# USABILITY ANALYSIS OF STABLE DIFFUSION-BASED GENERATIVE MODEL FOR ENRICHING BATIK BAKARAN PATTERN SYNTHESIS

**[1]Kornelius Aditya Septemedi, [2]Yonathan Purbo Santosa**
[1,2]Program Studi Teknik Informatika Fakultas Ilmu Komputer,
Universitas Katolik Soegijapranata
[2]yonathansantosa@unika.ac.id

## ABSTRACT

*The rapid development of technology today helps us in various fields of work. One of the fields that can utilize technology in helping their work is batik. Utilizing Deep Learning to manage data in the form of batik pattern images and typical bakaran batik patterns using the Generative Model method, namely Stable Diffusion which aims to produce better and more detailed batik pattern images by maintaining the original pattern of batik patterns and typical bakaran batik patterns. This research only uses datasets in the form of batik pattern images and typical bakaran batik patterns. The image data is processed augmentation first by performing the inverse on the image, resizing the image to 512x512, then randomly rotating the image, performing a random horizontal flip on the image, and performing the inverse again on the image. Pre-Training on image data to find the right parameters and conditions used in the training process. The result of this research is that the Stable Diffusion model version 1.4 and version 2.1 show good performance in processing and creating batik pattern images and batik patterns typical of Bakaran. In this study, the score calculation process for Stable Diffusion version 1.4 and version 2.1 was carried out using Inception Score and CLIP Score to calculate the images generated from the two versions. In the calculation using CLIP Score, the results obtained by version 1.4 are higher than version 2.1 for the same reason as Inception Score because the image produced by version 1.4 is more abstract. Of the two versions used is version 1.4 because the resulting image shows an abstract image that reflects a good batik pattern. Then, the version used to process batik patterns and batik patterns typical of Bakaran is Stable Diffusion version 1.4 which shows excellent performance in processing batik pattern images. The results of Stable Diffusion version 1.4 show good and abstract batik patterns in accordance with the characteristics of Bakaran batik.*

**Keywords:** Stable Diffusion, Generative Model, batik patterns and batik patterns typical of Bakaran, Deep Learning

## BACKGROUND

The rapid development of technology today helps us in various fields. By utilizing existing technology, it can make it easier for us to do work in various fields such as health, communication and others. One of the fields that can utilize technology in doing its work is batik or making batik, by utilizing existing technology, namely Deep Learning using the Generative Model. In the Generative Model there is a method that can be used, namely, Stable Diffusion.

Stable Diffusion is a method or algorithm in the Generative model that works iteratively in the latent representation space and then decodes or trains the dataset or representation used into a full image. Stable Diffusion is able to generate a detailed image from just a text description. Due to Stable Diffusion's good capabilities, it can generate images of various types of commands given. Stable Diffusion can be used for inpainting, outpainting, converting text to image and image to image, sound to image. Stable Diffusion can also be used to learn and reconstruct lost or damaged data, this makes Stable Diffusion widely used in various applications such as image and video processing and in the development of artificial intelligence systems. In the above case Stable Diffusion can also be used for batik making, by entering images in the form of batik patterns and typical bakaran batik patterns. The image is trained, and in one training, the image is created several times until the results are detailed by adding or replacing missing and damaged pixels.

This research aims to use Stable Diffusion for datasets in the form of batik pattern images and typical bakaran batik patterns in order to produce good and more detailed batik images by using input in the form of images and then replacing missing and damaged pixels. The results of this method are expected to be used in working or making batik, both ordinary batik patterns and typical bakaran batik patterns. Because by using Stable Diffusion, the results obtained have a good and quite high success rate.

## LITERATURE STUDY

Rombach et al.[1] Using Diffusion Model to process image shape composition becomes a sequential application of denoising autoencoders. Using latent space makes it possible to reach a near-optimal point between complexity reduction and detail preservation. This Latent Diffusion Model (LDM) gets new high scores in inpainting and class-conditional image synthesis as well as competitive performance. Then transformed the encoder for LDM for text-to-image image modeling, with training parameter 1.45B KL-regularized LDM and then conditioned the language cues on LAION-400M using BERT-tokenizer. Then evaluating text to image generation with MS-COCO validation set, this model improves AR method as well as GAN (Generative Adversarial Network) based. To analyze the flexibility of the cross-attention mechanism, training the model to synthesize images based on semantic-layouts in Open Images and improving COCO. The result is a simple and efficient LDM that significantly improves the training and sampling efficiency of the denoising diffusion model without degrading its quality. This result is suitable for use in task-specific architectures.

