

CESS

(Journal of Computer Engineering, System and Science)

Available online: <https://jurnal.unimed.ac.id/2012/index.php/cess>

ISSN: 2502-714x (Print) | ISSN: 2502-7131 (Online)



Penerapan *Low-Rank Adaptation (LoRA)* pada *Stable Diffusion* untuk Generasi Citra Motif Batik Indonesia

Application of Low-Rank Adaptation (LoRA) in Stable Diffusion for Indonesian Batik Motif Image Generation

Agung Riyadi^{1*}, M. Falih Hilmy²

^{1,2}Jurusan Teknik Informatika, Politeknik Negeri Batam
Jalan Ahmad Yani, Batam Kota, Kota Batam, Kepulauan Riau
Email: ¹agung@polibatam.ac.id, ²m.falih@gmail.com

ABSTRAK

Perkembangan model text-to-image berbasis difusi memberikan peluang dalam pengembangan motif batik Indonesia menggunakan kecerdasan buatan. Penelitian ini bertujuan menerapkan *Low-Rank Adaptation (LoRA)* pada *Stable Diffusion 1.5* untuk menghasilkan citra enam motif batik Indonesia, yaitu Parang, Ceplok, Kawung, Buketan, Lasem, dan Mega Mendung. Dataset yang digunakan terdiri atas 300 citra yang telah melalui tahap pra-pemrosesan sebelum dilatih menggunakan KohyaSS. Evaluasi dilakukan menggunakan *Frechet Inception Distance (FID)* dan *Contrastive Language-Image Pre-training (CLIP) Score*. Hasil penelitian menunjukkan bahwa motif Buketan memperoleh performa terbaik dengan nilai FID sebesar 24,81 dan CLIP Score sebesar 0,351, sedangkan motif Parang memperoleh nilai FID sebesar 41,52 dan CLIP Score sebesar 0,286. Hasil tersebut menunjukkan bahwa metode LoRA mampu menghasilkan citra motif batik dengan kualitas visual dan kesesuaian semantik yang baik pada beberapa kategori motif, meskipun performanya masih bervariasi pada setiap motif.

Kata Kunci: *Low-Rank Adaptation (LoRA)*; *Stable Diffusion*; Generasi Citra; Batik Indonesia; Kecerdasan Buatan.

ABSTRACT

The development of diffusion-based text-to-image models has created new opportunities for generating Indonesian batik motifs using artificial intelligence. This study aims to apply *Low-Rank Adaptation (LoRA)* to *Stable Diffusion 1.5* to generate images of six Indonesian batik motifs: Parang, Ceplok, Kawung, Buketan, Lasem, and Mega Mendung. The dataset consists of 300 images that underwent a preprocessing stage before being trained using Kohya SS. The model was evaluated using the *Frechet Inception Distance (FID)* and the *Contrastive*



Language–Image Pre-training (CLIP) Score. The results show that the Buketan motif achieved the best performance, with the lowest FID value of 24.81 and the highest CLIP Score of 0.351, while the Parang motif obtained the highest FID value of 41.52 and the lowest CLIP Score of 0.286. These findings indicate that the LoRA method is capable of generating Indonesian batik motif images with good visual quality and semantic alignment across several motif categories, although its performance still varies depending on the characteristics of each motif.

Keywords: Low-Rank Adaptation (LoRA); Stable Diffusion 1.5; Image Generation; Indonesian Batik; Artificial Intelligence.

1. PENDAHULUAN

Batik merupakan salah satu warisan budaya tak benda Indonesia yang telah diakui oleh UNESCO karena memiliki nilai sejarah, filosofi, dan identitas budaya yang tinggi. Setiap motif batik merepresentasikan makna tertentu yang dipengaruhi oleh karakteristik daerah asalnya, sehingga menghasilkan keragaman motif seperti Parang, Ceplok, Kawung, Buketan, Lasem, dan Mega Mendung. Namun, pelestarian motif batik menghadapi berbagai tantangan, antara lain berkurangnya jumlah pengrajin, terbatasnya regenerasi, serta belum optimalnya pemanfaatan teknologi digital dalam mendukung proses pengembangan motif batik [1][2].

