Experimental Exploration of Neural Style Transfer: Hyperparameter Impact and VGG Feature Dynamics in Batik Motif Generation

Happy Gery Pangestu[1], Andi Prademon Yunus[1*], Siti Khomsah[1], Yit Hong Choo[2], Takao Ito[3]
[1]Telkom University, Banyumas, Indonesia
[2]Deakin University, Australia
[3]Hiroshima University, Japan
Email: happygerypangestu@student.telkomuniversity.ac.id, andiay@telkomuniversity.ac.id,
sitijk@telkomuniversity.ac.id, y.choo@deakin.edu.au, itotakao@hiroshima-u.ac.jp
[*]Corresponding Author

Abstract

Innovating traditional batik designs while preserving their cultural essence remains a significant challenge in the intersection of heritage conservation and computational creativity. This study addresses this challenge by optimizing Neural Style Transfer (NST), a deep learning technique to synthesize batik motifs that harmonize structural fidelity and stylistic authenticity. Focusing on hyperparameter adjustments tailored to batik's abstract geometries, we systematically evaluate the impact of layer selection in VGG and pooling operations (max-pooling vs. average-pooling) on style-content synthesis. Experiments reveal that shallow layers (e.g., conv2, conv4) preserve explicit motifs and edge details (SSIM 0.85), while deeper layers (conv16, conv24) generate abstract textures. Average-pooling demonstrates superior stability, achieving smoother convergence (loss stabilized at 0.5 vs. 3.0 for max-pooling) and higher structural coherence (SSIM 0.6963 vs. 0.6634), whereas max-pooling introduces fragmented artifacts due to gradient explosion. The optimized framework, validated through quantitative metrics (MSE Loss 0.02, SSIM 0.82) and qualitative artisan evaluations, successfully transferring the style image into the image, while maintaining the original content of the image, with adjusted hyperparameters. This work advances AI-driven tools for batik preservation, offering a scalable methodology to sustain Indonesia's intangible heritage in the digital era..

Keywords: Batik preservation, Neural Style Transfer, hyperparameter exploration, VGG, pooling operations, cultural coherence.

1. Introduction

Batik, a traditional Indonesian textile art form, is renowned for its intricate motifs and cultural symbolism. Each pattern reflects regional philosophies, histories, and identities, making batik a cornerstone of Indonesia's intangible heritage [1]. Despite its UNESCO recognition in 2009 [2], batik designs remain challenging to innovate while preserving their cultural identity. Artisans and designers must carefully balance traditional values with contemporary aesthetics to create motifs that stay meaningful in a modern context while maintaining cultural authenticity. This delicate balance highlights the complexity of embracing innovation while safeguarding batik's rich heritage.[3].

The tension between tradition and modernity defines batik's creative process, as artisans strive to innovate motifs for contemporary markets while preserving their symbolic essence. Batik patterns encode cultural narratives and spiritual values, demanding careful reinterpretation to avoid diluting their heritage. Globalization and shifting aesthetics pressure designers to modernize motifs through abstract geometries or minimalist palettes yet rigid adherence to tradition risks stagnation. The lack of systematic frameworks forces reliance on trial-and-error methods, while digital tools prioritize efficiency over cultural depth, often sidelining traditional techniques like hand-drawn *canting* or natural dyes. Bridging this divide requires merging technological advancements, such as AI-driven design, with respect for traditional knowledge, ensuring batik evolves as a living art form rooted in cultural identity.[4]. Current practices often rely on manual reinterpretation of historical patterns, a time-intensive process constrained by limited technical guidance and creative inspiration [5]. While digital tools like graphic software have streamlined design workflows, they lack the capacity to autonomously generate culturally coherent patterns or synthesize traditional and contemporary elements [6]. Consequently, there remains a critical need for methodologies that augment human creativity while safeguarding batik's cultural integrity.

