Journal of Engineering, Project, and Production Management
2026, 16(6), 2026-0077

Color Transfer for Map Faceted Elements Based on an Enhanced Generative Adversarial Network with Attention Mechanisms

Bei Xie[1] and Dan Zeng[2]

[1] Associate Professor, The School of Visual Art, Hunan Mass Media Vocational and Technical College, Changsha, 410000, China, E-mail: dzengd@outlook.com (corresponding author).

[2] Professor, The School of Visual Art, Hunan Mass Media Vocational and Technical College, Changsha, 410000, China

Project Management
Received June 2, 2026; revised July 31, 2026; accepted August 8, 2026
Available online August 30, 2026

Abstract: Color transfer for map faceted elements in cartography has long relied on manual efforts, which are inefficient and highly subjective, making it difficult to balance color accuracy with structural fidelity. This study integrates generative adversarial networks with two types of attention mechanisms to propose a color transfer scheme tailored for map aerial elements. The method optimizes the generator using residual units to preserve spatial details and employs gradient normalization to enhance the discriminator's training stability. Channel attention governs the overall tonal distribution, whereas spatial attention focuses on regional boundaries, ensuring that the transferred results align closely with real maps in both color and shape. To evaluate performance, we used three metrics: mean Intersection over Union (mIoU), pixel error, and visual quality metrics. The mIoU measures semantic consistency in color distribution, while pixel error quantifies color reproduction accuracy. Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) assess visual quality. Experimental results on the Urban Green Space (UGS) dataset show strong performance. The proposed method achieves an mIoU of 0.81, a pixel error of 4.84%, and a processing time of 0.09 seconds per image. It also achieves a PSNR of 34.17 dB and an SSIM of 0.91. Comparable performance is observed on the EnvMonitor dataset. This technique effectively maintains semantic consistency in color distribution, spatial boundary precision, and cartographic standards while ensuring high transfer efficiency, offering a novel pathway for automated map color design that reconciles speed and accuracy.

Keywords: Improved generative adversarial networks; attention mechanisms; map faceted color transfer; residual networks; gradient normalization.

DOI 10.32738/JEPPM-2026-0077

1. Introduction

As a visual carrier of spatial information, the color design of maps bears a dual mission in information transmission and visual expression. At the cognitive level, color is a key visual variable that helps users quickly identify the categories, levels, and spatial relationships of geographic elements. For example, in administrative district maps, differentiated colors can intuitively distinguish administrative boundaries. In land-use maps, a standardized color scheme can clearly identify arable land, forest land, water, and other land types. Meanwhile, the aesthetic attributes of colors directly affect the user's acceptance and experience of the map. A coordinated color scheme can enhance the map's professionalism and artistry, whereas chaotic colors can increase cognitive load (Barua et al., 2024; Ma et al., 2025). Faceted elements (e.g., administrative divisions, land-use patches, natural geographic areas) have become the core objects of map color design due to their broad coverage and high visual appeal. However, there are significant pain points in its design process (Olivieri and Reichenbacher, 2024). First, the complexity is high, requiring consideration of element semantics, spatial topological relations, cartographic norms, and visual perception laws. Second, it is time-consuming. Large-scale maps often contain hundreds of surface areas. Artificial color matching needs to be debugged repeatedly to avoid confusion or visual conflict of neighboring colors, which can take hours or even days. The third is subjective dependence: ensuring consistent outputs is difficult because design quality heavily relies on the designer's experience and aesthetic sense (Thomas and Zacharias, 2026; Dey et al., 2026). With the rapid development of smart cities, digital twins and other fields, the demand for automated and intelligent mapping is becoming increasingly urgent. The traditional manual color-matching mode is ill-suited to high-frequency updates and large-scale batch production. The automatic color transfer technology extracts the color style from the reference image and applies it to the target map, thereby quickly producing stylized output while retaining the original geographic information.

Generative Adversarial Networks (GANs) combined with Attention Mechanisms (AMs) have been applied to a range of computer vision tasks, including image fusion, image inpainting, document denoising, fault diagnosis, and image translation. The following studies are representative examples in these areas. Tsai et al. (2026) addressed reduced data reliability caused by facial image occlusion by proposing a GAN framework that integrated occlusion segmentation and image restoration. The generator included a hybrid attention aggregation module to enhance global semantic consistency, and a multi-scale spatial attention module to capture fine texture information. On the CelebA-HQ dataset, this method achieved a PSNR of 35.01 dB and an SSIM of 0.931 at occlusion rates of 35%-45%. Yu et al. (2024) designed a priori-guided attention-gated convolutional GAN to address the semantic misalignment problem in large-area image restoration. This study constructed a spatial-channel attentional mechanism based on the image contour a priori and used an attentional upsampling module to refine the Feature Map (FM). The test dataset demonstrated that it was optimal across all indices, with a performance improvement of 3.21%- 6.52% over existing methods. It could effectively eliminate watermarks and blurring artifacts. Neji et al. (2024) addressed the challenges of background interference and contrast imbalance in historical document denoising by fusing a GAN with AM. Moreover, the deep network was used to generate an attention map to guide the generator to focus on the target region. The dataset validation revealed that the visual quality was significantly better than existing techniques, providing a new solution for degraded document restoration. Shao et al. (2023) faced the challenge of insufficient samples for fault diagnosis of rotating machinery under rotational speed fluctuation, and designed a dual-threshold attention-guided GAN. By fusing the Wasserstein distance and gradient penalty loss, this study extracted global features from infrared thermal imaging using the AM, which was optimized by a dual-threshold training mechanism. Experiments indicated that the quality of its generated images improved and that it outperformed the comparison model in rotor-bearing system fault diagnosis. Wang et al. (2025) addressed image quality degradation in single-pixel imaging caused by noise under conditions of insufficient single-photon detection and sampling. They proposed a CycleGAN denoising method that integrated an efficient attention module. Experimental results showed that this method outperformed traditional CycleGAN and Gaussian filtering methods in denoising performance, significantly improving imaging quality.

Good color matching can improve image readability, and researchers have developed systems that generate color schemes in real time to support dynamic content generation. Zhang et al. (2024) proposed the MapGPT framework to lower the barrier to entry for professional cartography. Integrating a large-scale language model with professional mapping tools, this research supported dialogic interactive adjustment by understanding user needs through natural language and invoking tools to control map elements (e.g., symbols/layouts). The open-source case verified that it could achieve user-friendly, element-level-controllable intelligent cartography. Si et al. (2023) designed an improved GAN for the identity distortion problem in Korean portrait headdress-style migration. It combined a landmark extractor with Gram-matrix-style correlation to preserve key identity features via facial landmark masking. Experiments demonstrated that it outperformed the facial feature retention ability of traditional methods while transferring texture. Jampour et al. (2023) proposed an advanced multi-GAN coloring framework. It innovatively introduced cluster-number-adaptive decision-making and a color-harmony mechanism to address color inconsistency in video coloring. The output realism and color coherence significantly outperformed traditional methods in image/video coloring tasks. Yang et al. (2026) addressed the issue that traditional coloring methods in GIS tools were not user-friendly for non-professionals and could not meet personalized requirements. They developed a system that used large language models to generate color schemes for surface maps. This system consists of three stages: data processing, color concept design, and color scheme design, and provides an interactive interface. User studies showed that this system had good usability, accuracy, and flexibility. Zhang et al. (2026) addressed the problem of inefficient existing methods for transferring fabric colors, which led to visual quality degradation due to color bleeding. They proposed a variational Wasserstein framework that included three modules: image decomposition, image segmentation, and image recoloring. Experiments showed that this framework significantly improved operational efficiency, effectively avoided color overflow and distortion, and retained the original geometric structure and fabric details.