Perkembangan kecerdasan buatan, khususnya model generatif berbasis diffusion seperti *Stable Diffusion*, telah menunjukkan kemampuan yang sangat baik dalam menghasilkan citra berdasarkan deskripsi tekstual (*text-to-image*)[3]. Meskipun demikian, model umum tersebut belum mampu menghasilkan motif batik yang sesuai dengan karakteristik budaya Indonesia tanpa proses adaptasi terhadap data yang lebih spesifik. Selain itu, proses *fine-tuning* secara penuh memerlukan sumber daya komputasi yang besar serta dataset berukuran besar sehingga kurang efisien untuk penelitian dengan data yang terbatas.

Low-Rank Adaptation (LoRA) merupakan salah satu metode parameter-efficient *fine-tuning* yang memungkinkan proses adaptasi model dilakukan dengan jumlah parameter yang jauh lebih sedikit dibandingkan *fine-tuning* penuh [4][5]. Melalui pendekatan ini, model dasar dapat disesuaikan dengan domain tertentu tanpa mengubah seluruh parameter model, sehingga lebih efisien dari sisi waktu pelatihan, kebutuhan memori, dan sumber daya komputasi.

Beberapa penelitian terdahulu telah menerapkan LoRA pada model difusi untuk berbagai domain citra khusus, namun dengan cakupan yang berbeda dari motif batik Indonesia. Yilmaz dkk. menerapkan *Stable Diffusion* dengan LoRA untuk menghasilkan citra sintetis kawasan industri guna keperluan augmentasi data [6], sedangkan Sutedy dan Qomariyah memanfaatkan pendekatan DreamBooth untuk generasi citra otomotif [7]. Pada domain yang relatif lebih dekat dengan motif tradisional, Liu dan Huang menerapkan LoRA untuk mendigitalisasi pola klasik Mongolia dan memperoleh hasil dengan fidelitas visual yang tinggi [8]. Ketiga penelitian tersebut menunjukkan bahwa LoRA efektif diterapkan pada objek dengan struktur visual yang relatif seragam dan berulang, tetapi belum menguji kemampuannya pada motif dengan variasi geometris, filosofi budaya, serta karakteristik regional yang jauh lebih beragam seperti pada batik Indonesia. Di sisi lain, penelitian mengenai citra batik yang telah ada masih berfokus pada pembangunan dan dokumentasi dataset motif,

serta belum mengarah pada upaya generasi citra baru menggunakan model difusi maupun evaluasi kuantitatif terhadap hasilnya. Kesenjangan ini menunjukkan bahwa efektivitas LoRA dalam mempelajari dan mereproduksi karakteristik visual sekaligus filosofis dari berbagai motif batik Indonesia belum banyak dikaji secara empiris, sehingga menjadi dasar pemilihan LoRA sebagai metode adaptasi yang diuji secara khusus pada enam motif batik dalam penelitian ini.

Berdasarkan kondisi tersebut, penelitian ini menerapkan metode LoRA pada *Stable Diffusion* 1.5 untuk menghasilkan citra enam motif batik Indonesia menggunakan dataset yang terdiri atas 300 citra. Kinerja model dievaluasi menggunakan metrik *Frechet Inception Distance (FID)* dan *CLIP Score*, serta penilaian oleh lima pakar batik. Hasil penelitian diharapkan dapat memberikan gambaran mengenai kemampuan LoRA dalam menghasilkan citra motif batik yang memiliki kualitas visual dan kesesuaian terhadap karakteristik motif batik Indonesia, sekaligus menjadi salah satu alternatif pemanfaatan kecerdasan buatan dalam mendukung pelestarian budaya.