Recent advances in computational creativity, particularly Artificial Intelligence (AI), present transformative opportunities for cultural heritage preservation. Among these, Neural Style Transfer (NST), first introduced by Gatys et al. in their foundational work "A Neural Algorithm of Artistic Style" [7], has emerged as a pivotal technique for artistic synthesis. NST algorithm leverages deep convolutional neural networks to decompose images into separable representations of content (structural features) and style (textural patterns), enabling the synthesis of novel artworks by recombining these elements across distinct sources. This framework, which optimizes a generated image to match the content of one input and the stylistic attributes of another via Gram matrix-based style loss, has since inspired adaptations in diverse artistic domains. In batik research, NST has been experimentally applied to generate hybrid designs by fusing traditional motifs with modern visual elements. For example, [8] propose a local style transfer model for batik patterns with enhanced edges and use the Stable Diffusion AI painting tool for style transfer Results: The local style transfer model generated images with good performance in detail texture and color space, achieving a peak signal-to-noise ratio (PSNR) of 25.3 dB and a structural similarity index measure (SSIM) of 0.85. However, the Stable Diffusion tool had limitations in inheriting style and content details. The results is local style transfer model generated images with good performance in detail texture and color space, achieving a peak PSNR of 25.3 dB and a structural similarity index measure (SSIM) of 0.85.. Similarly, [9] demonstrates the use of CNN for batik style transfer, showing that the technique can effectively combine the content of an image with the style of traditional batik, producing visually appealing and culturally coherent designs.

To enhance the application of NST for batik motif synthesis, this study focuses on systematically optimizing hyperparameters critical to style-content adaptation. Batik's abstract geometries and symbolic textures, which diverge from photorealistic imagery, necessitate tailored adjustments to standard NST workflows. Specifically, we investigate (1) selection layers in VGG to determine their impact on preserving batik's structural and stylistic nuances; and (2) pooling operations to evaluate their role in feature aggregation and texture synthesis. By testing these hyperparameters, we aim to develop an NST framework that aligns with batik's artistic principles. The efficacy of the optimized pipeline is validated through quantitative metrics MSE Loss and SSIM.

2. Methodology

The Neural Style Transfer framework begins with feature extraction, where a pretrained convolutional neural network (CNN) is employed to derive hierarchical representations from batik images. These features capture multi-scale attributes ranging from coarse patterns to fine-grained textures, collectively defining the visual lexicon of batik motifs. For style images, the Gram matrix a statistical measure of channel-wise feature correlations is computed from the extracted representations. The Gram matrix $\boldsymbol{G_{ij}^l}$ for layer $\boldsymbol{l}$ is calculated as:

$$G_{ij}^{l} = \sum_{k} F_{ik}^{l} F_{jk}^{l} \tag{1}$$

where $\boldsymbol{F_{ik}^l}$ and $\boldsymbol{F_{jk}^l}$ denote the activations of the $\boldsymbol{i}$-th and $\boldsymbol{j}$-th filters at spatial position $\boldsymbol{k}$ in layer $\boldsymbol{l}$. This matrix quantifies stylistic textures by capturing interdependencies between feature channels. To assess the impact of domain adaptation, this study compares feature extraction performance between generic pretrained CNNs (e.g., ImageNet-trained models) and counterparts fine-tuned on batik datasets, evaluating their ability to preserve culturally significant visual elements during style transfer.

The Gram matrix calculation serves as the cornerstone for encoding stylistic attributes. When a style image is processed through the CNN, the resulting feature tensor (dimensioned as batch size × channels × height × width) is reshaped into a two-dimensional matrix. The Gram matrix is derived by computing the inner product of this reshaped matrix with its transpose, producing a channel-wise correlation matrix. Each element is normalized by the total number of elements in the original tensor $\boldsymbol{b} \times \boldsymbol{c} \times \boldsymbol{h} \times \boldsymbol{w}$, ensuring scale invariance across varying image resolutions.

Loss computation integrates two components, a content loss and style loss. Content loss quantifies structural fidelity using the mean squared error (MSE) between feature maps of the generated image xx and the original content image $\vec{\boldsymbol{p}}$ expressed as :

$$L_{\text{content}}(\vec{p}, \vec{x}, l) = \frac{1}{2} \sum_{i,j} (F_{i,j}^{l} - P_{i,j}^{l})^{2} \tag{2}$$

where $F_{i,j}^l$ and $P_{i,j}^l$ represent activations of the i-th filter at position j in layer l for the generated and content images, respectively. Style loss evaluates stylistic alignment by comparing Gram matrices of the generated image x and reference style image a [10]. The layer-wise style error E_l is calculated as:

$$E_l = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} (G_{ij}^{l} - A_{ij}^{l})^{2} \tag{3}$$