Developing small data sets becomes more effective by proposing a new architecture Generative Adversarial Network (GAN) for semi-supervised augmentation of X-rays data for pneumonia and COVID-19 detection using Generative Model. Then the GAN used by Motamed et al.[2] effectively augments data and improves disease classification results in using X-rays for pneumonia and COVID-19. The results obtained were compared with Deep Convolutional GAN and traditional augmentation methods. Then the results of the comparison show that the GAN-based augmentation method surpasses other augmentation methods in training GANs for X-ray image detection. Then the results obtained from the study are with two different datasets of pneumonia and COVID-19 then calculating the ROC curve (AUC), with the pneumonia dataset being compared with a normal X-ray then the COVID-19 dataset with a positive test result is compared with the COVID-19 data with a negative result.

Presents a novel Deep Generative Model based on a non i.i.d. variational autoencoder to capture global dependencies between observations in an unsupervised manner. In contrast to the semi-supervised alternative, here Peis et al.[3] It combines a local space and a global Gaussian latent variable that obtains three results, namely an induced latent global space capturing a decomposed representation and then interpreted unregulated and user-defined. It then processes domain alignment to obtain correlation and interpolation between different datasets. It then learns the ability of the global space to distinguish between non-trivial group structures. The new Deep Generative Model found that by combining structured clustering in a local latent space with a Gaussian global prior and a family of structured variations, the model can interpret and infer based on a fully unsupervised global space.

Bosquet et al.[4] Developed full data augmentation using pipelines to detect small objects based on Generative Adversarial Network. It then combines the GAN object generator with object segmentation, inpainting, and image blending techniques to obtain high-quality synthetic data. The main components of this new DS-GAN architecture can generate realistic small objects compared to larger objects. The results show that this data augmentation method can improve the performance of the state-of-the-art model by 11.9% AP per-0.5 s in UAVDT and 4.7% AP per-0.5 s in iSAID both with small object subsets and when the number of training examples is limited.

In this study, Li et al.[5] performs gait recognition while carrying an object using Generative Adversarial Network alpha mixing where by taking an approach to generate a gait template without CO (carrying an object). Then given a gait template with CO for input using a conventional Generative Adversarial Network, so that no changes to unnecessary parts occur. An approximation of the gait template without CO is performed, and then merged with the original template using an alpha matte approximation that indicates the blending parameters. Then, an alpha-mixed template was created from the original template and the template without CO based on the estimated alpha matte. After that, using two independent generators to estimate the alpha matte and the resulting template without CO, and then input the alpha-mixed gait template into a sophisticated discrimination network for gait recognition. The results obtained from the study of existing gait datasets for the public with CO in real life show high performance.

Wang et al.[6] Create applications using Generative Adversarial Network for neuroimaging and clinical neuroscience by bridging Deep Learning and neurology methods to see how GAN is utilized to support clinical decision making. Then contribute to a better understanding of the structural and functional patterns of brain diseases. GAN research shows good performance in solving clinical tasks including disease diagnosis, anomaly and tumor detection, brain development modeling, Alzheimer's progression estimation, lesion dynamics prediction and brain tumor growth prediction. This method presents algorithm reproducibility, interpretability and fairness which are crucial in the deployment of potential machine learning models.

Developing a semi-supervised audio-visual domain adaptation of conditional Generative Adversarial Network. In this research, Athanasiadis et al.[7] using facial information from video to improve the awareness and tracking of emotion prediction in audio signals. This research uses the dacssGAN (Domain Adaptation Conditional Semi-Supervised Generative Adversarial Network) method to bridge two inherently different domains. After that, inputting the source domain (visual data) as well as some conditional information based on inductive conformal prediction, the proposed architecture provides results as close as possible to the target domain (audio data). The results show that the classification performance of the extended dataset using the original audio data and then enhanced using the samples from dacssGAN obtained results of 50.29% and 48.65% compared to the original audio samples of 49.34% and 46.90% for the two datasets used.