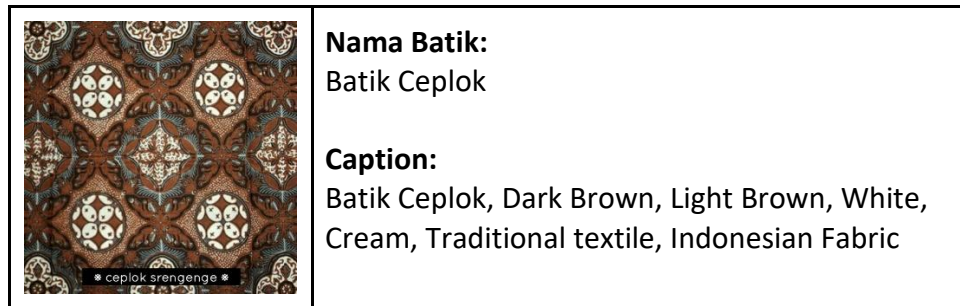
2. METODE PENELITIAN

Dataset sebanyak 300 citra motif batik dikompilasi melalui pemotretan langsung dan pemindaian flatbed kain batik autentik. Citra dikumpulkan dari koleksi dari berbagai sumber kain batik. Koleksi ini mencakup 6 gaya regional utama termasuk diantaranya Batik Parang, Ceplok, Kawung, Buketan, Lasem, Mega Mendung sehingga memastikan keragaman stilistik dalam dataset. Citra diambil menggunakan model kamera dengan resolusi minimum 2048 x 2048 piksel dalam format JPEG. Setiap sampel ditinjau oleh peneliti untuk memastikan pola motif terlihat jelas, bebas dari kerutan atau distorsi yang signifikan, dan merepresentasikan kualitas kerajinan batik yang autentik.

2.1 Pra-Pemrosesan Citra

Seluruh citra pada dataset terlebih dahulu melalui tahap pra-pemrosesan sebelum digunakan dalam proses pelatihan model LoRA. Tahap pertama adalah pengubahan ukuran (*resizing*), di mana seluruh citra disesuaikan menjadi resolusi 512 × 512 piksel agar sesuai dengan resolusi pelatihan model *Stable Diffusion* 1.5. Untuk meningkatkan keragaman data dan mengurangi risiko *overfitting*, dilakukan augmentasi data berupa pembalikan citra secara horizontal (*horizontal flip*) serta penyesuaian ringan pada hue dan saturasi.

Selanjutnya, setiap citra dilengkapi dengan caption yang mendeskripsikan karakteristik motif batik menggunakan format "*batik [nama motif] [warna primer], [warna sekunder], traditional textile, Indonesian fabric*". Caption tersebut berfungsi sebagai *conditioning prompt* selama proses pelatihan sehingga model dapat mempelajari hubungan antara representasi visual motif batik dan deskripsi tekstual yang sesuai.



Gambar 1. Sample dataset untuk batik ceplok dan caption

2.2 Konfigurasi Pelatihan

Proses pelatihan model dilakukan menggunakan Kohya SS dengan metode *DreamBooth* pada model dasar (base model) *Stable Diffusion* 1.5. Setiap model dilatih menggunakan kumpulan citra yang telah melalui tahap pra-pemrosesan dan dilengkapi dengan *caption* sebagai *conditioning text*. Pelatihan dilakukan secara terpisah untuk setiap kategori motif batik sehingga model dapat mempelajari dan merepresentasikan karakteristik visual yang khas dari masing-masing motif.

Table 1. Konfigurasi Parameter Pelatihan Model

	Value		Value
Base model	Stable Diffusion 1.5	LR Scheduler	Cosine
Metode fine-tuning	LoRA	Network rank	32
Framework	Kohya_ss	Precision	fp16
Image resolution	512 × 512	Seed	36
Batch size	3	Caption extension	.txt
Epoch	10	Shuffle caption	Ya
Learning rate	1.00E-04	Cache latents	Ya
Optimizer	AdamW8bit		

2.3 Pelatihan Model

Proses pelatihan model dilakukan menggunakan Kohya SS dengan metode *Low-Rank Adaptation (LoRA)* pada model dasar (base model) *Stable Diffusion* 1.5 [9]. Seluruh citra yang telah melalui tahap pra-pemrosesan dan dilengkapi dengan *training caption* digunakan sebagai data masukan selama proses pelatihan. *Caption* berfungsi sebagai *conditioning text* yang memungkinkan model mempelajari hubungan antara representasi visual motif batik dan deskripsi tekstual yang menyertainya.

