Journal of Informatics and Vocational Education (JOIVE)
Vol.9, No.1, January 2026, pp. 38~48
ISSN: 2746-7813

Generation of Batik Cual Bangka Belitung Motif Variations Using Stable Diffusion Models

Rakha Piadika[1], Zikri Wahyuzi[2], Arvi Pramudyantoro[3]
[123] Department of Computer Science, Universitas Muhammadiyah Bangka Belitung, Indonesia

Article Info

Article history:

Received Jan 17, 2026
Revised Jan 25, 2026
Accepted Jan 26, 2026

Corresponding Author:

Rakha Piadika,
Department of Computer Science,
Universitas Muhammadiyah Bangka Belitung,
Jl KH Ahmad Dahlan KM 4,
Kelurahan Keramat, Rangkui,
Pangkalpinang, Kepulauan Bangka Belitung 33134, Indonesia
Email:
rpiadika@gmail.com

ABSTRACT

Batik Cual Bangka Belitung is a visual cultural heritage characterized by repetitive motifs, fine line details, and high ornament complexity, presenting significant challenges for digital exploration without compromising its visual identity. This study aims to generate variations of Batik Cual motifs using a prompt-guided image-to-image (img2img) approach based on the Stable Diffusion model with a single-reference image. Furthermore, the research analyzes the influence of the strength parameter on the delicate balance between structural similarity and generative visual variation. The dataset consists of 13 independently collected Batik Cual motifs, pre-processed through RGB conversion and standardized to a resolution of 512×512 pixels. Controlled experiments were conducted using varied strength values of 0.4, 0.6, and 0.8, while maintaining other parameters constant. Quantitative evaluation utilized the Structural Similarity Index Measure (SSIM) to assess structural integrity and CLIP similarity to measure semantic alignment between the prompt and output image. The results indicate that increasing strength consistently decreases SSIM values, signifying greater structural deviation from the reference image, whereas CLIP similarity remains relatively stable across configurations. Quantitatively, a strength of 0.4 offers the optimal combination of structural similarity and semantic suitability. However, qualitative assessments reveal that for certain motifs, a strength of 0.6 produces a more balanced variation between pattern innovation and motif character preservation. These findings confirm a measurable trade-off between identity preservation and generative exploration, demonstrating the potential of Stable Diffusion as a controlled method for developing Batik Cual digital assets.

Keywords: Stable Diffusion, Image-to-Image Generation, Batik Cual, Diffusion Models, Cultural Heritage

1. INTRODUCTION

The development *of Artificial Intelligence* (AI) in recent years shows a strong shift towards generative modeling for visual content production, characterized by the increasing capabilities of models in generating coherent, realistic, and diverse new images. Advances *in text-conditioned image generation* allow natural language-based descriptions to be translated into meaningful visual outputs, so generative models not only serve as a tool for image synthesis, but also as an important component in the creative production chain and data-driven visual asset management (Rombach et al., 2022; Yang et al., 2024).

In the context of image generation, *diffusion models* combine high yield quality with training stability and generation control flexibility. Conceptually, diffusion models work through a gradual process that models the transformation from random patterns to images, making them easier to control than some previous generative approaches. One widely used implementation is Stable Diffusion which moves the diffusion process to a latent space for computational efficiency relevant for image variation exploration scenarios (Ho et al., 2020; Rombach et al., 2022).

Nonetheless, the generation of high-textured image variations still faces technical challenges, especially when objects have *repeating patterns*, fine lines, and ornamental details that are sensitive to change. Based on the general principle of image-to-image diffusion, inappropriate control can lead to noise, pattern structure instability, and degradation of ornament consistency due to excessive deviation from the reference image. In addition, the *guidance mechanism* can indeed improve the suitability of the prompt, but it has the potential to increase the risk of force distortion if it is not balanced with the right parameter settings (Ho & Salimans, 2022; Meng et al., 2022).

On the other hand, the digitization of visual-based cultural heritage is increasingly seen as strategic to support the preservation, documentation, and sustainable development of creative assets (Kurniawan et al., 2023). Batik as a cultural artifact has a rich visual character and has been extensively researched through the computational approach of imagery, both for pattern recognition and mapping of motif characteristics. In particular, Batik Cual Bangka Belitung has a complexity that stands out through the repetition of motifs, ornamental details, and tight lines, making it challenging to be reproduced consistently by generative systems that do not have adequate control (Sentosa et al., 2022; Mustofa & Triyono, 2024).