Research presented by Ates et al.[8] A conditional Generative Adversarial Network is used to model the fuel spray and GAN performs virtual spray simulation for conditions relevant to the combustion chamber of an aero engine. After that, it uses an autoencoder to transform the droplet trajectory at variable length to a fixed length as well as a lower dimensional representation. Then the Wasserstein GAN learns to mimic the latency of the vaporized droplet throughout its lifetime. The GANs are also conditioned with the injection location and droplet diameter to improve the generalization of the entire framework. The results obtained in the study show that GAN has great potential for low-order model approximation with high computational efficiency. Meanwhile, the autoencoder should be improved in order to maintain local dependency. The results of training for one day and on HPC with a conventional CFD approach showed a time of one week and performed in simulations of the order of seconds resulted in a droplet count of 200,000 passes.

Combines Generative and Discriminative models for Bayesian Semi-Supervised Learning. It then provides a framework that makes both models learn from labeled and unlabeled data. After that, it accounts for the uncertainty in the predictive distribution, and then provides the first Bayesian approach to semi-supervised learning with generative models. Gordon and Hernández-Lobato[9] obtained results showing this combined discriminative and generative model, outperforming the pure generative model on predictive performance and uncertainty calibration across several semi-supervised learning tasks.

Developing detail-preserving face sketch synthesis using Generative Adversarial Network. Wan et al.[10] provides a GAN framework for detail preservation of realistic face sketch synthesis. After that, a high-resolution network is modified as a generator to transform the face image from photo to sketch domain. Then, detail reduction is designed to force the synthesized facial sketch image to have details close to the corresponding image. This detail reduction is used to ensure that the synthesized facial sketch image still has clear sketch details like a hand-drawn sketch. The results show that the proposed model has high performance compared to existing ones, both in terms of visual perception and objective evaluation. This study achieved FSIM values of 0.7345 and 0.7080 and Scoot values of 0.5317 and 0.5091, which are higher than most of the comparison methods on the CUFS and CUFSF datasets, respectively.

Saeed et al.[11] Conducted research on bi-parametric prostate MR image synthesis using pathology and sequence-conditioned stable diffusion. They provided an image synthesis mechanism for text-conditioned multi-sequence prostate MR images to control the presence and order of lesions and to generate conditioned paired bi-parametric images. Then this mechanism proposes image-based conditioning for pairwise data generation. After that, they used 2D image slices of prostate cancer patients and synthesized image realization was performed and validated through radiologist evaluation to identify real and fake images. Afterwards, a radiologist with 4 years of experience reading urology MRs achieved only 59.4% accuracy across the tested sequences. The results obtained using this mechanism by training the model for lesion identification showed better performance and achieved an accuracy of 76.2% versus 70.4%, a statistically significant improvement.

This research presents how to improve Model-Agnostic Zero-Shot (MA-ZSC) classification using stable diffusion. MA-ZSC classification with reference to training is a non-specific classification architecture that is trained to classify real images without using real images during training. Shipard et al.[12] Results show that diffusion models provide the potential to address MA-ZSC, but current performance is still far from that of large-scale language-vision models. Subsequently, in this study using the diffusion model, it was shown that by providing initial insights, the performance of MA-ZSC can be improved by increasing the diversity of images in the generated dataset. Modifications were then made to the text-to-image generation process using the trained diffusion model. Thereafter, the results obtained show notable improvements in various classification architectures and are comparable to state-of-the-art models such as CLIP. A similar study was carried out on CIFAR10, CIFAR100, EuroSAT, ResNet and ViT which showed high difficulty for image-less classification due to their imagery domains.

Developing an interactive visualization tool for stable diffusion explanation by converting text commands to images. Lee et al.[13] used a diffusion-based generative model to tightly integrate visual depictions of the stable diffusion of complex components along with detailed explanations of the underlying operations. Afterwards, comparing the evolution of the image representation guided by two related text prompts during the refinement time step. The result is an interactive web tool that explains stable diffusion to generate high-resolution images with text commands. The tool can be accessed locally in the user's browser without installation or specialized hardware.

Of the thirteen literature studies above, this research uses one literature as a basis for research. Research conducted by Rombach et al.[1] concluded that the LDM is simple and efficient. The LDM model also significantly improves the training and sampling efficiency of the diffusion denoising model without degrading its quality.