Stable Diffusion 1.5 (SD 1.5) dipilih sebagai model dasar karena kinerjanya yang telah tervalidasi luas dalam sintesis gambar berbasis teks dan kesesuaiannya untuk *fine-tuning* LoRA [10] pada dataset domain-spesifik. SD1.5 beroperasi dalam ruang laten terkompresi yang dikodekan oleh *Variational Autoencoder (VAE)*, dengan *backbone denoising* berbasis U-Net yang dikondisikan pada embedding teks CLIP [11]. Model ini memiliki sekitar 860 juta parameter secara keseluruhan.

Pelatihan dilakukan secara terpisah untuk setiap kategori motif batik sehingga dihasilkan model LoRA yang mampu merepresentasikan karakteristik visual masing-masing motif. Selama proses pelatihan, hanya parameter tambahan pada lapisan LoRA yang diperbarui, sedangkan parameter pada model dasar tetap dipertahankan. Pendekatan ini memungkinkan proses *fine-tuning* berlangsung lebih efisien dari sisi komputasi dan penggunaan memori dibandingkan dengan melatih seluruh parameter model.

Parameter pelatihan, seperti learning rate [12], batch size, jumlah epoch, network rank (LoRA dimension) [13], network alpha, optimizer, dan learning rate scheduler, ditentukan menggunakan konfigurasi pada Kohya SS. Setelah proses pelatihan selesai, bobot LoRA disimpan dalam bentuk berkas model untuk digunakan pada tahap inferensi dan evaluasi.

2.4 Proses Inferensi

Setelah proses pelatihan selesai, model LoRA yang telah dihasilkan digunakan pada tahap inferensi untuk menghasilkan citra batik berdasarkan *prompt* yang sesuai dengan masing-masing kategori motif. Setiap model menerima *prompt* sebagai masukan untuk menghasilkan citra baru yang merepresentasikan karakteristik visual motif batik yang telah dipelajari selama proses pelatihan.

Pada penelitian ini, proses inferensi dilakukan sebanyak 100 kali untuk setiap kategori motif batik. Dengan enam kategori motif yang diuji, yaitu Parang, Ceplok, Kawung, Buketan, Lasem, dan Mega Mendung, total citra yang dihasilkan pada tahap inferensi adalah 600 citra. Hasil citra tersebut selanjutnya digunakan sebagai bahan evaluasi untuk menilai kemampuan model dalam menghasilkan motif batik yang sesuai dengan karakteristik masing-masing kategori.

2.5 Evaluasi

Evaluasi dilakukan untuk menilai kualitas citra batik yang dihasilkan oleh model LoRA berdasarkan aspek kualitas visual dan kesesuaian terhadap deskripsi tekstual (*prompt*). Pada penelitian ini digunakan dua metrik evaluasi kuantitatif, yaitu *Frechet Inception Distance (FID)* dan *CLIP Score*.

2.5.1 Frechet Inception Distance (FID)

Frechet Inception Distance (FID) digunakan untuk mengukur tingkat kemiripan distribusi antara citra hasil inferensi dan citra asli pada dataset[14]. Perhitungan FID dilakukan dengan membandingkan fitur yang diekstraksi dari kedua kelompok citra menggunakan jaringan Inception-v3. Pada penelitian ini, citra hasil inferensi untuk setiap kategori motif dibandingkan dengan citra asli dari kategori motif yang sama. Semakin rendah nilai FID yang diperoleh, semakin dekat distribusi citra hasil generasi dengan distribusi citra asli, sehingga menunjukkan bahwa model mampu menghasilkan citra dengan kualitas visual dan karakteristik yang menyerupai data referensi.