In conclusion, both current technical approaches to the automatic mapping of surface color migration have limitations. The traditional methods are unable to balance local semantic associations with mapping-norm constraints. When deep learning methods are applied directly to mapping scenarios, issues remain, including insufficient semantic understanding and poor retention of boundary details. Based on this, this paper focuses on three core issues in mapping color migration: the method for maintaining semantic consistency, the path for improving migration quality through AMs, and the comprehensive evaluation system for generating maps. It proposes a solution that integrates the improved GAN architecture with AM, innovatively reconfiguring the generator. The AM is embedded after the convolution module and focuses on key geographic elements, such as map boundary contours, through dynamic feature weighting. Second, the depth of the discriminator network is optimized to enhance its ability to discriminate local color distributions. Finally, the multi-scale perceptual loss function is designed to enforce color-transfer consistency.

The novelty of this study is mainly reflected in three aspects: (1) Most existing methods are oriented towards natural images and do not consider the spatial topology and semantic partition constraints of map elements. This study enhances the perception of administrative boundaries and regional outlines through a spatial attention module, avoiding color pollution across regions. (2) Existing methods usually adopt single pixel-level or perceptual-level losses. This study combines pixel loss, perceptual loss, and adversarial loss, and introduces multi-scale feature comparison to retain color accuracy while maintaining mapping norms. (3) Most existing AMs are directly applied to general FMs. This study embeds them in the encoder-decoder architecture, allowing global tone and local details to be controlled separately, thereby improving the conformity of the transfer results with map semantics.

2. Methods and Materials

The research first deeply optimizes the original GAN architecture, including improving the generator structure, designing a dual-resolution discriminator, and constructing a multi-scale adversarial loss function. This improves the accuracy of color transfer, boundary clarity, and training stability. Second, the multimodal AM is deeply coupled into this GAN framework. By dynamically strengthening feature attention in the channel and spatial dimensions (SD), the model's understanding of map boundary structure and local area semantics is enhanced.

2.1. Improved GAN Architecture with a Map Color Transfer Framework

Compared with traditional images, maps have strict requirements for geometric accuracy, element structure, and cartographic standardization. Although the traditional GAN model can achieve image-to-image conversion, it is prone to color distortion, boundary blurring, and cross-element color contamination when migrating map faceted colors. The fundamental reason is that the standard convolutional operation treats all pixels equally and cannot distinguish the map's semantic structure. Moreover, the deep network training is prone to gradient disappearance or pattern collapse (Chen et al., 2023; Purohit and Dave, 2023). Aiming to address the problems of traditional GANs in automatic map faceted color transfer, the study systematically optimizes the model across three dimensions: the generator, discriminator, and loss function. The fusion improvement GAN framework is constructed.

The generator utilizes the U-Net architecture, whose encoder-decoder structure with skip connections effectively preserves spatial details in the map. However, as the network deepens, the backpropagation path for the encoder's deep features becomes longer, making it prone to gradient attenuation. After introducing residual units, the gradient can be propagated directly to the shallow layers via the skip connections, effectively alleviating this problem. The mathematical expression is shown in Eq. (1).

$$y = F(x, \{W_i\}) + x \tag{1}$$

In Eq. (1), x is the input feature. F is the residual function (usually consisting of two 3×3 convolutional layers). W_i denotes the learnable parameters in the residual function. y is the output feature. Using a 2-step convolution instead of max pooling for downsampling is because pooling discards precise pixel positions, whereas convolution can reduce resolution while retaining more spatial information, which is crucial for the precise positioning of map boundaries. Its formal description is given by Eq. (2).

$$Z^{\ell} = \sigma(W^{\ell} * Z^{\ell-1} + b^{\ell}) \tag{2}$$

In Eq. (2), Z^{ℓ} and $Z^{\ell-1}$ denote the outputs of the lth and $l-1$th layers. W^{ℓ} denotes the weight parameter of the lth layer. b^{ℓ} is the bias term of the lth layer. $*$ is the Convolution Operation (CO). σ is the LeakyReLU Activation Function (AF). The generator encoding path progressively extracts multiscale features via residual units, and the decoding path performs upsampling via transposed convolution, as shown in Eq. (3).

$$Z^{\ell} = \sigma(W^{\ell} \odot Z^{\ell-1} + b^{\ell}) \tag{3}$$

In Eq. (3), $\odot$ is the transpose-CO. To improve faceted boundary reconstruction capabilities, the jump connection links the encoder's high-RFs to the corresponding layers of the decoder. The discriminator adopts the PatchGAN structure, and its receptive field is shown in Fig. 1. It divides the input map into N×N local blocks and independently discriminates the authenticity of each block, producing a probability matrix indicating whether each block is true. This structure can evaluate the local color distribution more precisely, but it is prone to instability due to parameter oscillations during training.

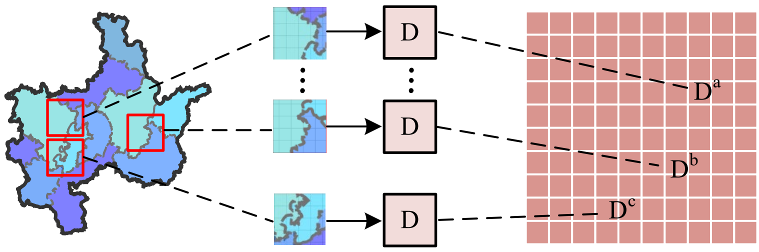

Fig. 1. PatchGAN receptive field (source: original)

The study uses gradient normalization instead of standard batch normalization after each convolutional layer of the discriminator. Gradient normalization is applied to each layer of the discriminator's convolutional parameters. It is executed after each round of backpropagation and before parameter update. The core idea is to uniformly constrain the L2 norm of each layer's gradient to the geometric mean of the L2 norms across all layers, thereby preventing the discriminator from becoming too powerful and causing generator training to stagnate. Its update rule is shown in Eq. (4).

$$\begin{cases} \tilde{\nabla} W_i = \dfrac{\nabla W_i}{\|\nabla W_i\|_2} \cdot \bar{G} \\ \bar{G} = \left(\prod_{i=1}^{N} \|\nabla W_i\|_2 \right)^{1/N} \\ \nabla W_i' = \nabla W_i \cdot \dfrac{\bar{G}}{\|\nabla W_i\|_2} \end{cases} \quad \begin{cases} \tilde{\nabla} W_i = \frac{\nabla W_i}{G_W} \cdot \bar{G} \\ G_W = \|\nabla W_i\|_2 \\ \bar{G} = exp(\frac{1}{n} \sum_{i=1}^{n} ln\, G_W) \end{cases} \tag{4}$$

In Eq. (4), after backpropagation, the gradient ∇W_i of the weights of the ith layer of the discriminator is obtained, and its L2 norm $\|\nabla W_i\|_2$ is calculated. The geometric mean $\bar{G}$ of the gradient norms of all layers is calculated. Each layer's gradient is normalized and updated, and the parameter update is performed using the normalized gradient $\nabla W_i^{'}$.