In conventional practice, the creation of motif variations generally relies on manual processes that require skill, time, and design iterations, making them less efficient for the needs of exploring variation on a large scale. This condition encourages the need for an automatic generation system that maintains the main visual characteristics of the motif, while providing creative space to produce new variations without losing their visual identity. A number of early efforts in the context of local textiles show that *diffusion-based generation* has the potential to be used to produce textile patterns, but still requires process design and structured parameter control to maintain aesthetic quality and pattern consistency (Ginantra et al., 2024) .

However, research on local batik generations that are specific to Batik Cual is still limited, especially on the use of Stable Diffusion in the single-reference img2img scheme as a mechanism to produce motif variations from one reference citr (Chrystian & Wahyono, 2023; Octadion et al., 2023). Then research that tests the influence of control parameters (e.g. *strength* as a control deviation from reference) in controlled experiments is still rare in the context of local batik. In addition, quantitative evaluation that directly reads the *trade-off* between the similarity of the motif to the reference and the semantic suitability to the prompt is still rarely applied to the study of batik motif generation, which is useful to avoid judgments that are only subjective (Ginantra et al., 2024).

The novelty of this research lies in the application of the Stable Diffusion model for the generation of variations of Batik Cual motifs in Bangka Belitung, which until now has been studied more from a qualitative and philosophical perspective without a generative approach based on artificial intelligence. Previous studies on diffusion-based archipelago textiles have generally focused on Balinese Endek fabrics through fine-tuning and text-to-image generation schemes, not on Batik Cual or an image-to-image approach based on a single reference image (Ginantra et al., 2024). In addition to the domain aspect, this study also presents methodological novelty through systematic testing of parameter strength in a single-reference prompt-guided image-to-image scheme, as well as an integrated quantitative evaluation using SSIM and CLIP similarity to read the trade-off between the structural similarity of the motif and the semantic suitability to the prompt, which have not been explicitly reported in previous studies of local batik motif generation (Meng et al., 2022).

Based on this gap, this study aims to produce a variation of Batik Cual Bangka Belitung motifs using *prompt-guided image-to-image generation* (img2img) with one reference image (*single-reference*) per sample. The research focused on *strength* as the main variable to test its effect on the quality and visual consistency of output. Quantitative evaluation was performed using SSIM to measure the similarity of the input and output image structures and CLIP similarity to measure the semantic suitability between the prompt and the generation result, so that the generation control behavior could be read more measurably (Radford et al., 2021). In line with this goal, this study contributes by applying the Stable Diffusion model in an image-to-image single-reference scheme to produce a variety of Batik Cual motifs in Bangka Belitung. In addition, this study analyzes the influence of parameter strength as a deviation control on the visual quality and consistency of the generation results, and conducts a quantitative evaluation using SSIM and CLIP similarity to measure the similarity of image structure and semantic suitability to prompts so as to provide a more measurable evaluation framework in the study of diffusion based batik motif generation.

2. RESEARCH METHOD

This research method is designed to evaluate the ability of the Stable Diffusion model in producing a controlled variation of Batik Cual motifs using a single-reference approach based on a prompt-guided image-to-image generation. The main focus of this study is to analyze the influence of parameter *strength* as a control of the level of visual deviation on the quality and consistency of the resulting motifs. The entire process was designed as a controlled experiment without retraining or *fine-tuning* the model, so the evaluation was focused on the generative behavior of the model at the inference stage. This research method is carried out through several stages arranged in order as shown in Figure 1.

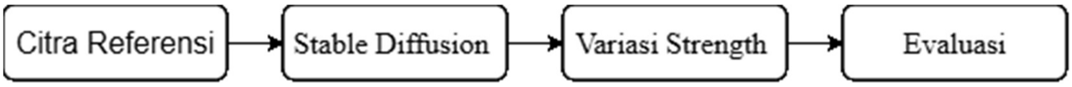

Figure 1. The flow of the research stages of the generation of variations of Batik Cual motifs in Bangka Belitung

2.1. DATASETS AND PRE-PROCESSING

The data used in this study is data on the image of Batik Cual motifs. Since Batik Cual's special dataset is not yet available in public repositories such as Kaggle, the entire data collection process was carried out independently by researchers to ensure adequate visual representation. Data collection of cual batik motif images was carried out through direct observation in three batik galleries in Bangka Belitung, namely Ishadi Cual Gallery (nine motifs), Santhi Cual Gallery (two motifs), and Destiani Gallery (two motifs), with a total of thirteen motifs. In the pre-processing stage, sampling is carried out by taking a maximum of 3 reference images per motif. Taking a maximum of three reference images per motif aims to capture intra-motif visual variation while reducing the bias that may arise when only one image is used. This number was chosen as a compromise between visual representativeness and experimental control, considering that this study did not involve a model training process. Then the image conversion to RGB format and the resolution uniformity to 512×512 pixels. This pre-processing aims to ensure compatibility with the Stable Diffusion model without changing the visual character of the Batik Cual motif. No additional visual augmentation or modification was made, as the image is only used as a reference in the image-to-image generation process, not for model training.