where N_l is the number of filters, M_l is the feature map size, and G_{ij}^l and A_{ij}^l are Gram matrices of the style reference and generated images. The total style loss aggregates contributions across layers:

$$L_{\text{style}}(\vec{a}, \vec{x}) = \sum_{l=0}^{L} w_l e_l \tag{4}$$

with w_l weighting layer-specific importance. The total loss combines these terms as:

$$L_{\text{total}}(\vec{p}, \vec{a}, \vec{x}) = \alpha L_{\text{content}}(\vec{p}, \vec{x}) + \beta L_{\text{style}}(\vec{p}, \vec{x}) \tag{5}$$

where α and β are hyperparameters balancing structural preservation and stylistic adaptation. A total variation regularization term L_{TV} is incorporated to enhance spatial smoothness.

Pixel optimization iteratively minimizes the total loss using the adam optimizer, which adaptively adjusts learning rates per parameter based on gradient momentum $\boldsymbol{m_t}$ and variance $\boldsymbol{v_t}$:

$$\theta_{t+1} = \theta_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \varepsilon}} \tag{6}$$

where η is the initial learning rate $\hat{m}_t$ and $\hat{v}_t$ are bias-corrected estimates of the first and second moments, and ϵ prevents division by zero. This approach accelerates convergence while mitigating oscillations common in gradient-based optimization. Over successive epochs, pixel values are updated via backpropagation, aligning the generated image with target content and style attributes.

Evaluation employs quantitative and qualitative metrics, the equation of MSE can be seen in the equation below:

$$MSE = \frac{1}{n}\sum_{i}^{n}(\hat{y}_i - y_i)^2 \tag{7}$$

Where $\hat{y}_i$ is generated pixel images and y_i is reference images [11]. The Structural Similarity Index (SSIM) assesses perceptual fidelity through luminance $(\mu_x\mu_y)$, contrast (σ_x, σ_y), and structural coherence (σ_{xy}):

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{8}$$

where C_1, C_2 stabilize the denominator. These metrics are complemented by expert-led qualitative assessments to ensure cultural relevance and aesthetic integrity in synthesized batik patterns.

3. Results and Discussions

3.1. *Testing Scenario*

The experimental framework evaluated NST performance across multiple scenarios involving diverse batik motifs, CNN architectures, and hyperparameter configurations. Content-style pairings were designed to test both representational fidelity. Key experimental configurations are summarized in Table 1.

Table 1. Testing Scenario of Neural Style Transfer.

Sub	Pooling	Layers		Weight	
		Content	Style	α	β
S1	Max	[2, 8, 16]	[2, 8, 16]	1	1e8
S2	Max	[2, 8, 16]	[2, 8, 16]	1	1e8
S2	Avg	16	8	1	1e8
S4	Max	16	8	1	1e8
S5	Avg	8	8	1	1e8
S6	Max	8	8	1	1e8
S7	[Avg, Max]	16	8	1	1e8
S8	[Avg, Max]	16	8	1	1e8
S9	[Avg, Max]	16	8	1	1e8

3.2. *Performance analysis*

In the implementation of Neural Style Transfer (NST), the selection of layers within the Convolutional Neural Network (CNN) model significantly influences the outcomes of image synthesis. This dependence arises because each layer governs the characteristics of the features extracted during the process. The content representations derived from these layers, which encode structural and textural attributes of the input image, are visually illustrated in Fig. 1, Fig. 2, and Fig. 3.

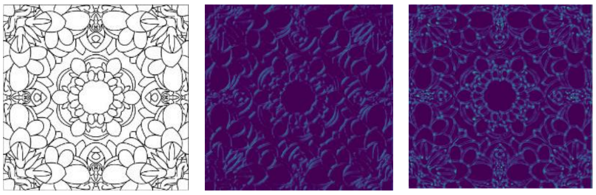

Figure 1. Compare feature map shallow layer of the pattern image.

When NST utilizes shallower layers, the resulting features tend to be simple and focus on the basic elements of the image, such as lines, edges, and basic geometric patterns. This makes the style transfer results more literal and visually recognizable.

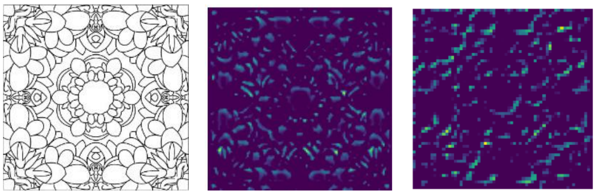

Figure 2. Compare feature map inner layer of the pattern image.