The generator is a symmetric U-shaped structure, with both the input and output image dimensions being 256×256×3. The encoder consists of 4 downsampling stages, each stage composed of 2 layers of 3×3 convolutions (with a stride of 2), LeakyReLU activation, batch normalization, and one residual unit, with channel numbers of 64, 128, 256, and 512. The bottleneck layer has a channel number of 1024. The decoder consists of 4 upsampling stages that gradually restore spatial resolution via transposed convolutions, with channel numbers of 512, 256, 128, and 64. The channel dimensions of the encoder's features at the same level are concatenated with those of the decoder's upsampled features through skip connections. Then, a 1-layer 3×3 convolution is used to fuse the features. The final output layer uses a 3×1 convolution + Tanh activation to map the features to the RGB color space. The discriminator adopts a PatchGAN structure with a 70×70 receptive field, consisting of 5 convolutional layers with channel numbers sequentially 64, 128, 256, 512, and 1, each with a kernel size of 4×4 and a stride of 2. Except for the first and last layers, the other layers use LeakyReLU activation and gradient normalization. The final output is a 30×30×1 probability matrix, with each value corresponding to the true/false judgment for the 70×70 local area in the input image $W_i i G_W \bar{G}$.

The study adopts a joint optimization strategy that combines L1 loss and perceptual loss to constrain the generator output along two dimensions: pixel-level accuracy and semantic-level style. The L1 loss L_{L1} constrains pixel-level color-value accuracy and prevents color shifts that may occur during adversarial training, as shown in Eq. (5).

$$L_{L1} = \frac{1}{H \times W \times C} \sum_{h=1}^{H} \sum_{w=1}^{W} \sum_{c=1}^{C} \left| I_{real}(h, w, c) - I_{gen}(h, w, c) \right| \tag{5}$$

In Eq. (5), $H \times W$ is the map size. C is the quantity of color channels. I_{real} and I_{gen} denote the pixel values of the real and generated maps, respectively. The perceptual loss L_{perc} fails to compare the feature space of the pre-trained network, forcing the generated results to retain the texture structure and visual style consistent with the reference image. Specifically, certain layers of the network's convolutional output are selected as perceptual features. The loss formula is shown in Eq. (6).

$$L_{\text{perc}} = \frac{1}{H_i \times W_i \times C_i} \sum_{h=1}^{H_i} \sum_{w=1}^{W_i} \sum_{c=1}^{C_i} \left(\phi_i(I_{\text{real}}) - \phi_i(I_{\text{gen}}) \right)^2 \tag{6}$$

In Eq. (6), ϕ_i denotes the feature extraction function of layer i. $H_i \times W_i \times C_i$ is the size of the FM of the ith layer. The feature-extraction backbone for the perceptual loss uses the pre-trained VGG19 network trained on the ImageNet dataset. The outputs of 4 convolutional layers, namely conv1_2, conv2_2, conv3_2, and conv4_2, are selected as the multi-scale perceptual features. The low-level features are used to constrain texture and color details, while the high-level features are used to constrain semantic style consistency. During training, the parameters of the VGG19 network are kept frozen and do not participate in gradient updates. They are only used as a fixed feature extractor. The feature losses of each layer are weighted equally and summed to balance multi-scale constraints. Generating adversarial loss L_{GAN} prompts the generator to produce output that is difficult to distinguish from the real map distribution, avoiding distortion of the results, as shown in Eq. (7).

$$L_{GAN} = E_{x \sim p_{data}}[log\, D\,(x)] + E_{z \sim p_z}[log(\,1 - D(G(z)))] \tag{7}$$

In Eq. (7), $E_{x \sim p_{data}}$ denotes the expectation of the true data distribution p_{data} and is the output of the discriminator D for the true sample x. $E_{z \sim p_z}$ denotes the expectation of the random noise distribution p_z, and is the output of the discriminator G on the noise sample z. The total loss consists of a weighted summation of the generated adversarial loss, the L1 loss, and the perceived loss, as shown in Eq. (8).

$$L_{\text{total}} = L_{\text{GAN}} + \lambda_1 L_{L1} + \lambda_2 L_{\text{perc}} \tag{8}$$

In Eq. (8), λ_1 and λ_2 are hyperparameters. The generator needs to simultaneously satisfy the discriminator's adversarial constraints, the pixel-level color-accuracy constraints, and the semantic-level stylistic-consistency constraints. Thus, automatic map-faceted color transfer produces accurate color, clear boundaries, and visual naturalness.

2.2. Integration of AMs

In automatic map faceted color transfer, the improved GAN model enhances feature representation and training stability through a residual network and gradient normalization. However, there are still problems, such as uneven coloring in local areas and unclear restoration of administrative area boundaries. Therefore, the study proposes an automatic map-faceted color transfer model that integrates the improved GAN with the AM, as shown in Fig. 2. By dynamically adjusting feature weights, the model strengthens its ability to perceive semantically critical regions in the input data. Thus, this improves the structural reasonableness and the richness of detail in the generated maps.

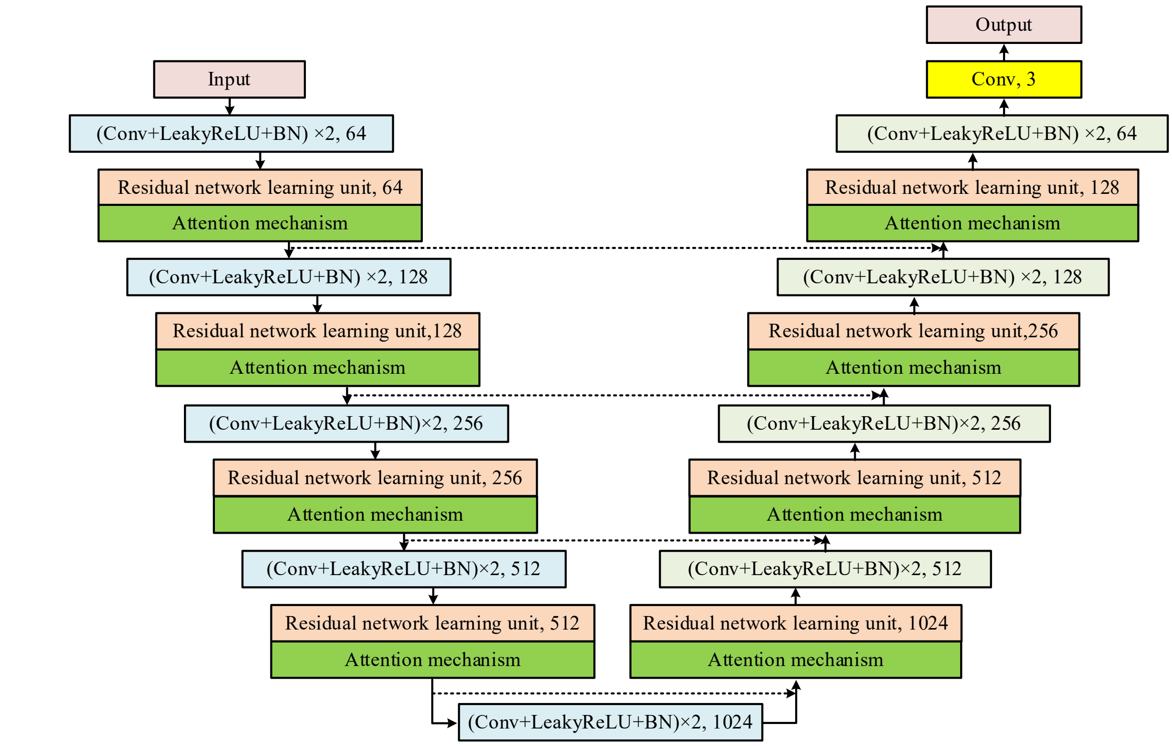

Fig. 2. Universal generator network architecture with AM (source: original)

In Fig. 2, while the attention module performs secondary feature filtering, the convolutional block extracts multi-scale spatial information. The Squeeze-and-Excitation Network (SENet) is a classical model of Channel Attention (CA), which implements channel-level feature recalibration through the "compression-excitation" mechanism. Its core value lies in explicitly modeling the interdependence among channels, which is well-suited to the task of map color transfer (Kong et al., 2023; Preethi et al., 2024). The SENet network structure is shown in Fig. 3.