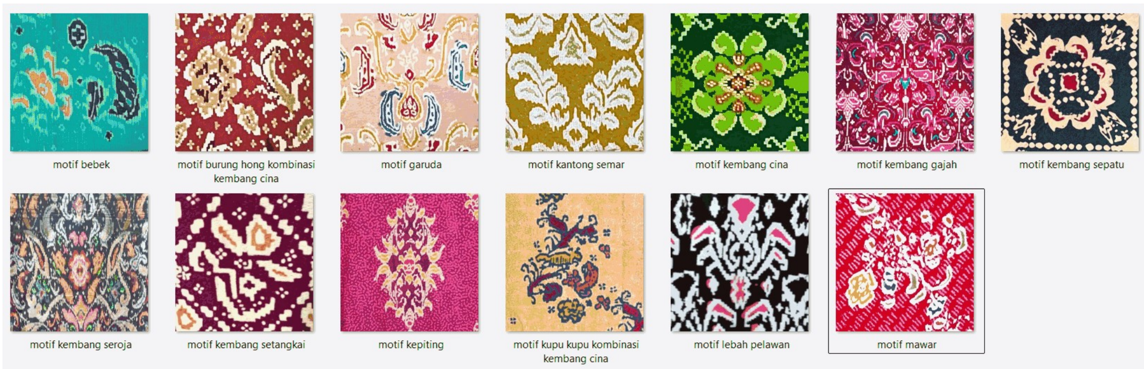

Figure 2. Image of reference image of Batik Cual motif Bangka Belitung

2.2. STABLE DIFFUSION AND IMG2IMG SCHEME

The generative model used in this study is a pretrained version of Stable Diffusion with the identity of the runwayml/stable-diffusion-v1-5 model, which is accessed through the Diffusers library. This study did not conduct retraining or fine-tuning models; The entire process is carried out at the inference stage by utilizing the ability of generative models that have been trained previously on a large scale. The generation scheme used is image-to-image (img2img), where a single reference image is used as the initial visual condition, while the text prompt serves to direct the visual characteristics of the output. Conceptually, img2img in the diffusion model works by adding controlled noise to the input image, then performing a directional denoising process based on the condition of the text (Yuan et al., 2024). The rate of change of the reference image can be controlled through the strength parameters and classifier-free guidance without the need for a change in the weight of the model (Meng et al., 2022; Rombach et al., 2022). This approach was chosen because it is in accordance with the purpose of the research, which is to produce a variation of Batik Cual motifs that still refer to the visual character of the original motif, instead of creating a completely new image without visual references.

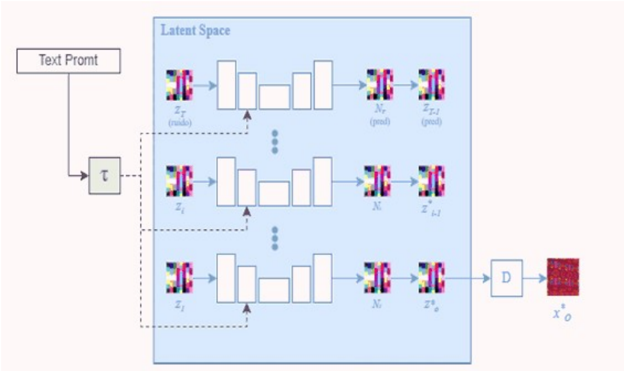

Figure 3. Image-to-image diffusion scheme in latent space

Figure 3 shows an image-to-image diffusion scheme in latent space, where the reference image is first mapped to a latent representation and degraded to a certain noise level before a prompt-based directional denoising process is performed. In this study, the strength parameter controls the initial noise level which determines how far the reference image is modified during the generation process.

2.3. EXPERIMENTAL DESIGN AND PARAMETER SETTINGS

This study was designed as a controlled study with *a prompt-guided image-to-image generation* experimental scheme with a single-reference scheme. Each reference image is processed using a combination of text prompts and *negative prompts* to generate a variation of the motif.

The main variable of the study is strength, which is a parameter that controls the level of output deviation from the reference image. In this study, strength was tested on three values *STRENGTH_SET* = [0.4, 0.6, 0.8]. Meanwhile, other parameters were maintained to maintain the validity of the comparison between treatments, namely the number of inference steps *STEPS* = 30, *guidance scale CFG_SET* = 5.5, image size 512×512, and random seed replication *SEEDS* = [111, 222] to ensure experimental repeatability. The use of *classifier-free guidance* as a generation control mechanism through *guidance scales* is a common practice in diffusion to improve the conformity of the output to the conditions of the text (Brooks et al., 2023).