## RESEARCH METHODOLOGY

### *Dataset Collection*

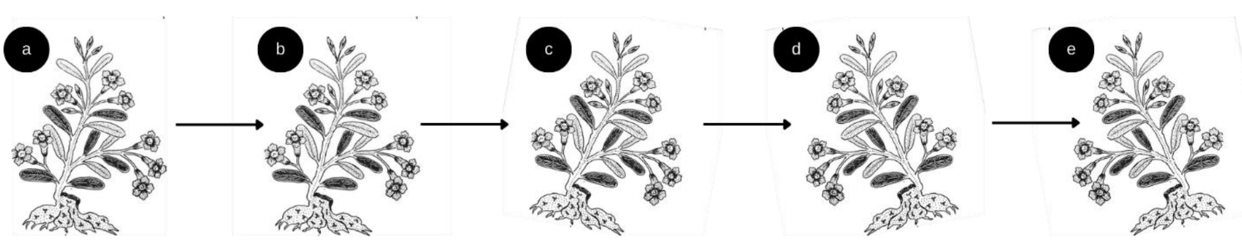
Dataset Collection is divided into the following 2 stages

a. The data that will be used in this research is a typical bakaran batik image data. This dataset is obtained by taking pictures or photos of several batik pattern designs on paper with typical bakaran batik patterns which are directly taken at a typical Bakaran batik producer in Pati Regency by a team from the Faculty of Architecture and Design.
b. The image data that has been obtained is processed into vectors and resized with a width of 512 and a length following. This image data is processed by the Architecture and Design faculty team.
c. In the second stage of data collection, interviews were conducted by the Faculty of Architecture and Design team to Bakaran batik producers in Pati Regency to collect prompt data used for the pre-training process of image data.

Then from the two stages of Dataset Collection above, the image and prompt datasets that have been obtained will be processed using Stable Diffusion.

### *Data Preprocessing*

Preprocessing the dataset is done before processing it using Stable Diffusion. In this research, data preprocessing is done by inverse image data and then resizing the input from the dataset by cropping the image to a size of 512x512, randomly rotating the image data, then doing a random horizontal flip on the image data, and inverse the input data again. After that, the dataset is saved in PNG format for processing using Stable Diffusion.




**Gambar 1.** Augmentation Process Steps

The augmentation process in the figure above shows the steps taken in the process. The first step of the augmentation process is the inverse image. Furthermore, the second step of the image is cropped to a size of 512x512. Then, for the third step the image is rotated randomly with a probability of 30%. The fourth step of the image is flipped horizontally randomly with a probability of 60%. The last augmentation process is done inverse back on the image. These steps are carried out 10 times for each image of a typical Bakaran batik pattern.

**Gambar 2.** Augmentation Result

The image above shows the results of the augmentation process performed for each image. The first image is the first image data before the augmentation process. Then, the images next to the first image are the results of augmentation starting from the inverse image, crop the image size to 512x512, rotate the image randomly, flip the horizontal image randomly and the image is inversed again.

### Pre-Training

Conduct pre-training to find the right parameters and conditions to be used in the training process. After getting the right parameters and conditions, process the dataset that has been obtained. Then do a further pre-training process to determine the Epoch and Checkpoint that will be used in the training process.

### Data Training

In the training process using two models, namely Stable Diffusion version 1.4[1] with Epoch 3000 and checkpoint 100 and Stable Diffusion version 2.1[1] with Epoch 5000 and checkpoints

200. Using version 1.4[1], The model that has been uploaded to Huggingface is called and processed by giving a text command, which is then converted into an image. The same process is also done for version 2.1[1], which is by calling the model that has been uploaded to Huggingface, then processing it by giving commands in the form of text that is converted into images.

### *Evaluation*

The evaluation is divided into two, namely qualitatively and quantitatively. Qualitatively, the use of Stable Diffusion for the Batik Bakaran pattern process is carried out by interviewing one of the batik craftsmen who has been working as a batik craftsman in Bakaran Village for a long time. Then quantitatively, the evaluation was carried out by calculating using Clip Score and Inception Score. Clip Score,[14] is a method used to calculate and evaluate the level of match between the command given to create an image and the image produced from the command. Inception Score, [15][16][17] is an algorithm used to determine or measure the quality of images created by generative AI by counting the images produced, the more diverse the results of the calculated images the higher the score obtained.