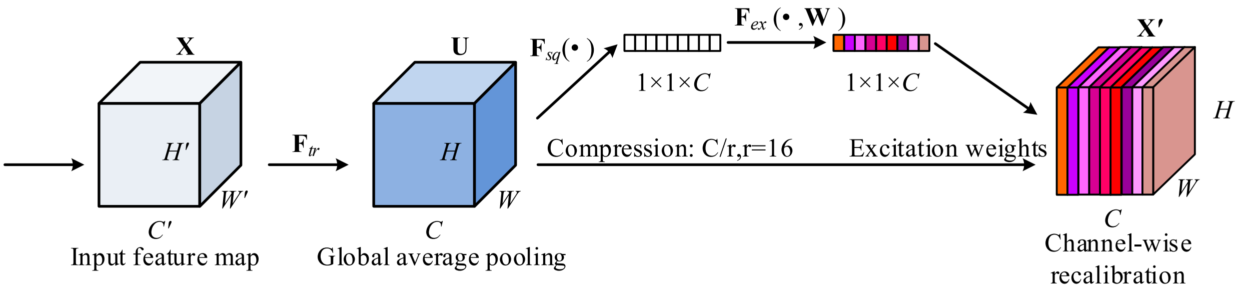

Fig. 3. SENet network structure (source: original)

SENet first performs global Average Pooling (AP) on the input FM $X \in R^{C \times H \times W}$ to compress the SD $H \times W$ into the channel-level global feature $z_c \in R^C$, as shown in Eq. (9).

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_c(i,j) \tag{9}$$

In Eq. (9), z_c denotes the global statistic of the cth channel, which captures the global distribution characteristics of this channel in the whole map by aggregating spatial information. Then, the Channel Weight (CW) $s \in R^C$ is generated by a two-layer fully connected network. The first layer compresses the quantity of channels to C/r (r is the dimensionality reduction factor, which is taken as 16 in the experiment), and the AF adopts ReLU. The second layer recovers to C channels, and the AF adopts a Sigmoid function, as shown in Eq. (10).

$$s = \sigma(W_2 \delta(W_1 z)) \tag{10}$$

In Eq. (10), δ displays the ReLU function. W_1 and $\begin{cases} F_{avg}^c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_c(i,j) \\ F_{max}^c = \underset{\substack{i=1 \\ H}}{max} \underset{\substack{j=1 \\ W}}{max} X_c(i,j) \end{cases}$ display the weight matrices. Finally, the CWs s are multiplied by the original FM to realize adaptive enhancement or suppression of each channel. Through the above operations, SENet can enhance the key channels related to map color, suppress irrelevant noise channels and improve the accuracy of color transfer. In the generator, the SENet module makes the network more focused on the global color distribution pattern by adjusting channel-level weights.

The CA and Spatial Attention (SA) in the Convolutional Block Attention Module (CBAM) respectively undertake the functions of "color grading" and "localization". CA models the dependency relationships between channels, selects the feature channels that contribute the most to the overall color tone distribution, and controls the global color tendency of the transfer results. SA enhances the response at boundaries and regional contours by computing importance weights for each spatial position, thereby strictly limiting color changes to the corresponding planar features. The two work together at different levels of the encoder and decoder to optimize migration quality from both global and local perspectives. The CBAM module structure is shown in Fig. 4.

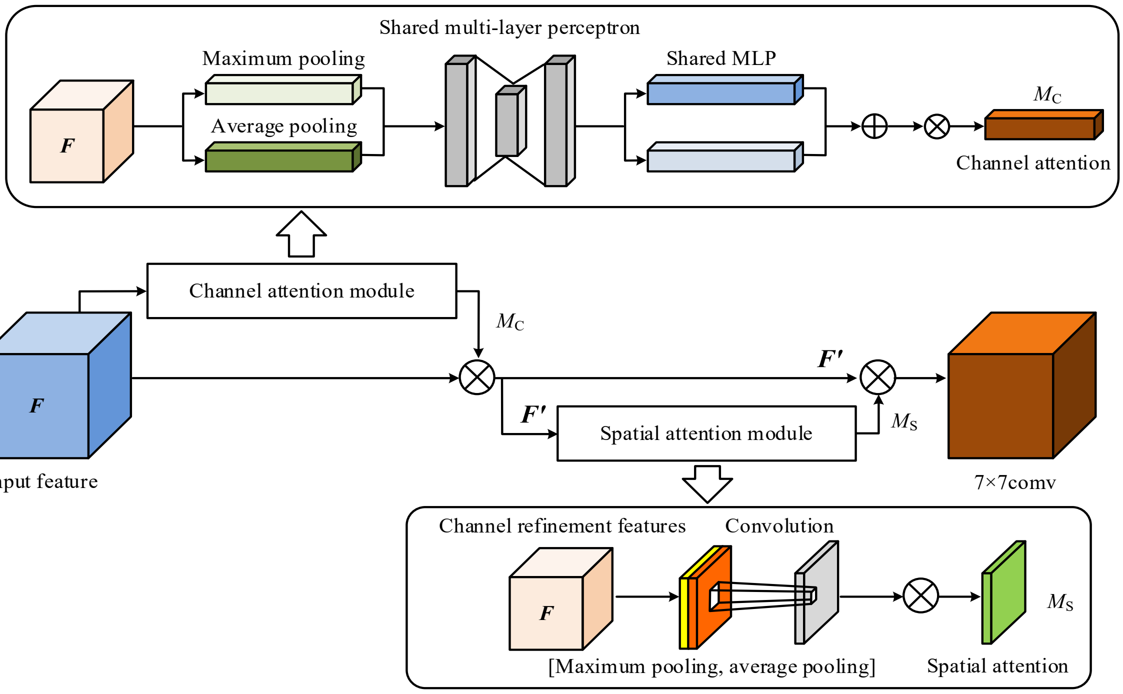

Fig. 4. CBAM module structure (source: original)

The CA module performs both AP and maximum pooling on the input FMs to generate channel-level statistics F_{avg}^c and F_{max}^c, as shown in Eq. (11).

$$\begin{cases} F_{avg}^c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_c(i,j) \\ F_{max}^c = \underset{\substack{i=1 \\ H}}{max} \underset{\substack{j=1 \\ W}}{max} X_c(i,j) \end{cases} \tag{11}$$