The prompts used consist of two cultural prompts (prompts) designed to guide the characteristics of batik textiles, such as repetitive patterns, symmetrical compositions, fine line details, and *wax-resistant* textured impressions. To reduce unwanted visual artifacts, such as watermarks, logos, cartoon/anime styles, *oversaturation*, and *blur*, a single *negative prompt* (BASE_NEG) is used as an additional control mechanism. The use of *classifier-free guidance* and prompt-based control is a common practice in directional generation in modern diffusion models (Brooks et al., 2023; Dhariwal & Nichol, 2021). The number of outputs can be calculated to ensure the experiment is scalable and auditable. With a total of 13 motifs, a maximum of 3 references/motifs, 3 strength values, 2 prompts, 2 seeds, and **1** CFG value, the maximum total output is 468.

Table 1. Eksperimental Parameters

Parameter	Value
Model	runwayml/stable-diffusion-v1-5
Image size	512 × 512
Steps	30
Guidance scale (CFG)	5.5
Strength	0.4, 0.6, 0.8
Seed	111, 222
Prompt 1	Traditional Batik Cual Bangka Belitung textile motif, repeating ornamental pattern, symmetrical composition, fine canting linework, wax-resist batik texture, high detail, seamless
Prompt 2	Batik Cual Bangka Belitung, repeating floral vine ornaments, delicate wax lines, natural dye look, symmetrical textile pattern, high detail, seamless
Negative Prompt	text, watermark, logo, signature, border, frame, anime, cartoon, 3d render, neon, oversaturated, blurry, low detail, glossy, plastic, metallic

2.4. METADATA GENERATION AND LOGGING PROCEDURES

All of the experiment's output is stored in a structured results directory on Google Drive storage for easy tracking and management of experiment results. Each execution results in a single session folder that serves to separate experiments and make tracking easier. The generation procedure is carried out as follows. First, the system scans the entire folder of the source motif containing the reference image of each motif and then retrieves a maximum of three reference images on each motif. Each reference image is pre-processed (RGB and resize 512×512), and then executed through the StableDiffusionImg2ImgPipeline with prompt parameters, negative_prompt, strength, guidance_scale, num_inference_steps, and a seed-based generator. The img2img approach to the diffusion model allows for directional transformation of the input image into a new output with a controllable rate of change, making it suitable for single reference-based motif variation scenarios (Meng et al., 2022).

Each output image is stored with a file naming that encodes the experiment variables (reference index, prompt, strength value, guidance scale, and seed) for easy browsing and analysis. To keep *experiment tracking*, each output is recorded in a structured metadata file (JSON Lines format) that stores key experimental attributes such as motif, ref_path, output_path, image size, steps, strength, CFG, seed, prompt_id, prompt, negative prompt, and input–output SSIM. In addition, the run configuration is also stored as a run_config.json so that all parameters can be restored for replication. Parameter *strength* plays an important role because it controls the intensity of changes from the reference image. The increase in strength value provides greater room for change in the *denoising process*, so that conceptually the structural similarity to the input image has the potential to decrease (Meng et al., 2022).

2.5. EVALUATION

a. SSIM (Input-output similarity)

The structural similarity between the reference image and the outcome image was evaluated using *the Structural Similarity Index Measure* (SSIM). SSIM is used to measure the level of similarity of the structure and texture of the image with a value range of 0–1, where higher values indicate a greater degree of structural similarity (Yang et al., 2024). Mathematically, the SSIM between the xxx reference image and the yyy result image is defined as:

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

Where μ_x and μ_γ express the mean values of intensity, σ_x^2 and σ_γ^2 variance, and $\sigma_{x\gamma}$ covariance between the two images. In implementation, the image is converted to grayscale and standardized in size, with the parameter data_range = 255 so that the calculation corresponds to the pixel intensity range (Yang et al., 2024).

b. CLIP Similarity (Prompt–Output Alignment)

The semantic similarity between the text prompt and the output image was evaluated using *CLIP similarity*, which measures the proximity of the semantic representation between text and image in a shared embedding space. This study used the openai/clip-vit-base-patch32 model to extract text embedding and image embedding, then calculate *cosine similarity* as a measure of semantic alignment (Radford et al., 2021). CLIP similarity is formulated as:

$$CLIPSim(I,T) = \frac{f_I(I) \bullet f_T(T)}{\|f_I(I)\|\|f_T(T)\|}$$

Where fi(I) represents image embedding and ft(T) text embedding. A higher cosine similarity value indicates a stronger degree of semantic compatibility between the prompt and the result image. This approach is commonly used as a reference-free evaluation metric on text-image-based generative tasks (Hessel et al., 2022; Radford et al., 2021).

