Fırat Üniversitesi Deneysel ve Hesaplamalı Mühendislik Dergisi

Manga için U-Net ve ResNet'in Karşılaştırmalı Analizi Renklendirme: Tutarlılık, Detay ve Hesaplamalı Ödünleşimler

Sami CAKIRAL[1], MD Imran HOSEN[2*], Tarkan AYDIN[3]

[1,2]Yazılım Mühendisliği Bölümü, Mühendislik Fakültesi, Istanbul Ticaret Üniversitesi, İstanbul, Türkiye.
[3]Bilgisayar Mühendisliği Bölümü, Mühendislik ve Doğa Bilimleri Fakültesi, Bahçeşehir Üniversitesi, İstanbul, Türkiye.
[1]Bilgisayar Bilimi ve Yapay Zeka Bölümü, IMT Mines Alès Üniversitesi, Alès, France
[1]sami.cakiral@gmail.com, [2]mdihosen@ticaret.edu.tr, [3]tarkan.aydin@bau.edu.tr

Geliş Tarihi: 15.4.2025
Kabul Tarihi: 31.10.2025

Düzeltme Tarihi: 03.9.2025

doi: https://doi.org/10.62520/fujece.1676853
Araştırma Makalesi

Alıntı: S. Çakıral, MD. I. Hosen ve T. Aydın, “Manga için U-Net ve ResNet'in karşılaştırmalı analizi renklendirme: tutarlılık, detay ve hesaplamalı ödünleşimler”, Fırat Üni. Deny. Müh. Derg., vol. 5, no 1, pp. 119-134, Şubat 2026.

Öz

Manganın otomatik olarak renklendirilmesi, kendine özgü sanatsal tarzı ve karmaşık görselliği nedeniyle benzersiz zorluklar ortaya koymaktadır. Derin öğrenme görüntü renklendirmede umut vaat etse de, mevcut yaklaşımlar genellikle manga sanat eserlerinin tutarlılığı ve sanatsal bütünlüğü ile mücadele etmektedir. Bu makale, manga renklendirme için iki derin öğrenme mimarisinin karşılaştırmalı bir analizini sunmaktadır: aşamalı bırakma özelliğine sahip değiştirilmiş bir U-Net ve uyarlanabilir atlama bağlantılarına sahip ResNet tabanlı bir otomatik kodlayıcı. Yapısal ve algısal bileşenleri bir araya getirerek mangaya özgü zorlukları özellikle ele alan yeni bir bileşik kayıp fonksiyonu sunuyoruz. Farklı bir manga veri kümesi üzerinde yapılan deneyler, ResNet tabanlı modelin daha yüksek renk tutarlılığı ve daha iyi kararlılık sağladığını ve tek tip alanlarda daha az yapaylık ürettiğini gösteriyor. Bununla birlikte, U-Net ince ayrıntıları daha etkili bir şekilde korur. Bu sonuçlar, manga renklendirme sistemlerinin pratik uygulamalarına rehberlik ederek mimariler arasındaki değiş tokuşlar hakkında fikir vermektedir.

Anahtar kelimeler: Manga renklendirme, Derin öğrenme, U-Net, ResNet, Otomatik sanatsal işleme

*Yazışılan yazar

İntihal Kontrol: Evet – Turnitin
Şikayet: fujece@firat.edu.tr

Firat University Journal of Experimental and Computational Engineering

FUJECE

Comparative Analysis of U-Net and ResNet for MangaColorization: Consistency, Detail, and Computational Trade-offs

Sami CAKIRAL[1], MD Imran HOSEN[2*], Tarkan AYDIN[3]

[1,2]Software Engineering Department, Engineering Faculty, Istanbul Ticaret University, Istanbul, Türkiye.
[3]Computer Engineering Department, Faculty of Engineering and Natural Sciences, Bahçeşehir University, Istanbul, Türkiye.
[1]Computer Science and Artificial Intelligence Department, IMT Mines Alès University, Alès, France
[1]sami.cakiral@gmail.com, [2]mdihosen@ticaret.edu.tr, [3]tarkan.aydin@bau.edu.tr

Received: 15.4.2025
Accepted:31.10.2025

Revision:03.9.2025

doi: https://doi.org/10.62520/fujece.1676853
Research Article

Citation: S. Çakıral, MD. I. Hosen and T. Aydın, "Comparative analysis of U-Net and ResNet for mangacolorization: consistency, detail, and computational trade-offs", Firat Univ. Jour.of Exper. and Comp. Eng., vol. 5, no 1, pp. 119-134, February 2026.

Abstract

The automated colorization of manga presents unique challenges due to its distinctive artistic style and complex visuals. While deep learning has shown promise in image colorization, existing approaches often struggle with consistency and artistic integrity of manga artwork. This paper presents a comparative analysis of two deep learning architectures for manga colorization: a modified U-Net with progressive dropout and a ResNet-based autoencoder with adaptive skip connections. We introduce a novel composite loss function that specifically addresses manga-specific challenges by incorporating structural and perceptual components. Experiments on a diverse manga dataset show that the ResNet-based model achieves higher color consistency and better stability, producing fewer artifacts in uniform areas. However, U-Net preserves fine details more effectively. These results provide insights into trade-offs between architectures, guiding practical implementations of manga colorization systems.

Keywords: Manga colorization, Deep learning, U-Net, ResNet, Automated artistic processing

*Corresponging author

Plagiarism Checks: Yes – Turnitin
Complaints: fujece@firat.edu.tr

1. Introduction

Manga, a distinctive form of Japanese comic art, has gained substantial global popularity, with the industry reaching a market value of $5.77 billion in 2020 [1]. While traditionally published in black and white, the growing demand for colored versions presents a significant challenge for publishers and artists. Manual colorization remains a time-intensive process, typically requiring 8-12 hours per page for professional artists. This creates a bottleneck in manga production and distribution, particularly for international markets, where colored editions are increasingly preferred.