The two are spliced together to generate the CWs M_c through a two-layer fully connected network (parameter sharing), as shown in Eq. (12).

$$M_c = \sigma(W_1(F_{avg}^c + F_{max}^c)) \tag{12}$$

The SA module performs global AP and maximum pooling of SDs on the FM X_c' output from the CA to generate spatial features F_{avg}^s and F_{max}^s, as shown in Eq. (13).

$$\begin{cases} F_{avg}^s = \frac{1}{C} \sum_{c=1}^{C} X_c'(i,j) \\ F_{max}^s = \underset{\substack{c=1 \\ C}}{max} X_c'(i,j) \end{cases} \tag{13}$$

The two are spliced together to generate the Spatial Weights (SWs) M_s by means of a 7×7 convolutional layer, as shown in Eq. (14).

$$M_s = \sigma\left(f^{7\times7}\left([F_{avg}^s; F_{max}^s]\right)\right) \tag{14}$$

In Eq. (14), $f^{7\times7}$ is the 7×7 CO and $[F_{avg}^s; F_{max}^s]$ is the channel splicing operation. The CWs and SWs are applied sequentially to the FM to realize adaptive enhancement or suppression of each channel.

3. Results

By contrasting the suggested approach with other techniques in terms of alignment time and accuracy, the study examined the effectiveness of the suggested algorithm and the impact of automatically transferring faceted colors to maps.

3.1. Performance Analysis of Improved GAN

Research hardware came with an NVIDIA A100 GPU with 40GB of video memory and an Intel Xeon Platinum 8268 CPU. RAM was 256GB DDR4. On the software side, PyTorch 1.10 was used as the Deep Learning framework, and CUDA 11.3 was used. The Urban Green Space (UGS) Dataset for Cartographic Stylization (https://liumency.github.io/ugs-1m/) and Blue Map China Environmental Monitoring Dataset (EnvMonitor) (https://www.ipe.org.cn/) were selected to validate the proposed method. The UGS, developed by a team from Sun Yat-sen University, provided high-resolution, fine-grained green space maps for 31 major cities in China and 2,179 faceted green space labeling data in GeoTIFF format. Its training data contained 4554 samples of diverse green spaces (e.g., parks, woods and lands). The pre-processing was enhanced by rotating and flipping, and the images were converted to the Lab color space to reduce gamut distortion. It was suitable for verifying the effect of boundary optimization in the color transfer of faceted green spaces. The EnvMonitor was released by the Center for Public Environmental Research (CPER) via the Azure Map platform, integrating air quality (PM2.5, O3), water quality, and pollution-source surface-monitoring data from 2013 to the present. It was visualized using the color-level mapping approach, with administrative divisions as units.

The evaluation metrics employed in this study corresponded to distinct quality dimensions of the color transfer task. Mean Intersection over Union (mIoU) quantified the geometric accuracy of polygon boundaries in the transferred results, reflecting the model’s capacity to preserve spatial structures. Percentage Error (PE) directly assessed the accuracy of color value restoration at each pixel, indicating the numerical precision of color mapping. Peak Signal-to-Noise Ratio (PSNR) evaluated overall image signal fidelity, while Structural Similarity Index (SSIM) measured similarity between generated and ground-truth images across luminance, contrast, and structure, jointly characterizing the visual naturalness of the output.

In this task, each input grayscale image was accompanied by a color reference style image of the same region and semantic category. This reference image was generated by professional graphic designers using standard color matching, which served as the "real annotation" for color transfer. All quantitative indicators were calculated by comparing the model output with the corresponding reference graph, and the average was taken after 3 repeated experiments. The dataset was divided into strata by geographical units (cities/administrative regions), with training, validation, and testing ratios of 70%, 10%, and 20%, respectively, to ensure that test areas did not appear during training. Adopting an early stop strategy, the training was terminated when the mIoU on the validation set did not improve for 15 consecutive training cycles. All experiments used a fixed random seed of 2024 to ensure reproducibility. The learning rate of 3×10^{-4} was the default recommended starting value for the AdamW optimizer. After a preliminary search, it was confirmed that the performance was stable within the range of 1×10^{-4} to 5×10^{-4}. The loss weights λ_1=100 and λ_2=0.01 were determined through grid search and fine-tuned on the validation set for confirmation. The batch size of 32 was limited by GPU memory (40GB) and was the maximum value that could be carried.

When evaluating, mIoU and EE first performed semantic category mapping on the output color and then calculated it to measure the correctness of color classification. PE and RMSE were calculated within the effective area of the Lab space, which was the non-background pixel area. PSNR and SSIM were globally calculated in RGB space. All comparison algorithms were trained and tested under identical software and hardware environments, data partitioning, and training strategies, with evaluation metrics uniformly calculated on the test set using the same code. The study calculated the forward inference time for a single 256×256 image, averaged 100 repeated tests, excluded data loading and post-processing time, and ensured fair comparisons.

Fig. 5 compares the mIoU curves of the proposed method for the UGS and the EnvMonitor. In Fig. 5(a), the mIoU metric of the study method showed a steep upward trend over the first 50 training cycles. With deeper training, the mIoU of the research method eventually stabilized at 0.85, which was higher than the 0.78 achieved by the Hybrid Attention Generative Adversarial Network (Hya-GAN; Zhang et al., 2026). In Fig. 5(b), the research method stabilized at 0.78 after 200 training cycles, significantly improving from the comparison method's 0.73. It was fully demonstrated that the improved GAN architecture incorporating the AM enhanced the consistency of feature extraction for map-faceted elements, ultimately achieving a synergistic enhancement in color transfer accuracy and training stability.

Table 1 compares the performance of the proposed method with six baseline models on the UGS and EnvMonitor datasets. The comparative methods included Graph Recurrent Neural Network (GraphRNN), Hya-GAN, Cycle-Consistent Generative Adversarial Network (CycleGAN), Pixel-to-Pixel Generative Adversarial Network (Pix2Pix), Unsupervised Generative Attentional Networks with Adaptive Layer-Instance Normalization (U-GAT-IT), and a Transformer-based architecture. Experimental results demonstrated that the proposed method achieved mIoU scores of 84.03% and 83.61%, PE valued as low as 4.84% and 5.23%, and the highest Peak Signal-to-Noise Ratio (PSNR) and SSIM across both datasets, outperforming all compared methods in every metric. On the UGS dataset, the proposed method improved mIoU by 2.73, 2.47, 3.51, 5.68, and 4.70 percentage points over Hya-GAN, U-GAT-IT, GraphRNN, Transformer, and Pix2Pix, respectively. On the EnvMonitor dataset, it achieved mIoU gains of 1.56 and 1.28 percentage points over Hya-GAN and U-GAT-IT, and a 6.33-

percentage-point improvement over CycleGAN. Regarding the two core metrics, mIoU and PE, the improvements of the proposed method over all six baselines were statistically significant at $p<0.01$ based on paired t-tests, confirming the robustness of the reported performance advantages. The FLOPs of this method were 68.4G, which was slightly higher than those of CycleGAN (38.7G) and Pix2Pix (45.2G). However, compared to the second-best performing U-GAT-IT, it was 8.4G lower, indicating that this method achieved a better balance between accuracy and efficiency.