2.6. SOFTWARE AND EXECUTION ENVIRONMENT

The experiment was run on a Google Colab environment with storage via Google Drive mount. Implementation uses the Python language with the main libraries: diffusers, transformers, accelerate, safetensors, xformers, scikit-image, pandas, tqdm, and matplotlib. Execution utilizes GPU Colab (CUDA) when available to accelerate diffusion inference and extraction of CLIP embedding. Modern diffusion implementation practices are heavily supported by the PyTorch ecosystem for tensor computing and GPU acceleration (Paszke et al., 2019).

3. RESULT AND ANALYSIS

3.1. RESULT

3.1.1. GENERATION RESULTS

The experiment of generating variations of Batik Cual motifs was carried out using a single-reference prompt-guided image-to-image (img2img) scheme. All outputs are stored in a run folder structure for easy tracking and auditing, including experimental configuration (run_config.json) and metadata records per sample. This study produced 192 output images through a series of generative experiments. Theoretically, the maximum number of outputs in the experimental design is 468 images (13 motifs × max. 3 references/motifs × 3 strength × 2 prompts × 2 seeds × 1 CFG). The difference between the actual and maximum outputs indicates that in this run, the number of actual references used does not always reach 3 images per motif. This is due to the limited availability of reference images that meet the visual quality criteria for several motifs, so the number of references per motif is adjusted to maintain the consistency of the experiment.

3.1.2. THE EFFECT OF STRENGTH ON SSIM

The quality of visual proximity between the reference and output images was analyzed using SSIM (input–output). The summary of SSIM per strength and prompt_id shows a consistent trend: SSIM decreases as strength increases. This indicates that the greater the strength, the farther the variation in output from the structure/texture of the reference image.

Table 2. SSIM (mean ± std) per strength dan prompt

strength	prompt_id	SSIM mean	SSIM std
0.4	1	0.4903	0.1200
0.4	2	0.4892	0.1181
0.6	1	0.3133	0.1107
0.6	2	0.3195	0.1152
0.8	1	0.1795	0.0800
0.8	2	0.1981	0.0737

Specifically, at strength 0.4 the SSIM is in the range of about 0.489–0.490, then drops to about 0.313–0.320 at strength 0.6, and drops further to about 0.179–0.198 at strength 0.8. This pattern is in line with the role of strength in img2img: the higher the strength, the greater the allowed deviation of the input image.

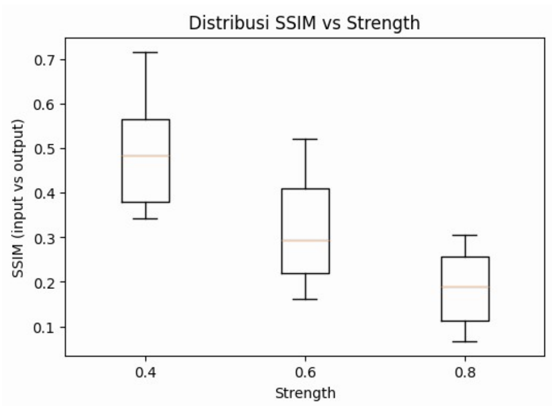

Figure 4. Boxplot SSIM vs strength.

3.1.3. THE EFFECT OF STRENGTH ON CLIP SIMILARITY

The semantic compatibility of the output to the prompt is analyzed using CLIP similarity (prompt–output). The summary per strength and prompt_id shows that the similarity CLIP is relatively stable and does not experience major changes as strength increases.

Table 3. CLIP similarity (mean ± std) per strength dan prompt

strength	prompt_id	CLIP mean	CLIP std
0.4	1	0.2905	0.0107
0.4	2	0.2949	0.0109
0.6	1	0.2866	0.0133
0.6	2	0.2926	0.0114
0.8	1	0.2855	0.0125
0.8	2	0.2861	0.0113

The mean CLIP value is in the range of about 0.285–0.295 on all configurations. Thus, in this study configuration, the increase in strength strongly affects structural similarity (SSIM) compared to the increase in prompt conformity (CLIP).