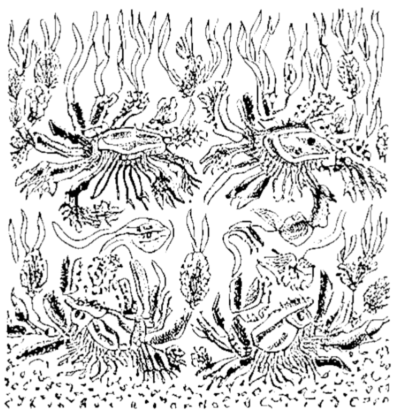
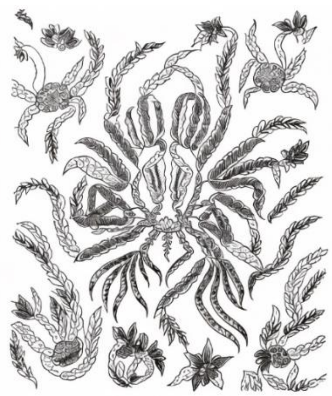
## RESULTS

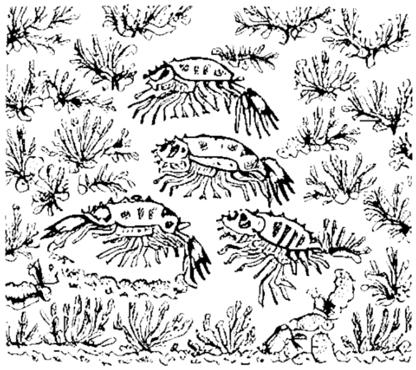
Qualitatively conducted by interviewing one of the Batik Bakaran craftsmen obtained positive results. The batik craftsman said that the resulting batik pattern really showed that it was a good batik pattern, because the results obtained showed the abstract value that a batik pattern must have. The version used for batik craftsmen is Stable Diffusion version 1.4.

**Tabel 1.** Image Results of Stable Diffusion Version 1.4 and Version 2.1

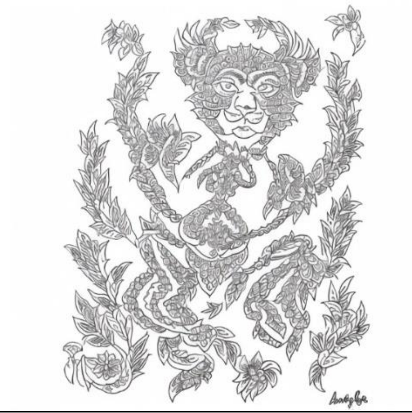
| **Prompt** | **Versi 1.4** | **Versi 2.1** |
|---|---|---|
| tiger head |  |  |

| | | |
|---|---|---|
| |  |  |
| |  |  |
| |  |  |

| | | |
|---|---|---|
| crab and seaweed |  |  |
| |  |  |
| |  |  |

| | | |
|---|---|---|
| |  |  |
| panda and bamboo |  |  |
| |  |  |

| | | |
|---|---|---|
| |  |  |
| |  |  |
| lion |  |  |

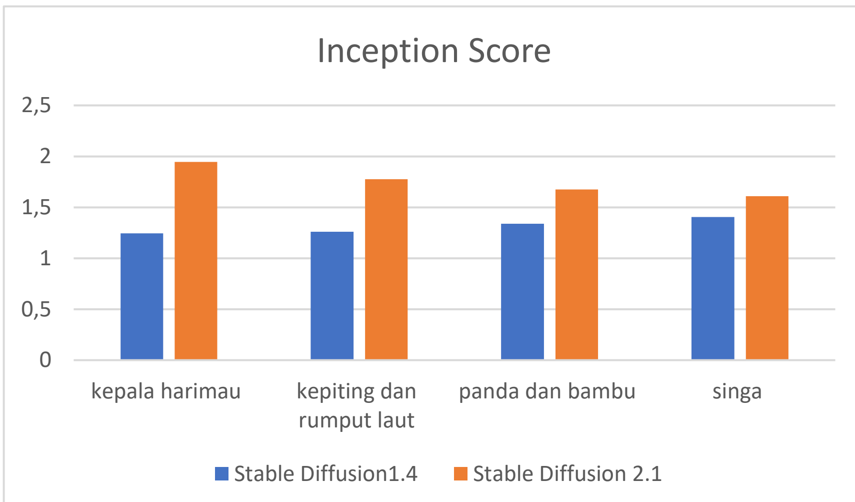
| | | |
|---|---|---|
| |  |  |
| |  |  |
| |  |  |

Quantitatively, the calculation is done using the Inception Score and CLIP Score. The table above shows the results of making Batik Bakaran patterns using Stable Diffusion versions 1.4 and 2.1. From the results of the image, the score calculation is done by calculating the score of the image based on the prompt used to create the image.