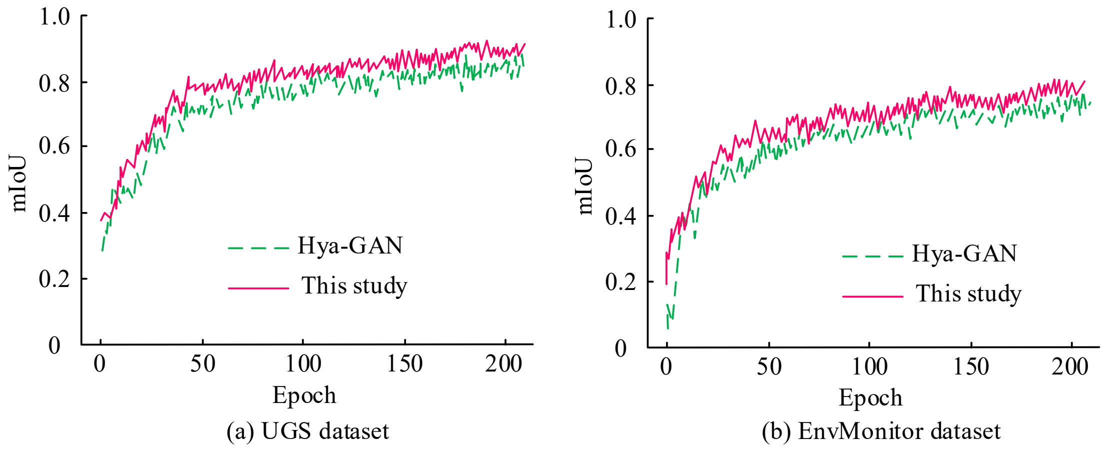

Fig. 5. Comparison of mIoU curves (source: original)

3.2. Automatic Map Faceted Color Transfer Validation

The study used two objective metrics, PSNR and SSIM, to quantify the generation effect. The comparative outcomes of the four distinct methods on the UGS and EnvMonitor datasets were displayed in Fig. 6. In Fig. 6(a), for the UGS dataset, the PSNR and SSIM values of the study method were ranked first at 34.17 dB and 0.91, respectively. In Fig. 6(b) of EnvMonitor, the research method still had the highest PSNR and SSIM values. These data confirmed the advantages of the improved model in reducing the color transfer noise and maintaining the structural integrity of the map boundary by incorporating the dual AM. This made the generated results more closely match the visual cognitive demands of real maps.

The ablation experiment results in Table 2 confirmed the effectiveness of each core module. The complete model, combined with residual networks, gradient normalization, and dual AMs, achieved the best performance on both datasets. On the UGS dataset, the proposed method achieved an mIoU of 84.03%, a PE of 4.84%, a PSNR of 34.17 dB, and an SSIM of 0.91. On the EnvMonitor dataset, the corresponding values were 83.61%, 5.23%, 33.05 dB, and 0.91, indicating excellent performance in terms of clarity, structural fidelity, and visual quality. If the complete attention module were removed, the performance would drop sharply. On the UGS dataset, the mIoU dropped to 76.45%, a decrease of 7.58 percentage points, further verifying the necessity of the synergy between channel and spatial attention. If CBAM was omitted, the mIoU on the UGS dataset would drop to 80.12%, indicating that spatial attention was crucial for boundary preservation. Furthermore, if the residual network was removed, the PE on the UGS dataset would increase to 6.32%.

Fig. 7 demonstrates the effect of applying the improved GAN model to automatically convert a grayscale map into a color map. Fig. 7(a) shows the original color style map as a color reference sample. Fig. 7(b) shows the input grayscale map that needs to be color-filled. Fig. 7(c) shows the color output map generated after processing with the improved GAN model. The output closely matched the reference in hue and regional coherence, especially within faceted areas, demonstrating that the AM effectively captured and transferred color styles. The generated maps exhibited clearer boundaries, richer textures, and greater structural similarity, confirming the model's ability to automatically produce high-quality, style-consistent color maps.

Table 3 further validates the generalization capability of the proposed method on unseen datasets and in cross-domain scenarios. The MassGIS dataset was sourced from the Massachusetts Geographic Information System (MassGIS; https://www.mass.gov/info-details/massgis-data-layers) and comprised standard cartographic polygon features for the state, including administrative boundaries, land-use types, and hydrographic networks. It employed an official, standardized color scheme for thematic mapping, aligning closely with the color transfer task addressed in this study. On the MassGIS dataset, the proposed method achieved an mIoU of 83.12%, a PE of 4.96%, a PSNR of 33.78 dB, and an SSIM of 0.91, outperforming all comparative methods across these four metrics. Specifically, its mIoU surpassed the second-best performer, U-GAT-IT, by 2.67 percentage points and exceeded Hya-GAN and GraphRNN by 3.00 and 4.23 percentage points, respectively. In cross-domain zero-shot transfer experiments, the proposed method still attained an mIoU of 79.84% and a PE of 6.01%. This represented a 9.72 percentage point improvement over CycleGAN (70.12%) and outperformed GraphRNN and Hya-GAN by 5.16 and 3.31 percentage points, respectively. Regarding the two core metrics, mIoU and PE, the enhancements achieved by the proposed method over all baselines in both generalization scenarios were statistically significant at $p<0.01$ based on paired t-tests. These results demonstrated that the proposed method effectively captured the shared structural characteristics of map polygons and maintained stable transfer quality despite variations in regional cartographic styles and data distributions, confirming its robust generalization ability.

Table 1. Comparative results of quantitative analysis

Dataset	Method	mIoU (%)	PE (%)	PSNR (dB)	SSIM	FLOPs (G)
UGS	Transformer	78.35±1.15	6.99±0.32	30.52±0.44	0.84±0.018	156.7
	GraphRNN	80.52±0.95	6.33±0.26	31.68±0.38	0.86±0.015	42.3
	Hya-GAN	81.30±0.88	5.69±0.23	33.21±0.35	0.88±0.012	89.5
	CycleGAN	76.48±1.24	8.12±0.35	31.24±0.42	0.83±0.021	38.7
	Pix2Pix	79.33±0.98	6.87±0.28	32.05±0.38	0.86±0.016	45.2
	U-GAT-IT	81.56±0.87	5.96±0.22	32.87±0.35	0.88±0.013	76.8
	This study	84.03±0.75	4.84±0.19	34.17±0.30	0.91±0.009	68.4
EnvMonitor	Transformer	78.70±1.18	7.01±0.33	29.12±0.46	0.81±0.022	156.7
	GraphRNN	80.20±0.98	6.49±0.27	30.05±0.40	0.84±0.018	42.3
	Hya-GAN	82.05±0.90	5.95±0.24	32.04±0.36	0.87±0.014	89.5
	CycleGAN	77.28±1.31	8.18±0.40	30.88±0.45	0.82±0.024	38.7
	Pix2Pix	80.15±1.05	6.95±0.32	31.79±0.41	0.85±0.019	45.2
	U-GAT-IT	82.33±0.92	5.77±0.25	32.54±0.37	0.89±0.011	76.8
	This study	83.61±0.80	5.23±0.21	33.05±0.32	0.91±0.008	68.4