Recent advances in deep learning have opened new possibilities for automated image colorization [2,3]. Traditional approaches to manga colorization have faced unique challenges arising from the specific characteristics of manga artwork. These include distinct line art, screen tone patterns, and artistic conventions that differ significantly from those of natural images [4]. Although several automated colorization systems have been developed, they often struggle to maintain consistency across panels and preserve the artistic integrity of the original work [5,6].

Current state-of-the-art approaches in manga colorization primarily employ single-stream architectures, typically based on Generative Adversarial Networks (GANs) [7] or U-Net variants [5], and newer versions use diffusion models [8].

Although these methods show promise for manga colorization, they struggle with complex screen tones and intricate textures. In addition, maintaining consistent color across large panels remains a significant challenge. Furthermore, existing methods are computationally too expensive and, therefore, not ideal for real-world applications in the manga industry.

The limitation of current methodologies can be outlined in several perspectives. The first is that single-stream architecture needs to maintain both fine details as well as global color coherence, which is often done at the expense of one for another. Secondly, existing methods treat all regions of the image equally when, in fact, different regions have different complexities and are more important than others in the same manga artwork. Third, current solutions often fail to adequately capture the unique artistic conventions of manga, resulting in colorization that may appear technically correct but artistically inappropriate.

This paper presents a comparative study of two distinct deep learning approaches for manga colorization: a modified U-Net architecture and a ResNet-based autoencoder. Our research makes several key contributions to the field.

- We introduce a composite loss function that specifically addresses the unique challenges of manga colorization, incorporating both structural and perceptual components.
- We provide a comprehensive analysis of the strengths and limitations of each architectural approach, supported by extensive quantitative and qualitative evaluations.
- We demonstrate practical applications of our findings through detailed case studies, providing insight that can guide future research and development in this domain.

The remainder of this paper is organized as follows. Section 2 briefly discusses existing manga image colorization techniques. In Section 3, we present our methodology, including detailed descriptions of both architectural approaches and our novel loss function. In section 4, we provide details of experimental setup, dataset preparation, and evaluation criteria. The experimental results and discussion including potential future applications are presented in Sections 5 and 6, respectively.

2. Related Work

Image colorization is the process of transforming grayscale images into their corresponding RGB representations, thereby improving both perceptual fidelity and aesthetic appeal [9]. The rapid advancement of deep learning methods in recent years has significantly expanded the possibilities for automating colorization [3,6]. However, conventional automated colorization approaches are not directly applicable to manga colorization. This is primarily because they must preserve the artistic integrity of the original work while maintaining visual consistency across sequential panels.

The evolution of manga colorization techniques reflects the larger advancement of deep learning in computer vision [10, 11, 12]. This section presents a comprehensive overview of significant developments in the field, from foundational approaches to current state-of-the-art methods.

Manga-specific colorization systems

The application of deep learning to image colorization was pioneered by Zhang et al. [13], who demonstrated the potential of convolutional neural networks for this task. Their work established a framework for learning color distributions from large-scale datasets. Building on this foundation, Iizuka et al. [14] introduced a significant advancement with their two-stream architecture that combined local and global features. This marked an important step toward context-aware colorization.

Manga colorization presents unique challenges that have spurred the development of specialized approaches. For instance, Furusawa et al. [15] introduced Comicolorization, a semi-automatic system designed to maintain color consistency across multiple panels by using reference images. This work tackled one of the key issues in manga colorization: ensuring that character colors remain consistent throughout the narrative. In a similar vein, Zhang et al. [16] proposed a two-stage colorization method focused on preserving the artistic style of manga and anime. Their approach demonstrated superior stylistic consistency, offering more natural and coherent colorization results.

In recent years, significant architectural innovations have emerged in manga colorization [17, 18, 19, 20, 21, 22]. Kim et al. [18] presented a two-generator GAN-based method for line art colorization that preserves the artist's intent. Their system introduced the use of semantic color tags, providing a more intuitive interface for artists while still maintaining high quality results.

Furthermore, the incorporation of style transfer techniques has enriched manga colorization. Li et al. [17] proposed the Inkn'Hue framework, which leverages a multi-encoder VAE approach to harmonize models for shading and vibrant coloring. This model adds a more sophisticated method for color application while respecting the original artistic style of the manga.

The most recent advancement in the field has been the adoption of diffusion models. Lin et al. [23] introduced Sketch2Manga, a framework that uses diffusion models to generate high-quality color illustrations from sketches. This model excels at handling the complex screen tone patterns that are a hallmark of manga artwork.

Finally, Xiang et al. [19] advanced the use of diffusion models with their AnimeDiffusion system, which demonstrated superior performance in colorizing facial features compared to traditional GAN-based methods. Their work underscores the potential of diffusion models to capture subtle color variations while maintaining the stylistic integrity of the original artwork.

Loss functions and training strategies

The development of specialized loss functions has been instrumental in enhancing the quality of manga colorization. Johnson et al. [24] introduced perceptual losses, which significantly improved the visual fidelity of generated images by focusing on high-level feature matching rather than pixel-wise accuracy. Gatys et al. [25] further refined this concept with style-aware loss functions, which emphasize the preservation of artistic intent and style during the colorization process. As a result, the final output aligns more closely with the original artwork.

These advancements have contributed to the progression of manga colorization, each addressing unique challenges in the process. However, there remains an ongoing debate about the optimal balance between architectural choices and the impact of these choices on colorization quality. This study builds on these previous works by providing a comparative analysis of U-Net [26] and ResNet [27] architectures, offering a deeper understanding of their strengths and limitations specifically within the context of manga colorization.

3. Materials and Methods

In this section, we describe the neural architectures employed for automated manga colorization. We have investigated two deep learning models: a modified U-Net and a ResNet-based autoencoder.
We designed each architecture to address key challenges in manga colorization, such as the preservation of structural details, color consistency, and computational efficiency.