**Tabel 2.** Inception Score Results

| Prompt | Inception Score | |
|---|---|---|
| | ***Stable Diffusion1.4*** | ***Stable Diffusion 2.1*** |
| tiger head | 1.243895488 | 1.946171042 |
| crab and seaweed | 1.259745643 | 1.775231502 |
| panda and bamboo | 1.338376894 | 1.673492992 |
| lion | 1.405875089 | 1.608853808 |

The score obtained in the calculation using Inception Score for version 1.4 with the tiger head prompt scored 1.2438954877681947, crab and seaweed scored 1.2597456425691815, panda and bamboo scored 1.338376894409098, and lion scored 1.4058750886054834. After that, the calculation results for version 2.1 with the tiger head prompt scored 1.9461710421110832, crab and seaweed scored 1.7752315018654592, panda and bamboo scored 1.6734929924539454, and lion scored 1.6088538077654906.



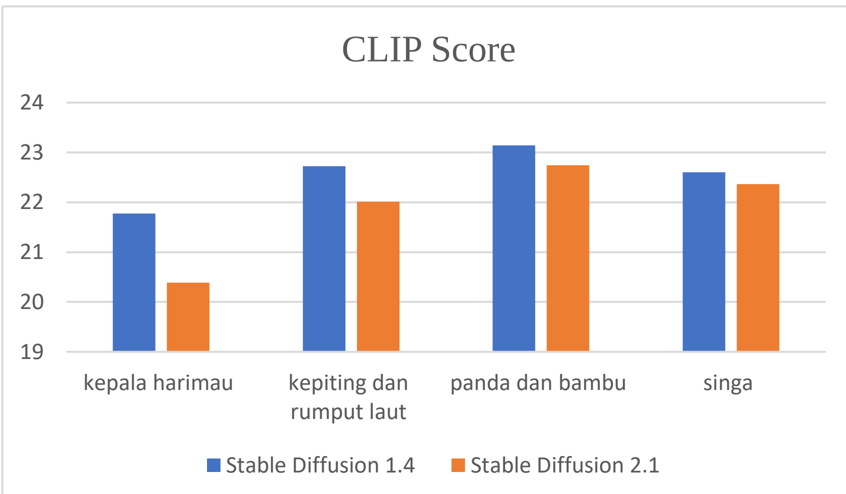

**Gambar 3.** Inception Score Diagram

It can be seen from the calculation diagram using Inception Score above that the tiger head prompt for version 1.4 obtained a smaller score compared to version 2.1. The crab and seaweed prompt for version 1.4 obtained a smaller score compared to version 2.1. Then, the panda and bamboo prompt for version 1.4 also obtained a smaller score compared to version 2.1, and for the lion prompt in version 1.4 also obtained a smaller score compared to version 2.1. From these results, it can be seen that Stable Diffusion version 2.1 obtained a higher score for each of its prompts.

**Tabel 3.** CLIP Score Results

| Prompt | CLIP Score | |
|---|---|---|
| | ***Stable Diffusion 1.4*** | ***Stable Diffusion 2.1*** |
| tiger head | 21.7768 | 20.3892 |
| crab and seaweed | 22.7241 | 22.0098 |
| panda and bamboo | 23.1392 | 22.7417 |
| lion | 22.6044 | 22.365 |

From the table above the results of calculations using CLIP Score, for version 1.4 with the tiger head prompt scored 21.7768, crabs and seaweed scored 22.7241, pandas and bamboo scored 23.1392, and lions scored 22.6044. In version 2.1 with the tiger head prompt, it scored 20.3892, crab and seaweed scored 22.0098, panda and bamboo scored 22.7417, and lion scored 22.365.




**Gambar 4.** CLIP Score Diagram

The calculation diagram using CLIP Score above shows that the tiger head prompt for version 1.4 obtained a smaller score compared to version 2.1. The crab and seaweed prompts for version 1.4 scored smaller compared to version 2.1. Then, the panda and bamboo prompt for version 1.4 also obtained a smaller score compared to version 2.1, and for the lion prompt in version 1.4 also obtained a smaller score compared to version 2.1. From these results, it can be seen that Stable Diffusion version 2.1 obtained a higher score for each of its prompts.