In contrast, the use of deeper layers produces more abstract and complex features. Deep layers capture the representation of the image, such as shape, texture, and deeper concepts. To directly determine the effect of layer depth on the results of style transfer, we also conducted experiments with various layers that would be applied to 2 batik images. The results of the combination of batik motifs can be seen in the following image:

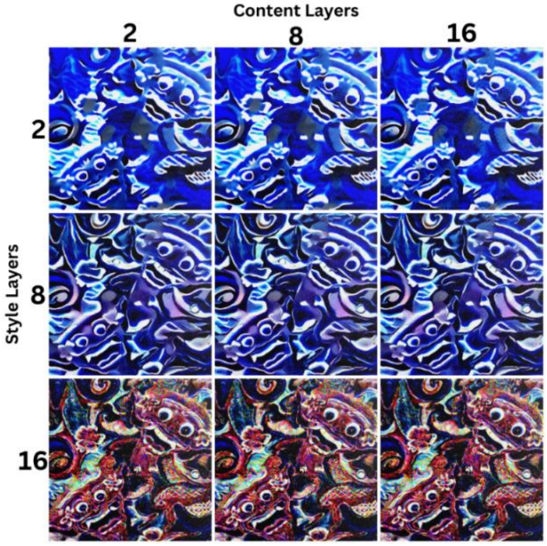

Figure 3. Comparison of layer difference results.

As can be seen in Fig. 3 (Sub 1) above, the effect of layer depth variation on neural style transfer, with a focus on the difference in influence between the content layer and the style layer. The test results show that the variation of the content layer depth (2, 8, 16) does not significantly affect the visual results, indicating that the content representation can be maintained even when using deeper layers. In contrast, the variation of the style layer depth shows a significant effect on the final result, where increasing the style layer depth produces increasingly complex patterns. This can happen because the ability of deeper layers in CNN can capture more complex style features. Therefore, determining the weight value in the loss function is an important factor to balance between maintaining the content structure and applying details to the style.

In the second experiment, we used other batik motifs to test the results of the influence of different layers. The batik used for this test was peacock batik as content and also megamendung batik as style. The batik image and style transfer results can be seen in the image below:

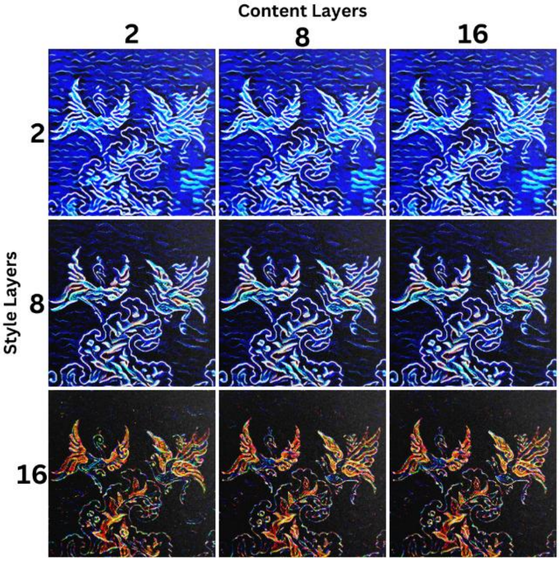

Figure 4. Comparison of difference layer results.

It can be seen in Fig. 4 (Sub 2), the results of combining batik motifs in the second experiment to determine the effect of different layers on the features extractor. The results show that selecting the initial layer up to the first 1/4 of the network provides more visible or more representative transfer results, conversely if the feature map taken from the layer is too deep, the features captured will also tend to be more abstract. The test results show that the selection of layers needs to be adjusted to the desired artistic goals. Shallow layers are more suitable for producing batik with explicit patterns and more representative style transfer results, while deep layers will create designs with a style that tends to be abstract.

We performed a thorough test by comparing the SSIM values for each generated image when utilizing average pooling and max pooling strategies in order to assess the effect of pooling selection on the performance of NST. The purpose of this analysis was to determine how the pooling decision impacts the integrity and quality of the stylized outputs. Our selected batik pattern is megamendung for the style input and barong images for the content input in this experiment. The images below serves as a visual reference for the components involved in the transfer process and displays the content and style images utilized in this experiment:

Figure 5. Comparison of difference layer results.