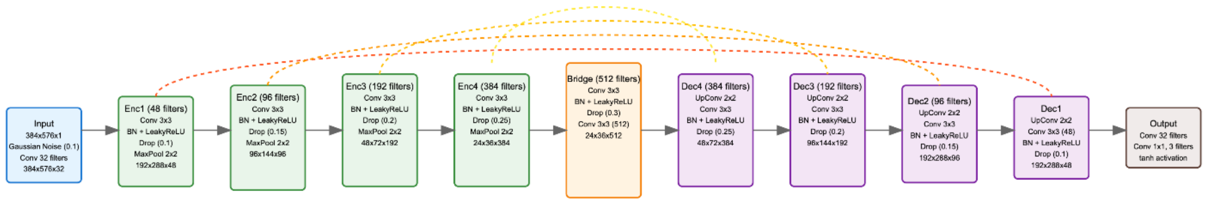

Figure 1. Modified U-Net architecture highlighting the progressive dropout, skip connections and gaussian noise layer

Neural architectures

Modified U-Net for manga colorization

The first architecture is a modified U-Net as shown in Figure 1, specifically tailored for manga colorization by integrating enhancements that improve spatial feature propagation, generalization, and stability.

Architectural Modifications: The U-Net architecture is enhanced with the following key modifications:

- Progressive Dropout Mechanism: Dropout rates are incrementally increased in the encoder path, ranging from 0.1 in the early layers to 0.3 in the deeper layers. This progressive dropout strategy helps mitigate overfitting while allowing robust feature extraction.
- Enhanced Skip Connections: Unlike the traditional U-Net, our model incorporates additional convolutional layers in the skip pathways. These layers refine transferred feature maps, reducing artifacts and improving local consistency in colorized outputs.
- Regressive Dropout in the Decoder: A decreasing dropout rate is applied in the decoder path to facilitate smoother reconstruction and reduce artifacts, ensuring better spatial coherence in colorized outputs.
- Input Gaussian Noise Layer: A Gaussian noise layer is introduced at the input stage to enhance model generalization, preventing over-reliance on specific textures and enabling more robust colorization across diverse manga styles.

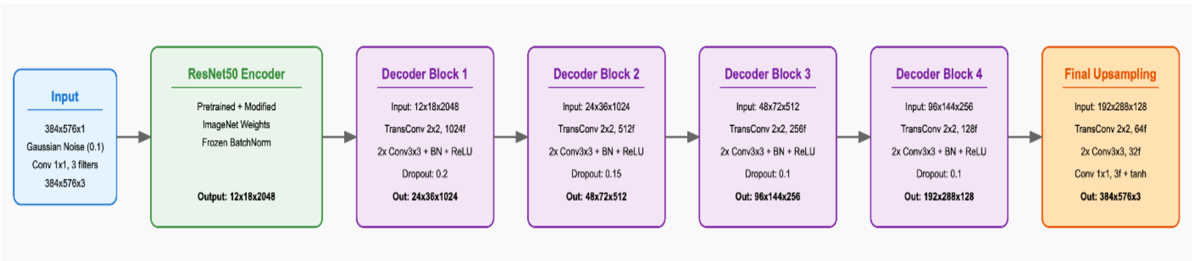

Figure 2. ResNet Autoencoder architecture showing the integration of pretrained components

ResNet-Based autoencoder

The second model, a ResNet-based autoencoder, utilizes a deep residual learning framework to enhance color consistency and feature abstraction, as illustrated in Figure 2.

Architectural features: The ResNet-based autoencoder incorporates the following key enhancements:

- Pretrained ResNet50 Backbone: The encoder is built upon a ResNet50 network pretrained on ImageNet, which allows the model to leverage robust, hierarchical feature representations learned from large-scale natural image datasets. This aids in capturing high-level semantic features that are crucial for manga image colorization.
- Adaptive Skip Connections: To ensure consistent spatial dimensions and prevent distortions in the colorized outputs, feature maps from the encoder are selectively passed to the decoder via adaptive skip connections. These connections employ feature alignment techniques to maintain spatial coherence.
- Progressive Decoder with Deconvolution Blocks: The decoder reconstructs the colorized manga images using a series of deconvolution layers. This progressive refinement process helps maintain the structural integrity of the image while improving colorization quality.
- Dynamic Resizing for Output Consistency: A dynamic resizing mechanism ensures that the final output resolution remains consistent, even when the input images have varying dimensions. This approach is particularly valuable when working with large-scale manga datasets that contain pages with different layouts.

Our modified ResNet architecture incorporates skip connections between the encoder and decoder. We refer to it interchangeably as ResUNet or ResNet.

Composite loss function

To ensure high-quality manga colorization, we designed a composite loss function that simultaneously optimizes multiple aspects of colorization accuracy. Our objective is to balance color consistency, structural preservation, and perceptual quality. The total loss function (equation 1) is formulated as follows:

$$\mathcal{L}_{total} = \alpha\mathcal{L}_{MSE} + \beta\mathcal{L}_{SSIM} + \gamma\mathcal{L}_{structure} + \sum_{i\in\{H,S,V\}} \omega_{i\mathcal{L}_i} \tag{1}$$

where each term plays a specific role in guiding the learning process:

- Mean Squared Error ($\mathcal{L}_{MSE}$): This term minimizes the pixel-wise difference between the predicted and ground truth images, ensuring overall color accuracy.
- Structural Similarity Index ($\mathcal{L}_{SSIM}$): SSIM is incorporated to enhance perceptual fidelity by prioritizing structural relationships within the image rather than just pixel-wise errors.
- Edge-Preservation Loss ($\mathcal{L}_{structure}$): This component encourages the model to maintain sharp contours and fine details, which are critical for preserving the artistic integrity of manga line art.
- HSV Color Space Loss ($\mathcal{L}_{H,S,V}$): To improve color stability and saturation balance, we introduce losses for each channel (Hue, Saturation, and Value). This helps prevent unnatural color shifts and maintains stylistic consistency.

The weighting coefficients $(\alpha, \beta, \gamma, \omega_i)$ were determined through extensive ablation studies, ensuring an optimal balance between these competing objectives.