### *Discussion*

From the results obtained through interviews with craftsmen, the Batik Bakaran craftsmen are very happy and enthusiastic about the batik pattern images produced by Stable Diffusion. The Batik Bakaran craftsmen said that the resulting image shows how the batik pattern should be, where the batik pattern should have an abstract pattern that shows the uniqueness of the batik.

The results of the score calculation using Inception Score and CLIP Score which calculates the score of each prompt with four images from each prompt for Stable Diffusion version 1.4 and version 2.1 show interesting results. In the score calculation using Inception Score, the score of the tiger head prompt for version 1.4 shows smaller results compared to version 2.1. Then for the crab and seaweed prompt, version 1.4 gets smaller results compared to version 2.1. At the panda and bamboo prompts for version 1.4 also get smaller results compared to version 2.1. Then for the lion prompt, version 1.4 still gets smaller results compared to version 2.1. From the four prompt scores, it can be seen that Stable Diffusion version 2.1 obtained a higher score. These scores are obtained by calculating the Average score of each image, then calculating the Mean of the average score to obtain the Inception Score of each prompt. The score results from both versions are still very small because the scores obtained still do not reach 20 and above.

Then in the calculation of scores using CLIP Score, the results of the tiger head prompt version 1.4 get a higher score than version 2.1. Then for crab and seaweed prompts also get higher scores than version 2.1. At the panda and bamboo prompt, the score results obtained for version 1.4 are still higher than version 2.1. After that, at the lion prompt version 1.4 the score results obtained are also still higher than version 2.1. Of the four scores, it shows that in the calculation using CLIP Score, the Stable Diffusion version 1.4 score results obtained a higher score. The score calculation is done by calculating the score of each image and matching it with the prompt list which then shows the score results for each prompt.

The results of the score calculation using Inception Score and CLIP Score show that if the calculation uses Inception Score, version 1.4 gets smaller results compared to version 2.1. The score calculation results using Inception Score are still very small because the score does not reach 20 and above. Then, if the calculation using CLIP Score shows that the results obtained by version 1.4 are higher than version 2.1.

The limitation of this research is the data used. The data used is limited to Batik Bakaran pattern data from one of the batik craftsmen in Bakaran Village. Then this research is also still limited in that it can only create batik pattern images in the form of plants, marine life, and land animals only because of the limited dataset and has not been able to process and create images with Wayang patterns.

## CONCLUSION

From this research it can be concluded that Stable Diffusion has excellent performance in processing batik pattern data and batik patterns typical of Bakaran. In this study, what is used by batik makers to process batik patterns is Stable Diffusion version 1.4 because in version 1.4 the resulting batik patterns show good batik pattern results. From the results obtained, Stable Diffusion version 1.4 and version 2.1 show very different results in each image created by the two versions.

Then the calculation results obtained using Inception Score and CLIP Score show the difference in the results obtained. In calculations using Inception Score version 1.4 tends to get smaller results than version 2.1 for each prompt used. The Inception Score results are also overall from both versions are still very small. The calculation using CLIP Score shows different results from Inception Score, CLIP Score shows the result that Stable Diffusion version 1.4 gets higher results than version 2.1 for each prompt. The results of calculations using Inception Score also show that the score results from version 1.4 are smaller than version 2.1, this is because the Inception Score calculation is done by calculating the Average score of each image and then calculating the Mean of all image scores.

From the results of the interview and the results of these calculations, it shows that Stable Diffusion has excellent performance in processing batik patterns and batik typical of Bakaran which of the two existing versions, Stable Diffusion version 1.4 shows more abstract results compared to version 2.1 which is more realistic. In future research, if you want to process batik patterns, you should prepare more batik pattern data so that you can process a variety of existing batik patterns. Then, the Stable Diffusion version used in future research can use a different version so that the difference can be seen. If using the same version, then the dataset used is prepared more and more varied.

## DAFTAR PUSTAKA


[1] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, “High-Resolution Image Synthesis with Latent Diffusion Models.” arXiv, Apr. 13, 2022. Accessed: May 09, 2023. [Online]. Available: http://arxiv.org/abs/2112.10752

[2] S. Motamed, P. Rogalla, and F. Khalvati, “Data augmentation using Generative Adversarial Networks (GANs) for GAN-based detection of Pneumonia and COVID-19 in chest X-ray images,” *Informatics in Medicine Unlocked*, vol. 27, p. 100779, 2021, doi: 10.1016/j.imu.2021.100779.