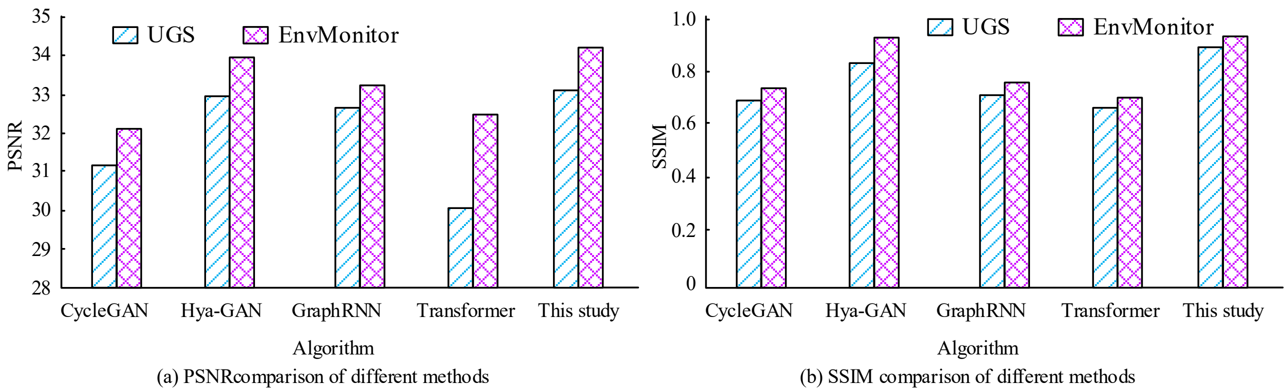

Fig. 6. Results of objective evaluation metrics on different datasets (source: original)

The residual unit alleviated the gradient disappearance, enabling the generator to learn deeper features. Gradient normalization balanced the learning rates of each discriminator to avoid mode collapse and enhance training stability. Channel attention filtered and selected features related to hue distribution, controlling the global color tendency. Spatial attention strengthened the boundary response and constrained color changes within the surface elements. Abandoning the CBAM resulted in an IoU drop from 84.03% to 80.12%, a decrease of 3.91 percentage points, confirming the crucial role of spatial attention in maintaining boundaries. Removing the residual network increased PE from 4.84% to 6.32%, underscoring the importance of the residual structure for color accuracy. Completely removing the attention module resulted in an IoU of 76.45%, verifying the necessity of the collaborative optimization of dual attention.

There were essential differences between the migration of map surface element colors and the migration of natural image style. Natural image style migration focused on global statistical feature matching of texture and color, whereas map color not only carried visual information but was also related to the semantic identification of land features. The spatial attention module made a significant contribution to boundary preservation by effectively suppressing color crossing semantic boundaries, thereby underscoring the necessity of semantic perception in map generation. At the same time, the AM's migration effect also exposed certain limitations. Its effectiveness depended on the semantic category system covered by the training data. When the target map contained unseen categories, the allocation of attention weights might be biased.

Table 2. Ablation experiment

Dataset		Model variant	IoU (%)	PE (%)	PSNR (dB)	SSIM	Time (s)
UGS	1	This study	84.03±0.75	4.84±0.19	34.17±0.30	0.91±0.009	0.09±0.008
	2	No residual network	79.21±1.08	6.32±0.28	32.54±0.38	0.85±0.016	0.08±0.004
	3	No gradient normalization	81.56±0.92	5.51±0.24	33.02±0.35	0.88±0.013	0.09±0.008
	4	No SENet	82.37±0.85	5.18±0.22	33.52±0.33	0.89±0.011	0.09±0.010
	5	No CBAM	80.12±1.02	5.73±0.26	32.89±0.36	0.87±0.014	0.08±0.007
	6	No AM	76.45±1.21	7.21±0.32	31.25±0.42	0.82±0.020	0.07±0.006
EnvMonitoro	1	This study	83.61±0.80	5.23±0.21	33.05±0.32	0.91±0.008	0.10±0.005
	2	No residual network	78.94±1.12	6.58±0.30	31.46±0.40	0.86±0.018	0.09±0.008
	3	No gradient normalization	81.03±0.95	5.97±0.26	32.11±0.36	0.89±0.014	0.10±0.007
	4	No SENet	82.05±0.88	5.52±0.23	32.48±0.34	0.90±0.010	0.10±0.010
	5	No CBAM	79.83±1.05	6.19±0.27	31.74±0.37	0.88±0.015	0.09±0.006
	6	No AM	75.32±1.28	7.89±0.35	30.12±0.45	0.83±0.022	0.08±0.007

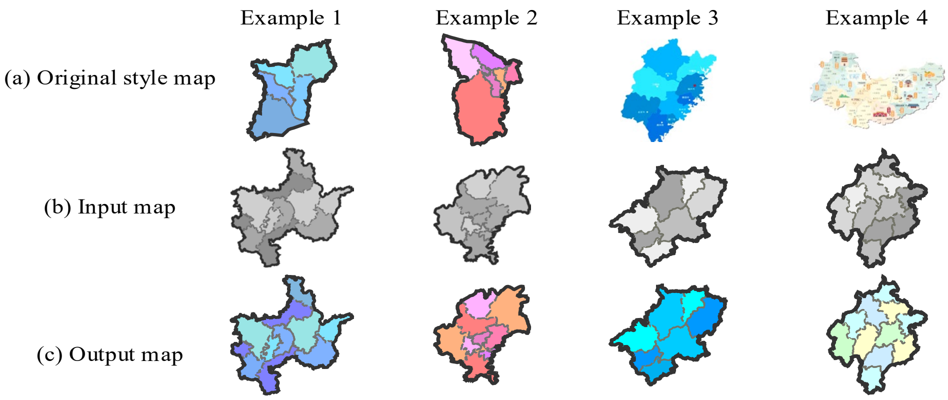

Fig. 7. Results of map conversion to color maps (source: original)

4.2. Practical Application and Deployment Potential

Compared with the traditional manual color assignment cycle of several hours to several days, this method requires only 0.09 seconds per image, significantly improving the efficiency of map production. It had the potential to be integrated into mainstream GIS platforms as an automatic color assignment plugin. However, there were still several challenges from the research prototype to actual deployment. Model training needed to be completed on a high-performance GPU with 40GB of memory, which was high for ordinary users. The current color assignment specifications were derived from customized mapping standards and needed retraining or fine-tuning when migrating to other thematic maps or different countries' color assignment systems. In addition, GIS platforms usually require algorithms to have an interpretable parameter adjustment interface, but the black-box nature of deep models and this requirement created tension, and further development of interactive color parameter control mechanisms was needed.

4.3. Limitations

This paper still has several limitations. UGS and EnvMonitor mainly consist of urban green space and environmental monitoring maps of China, not covering topographic maps, nautical maps, etc., and use different styles and color assignment systems for different countries. The cross-domain generalization ability needs to be verified on a wider range of data. For small-area surface elements, the attention advantages learned by the model on large-scale elements cannot be fully transferred, leading to reduced color transfer accuracy in small areas. Metrics such as mIoU and PE rely on the category-mapping threshold and the method for calculating color difference in Lab space. Different settings may affect the ranking of the results. Further investigation into how these factors influence the evaluation results is possible in the future. At the same time, a large-scale map-color dataset covering multiple regions and topics will be constructed. Combined with domain adaptation and meta-learning strategies, the cross-domain generalization ability and the processing accuracy of small elements will be improved.