Training strategy

The training process employs a cosine learning rate scheduler with restarts (equation 2):

$$\eta_t = \eta_{min} + \frac{1}{2}(\eta_{max} - \eta_{min})\left(1 + \cos\left(\frac{T_{cur}}{T_i}\pi\right)\right) \quad (2)$$

We use the AdamW optimizer which is the updated version of Adam. AdamW combines adaptive moment estimation decoupled with weight decay to improve generalization. The optimizer is configured with an initial learning rate of 4×10^{-4}, weight decay of 1×10^{-4}, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 1 \times 10^{-8}$. To further refine training, we apply a cosine learning rate scheduler with warm restarts [28]. This approach periodically resets the learning rate to promote better convergence. Training was conducted over 50 epochs with dynamic batch sizing based on available GPU memory.

Color accuracy metric

To effectively evaluate colorization quality with a focus on human perception, we introduce a custom color accuracy metric that operates in HSV color space. This metric provides a nuanced evaluation of color similarity by independently assessing hue, saturation, and value components with different tolerance thresholds (equation 3):

$$\text{ColorAccuracy} = \frac{1}{N}\sum_i [(h_i < 0.15) \wedge (s_i < 0.25) \wedge (v_i < 0.3)] \quad (3)$$

where h_i, s_i, and v_i represent the differences in hue, saturation, and value components respectively. It is important to note that this metric serves purely as a monitoring tool during training to evaluate how well our models approach proper color reproduction, and is not used in the training process itself, which relies on our composite loss function described in Section 3 (Composite Loss Function).

The metric is implemented as a custom metric class that performs several key operations:

1. Normalization of input tensors from [-1, 1] to [0, 1] range
2. Conversion from RGB to HSV color space
3. Computation of circular difference for hue values
4. Application of component-specific thresholds:
    - Hue: 0.15 (strict threshold to ensure color accuracy)
    - Saturation: 0.25 (balanced to prevent dull colors)
    - Value: 0.3 (more tolerant for lighting variations)

The thresholds for hue, saturation, and value were determined empirically through a series of preliminary experiments. A stricter hue threshold (0.15) was adopted to ensure accurate representation of character colors. In contrast, a more lenient value threshold (0.3) was chosen to accommodate the natural variations in shading and lighting that frequently occur in manga artwork. The saturation threshold (0.25) serves as a balance to prevent dull colors. By monitoring this metric during training, we can better understand how our models progress in terms of perceptual color accuracy, even as they optimize for the composite loss function.

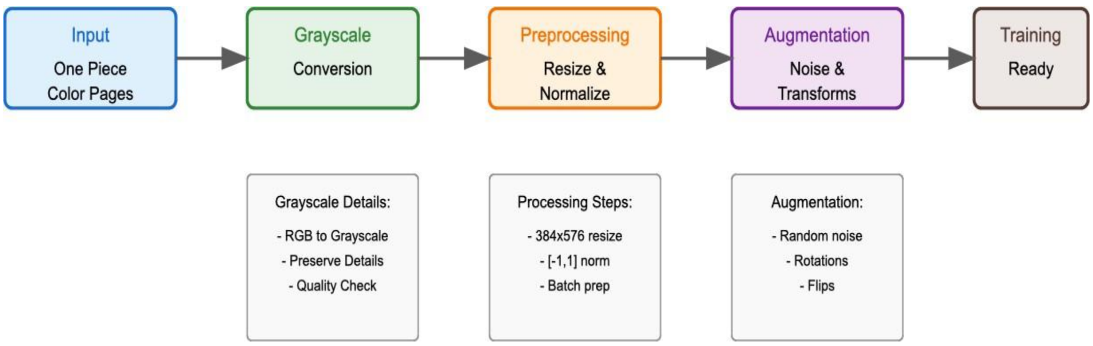

Figure 3. Data preprocessing pipeline, illustrating the transformation stages from raw manga images to training-ready inputs

4. Dataset And Experimental Setup

Dataset construction

Our dataset is derived from the colored edition of *One Piece* manga, with training samples drawn from the first 50 volumes and testing samples from the subsequent 50 volumes. Both training and testing set consists of 7237 images of size (768×1152). This dataset provides a diverse and high-quality source of training data. This selection is motivated by its consistent artistic style, the availability of professionally colored versions. Its wide range of scenes and character interactions serve as rich training examples.

To construct paired training data, we employed a systematic pipeline, illustrated in Figure 3:

i. Extracting high-resolution colored images from official *One Piece* colored editions.
ii. Converting colored images to grayscale using standardized desaturation techniques.
iii. Ensuring precise alignment between grayscale and color images through automated quality verification.
iv. Extracting metadata (e.g., chapter and volume information) for structured dataset management.

This process eliminates common dataset inconsistencies where grayscale versions may not perfectly align with their colorized counterparts, ensuring reliable training pairs for learning-based colorization.

Data preprocessing

To prepare the dataset for training, we implemented a preprocessing pipeline that transforms raw images into a standardized format suitable for deep learning models. The pipeline is depicted in Fig. 3.
The key preprocessing steps include:

- Resizing: All images are scaled to a uniform resolution of 384×576 pixels to maintain consistency across the dataset.
- Normalization: Pixel values are normalized to the $[-1, 1]$ range to facilitate stable neural network training.
- Data Augmentation: To improve generalization, various augmentation techniques are applied:
  i) Adaptive noise injection to simulate natural variations in manga printing and scanning.
  ii) Geometric transformations such as random flips and rotations to enhance model robustness.
  iii) Color space augmentations applied to the colored images to introduce variations in hue and saturation.
- Memory Optimization: We employ dynamic batching and efficient data prefetching to optimize training performance on limited GPU resources.

Experimental setup

Experiments were conducted on a GPU-accelerated environment using Google Colab with TensorFlow. To enhance computational efficiency, we enabled GPU memory optimizations and activated XLA Just-In-Time (JIT) compilation. The implementation follows modern deep learning best practices to ensure reproducibility and optimal performance.