[3] I. Peis, P. M. Olmos, and A. Artés-Rodríguez, “Unsupervised learning of global factors in deep generative models,” *Pattern Recognition*, vol. 134, p. 109130, Feb. 2023, doi: 10.1016/j.patcog.2022.109130.

[4] B. Bosquet, D. Cores, L. Seidenari, V. M. Brea, M. Mucientes, and A. D. Bimbo, “A full data augmentation pipeline for small object detection based on generative adversarial networks,” *Pattern Recognition*, vol. 133, p. 108998, Jan. 2023, doi: 10.1016/j.patcog.2022.108998.

[5] X. Li, Y. Makihara, C. Xu, Y. Yagi, and M. Ren, “Gait recognition invariant to carried objects using alpha blending generative adversarial networks,” *Pattern Recognition*, vol. 105, p. 107376, Sep. 2020, doi: 10.1016/j.patcog.2020.107376.

[6] R. Wang *et al.*, “Applications of generative adversarial networks in neuroimaging and clinical neuroscience,” *NeuroImage*, vol. 269, p. 119898, Apr. 2023, doi: 10.1016/j.neuroimage.2023.119898.

[7] C. Athanasiadis, E. Hortal, and S. Asteriadis, “Audio–visual domain adaptation using conditional semi-supervised Generative Adversarial Networks,” *Neurocomputing*, vol. 397, pp. 331–344, Jul. 2020, doi: 10.1016/j.neucom.2019.09.106.

[8] C. Ates, F. Karwan, M. Okraschevski, R. Koch, and H.-J. Bauer, “Conditional Generative Adversarial Networks for modelling fuel sprays,” *Energy and AI*, vol. 12, p. 100216, Apr. 2023, doi: 10.1016/j.egyai.2022.100216.

[9] J. Gordon and J. M. Hernández-Lobato, “Combining deep generative and discriminative models for Bayesian semi-supervised learning,” *Pattern Recognition*, vol. 100, p. 107156, Apr. 2020, doi: 10.1016/j.patcog.2019.107156.
[10] W. Wan, Y. Yang, and H. J. Lee, “Generative adversarial learning for detail-preserving face sketch synthesis,” *Neurocomputing*, vol. 438, pp. 107–121, May 2021, doi: 10.1016/j.neucom.2021.01.050.
[11] S. U. Saeed *et al.*, “Bi-parametric prostate MR image synthesis using pathology and sequence-conditioned stable diffusion.” arXiv, Mar. 03, 2023. Accessed: May 15, 2023. [Online]. Available: http://arxiv.org/abs/2303.02094
[12] J. Shipard, A. Wiliem, K. N. Thanh, W. Xiang, and C. Fookes, “Diversity is Definitely Needed: Improving Model-Agnostic Zero-shot Classification via Stable Diffusion.” arXiv, Apr. 16, 2023. Accessed: May 15, 2023. [Online]. Available: http://arxiv.org/abs/2302.03298
[13] S. Lee *et al.*, “Diffusion Explainer: Visual Explanation for Text-to-image Stable Diffusion.” arXiv, May 08, 2023. Accessed: May 15, 2023. [Online]. Available: http://arxiv.org/abs/2305.03509
[14] A. Nichol *et al.*, “GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models.” arXiv, Mar. 08, 2022. Accessed: Nov. 01, 2023. [Online]. Available: http://arxiv.org/abs/2112.10741
[15] A. Grover, M. Dhar, and S. Ermon, “Flow-GAN: Combining Maximum Likelihood and Adversarial Learning in Generative Models,” *AAAI*, vol. 32, no. 1, Apr. 2018, doi: 10.1609/aaai.v32i1.11829.
[16] S. Barratt and R. Sharma, “A Note on the Inception Score.” arXiv, Jun. 21, 2018. Accessed: Nov. 01, 2023. [Online]. Available: http://arxiv.org/abs/1801.01973
[17] K. G. Hartmann, R. T. Schirrmeister, and T. Ball, “EEG-GAN: Generative adversarial networks for electroencephalograhic (EEG) brain signals.” arXiv, Jun. 05, 2018. Accessed: Nov. 01, 2023. [Online]. Available: http://arxiv.org/abs/1806.01875