef\]](#)
12. Li, X.; Li, Y.; Ai, J.; Shu, Z.; Xia, J.; Xia, Y. Semantic segmentation of UAV remote sensing images based on edge feature fusing and multi-level upsampling integrated with Deeplabv3+. *PLoS ONE* **2023**, *18*, e0279097. [\[CrossRef\]](#)
13. Chen, Y.; Fang, P.; Zhong, X.; Yu, J.; Zhang, X.; Li, T. Hi-ResNet: Edge Detail Enhancement for High-Resolution Remote Sensing Segmentation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 15024–15040. [\[CrossRef\]](#)
14. Hosseiny, B.; Mahdianpari, M.; Hemati, M.; Radman, A.; Mohammadimanesh, F.; Chanussot, J. Beyond Supervised Learning in Remote Sensing: A Systematic Review of Deep Learning Approaches. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 1035–1052. [\[CrossRef\]](#)
15. Gütter, J.; Kruspe, A.; Zhu, X.X.; Niebling, J. Impact of Training Set Size on the Ability of Deep Neural Networks to Deal with Omission Noise. *Front. Remote Sens.* **2022**, *3*, 932431. [\[CrossRef\]](#)

16. Gatys, L.A.; Ecker, A.S.; Bethge, M. Image Style Transfer Using Convolutional Neural Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2414–2423. [\[CrossRef\]](#)
17. Chen, X.; Xu, C.; Yang, X.; Song, L.; Tao, D. Gated-GAN: Adversarial Gated Networks for Multi-Collection Style Transfer. *arXiv* **2019**, arXiv:1904.02296. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Deng, Y.; Tang, F.; Dong, W.; Ma, C.; Pan, X.; Wang, L.; Xu, C. StyTr2: Image Style Transfer with Transformers. *arXiv* **2022**, arXiv:2105.14576. [\[CrossRef\]](#)
19. Wei, H.P.; Deng, Y.Y.; Tang, F.; Pan, X.J.; Dong, W.M. A Comparative Study of CNN- and Transformer-Based Visual Style Transfer. *J. Comput. Sci. Technol.* **2022**, *37*, 601–614. [\[CrossRef\]](#)
20. Wang, J.; Yang, H.; Fu, J.; Yamasaki, T.; Guo, B. Fine-Grained Image Style Transfer with Visual Transformers. *arXiv* **2022**, arXiv:2210.05176. [\[CrossRef\]](#)
21. Yao, G.; Abindinhazir, Z. A Study and Appreciation of Claude Monet’s Artistic Creation and Life Experience. *Ideology J.* **2023**, *8*, 31–36. [\[CrossRef\]](#)
22. Schapiro, M. *Van Gogh*; Harry N. Abrams: New York, NY, USA, 1994.
23. Elyse Dodge. Available online: <https://www.elysedodge.com> (accessed on 15 January 2026).
24. Smith, K.A. Egon Schiele’s Treescapes. Work and world: Unframing the autonomous landscape. *Art Hist.* **2000**, *23*, 233–261. [\[CrossRef\]](#)
25. Chung, J.; Hyun, S.; Heo, J.P. Style Injection in Diffusion: A Training-free Approach for Adapting Large-scale Diffusion Models for Style Transfer. *arXiv* **2024**, arXiv:2312.09008. [\[CrossRef\]](#)
26. Huang, X.; Belongie, S. Arbitrary Style Transfer in Real-Time with Adaptive Instance Normalization. *arXiv* **2017**, arXiv:1703.06868. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.