The left image shows the content source featuring a detailed barong motif, chosen for its intricate textures to test neural style transfer capabilities. The right image displays the style source with a traditional megamendung batik pattern, known for its flowing cloud-like design. The results can be seen in the image below:

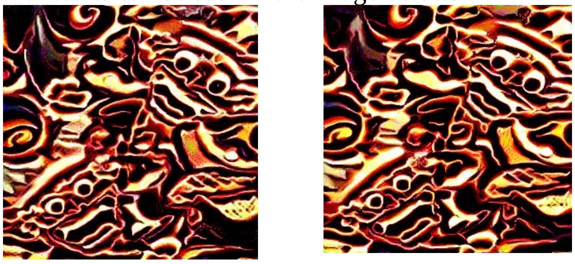

Figure 6. Comparison of difference layer results.

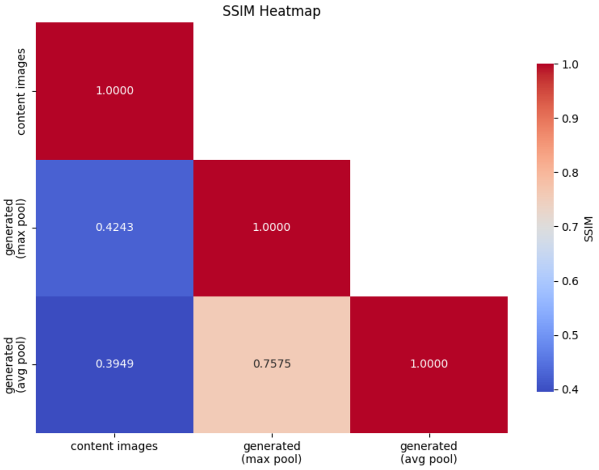

Figure 7. Comparison SSIM of layer difference results.

Fig. 6 and Fig. 7 (Sub 3 and 4) present the experimental results evaluating the impact of distinct pooling operations (max-pooling vs. average-pooling) on style transfer outcomes. The content image, depicting a barong batik motif, and the style image, featuring a megamendung pattern, were processed through the CNN architecture. Synthesized outputs using max-pooling and average-pooling exhibited measurable differences in visual characteristics, quantified via the Structural SSIM. Quantitative analysis revealed that the max-pooling-derived output achieved an SSIM score of 0.4243 relative to the original content, while the average-pooling output scored 0.3949. Notably, the SSIM between the two synthesized outputs was 0.7575, indicating substantial structural similarity despite differing pooling strategies. This suggests that both techniques preserve analogous visual traits in the final stylized image.

Next, we conducted testing using bird-patterned batik and blue megamendung batik. The content and style images can be seen in the image below:

Figure 8. Content (left) and style (right) images.

In this test, the layers used were shallow layers, namely layers 2 and 4. The results presented in the image below:

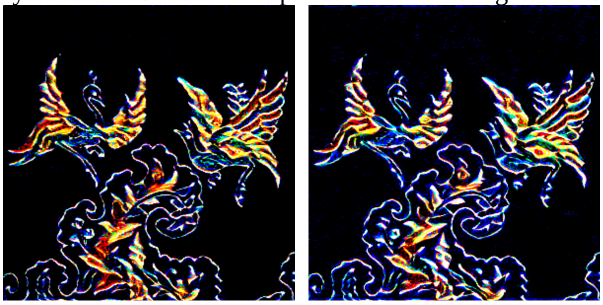

Figure 9. The results generated with max pooling (left) average pooling (right) with epochs 100.

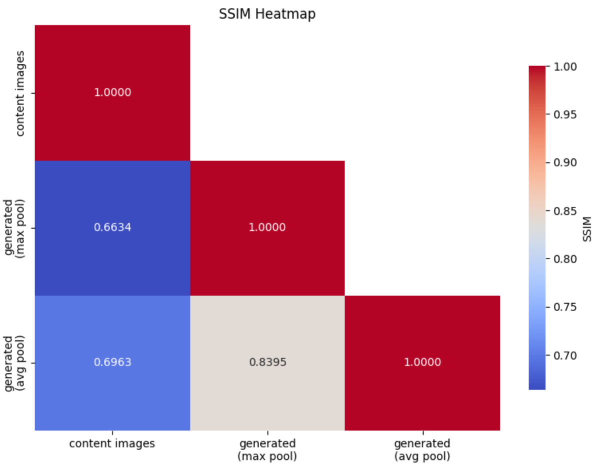

Figure 10. SSIM heatmap of test result image.