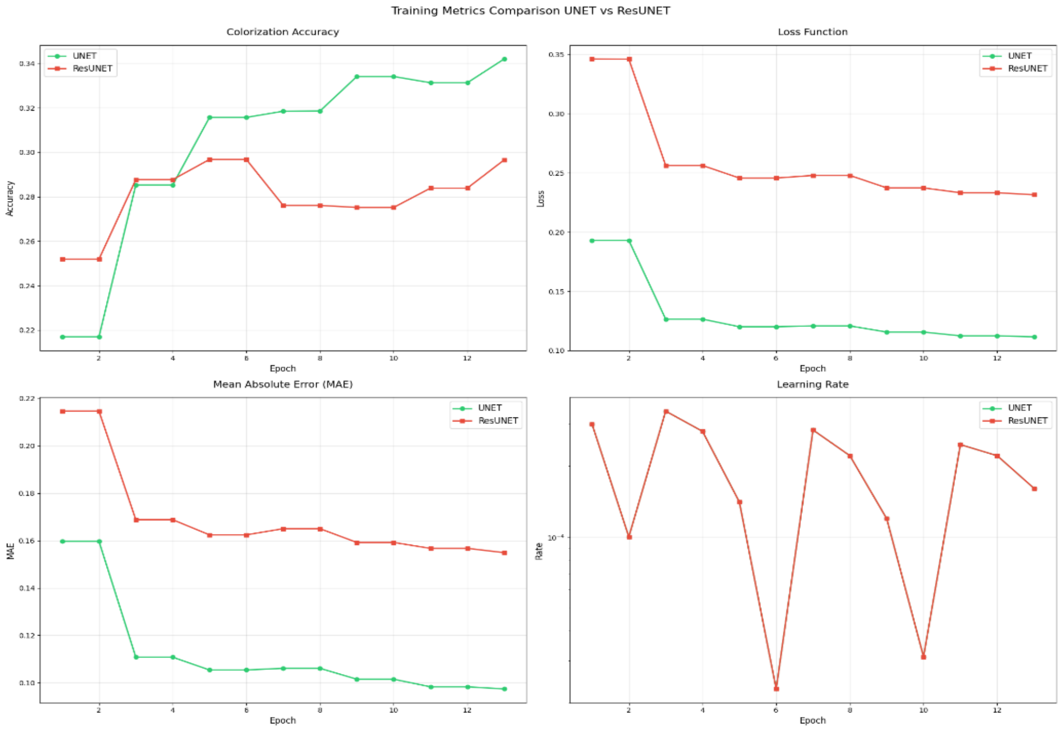

Figure 4. Training metrics evolution showing convergence patterns and performance indicators

5. Experimental Results

Quantitative analysis

Our quantitative evaluation examines four key performance metrics: colorization precision, training loss, Mean Absolute Error (MAE), and learning rate evolution, as shown in Figure 4. The training process was automatically terminated at epoch 13 for both architectures due to early stopping criteria. These metrics provide valuable insights into learning dynamics and overall model performance.

The colorization precision curve (top left plot in Fig. 4) reveals a consistent improvement in U-Net's performance, rising from 0.22 to 0.34 by epoch 12. In contrast, ResUNet exhibits a sharp initial gain, reaching 0.29 by epoch 4, but stabilizes at a lower level around 0.28. This suggests that U-Net progressively refines its colorization ability throughout training, while ResUNet plateaus earlier.

The training loss curve (top right plot) provides further insights. ResUNet initially exhibits a higher loss (0.35), whereas U-Net starts lower at approximately 0.19. Both architectures demonstrate rapid early improvement, yet U-Net maintains consistently lower loss values throughout training.

By epoch 12, U-Net stabilizes at 0.11, while ResUNet at 0.23, indicating better optimization characteristics in U-Net.

The Mean Absolute Error (bottom left plot) reinforces these findings. U-Net achieves lower MAE, stabilizing around 0.10, compared to ResUNet's 0.16. This 0.06 reduction in MAE suggests that U-Net produces more precise pixel-wise color predictions, capturing finer details more accurately.

In terms of computational efficiency, we conducted benchmarking on inference speed and memory consumption. The results indicate that U-Net demonstrates superior efficiency compared to ResUNet. Specifically, U-Net achieves 4.95 frames per second (FPS), while ResUNet reaches 2.65 FPS. Regarding memory usage, U-Net requires 2959.7 MB of CPU memory, whereas ResUNet consumes 3044.6 MB. Both models exhibit comparable GPU memory usage of approximately 617 MB.

The learning rate adaptation (bottom right plot) highlights distinct training dynamics between the architectures. ResUNet exhibits pronounced cyclical variations due to the cosine annealing schedule, while U-Net follows a more stable learning rate trajectory. This suggests that the two architectures respond differently to learning rate scheduling, influencing their respective optimization behaviors. The stabilization of these trends contributed to the early stopping decision at epoch 13.

These results challenge initial expectations regarding ResUNet's superiority. U-Net demonstrates consistently better performance across key metrics, suggesting that architectural complexity does not necessarily correlate with improved results in manga colorization. This finding underscores the importance of empirical evaluation in model selection, highlighting U-Net as a more effective choice for this task.

To further contextualize these findings within the existing literature, we adapted a similar architecture proposed in [29], due to the limited availability of source code for direct manga colorization. Under the same experimental setup, Hosen's method achieved a color accuracy of 18.17%, an MAE of 0.7423, and a loss of 0.6088. In terms of computational efficiency, Hosen's model exhibited a higher inference speed, approximately 13.1% faster than our ResUNet (2.51 FPS), while consuming about 0.25% less CPU memory; GPU memory usage was identical. Although Hosen's method demonstrates marginally better efficiency, our ResUNet achieves substantially higher accuracy and lower error rates, providing a more effective solution for manga colorization.