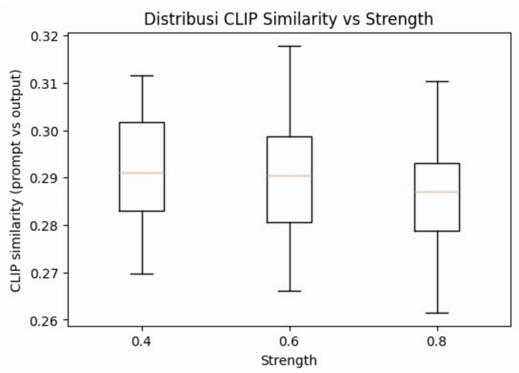

Figure 5. Boxplot CLIP similarity vs strength

3.1.4. TRADE-OFF SSIM VS CLIP SIMILARITY

The quantitative results show a clear trade-off: as strength increases, SSIM decreases, while CLIP tends to stabilize. The implication of this pattern is that increased variation (through strength) is primarily compensated for by a decrease in similarity to references, rather than by a significant increase in prompt alignment. In this trade-off space, a lower strength configuration is potentially more appropriate when the primary target is to maintain the visual character of the motif (higher SSIM), while higher strength can be considered if a further variation of the reference is desired, with lower SSIM consequence.

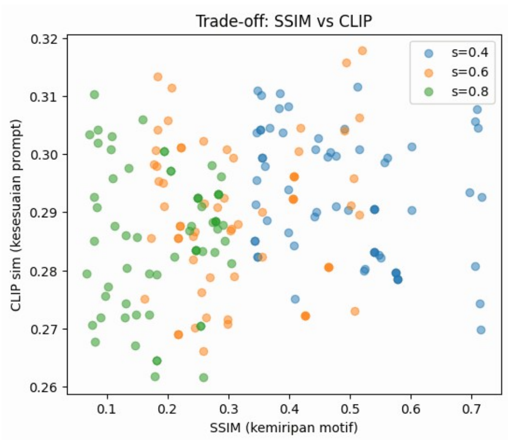

Figure 6. Scatter SSIM vs CLIP similarity per strength

3.1.5. SELECTION OF THE BEST SETTINGS

The selection of the best configuration was made using a mix_mean (combined score) calculated as a standardized average of SSIM and CLIP similarity to represent the trade-off between motif similarity and prompt suitability. In a global summary, configurations with a strength of 0.4 occupy the top position, especially at prompt_id = 2.

Table 4. Top global settings by mix_mean

peringkat	strength	prompt_id	SSIM mean	CLIP mean	mix_mean
1	0.4	2	0.4892	0.2949	0.6210
2	0.4	1	0.4903	0.2905	0.5829
3	0.6	2	0.3195	0.2926	0.4696
4	0.6	1	0.3133	0.2866	0.4121
5	0.8	2	0.1981	0.2861	0.3184

Furthermore, a summary of the best setting per motif shows that the majority of motifs have the best configuration at strength 0.4, prompt_id 2, with a few exceptions.

Table 5. Best setting per motif (13 motif)

Motif	strength	prompt_id	SSIM mean	CLIP mean	mix_mean
motif bebek	0.4	2	0.3484	0.2947	0.5106
motif burung hong	0.4	2	0.5017	0.3080	0.7468
motif garuda	0.4	2	0.5381	0.2812	0.5368
motif kantong semar	0.4	2	0.4033	0.2891	0.5032
motif kembang cina	0.4	2	0.5797	0.3000	0.7358
motif kembang gajah	0.4	2	0.3919	0.3094	0.6747
motif kembang sepatu	0.4	2	0.5581	0.2851	0.5871
motif kembang seroja	0.4	2	0.3539	0.2967	0.5329
motif kembang setangkai	0.6	2	0.5114	0.3011	0.6927
motif kepiting	0.4	2	0.3579	0.3077	0.6338
motif kupu kupu	0.4	2	0.4453	0.2960	0.5969
motif lebah pelawan	0.4	2	0.7079	0.3067	0.8939
motif mawar	0.4	1	0.4733	0.3000	0.6541

Overall, these results indicate that in this experimental configuration, low strength (0.4) tends to provide the best combination of maintaining the character of the motif and maintaining alignment with the prompt. Meanwhile, there are certain motifs (e.g. the stalk flower motif) that show the best indication at strength 0.6, so the configuration selection can be motif-dependent.

3.1.6. QUALITATIVE VISUALIZATION (GRID BY MOTIF)

Quantitative evaluation using SSIM and CLIP similarity has not fully captured the visual aesthetic aspects and legibility of batik ornaments, so qualitative evaluation is needed as a complement. Therefore, the generation results are visualized in the form of a per-motif table that displays a comparison between the reference image and the result image at various strength values (0.4, 0.6, and 0.8) for each prompt. This presentation allows for a more focused visual inspection of changes in pattern structure, consistency of ornament details, and the stability of the motif composition.