Based on the results of the second experiment shown in Fig. 9 and 10 (Sub 5 and 6), the effect of using different pooling layers in the CNN architecture for the style transfer process can be seen. The content image showing the peacock batik motif and the style image with the mega mendung motif are used as input. The results generated using max pooling and average pooling with epochs 100 show slightly different visual characteristics in terms of color. The results generated with max pooling have a structural similarity level of 0.6634 to the original image, while the results with average pooling show an SSIM value of 0.6963. Both generated results have a fairly high similarity to each other with an SSIM value of 0.8395, indicating good consistency of results between the two pooling techniques. Compared to the previous experiment, these results show an increase in the SSIM value indicating better style transfer in peacock batik.

It can be seen in Fig. 4, the results of combining batik motifs in the second experiment to determine the effect of different layers on the features extractor. The results show that selecting the initial layer up to the first 1/4 of the network provides more visible or more representative transfer results, conversely if the feature map taken from the layer is too deep, the features captured will also tend to be more abstract. The test results show that the selection of layers needs to be adjusted to the desired artistic goals. Shallow layers are more suitable for producing batik with explicit patterns and more representative style transfer results, while deep layers will create designs with a style that tends to be abstract.

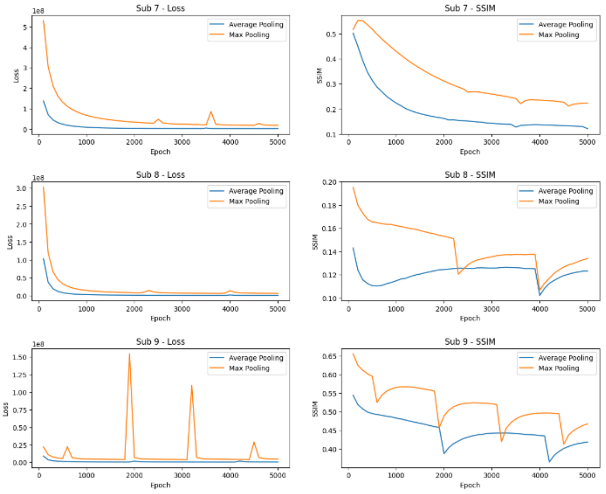

Figure 11. Comparison of loss and SSIM max pooling and average pooling.

The comparative evaluation of max-pooling and average-pooling in NST revealed significant differences in stability and performance (Sub 7, 8, and 9). Max-pooling exhibited exploding gradients, evidenced by erratic loss trajectories (e.g., initial loss of 3.0 dropping sharply to near-zero by epoch 5,000 in Image 6), attributed to its selective gradient amplification through dominant activations. In contrast, average-pooling demonstrated faster and smoother convergence (e.g., loss stabilizing at 0.5 for Image 6), owing to uniform gradient distribution across pooled regions. While max-pooling occasionally achieved marginally higher Structural Similarity Index (SSIM) scores (e.g., Image 22: 0.65 vs. 0.60), it introduced fragmented artifacts, compromising visual coherence. Average-pooling consistently prioritized holistic texture blending, achieving superior stability (e.g., Image 44 SSIM 0.16 vs. max-pooling's 0.12) and mitigating high-frequency distortions. These results underscore average-pooling's robustness for stable training and culturally nuanced batik synthesis, balancing edge preservation with reliable convergence.

4. Conclusion

The experimental investigation into NST for batik motif synthesis underscores the critical interplay between layer selection and pooling strategies in balancing artistic fidelity and computational stability. Shallow layers (e.g., layers 2, 4, 8) excel at preserving explicit patterns and structural details, producing literal style transfers aligned with traditional batik aesthetics. In contrast, deeper layers (e.g., layers 16, 24) capture abstract textures and complex geometries, enabling creative reinterpretations but risking loss of motif recognizability. Notably, style layer depth significantly influences output complexity, while content layer depth exhibits minimal impact, highlighting the importance of prioritizing style representation in parameter tuning. Pooling operations further delineate performance trade-offs: max-pooling, despite occasional edge sharpness (e.g., SSIM 0.65 in structured motifs), suffers from exploding gradients and fragmented artifacts due to selective gradient amplification. Average-pooling ensures stable convergence, smoother texture blending (e.g., SSIM 0.6963 in peacock batik), and superior preservation of holistic patterns, making it ideal for culturally nuanced applications. These findings advocate for average-pooling coupled with shallow style layers to harmonize stability, aesthetic coherence, and computational efficiency in batik-inspired NST algorithm.