Table 3. Validation of generalization performance across datasets

Dataset	Method	mIoU (%)	PE (%)	PSNR (dB)	SSIM
MassGIS	Transformer	76.89±1.20	7.34±0.35	29.88±0.48	0.82±0.026
	GraphRNN	78.13±1.02	6.98±0.29	31.02±0.42	0.84±0.022
	Hya-GAN	79.56±0.92	6.12±0.25	32.41±0.38	0.87±0.016
	CycleGAN	74.23±1.45	8.87±0.42	30.11±0.48	0.80±0.031
	Pix2Pix	77.56±1.12	7.34±0.35	31.45±0.40	0.83±0.024
	U-GAT-IT	79.88±0.96	6.45±0.28	32.10±0.36	0.86±0.018
	This study	82.47±0.82	5.07±0.20	33.52±0.31	0.90±0.010
Zero-shot transfer (UGS→EnvMonitoro)	Transformer	72.15±1.35	8.52±0.40	28.45±0.52	0.78±0.033
	GraphRNN	74.68±1.15	7.89±0.35	29.72±0.45	0.80±0.028
	Hya-GAN	76.53±1.00	6.98±0.28	30.85±0.40	0.83±0.022
	CycleGAN	70.12±1.52	9.23±0.48	29.45±0.50	0.77±0.036
	Pix2Pix	73.45±1.20	8.01±0.40	30.56±0.42	0.80±0.030
	U-GAT-IT	76.23±1.05	7.12±0.32	31.23±0.38	0.83±0.024
	This study	79.84±0.95	6.01±0.25	32.41±0.35	0.87±0.015

Author Contributions

Bei Xie contributed to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation and manuscript editing. Dan Zeng contributed to draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

Declaration of Artificial Intelligence (AI) Tools

The authors used DeepL Write only for language editing and formatting assistance. The authors reviewed and take full responsibility for all content.

References

Barua, G., Pingel, T., and Lim, T. (2024). Urban thermal map design considerations: color, shading, and resolution. *Cartography and Geographic Information Science*, 51(4), 549-564.

Chen, M., Dai, H., Wei, S., and Hu, Z. Z. (2023). Linear-ResNet GAN-based anime style transfer of face images. *Signal, Image and Video Processing*, 17(6), 3237-3245.

Dey, A., Biswas, S., Gupta, A., and Debnath, M. (2026). Multi-head spectral-attentive residual generative adversarial network: A high-fidelity generative adversarial network based model for image haze removal. *International Journal of Engineering Transactions B: Applications*, 39(8), 1812-1820.

Jampour, M., Zare, M., and Javidi, M. (2023). Advanced multi-GANs towards near to real image and video colorization. *Journal of Ambient Intelligence and Humanized Computing*, 14(9), 12857-12874.

Kong, F., Pu, Y., Lee, I., Nie, R., Zhao, Z., Xu, D., Qian, W. H., and Liang, H. (2023). Unpaired artistic portrait style transfer via asymmetric double-stream GAN. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, 34(9): 5427-5439.

Ma, Y., Zhang, C., He, C., and Li, X. (2025). Radio map estimation using a CycleGAN-based learning framework for 6G wireless communication. *Digital Communications and Networks*, 11(6), 1822-1830.

Neji, H., Ben Halima, M., Nogueras-Iso, J., Hamdani, T.M., Lacasta, J., Chabchoub, H., and Alimi, A. M. (2024). Doc-Attentive-GAN: Attentive GAN for historical document denoising. *Multimedia Tools and Applications*, 83(18), 55509-55525.

Olivieri, M. and Reichenbacher, T. (2024). A study on the aptitude of color hue, value, and transparency for geographic relevance encoding in mobile maps. *Cartography and Geographic Information Science*, 51(5), 674-686.

Preethi, T., Anjum, A., Ahmad, A. A., Kaur, C., Rao, V. S., El-Ebiary, Y. A. B., and Taloba, A.I. (2024). Advancing healthcare anomaly detection: Integrating GANs with attention mechanisms. *International Journal of Advanced Computer Science and Applications*, 15(6), 1113-1123.

Purohit, J. and Dave, R. (2023). Leveraging deep learning techniques to obtain efficacious segmentation results. *Archives of Advanced Engineering Science*, 1(1), 11-26.
Shao, H., Li, W., Cai, B., Wan, J., Xiao, Y., and Yan, S. (2023). Dual-threshold attention-guided GAN and limited infrared thermal images for rotating machinery fault diagnosis under speed fluctuation. *IEEE Transactions on Industrial Informatics*, 19(9), 9933-9942.
Si, J. and Kim, S. (2023). PP-GAN: Style transfer from Korean portraits to ID photos using landmark extractor with GAN. Computers, *Materials and Continua*, 77(12), 3119-3138.
Thomas, L. and Zacharias, J. (2026). Enhanced slicing adversarial network with attention and multi-resolution generators for high-fidelity MRI reconstruction. *International Journal of Biomedical Engineering and Technology*, 50(4), 347-373.
Tsai, C. S., Wu, H. C., and Chen, Y. Y. (2026). A robust facial image restoration support system for human environment security based on generative adversarial networks with hybrid and multi-scale spatial attention mechanisms. *Sensors and Materials*, 38(6), 3315-3336.
Wang, Y., Yang, L., and Wu, X. (2025). Image denoising for single-pixel imaging based on cycle generative adversarial networks with attention mechanisms. *Applied Optics*, 64(22), 6341-6347.
Yang, N., Wang, Y., Wei, Z., and Wu, F. (2026). MapColorAI: Designing contextually relevant choropleth map color schemes using a large language model. *Cartography and Geographic Information Science*, 53(4), 479-497.
Yu, X., Dai, L., Chen, Z., and Sheng, B. (2024). AGG: Attention-based gated convolutional GAN with prior guidance for image inpainting. *Neural Computing and Applications*, 36(20), 12589-12604.
Zhang, X., Xu, C., Han, Y., Baciu, G., and Zhang, W. (2026). A novel Wasserstein color transfer variational framework on fabric images. *IEEE Transactions on Consumer Electronics*, 72(1), 192-210.
Zhang, Y., Gan, Y., Tang, M., and Gan, X. (2026). A super-resolution generative adversarial network for remote sensing images based on improved residual module and attention mechanism. *Computers, Materials and Continua*, 86(2), 689-707.
Zhang, Y., He, Z., Li, J., Lin, J., Guan, Q., and Yu, W. (2024). MapGPT: An autonomous framework for mapping by integrating large language model and cartographic tools. *Cartography and Geographic Information Science*, 51(6), 717-743.

Bei Xie received her master’s degree from Huazhong University of Science and Technology in 2012. Currently, she serves as an Associate Professor at the School of Visual Arts, Hunan Mass Media Vocational and Technical College. She has been awarded the title of Hunan Provincial Young Backbone Teacher. She has published more than 20 academic papers and edited six textbooks as the chief editor. Her research interests cover landscape design, vocational education, garden plants and other related fields.

Dan Zeng earned her master's degree from Tongji University in 2011. She currently serves as a Professor at the School of Visual Arts, Hunan Mass Media Vocational and Technical College. Invited as a special expert, she has delivered numerous academic lectures on topics such as vocational education curriculum development, teaching competitions, and image processing. She has published over 20 academic papers and led the completion of six provincial and departmental-level research projects. Her research focuses on vocational education, art education, image processing, and curriculum development.