2.5.2 CLIP Score

CLIP Score digunakan untuk mengukur kesesuaian semantik antara citra hasil inferensi dan *prompt* yang digunakan selama proses generasi [15]. Perhitungan dilakukan dengan menghitung nilai *cosine similarity* antara representasi (*embedding*) citra dan representasi teks menggunakan model CLIP (*Contrastive Language–Image Pre-training*). Semakin tinggi nilai *CLIP Score*, semakin baik tingkat kesesuaian antara deskripsi tekstual dan citra yang

dihasilkan, yang menunjukkan bahwa model mampu menghasilkan motif batik sesuai dengan *prompt* yang diberikan.

Tabel 2. Skenario Evaluasi Model

Metrik	Data Referensi	Data Uji	Tujuan Evaluasi	Interpretasi
Frechet Inception Distance (FID)	Citra asli pada dataset	Citra hasil inferensi	Mengukur kemiripan distribusi antara citra hasil generasi dan citra asli	Semakin rendah nilai FID, semakin baik kualitas dan kemiripan citra terhadap dataset
CLIP Score	Prompt teks	Citra hasil inferensi	Mengukur kesesuaian semantik antara prompt dan citra yang dihasilkan	Semakin tinggi nilai CLIP Score, semakin baik kesesuaian antara prompt dan citra

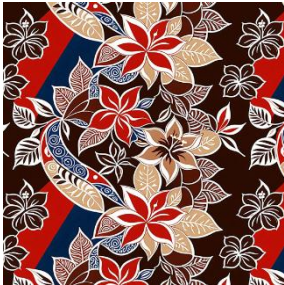
3. HASIL DAN PEMBAHASAN

Setelah proses pelatihan selesai, model *Stable Diffusion* 1.5 yang telah di-*fine-tuning* menggunakan *Low-Rank Adaptation (LoRA)* digunakan untuk menghasilkan citra motif batik berdasarkan prompt tekstual. Setiap prompt disusun sesuai dengan kategori motif yang dilatih, kemudian digunakan sebagai masukan pada proses inferensi untuk menghasilkan citra baru yang memiliki karakteristik visual sesuai dengan deskripsi yang diberikan.

Tabel 3 menyajikan contoh hasil inferensi untuk setiap kategori motif batik dengan membandingkan prompt yang digunakan, citra pada dataset sebagai referensi, dan citra yang dihasilkan oleh model. Perbandingan ini bertujuan untuk memberikan gambaran visual mengenai kemampuan model dalam merepresentasikan karakteristik motif batik berdasarkan deskripsi tekstual yang diberikan.

Selain melalui pengamatan visual terhadap hasil inferensi, performa model juga dievaluasi secara kuantitatif untuk mengukur kualitas citra yang dihasilkan secara objektif. Evaluasi kuantitatif dilakukan menggunakan *Frechet Inception Distance (FID)* untuk mengukur tingkat kemiripan distribusi antara citra hasil inferensi dan citra pada dataset, serta *CLIP Score* untuk menilai kesesuaian semantik antara prompt dan citra yang dihasilkan. Hasil evaluasi ini memberikan gambaran mengenai kemampuan model dalam menghasilkan motif batik yang tidak hanya menyerupai data referensi, tetapi juga sesuai dengan deskripsi tekstual yang diberikan.

Tabel 3. Perbandingan image dataset dan hasil generative

Motif Batik	Dataset	Image Generatif dengan Model LORA
Batik Buketan		
Batik Ceplok		

3.1 Hasil Evaluasi

Hasil evaluasi menggunakan *Frechet Inception Distance (FID)* dan *CLIP Score* menunjukkan adanya perbedaan performa model dalam menghasilkan citra untuk setiap motif batik. Nilai FID digunakan untuk mengukur kemiripan distribusi antara citra hasil inferensi dengan citra asli, di mana nilai yang lebih rendah menunjukkan kualitas visual yang semakin mendekati dataset referensi. Sementara itu, *CLIP Score* digunakan untuk mengukur kesesuaian semantik antara *prompt* dan citra yang dihasilkan, dengan nilai yang lebih tinggi menunjukkan keselarasan yang lebih baik.