Table 6. Qualitative grid of motifs

Motif	Prompt	Reference image	Strength 0.4	Strength 0.6	Strength 0.8
Burung Hong Kembang Cina	1				

	2				
Kembang Sepatu	1				
	2				
Kembang Setangkai	1				
	2				

Although configurations with low strength (0.4) showed the highest SSIM values, qualitative inspection showed that in some motifs, intermediate strength values (0.6) resulted in more balanced visual variation between pattern change and reference structure maintenance. This shows that the optimal configuration is determined not only by quantitative metrics alone, but also by the desired visual design objectives.

3.2. ANALYSIS

Based on the results of the quantitative and qualitative evaluation obtained, it can be concluded that the strength parameter plays a central role in controlling the balance between the preservation of the visual character of the motif and the exploration of generative variation. The consistent decrease in SSIM values with increased strength suggests that the img2img mechanism in the diffusion model progressively enlarges the structural deviation from the reference image. This phenomenon is in harmony with the basic principle of diffusion, where an increase in noise intensity in the early stages of inference provides a wider space for transformation, but with the consequent reduction of attachment to the original structure of the input image.

Interestingly, the change in strength was not followed by a significant increase in the CLIP similarity value. The stability of the CLIP score on various configurations indicates that the semantic fit between the prompt and the output image is relatively maintained, although the visual structure undergoes considerable changes. This shows that in this experimental configuration, the influence of strength is more dominant on the spatial and textural aspects of the image than on text-based semantic alignment. In other words, the visual variation produced through increased strength is more manifested in the form of changes in patterns and ornaments, rather than in shifts in meaning or visual themes directed by prompts.

The observed trade-off pattern between SSIM and CLIP similarity emphasizes that the selection of strength values is a contextual design decision. Low-strength configurations tend to be more suitable for scenarios that emphasize the preservation of the visual identity of the motif, such as digital documentation or the exploration of limited variations that remain faithful to the original character of Batik Cual. In contrast, higher strength values open up opportunities for more aggressive visual exploration, albeit with the consequence of reduced structural similarity to references. These findings show that the variations produced by the diffusion model are not uncontrollably free, but can be directed measurably through parameter settings.

The results of selecting the best configuration using a combined score (mix_mean) reinforce the findings. The dominance of the configuration with strength = 0.4 on most motifs suggests that, in the context of the dataset and the prompts used, the best balance between structural stability and semantic fit is quantitatively achieved at a relatively low level of deviation. However, the presence of several motifs that exhibit optimal configuration at medium strength indicates that the visual characteristics of the motif also influence the model's generative response, so that the motif variation generation approach is not entirely universal, but can be motif-dependents.

Furthermore, qualitative inspection of the visual grid shows that configurations with strength = 0.4 often produce outputs very close to the reference image, with a relatively limited degree of visual variation. On the other hand, at strength = 0.6, the output shows a clearer balance between the preservation of the main structure of the motif and the emergence of new variations of ornaments and compositions that can still be recognized as derivatives of the original motif. These findings indicate that, although quantitatively low strength obtained the highest score, medium strength values are potentially more relevant for motif design exploration scenarios, where visual innovation is desired without eliminating the association with the identity of the Batik Cual motif. This emphasizes the importance of combining quantitative and qualitative evaluations in assessing the output of diffusion-based generative systems, particularly in the visual domain of cultures rich in detail and repetitive structure.

Overall, this study places the strength parameter as the main lever in regulating the behavior of the generation of variations of Batik Cual motifs using Stable Diffusion. The single-reference img2img approach used has been proven to provide a controlled space for visual exploration, where parameter adjustments allow for a measurable compromise between the preservation of motif character and design innovation. Thus, the results of this study not only provide an empirical understanding of the behavior of diffusion models in the local batik domain, but also offer a practical framework for the use of generative models in the context of the preservation and development of cultural visual assets.

4. CONCLUSION

In the configuration of this study, the Stable Diffusion img2img single-reference approach guided by prompt showed the ability to produce variations of Batik Cual Bangka Belitung motifs from the dataset used. The results of the evaluation showed that the SSIM value tended to decrease with the increase in strength, which indicated an increase in structural and textural deviations from the reference image. In contrast, the CLIP similarity value is relatively stable between strength configurations, so the increase in strength does not directly increase the semantic fit of the output to the prompt. Based on a trade-off analysis of motif similarity and prompt suitability, low-strength configurations tend to provide a better balance in maintaining the visual character of the motif while maintaining semantic alignment. These findings suggest that conservative strength settings are more relevant for the exploration of controlled variations of batik motifs, especially in the context of the development of cultural heritage-based digital visual assets. This study has several limitations, including the limited scope of datasets and prompt variations, as well as the use of automatic evaluation metrics that do not fully represent the aesthetic aspects and cultural significance of batik. Follow-up research is suggested to expand the data and prompt strategies, explore other model alignment approaches, and involve validation based on the assessment of batik experts or artisans through a more systematic evaluation procedure.