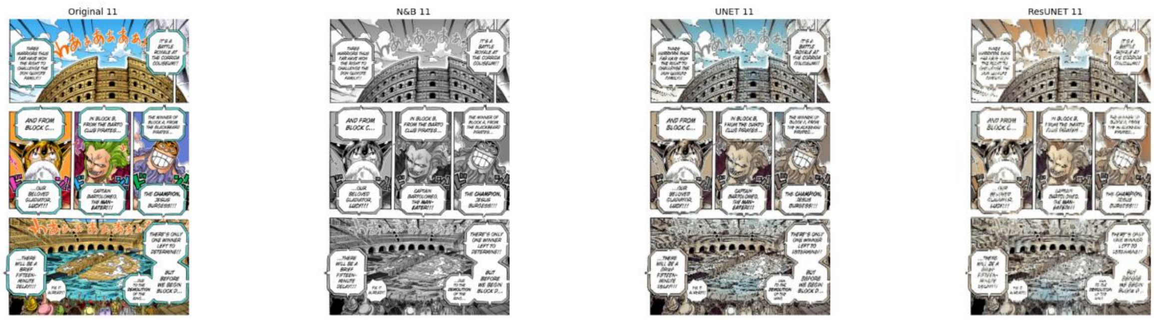

Figure 5. Comparative visualization of colorization results across different manga panels

Qualitative assessment

To evaluate the visual quality of our colorization results, we conducted a detailed comparison of both architectures against the original-colored manga panels. Figure 5 displays four versions of the same panel: the original-colored image, the grayscale input, and the outputs from U-Net and ResUNet. This layout facilitates a direct comparison of each model's ability to reconstruct color information and artistic detail.

Both models preserve the fundamental structure of the original artwork, yet they exhibit distinct approaches to colorization. The top panel, depicting the Corrida Coliseum against a sky background, highlights each model's ability to render large architectural elements and atmospheric effects. U-Net excels in maintaining contrast between the stone textures and sky, whereas ResUNet offers a more uniform but subdued interpretation.

The middle section, featuring three-character portraits, serves as a key test for facial detail preservation and emotional expressiveness. Each portrait presents distinct characteristics—Lucy's determined expression, Bartolomeo's sharp features, and Jesus Burgess's dynamic pose. U-Net demonstrates superior preservation of facial nuances and emotional depth, particularly in fine-grained details. While ResUNet produces coherent results, it tends to slightly flatten these expressive details, reducing the overall impact.

The bottom panel, which showcases a wide view of the coliseum with water elements and crowd scenes, illustrates how each model handles complex scene composition. U-Net retains more of the original vibrancy, particularly in water reflections and architectural intricacies, whereas ResUNet, though maintaining consistency, renders a slightly less dynamic interpretation of the scene.

Notably, both models successfully preserve manga-specific elements such as speech bubbles, text overlays, and panel boundaries. However, U-Net exhibits superior adherence to the original artistic style, particularly in background effects and environmental textures. While ResUNet ensures consistency, it occasionally diminishes the artistic dynamism of the original panels, leading to a more uniform but less expressive rendering.

Table 1. Post-processing enhancement parameters

Parameter	Temperature	Tint	Vibrance	Saturation	Exposure	Contrast	Highlights	Shadows	Whites
Value	-7	+4	+50	+30	+0.1	+10	+5	+5	+20

6. Analysis and Discussion

Post-processing enhancement analysis

While initial colorization outputs may appear subdued, the application of HSV adjustments in post-processing reveals the true potential of our models. Table 1 details the color grading parameters we applied to enhance the models' outputs.

As demonstrated in Figure 6, these adjustments significantly improve the visual appeal of both models' outputs. The enhanced results reveal that our architecture successfully captures underlying spatial and contextual relationships necessary for meaningful colorization, even if their direct output appears conservative. Particularly notable improvements can be observed in the sky and water renderings, where enhanced tones reveal the models' inherent understanding of scene composition. This enhancement process not only improves the aesthetic quality but also demonstrates that our models generate coherent base colorization that can be effectively refined through post-processing.

Figure 6. Comparison of original model outputs (left) with post-processed results (right) for both U-Net and ResUNet architectures

Interpretation of results

Our comparative analysis of U-Net and ResNet architecture reveals significant insights into manga colorization approaches. The U-Net implementation demonstrated notably superior performance, achieving a final accuracy of 0.342 compared to ResNet's 0.297. This substantial difference challenges our initial assumptions about the potential advantages of ResNet's more complex architecture. Particularly noteworthy is U-Net's improvement trajectory, showing a 57.6% increase in accuracy over the training period, compared to ResNet's more modest 17.7% improvement.

The performance disparity becomes even more evident when examining the loss metrics, where U-Net achieved a final loss of 0.112 versus ResNet's 0.232. This difference suggests that U-Net's simpler architecture may be better suited for the specific challenges of manga colorization, particularly in capturing the nuanced details characteristic of this artistic medium.

Loss function weight optimization

Our initial experiments with balanced loss components (MSE 40%, SSIM 20%, HSV 20%) revealed significant convergence issues during the training process. As illustrated in Figure 7, this configuration led to persistent oscillations that prevented the model from reaching a stable optimum. Despite achieving relatively low loss values around 0.35, the resulting colorization exhibited inconsistent quality and frequent grayscale bias.

These observations led us to fundamentally reconsider our approach to loss weighting. After several tests, we discovered that heavily prioritizing the MSE component ($\alpha = 0.9$) while maintaining a smaller but crucial structural term ($\beta = 0.1$) produced significantly more stable training dynamics. This revised configuration addresses two critical aspects: the dominant MSE term ensures consistent and accurate color reproduction, while the SSIM component maintains structural coherence without interfering with color stability.

The success of this asymmetric weighting strategy challenges the intuitive assumption that balanced loss components would yield optimal results. Our findings suggest that in the specific context of manga colorization, strong color accuracy constraints combined with subtle structural guidance provide a more effective learning framework than attempting to simultaneously optimize multiple competing objectives with equal importance. Therefore, we used for the training the following Loss Function (equation 4):

$$\mathcal{L}_{total} = 0.9\mathcal{L}_{MSE} + 0.1\mathcal{L}_{SSIM} \tag{4}$$

The disparity in computational requirements between the architecture presents practical implications. U-Net's more efficient resource utilization (12.4 GB GPU memory versus ResNet's 13.8 GB) and faster inference time (as discussed in Quantitative section) suggests it may be more suitable for production environments where computational resources are constrained.