Tabel 4. Hasil Evaluasi FID dan CLIP Score

Motif Batik	FID	CLIP Score
Batik Parang	41.52	0.286
Batik Ceplok	35.18	0.314
Batik Kawung	29.77	0.337
Batik Buketan	24.81	0.351
Batik Lasem	31.24	0.326
Mega Mendung	38.66	0.301

Berdasarkan hasil pengujian, motif Buketan memperoleh nilai FID terendah sebesar 24,81 dan *CLIP Score* tertinggi sebesar 0,351. Hasil ini menunjukkan bahwa model mampu

menghasilkan citra Buketan dengan kualitas visual yang paling mendekati citra asli sekaligus memiliki tingkat kesesuaian yang tinggi terhadap *prompt* yang diberikan. Hal ini mengindikasikan bahwa karakteristik visual motif Buketan lebih mudah dipelajari oleh model dibandingkan motif lainnya.

Motif Kawung juga menunjukkan performa yang baik dengan nilai FID sebesar 29,77 dan CLIP Score sebesar 0,337. Nilai tersebut menunjukkan bahwa model mampu mempertahankan karakteristik visual dan kesesuaian semantik motif Kawung dengan baik selama proses generasi citra.

Motif Lasem dan Ceplok memperoleh hasil yang berada pada kategori menengah. Lasem memiliki nilai FID sebesar 31,24 dengan CLIP Score 0,326, sedangkan Ceplok memperoleh nilai FID sebesar 35,18 dan CLIP Score 0,314. Hasil tersebut menunjukkan bahwa model masih mampu menghasilkan citra yang sesuai dengan *prompt*, meskipun kualitas visualnya belum sebaik Buketan maupun Kawung.

Sebaliknya, motif Mega Mendung dan Parang memperoleh nilai FID yang relatif lebih tinggi, masing-masing sebesar 38,66 dan 41,52, disertai CLIP Score yang lebih rendah yaitu 0,301 dan 0,286. Nilai tersebut mengindikasikan bahwa citra yang dihasilkan masih memiliki perbedaan distribusi yang cukup besar dibandingkan citra asli serta tingkat kesesuaian terhadap *prompt* yang lebih rendah dibandingkan motif lainnya. Hal ini menunjukkan bahwa model mengalami kesulitan dalam mempelajari karakteristik visual kedua motif tersebut.

Secara keseluruhan, hasil evaluasi menunjukkan adanya kecenderungan bahwa motif dengan nilai FID yang lebih rendah juga memperoleh CLIP Score yang lebih tinggi. Dengan kata lain, semakin baik kualitas visual citra yang dihasilkan, semakin tinggi pula kesesuaiannya terhadap deskripsi tekstual yang digunakan selama proses inferensi. Berdasarkan kedua metrik tersebut, urutan performa model dari yang terbaik hingga terendah adalah Buketan, Kawung, Lasem, Ceplok, Mega Mendung, dan Parang.

3.2 Pembahasan

Hasil evaluasi kuantitatif menunjukkan bahwa kemampuan model dalam menghasilkan motif batik berbeda pada setiap kategori motif. Berdasarkan hasil pengujian, motif Buketan memperoleh nilai FID terendah sebesar 24,81 dan CLIP Score tertinggi sebesar 0,351. Nilai tersebut menunjukkan bahwa citra hasil inferensi memiliki kualitas visual yang paling mendekati dataset asli serta tingkat kesesuaian semantik yang paling tinggi terhadap *prompt* dibandingkan motif lainnya. Sebaliknya, motif Parang memperoleh nilai FID tertinggi sebesar 41,52 dan CLIP Score terendah sebesar 0,286, yang mengindikasikan bahwa kualitas visual hasil generasi serta kesesuaian dengan *prompt* masih lebih rendah dibandingkan motif lainnya.