ACKNOWLEDGEMENTS

This research was funded by the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia through the Belmawa Student Creativity Program (PKM), which supported the research activities until the National Student Scientific Week (PIMNAS). The authors would also like to thank Universitas Muhammadiyah Bangka Belitung for providing institutional support and research facilities, particularly the Computer Laboratory of the Computer Science Program. In addition, the authors express their sincere gratitude to Zikri Wahyuzi, S.T., M.Kom and Yudistira Bagus Pratama, M.Kom for their valuable guidance and constructive feedback throughout the research process.

REFERENCES

Brooks, T., Holynski, A., & Efros, A. A. (2023). *InstructPix2Pix: Learning to follow image editing instructions* [Preprint]. arXiv. https://arxiv.org/abs/2211.09800

Chrystian, & Wahyono. (2023). *SP-BatikGAN: An efficient generative adversarial network for symmetric pattern generation* [Preprint]. arXiv. https://arxiv.org/abs/2304.09384

Dhariwal, P., & Nichol, A. (2021). *Diffusion models beat GANs on image synthesis* [Preprint]. arXiv. https://arxiv.org/abs/2105.05233

Ginantra, N. L. W. S. R., Hendrawati, T., & Wulandari, D. A. P. (2024). Penerapan metode stable diffusion dengan fine tuning untuk pola endek Bali. *TEMATIK*, *11*(2), 141–147. https://doi.org/10.38204/tematik.v11i2.2069

Hessel, J., Holtzman, A., Forbes, M., Le Bras, R., & Choi, Y. (2022). *CLIPScore: A reference-free evaluation metric for image captioning* [Preprint]. arXiv. https://arxiv.org/abs/2104.08718

Ho, J., Jain, A., & Abbeel, P. (2020). *Denoising diffusion probabilistic models* [Preprint]. arXiv. https://arxiv.org/abs/2006.11239

Ho, J., & Salimans, T. (2022). *Classifier-free diffusion guidance* [Preprint]. arXiv. https://arxiv.org/abs/2207.12598

Kurniawan, T. S., Alwi, M. N. H., & Purnawirawan, O. (2023). Systematic literature review: Strategi pelestarian budaya batik di era Revolusi Industri 4.0 dengan memanfaatkan perkembangan perangkat teknologi informasi. *Prosiding Seminar Nasional Batik (SNBK)*. https://proceeding.batik.go.id/index.php/SNBK/article/view/240

Meng, C., He, Y., Song, Y., Song, J., Wu, J., Zhu, J.-Y., & Ermon, S. (2022). *SDEdit: Guided image synthesis and editing with stochastic differential equations* [Preprint]. arXiv. https://arxiv.org/abs/2108.01073

Mustofa, S. S., & Triyono. (2024). Analisis filosofi motif kain tenun cual dalam upacara adat Ngarak Telok Seroja Bangka Belitung. *Cilpa: Jurnal Ilmiah Pendidikan Seni Rupa*, *9*(2). https://doi.org/10.30738/cilpa.v9i2.17885

Octadion, O., Yudistira, N., & Kurnianingtyas, D. (2023). *Synthesis of batik motifs using a diffusion–generative adversarial network* [Preprint]. arXiv. https://arxiv.org/abs/2307.12122

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., … Chintala, S. (2019). *PyTorch: An imperative style, high-performance deep learning library* [Preprint]. arXiv. https://arxiv.org/abs/1912.01703

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). *Learning transferable visual models from natural language supervision* [Preprint]. arXiv. https://arxiv.org/abs/2103.00020

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). *High-resolution image synthesis with latent diffusion models* [Preprint]. arXiv. https://arxiv.org/abs/2112.10752

Sentosa, E., Mulyana, D. I., Cahyana, A. F., & Pramuditasari, N. G. (2022). Implementasi image classification pada batik motif Bali dengan data augmentation dan convolutional neural network. *Jurnal Pendidikan Tambusai*, *6*(1), 1451–1463. https://doi.org/10.31004/jptam.v6i1.3137

Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., Zhang, W., Cui, B., & Yang, M. H. (2024). Diffusion models: A comprehensive survey of methods and applications. *ACM Computing Surveys*, *56*(4). https://doi.org/10.1145/3626235

Yuan, L., Yan, D., Saito, S., & Fujishiro, I. (2024). DiffMat: Latent diffusion models for image-guided material generation. *Visual Informatics*, *8*(1), 6–14. https://doi.org/10.1016/j.visinf.2023.12.001