Acknowledgements

This work was supported by Telkom University.

References

1. A. Amanzholova And I. Wirasari, “Batik Innovation: A Harmony Of Indonesian Batik Motifs And Kazakh Motifs,” *International Journal Of Multicultural And Multireligious Understanding* , Vol. 11, Pp. 356–362, 2024.
2. R. Febriani, L. Knippenberg, And N. Aarts, “The Making Of A National Icon: Narratives Of Batik In Indonesia,” *Cogent Arts Humanit*, Vol. 10, No. 1, 2023.
3. M. Xiao, “Innovative Applications And Market Impact Of Indonesian Batik In Modern Fashion,” Studies In Art And Architecture, Vol. 3, No. 2, Pp. 62–66, Jun. 2024.
4. N. M. Faizah, “Transformation Of Batik As A Symbol Of Nationalism And National Identity,” Kamara Journal, Vol. 1, Pp. 24–35, 2024.
5. V. Jenny Basiroen, M. Purbasari Wahidiyat, F. Marcelinna Suhendra, And D. Carolina, “Bridging Tradition And Innovation: Exploring Design Thinking For Lasem Batik Tulis Motif Creation,” In E3s Web Of Conferences, Edp Sciences, Sep. 2023.
6. K. W. Mas’udah And R. D. Issafira, “Improving Skills In Batik Design Through Digital Application For A Pesantren Community In Jombang,” Asean Journal Of Community Engagement, Vol. 6, No. 2, Dec. 2022.
7. L. A. Gatys, A. S. Ecker, And M. Bethge, “A Neural Algorithm Of Artistic Style,” Aug. 2015.
8. J. Zhang And J. Yan, “Style Transfer Technology Of Batik Pattern Based On Deep Learning,” Journal Of Fiber Bioengineering And Informatics, Vol. 16, No. 1, Pp. 57–67, 2023.
9. G. Guntur, P. Ponimin, And M. A. J. Purnomo, “Innovation And Creativity In Batik Motif Design: A Study Of Students’ Art Theses,” *Creativity Studies*, Vol. 16, No. 2, Pp. 668–681, Oct. 2023.
10. L. A. Gatys, A. S. Ecker, And M. Bethge, “Image Style Transfer Using Convolutional Neural Networks.”
11. S. Sima, T. Neda, And Akbar Siami, *The Performance Of Lstm And Bilstm In Forecasting Time Series*. 2019.

Authors Introduction

Happy Gery Pangestu

He is currently pursuing the Bachelor of Data Science in Telkom University, Indonesia. His research interests are machine learning, deep learning, Computer Vision, Natural Language Processing.

Andi Prademon Yunus, Ph.D

He is an Assistant Professor at Telkom University and he received his PhD in Engineering from Mie Unviersity, Japan. His research focuses on applied and fundamental machine learning for motion and behavior computing. He also collaborates with industry partners to develop AI-based tools for language modeling and image analytics.

Siti Khomsah, M.Cs

She is an Lecturer in Faculty of Informatics at Telkom University in Indonesia. She received her Master in Computer Science from Universitas Gajah Mada. Her research interests are Text Mining, Data Analytics, and Big Data Analytics.

Dr. Yit Hong Choo

He has completed his PhD and is now a Research Fellow in Operations Analytics at Deakin University's Institute for Intelligent Systems Research and Innovation (IISR)． His research focuses on advanced multi-objective optimisation algorithms for complex maintenance scheduling in rolling stock. He also collaborates with transportation industry partners to develop AI-based tools for video and image analytics.

Dr. Takao Ito

He is Professor of Management of Technology (MOT) in Graduate School of Advanced Science and Engineering at Hiroshima University. His current research interests include automata theory, artificial intelligence, systems control, quantitative analysis of interfirm relationships using graph theory, and engineering approach of organizational structures using complex systems theory.