Hasil evaluasi juga menunjukkan kecenderungan bahwa semakin rendah nilai FID, semakin tinggi nilai CLIP Score yang diperoleh. Kondisi tersebut terlihat pada motif Kawung dan Lasem yang memiliki nilai FID relatif rendah, yaitu 29,77 dan 31,24, disertai CLIP Score masing-masing sebesar 0,337 dan 0,326. Hasil tersebut mengindikasikan bahwa model mampu menghasilkan citra yang tidak hanya memiliki distribusi visual yang lebih mendekati dataset asli, tetapi juga memiliki representasi semantik yang sesuai dengan deskripsi *prompt*.

Meskipun demikian, hubungan antara nilai FID dan CLIP Score tidak selalu menunjukkan peningkatan yang proporsional. Sebagai contoh, motif Ceplok memperoleh nilai FID sebesar

35,18 dengan CLIP Score sebesar 0,314. Hasil ini menunjukkan bahwa kualitas visual yang diukur menggunakan FID tidak selalu berbanding lurus dengan tingkat kesesuaian semantik yang diukur menggunakan CLIP Score. Hal tersebut dapat terjadi karena kedua metrik mengevaluasi aspek yang berbeda, di mana FID berfokus pada kemiripan distribusi fitur visual terhadap dataset referensi, sedangkan CLIP Score mengukur tingkat kesesuaian antara citra yang dihasilkan dan prompt teks.

Secara keseluruhan, penggunaan FID dan CLIP Score memberikan evaluasi yang saling melengkapi terhadap performa model. FID digunakan untuk mengukur kualitas visual citra hasil generasi berdasarkan kemiripannya dengan dataset asli, sedangkan CLIP Score digunakan untuk mengevaluasi kesesuaian semantik antara citra dan prompt. Dengan demikian, kombinasi kedua metrik tersebut mampu memberikan gambaran yang lebih komprehensif mengenai kemampuan model dalam menghasilkan citra motif batik Indonesia.

3. KESIMPULAN

Penelitian ini berhasil menerapkan metode pada model *Stable Diffusion* 1.5 untuk menghasilkan citra enam motif batik Indonesia, yaitu Parang, Ceplok, Kawung, Buketan, Lasem, dan Mega Mendung.

Berdasarkan hasil evaluasi kuantitatif menggunakan *Frechet Inception Distance (FID)* dan CLIP Score, model menunjukkan performa yang berbeda pada setiap kategori motif. Motif Buketan memperoleh hasil terbaik dengan nilai FID terendah sebesar 24,81 dan CLIP Score tertinggi sebesar 0,351, sedangkan motif Parang memperoleh nilai FID tertinggi sebesar 41,52 dan CLIP Score terendah sebesar 0,286. Hasil tersebut menunjukkan bahwa kualitas visual dan kesesuaian semantik citra yang dihasilkan dipengaruhi oleh karakteristik masing-masing motif batik.

Secara keseluruhan, hasil penelitian menunjukkan bahwa metode LoRA mampu mengadaptasi model *Stable Diffusion* 1.5 untuk menghasilkan citra motif batik Indonesia dengan kualitas visual yang baik dan kesesuaian semantik terhadap prompt. Meskipun demikian, performa model masih bervariasi antar motif, yang mengindikasikan bahwa kompleksitas pola dan karakteristik visual setiap motif memengaruhi kemampuan model dalam mempelajari representasi batik. Oleh karena itu, penelitian selanjutnya dapat difokuskan pada penambahan jumlah dan keragaman dataset, optimasi parameter pelatihan, serta pengujian menggunakan model difusi yang lebih mutakhir untuk meningkatkan kualitas hasil generasi.