To further assess the applicability of our proposed loss function, we conducted additional experiments on another publicly available dataset [30]. We utilized a comparable number of samples (approximately 7K) to those in our original dataset. The dataset provides images at an original resolution of 1024×1024. We applied the same preprocessing steps and experimental setup as before, with the exception of batch size (set to 2 instead of 4 due to memory constraints). Our ResUNet model achieved a color accuracy of 29.78%, an MAE of 0.1802, and a loss of 0.1374. These results demonstrate the broader applicability and robustness of the proposed loss function across different datasets.

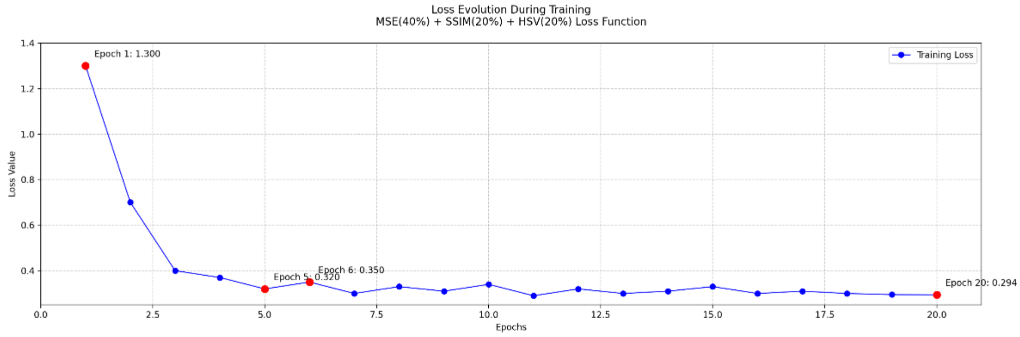

Figure 7. Training instability observed with initial balanced loss weights showing persistent oscillations and failure to converge

Effect of image size

To balance computational efficiency and model performance, we downscaled the original images from 768×1152 to 384×576. While standard ResNet architectures typically accept input dimensions of 224×224, we did not employ a pretrained ResNet; instead, we implemented a custom ResUNet architecture that allowed us to experiment with varying image resolutions. To evaluate the effect of image size, we trained our model with both 224×224 and 384×576 inputs. At 224×224, the model achieved a color accuracy of 29.25%, an MAE of 0.3096, and a loss of 0.1920. In contrast, training with 384×576 inputs yielded superior results, with a color accuracy of 29.66%, an MAE of 0.2315, and a loss of 0.1549. It demonstrates that larger input dimensions (384×576) lead to better overall performance in our ResUNet model due to rich spatial details during training.

Limitations and challenges

Despite the strong quantitative results, several limitations warrant consideration. First, while U-Net shows superior metrics overall, both architectures still struggle with maintaining consistent coloring across complex multi-panel layouts. This limitation becomes particularly apparent in scenes with varying lighting conditions or complex character interactions.

The Mean Absolute Error results (U-Net: 0.097, ResNet: 0.155) indicate room for improvement in pixel-level accuracy for both architectures. Additionally, we observed that both models occasionally struggle with preserving the distinctive artistic techniques common in manga, particularly in areas with complex screen tone patterns or detailed background elements.

Practical implications

Our findings have direct implications for real-world manga colorization pipelines. U-Net's combination of superior performance metrics and lower resource requirements makes it particularly attractive for commercial applications. The 39.0% improvement in MAE for U-Net, compared to ResNet's 27.8%, suggests it would provide more reliable results in production environments.

The observed training dynamics also inform practical deployment considerations. U-Net's more stable learning progression and better final metrics indicate it might require less fine-tuning and post-processing in production scenarios, potentially reducing the overall computational and human resource requirements.

Future research directions

This study opens several promising avenues for future research. One of the primary challenges remaining is improving color consistency across multiple panels, while also preserving U-Net's ability to capture fine details effectively. Addressing this could enhance the quality and coherence of the colorization process in longer manga sequences. Additionally, further optimization of the ResNet architecture to reduce its computational overhead, while retaining its advantages in global color consistency, could offer more efficient solutions without compromising performance.

Future research efforts may also explore the following directions:

- Combining the strengths of U-Net's efficient feature extraction and ResNet's robust feature representation in hybrid architectures.
- Fine-tuning the loss function weighting strategy to improve the handling of edge cases and more complex scenes.
- Investigating the application of attention mechanisms specifically tailored for manga-style artwork to enhance the model's focus on relevant image features.
- Leveraging transfer learning techniques to better accommodate diverse artistic styles, ensuring broader applicability of colorization models.

7. Conclusion

In this paper, we performed a comparative analysis of U-Net and ResNet architectures for manga colorization, highlighting their strengths and weaknesses. While the results show clear performance trends, additional research is necessary to draw more definitive conclusions. The proposed composite loss function shows potential for enhancing colorization quality.

The findings offer valuable insights for advancing manga colorization systems, with the evaluation of performance metrics and resource requirements providing a foundation for optimizing model selection and training strategies. However, real-world applications will necessitate more extensive testing across diverse manga styles. Manga colorization remains a complex challenge, requiring improvements in computational efficiency, architectural design, and style preservation. Future research should focus on novel deep learning approaches, enhanced loss functions, and dataset diversification to improve generalization. As demand for high-quality colored manga increases, continued innovation will be critical to bridge the gap between automated and artist-driven colorization.

8. Author Contribution Statement

In the study, Author 1 contributed to forming the idea, making the design, literature review, and the analysis of the results; Author 2 contributed to provision of the materials and examination of the results, checking the spelling and checking the article in terms of content; Author 3 contributed to the organization of the research.