DAFTAR PUSTAKA

- [1] A. E. Minarno, I. Soesanti, and H. A. Nugroho, "Dataset of Batik Nitik Sarimbit 120," *Data Br.*, p. 110671, 2024, doi: 10.1016/j.dib.2024.110671.
- [2] V. J. Basiroen, "Batik Lasem motifs dataset: Gunung Ringgit, Kricak, Latohan, Nyuk Pitu, Seritan," *Data Br.*, 2025, doi: 10.1016/j.dib.2025.110671.
- [3] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *CVPR 2022*, 2022, pp. 10684–10695. [Online]. Available: <https://arxiv.org/abs/2112.10752>
- [4] E. J. Hu *et al.*, "LoRA: Low-rank adaptation of large language models," in *ICLR 2022*,

2022. [Online]. Available: <https://arxiv.org/abs/2106.09685>
- [5] Y. Mao, Y. Ge, Y. Fan, and Y. Gao, "A Survey on LoRA of Large Language Models," *Front. Comput. Sci.*, vol. 19, no. 7, 2024, doi: 10.1007/s11704-024-40663-9.
 - [6] M. T. Yilmaz, A. K. Erdogmus, D. Bayar, and G. Çökünlü, "Synthetic Image Generation with *Stable Diffusion* and Trained LoRA Model for Industrial Areas," in *2024 International Conference on Artificial Intelligence and Signal Processing (ASYU)*, 2024. doi: 10.1109/ASYU62119.2024.10757052.
 - [7] M. F. Sutedy and N. N. Qomariyah, "Text to Image Latent Diffusion Model with DreamBooth Fine Tuning for Automobile Image Generation," in *2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, Yogyakarta, Indonesia, 2022, pp. 343–348. doi: 10.1109/ISRITI56927.2022.10052908.
 - [8] J. Liu and Y. Huang, "LoRA-Fine-Tuned Latent Diffusion for High-Fidelity Digitization of Classic Mongolian Patterns," *Appl. Sci.*, vol. 16, no. 1, p. 11, 2026, doi: 10.3390/app16010011.
 - [9] A. Lv and J. Yen, "A Comparative Analysis of Different Versions of *Stable Diffusion*," in *2025 IEEE International Conference on Advanced Visual and Signal-Based Systems (AVSS)*, Tainan, Taiwan, 2025. doi: 10.1109/AVSS65446.2025.11149921.
 - [10] I. Adegoke, D. Izenyi, and L. Dilli, "Efficiently Fine-tuning Large Language Model: LoRA Approach," 2024, *Zenodo*. doi: 10.5281/zenodo.11312792.
 - [11] K. Schall, K. U. Barthel, N. Hezel, and A. Moelle, "A comprehensive approach to improving CLIP-based image retrieval while maintaining joint-embedding alignment," *Inf. Syst.*, vol. 134, p. 102581, 2025, doi: 10.1016/j.is.2025.102581.
 - [12] N. Chen, S. Villar, and S. Hayou, "Learning Rate Scaling across LoRA Ranks and Transfer to Full Finetuning," in *Proceedings of the 43rd International Conference on Machine Learning (ICML)*, 2026.
 - [13] J. V. M. Marques, W. S. Santos, A. I. M. Luz, R. L. Araújo, A. D. de Sousa, and J. L. R. Sarmiento, "U-Net + Transformer Encoder and LoRA Adaptation for Rural Sheep Segmentation," in *2025 38th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, Salvador, Brazil, 2025. doi: 10.1109/SIBGRAPI67909.2025.11223404.
 - [14] D. A. Chan and S. P. Sithungu, "Evaluating the Suitability of Inception Score and Fréchet Inception Distance as Metrics for Quality and Diversity in Image Generation," in *Proceedings of the 2024 7th International Conference on Computational Intelligence and Intelligent Systems (CIIS '24)*, 2025, pp. 79–85. doi: 10.1145/3708778.3708790.
 - [15] J. Hessel, A. Holtzman, M. Forbes, R. Le Bras, and Y. Choi, "CLIPScore: A Reference-free Evaluation Metric for Image Captioning," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2021, pp. 7514–7528. doi: 10.18653/v1/2021.emnlp-main.595.