9. Ethics Committee Approval and Conflict of Interest

There is no need for an ethics committee approval in the prepared article. There is no conflict of interest with any person/institution in the prepared article.

10. Ethical Statement Regarding the Use of Artificial Intelligence

No artificial intelligence-based tools or applications were used in the preparation of this study. The entire content of the study was produced by the author(s) in accordance with scientific research methods and academic ethical principles.

11. References

[1] Anime News Network, "Manga market in Japan hits record 612.6 billion yen in 2020," Feb. 2021, archived Nov. 14, 2022, accessed Nov. 14, 2021. [Online]. Available: https://www.animenewsnetwork.com/news/2021-02-26/manga-market-in-japan-hits-record-612.6-billion-yen-in-2020/.169987.

[2] S. Guadarrama, R. Dahl, D. Bieber, M. Norouzi, J. Shlens, and K. Murphy, "PixColor: Pixel recursive colorization," arXiv:1705.07208, 2017.

[3] M. Kumar, D. Weissenborn, and N. Kalchbrenner, "Colorization transformer," arXiv:2102.04432, 2021.

[4] T. Jiramahapokee, "Inkn'hue: Enhancing manga colorization from multiple priors with alignment multi-encoder VAE," arXiv:2311.01804, 2023.

[5] L. Zhang, Y. Ji, X. Lin, and C. Liu, "Style transfer for anime sketches with enhanced residual U-Net and auxiliary classifier GAN," in *Proc. 4th IAPR Asian Conf. Pattern Recognit. (ACPR)*, 2017, pp. 506–511.

[6] Y. Ci, X. Ma, Z. Wang, H. Li, and Z. Luo, "User-guided deep anime line art colorization with conditional adversarial networks," in *Proc. 26th ACM Int. Conf. Multimedia*, 2018, pp. 1536–1544.

[7] Y. Wu, X. Wang, Y. Li, H. Zhang, X. Zhao, and Y. Shan, "Towards vivid and diverse image colorization with generative color prior," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2021, pp. 14377–14386.

[8] N. Zabari, A. Azulay, A. Gorkor, T. Halperin, and O. Fried, "Diffusing colors: Image colorization with text guided diffusion," in *SIGGRAPH Asia Conf. Papers*, 2023, pp. 1–11.

[9] S. Anwar, M. Tahir, C. Li, A. Mian, F. S. Khan, and A. W. Muzaffar, "Image colorization: A survey and dataset," arXiv:2008.10774, 2020.

[10] M. B. Islam and M. I. Hosen, "Emotion, age and gender prediction through masked face inpainting," in *Int. Conf. Pattern Recognit.*, 2022, pp. 37–48.

[11] M. I. Hosen, T. Aydin, and M. B. Islam, "WNet: A dual-encoded multi-human parsing network," *IET Image Process.*, vol. 18, no. 12, pp. 3316–3328, 2024.

[12] M. I. Hosen and S. Kahraman, "Deep learning-based efficient drone detection and tracking in real-time," in *Proc. 33rd Signal Process. Commun. Appl. Conf. (SIU)*, 2025, pp. 1–4.

[13] R. Zhang, P. Isola, and A. A. Efros, "Colorful image colorization," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 649–666.

[14] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Let there be color! Joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification," *ACM Trans. Graph.*, vol. 35, no. 4, pp. 1–11, 2016.

[15] C. Furusawa, K. Hiroshiba, K. Ogaki, and Y. Odagiri, "Comicolorization: Semi-automatic manga colorization," in *SIGGRAPH Asia Tech. Briefs*, 2017, pp. 1–4.

[16] L. Zhang, C. Li, T.-T. Wong, Y. Ji, and C. Liu, "Two-stage sketch colorization," *ACM Trans. Graph.*, vol. 37, no. 6, pp. 1–14, 2018.

[17] Y. Li, M.-Y. Liu, X. Li, M.-H. Yang, and J. Kautz, "A closed-form solution to photorealistic image stylization," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 453–468.

[18] H. Kim, H. Y. Jhoo, E. Park, and S. Yoo, "Tag2Pix: Line art colorization using text tag with SECat and changing loss," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2019, pp. 9056–9065.

[19] Y. Cao, X. Meng, P. Mok, X. Liu, T.-Y. Lee, and P. Li, "AnimeDiffusion: Anime face line drawing colorization via diffusion models," arXiv:2303.11137, 2023.

[20] S. Chen, D. Li, Z. Bao, Y. Zhou, L. Tan, Y. Zhong, and Z. Zhao, "Manga generation via layout-controllable diffusion," arXiv:2412.19303, 2024.

[21] A. Maejima *et al.*, "Continual few-shot patch-based learning for anime-style colorization," *Comput. Vis. Media*, vol. 10, no. 4, pp. 705–723, 2024.

[22] Y. Cao, X. Meng, P. Mok, T.-Y. Lee, X. Liu, and P. Li, "AnimeDiffusion: Anime diffusion colorization," *IEEE Trans. Vis. Comput. Graph.*, vol. 30, no. 10, pp. 6956–6969, 2024.

[23] J. Lin, X. Liu, C. Li, M. Xie, and T.-T. Wong, "Sketch2Manga: Shaded manga screening from sketch with diffusion models," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2024, pp. 2389–2395.

[24] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 694–711.
[25] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 2414–2423.
[26] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv. (MICCAI)*, 2015, pp. 234–241.
[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.
[28] I. Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with warm restarts," arXiv:1608.03983, 2016.
[29] M. I. Hosen and M. B. Islam, "Masked face inpainting through residual attention U-Net," in *Proc. Innov. Intell. Syst. Appl. Conf. (ASYU)*, 2022, pp. 1–5.
[30] MichaelP84, "Manga-colorization-dataset," Hugging Face Dataset, 2024, accessed Sep. 20, 2025. [Online]. Available: https://huggingface.co/datasets/MichaelP84/manga-colorization-dataset.