Generative Artificial Intelligence for Visualizing Indonesian Historical Figures in the Majapahit Era

Peter Ardhianto, *Member, IAENG,* Yonathan Purbo Santosa, *Member, IAENG,* Gustav Anandhita, *Member, IAENG,* Sayid Mataram, Christian Moniaga, Gita Claudia, and Stefani Lily Indarto.

***Abstract*— The creative industry is making significant progress due to the rise of generative artificial intelligence that can execute text-based commands. One of the often-used models is Stable Diffusion because of its ability to produce results that closely resemble the original. However, Stable Diffusion needs to be adapted to generate local images of Indonesia, especially of historical figures from the Majapahit Kingdom era. The difficulty in obtaining references that visually represent historical figures, which are close to historical sources and evidence, is an obstacle to the use of visual assets in the creative industry and history learning media. This study uses a data collection approach from Majapahit archaeologists, literature sources, and historical artifacts as sources of visual datasets. The dataset will be trained using a Stable Diffusion model and LoRA validated using CLIP score and VQA to determine image performance, and compared with a survey of respondents in the field of history related to generative image results. The results of the study showed that the VQA test aligned, and the human evaluation test showed that the Anithing_V30Pruned model was successful in visualizing the faces of the king and queen according to the sketch statue. The RealisticVisionB1_v51VAE model is successful in visualizing the attributes of the kings and queens of the Majapahit era. This differs from the results of the CLIP Score evaluation, as the CLIP score tends to identify the entire image, including the background of the sample image.**

***Index Terms*— Stable diffusion, LoRA, Generative AI, Majapahit kingdom, Historical figure.**

Manuscript received October 1, 2024; revised April 18, 2025.

This work was supported by a grant from the Directorate of Research and Community Service, Deputy for Research and Development, Indonesian Ministry of Research and Technology BIMA 2024.

Peter Ardhianto is an assistant professor at the Department of Visual Communication Design of the Soegijapranata Catholic University, Semarang 50234 Indonesia (phone: +62248505003; e-mail: peter.ardhianto@unika.ac.id).

Yonathan Purbo Santosa is an assistant professor at the Department of Informatics Engineering of the Soegijapranata Catholic University, Semarang 50234 Indonesia (e-mail: yonathansantosa@unika.ac.id).

Gustav Anandhita is an assistant professor at the Department of Architecture of the Soegijapranata Catholic University, Semarang 50234 Indonesia (e-mail: gustav@unika.ac.id).

Sayid Mataram is an assistant professor at the Department of Visual Communication Design of the Sebelas Maret University, Surakarta 57121 Indonesia (e-mail: sayidmataram@staff.uns.ac.id).

Christian Moniaga is an assistant professor at the Department of Architecture of the Soegijapranata Catholic University, Semarang 50234 Indonesia (e-mail: christianmoniaga@unika.ac.id).

Gita Claudia is an assistant professor at the Department of Accounting of Soegijapranata Catholic University, Semarang 50234 Indonesia (e-mail: gita@unika.ac.id).

Stefani Lily Indarto is an associate professor at the Department of Accounting of the Soegijapranata Catholic University, Semarang 50234 Indonesia (e-mail: sli@unika.ac.id).

I. Introduction

The Majapahit Kingdom was an important historical entity in Indonesian civilization that lasted from around the 13th to the 16th centuries AD [1]. The presence of Majapahit was an important milestone in the history of the archipelago in the political, economic, and cultural fields. Efforts to reconstruct the faces and characters of important figures in Majapahit history are still limited. [2].

Majapahit was an Indonesian kingdom in the classical era, during which its cultural development. This development not only emerged from the progress of domestic community thinking in the region, but also received influence from various foreign cultures, including South Asia and Southeast Asia [3]. The common thread in these cultural relations is the religious side, in the form of Hindu and Buddhist influences, which developed in these regions. Manifestations of this cultural influence include statues and literary artifacts, which ultimately became the main source for understanding the physical appearance of Majapahit [4]. However, the interpretation of these artifacts is often still speculative and inadequate for producing illustrations that are close to the source.

These statues were made to depict great figures from within the palace, and not all figures from the palace could be honored, only those who were recorded by the king or queen. The problem that arises when studying people's lives in the Majapahit era is that researchers want to reproduce the visuals of society, from kings to ordinary people. Books, such as Pararaton and Tantu Panggelaran [5], as well as inscriptions which serve as sources of written records during the kingdom period, do not provide detailed descriptions of people's lives.

Researchers can observe a variety of clothing and attributes from what is worn on statues and temple reliefs, as well as terracotta artifacts, but there is a limited record [6]. This condition is a gap and a challenging problem for this study, to reconstruct the visuals, reproduce visuals, to re-inventory so that they can slowly describe the clothing worn during the Majapahit kingdom.

Many studies have been carried out regarding the history of Majapahit, such as comparative studies with contemporary kingdoms to examine the influences and relationships of these kingdoms, allowing us to predict the cultural perceptions that existed in Majapahit society and Java in the 14th to 15th centuries [7]. This research serves as a reference regarding perception and visual projection, as well as building images of visualizations of figures in the Majapahit era, in the research that will be carried out.

Studies on the visual character forms of the Majapahit era have examined terracotta artifacts from the perspective of objects used in everyday life [8]. Statue artifacts, such as Harihara from several museum collections, focusing on clothing and the details of patterned jewelry, one of which is Kalamakara [9, 10]. This research has discussed various artifacts used in the daily lives of the Majapahit people; however, a concrete depiction of the activities involving the use of these artifacts has not yet been visualized. For this reason, this research will visualize activities involving the use of these artifacts, making it easier for today's society to study the activities of the Majapahit society in the past.

Apart from being found physically, these artifacts are also depicted in temple reliefs. Research on iconography has been carried out, including on the iconography of Javanese women in the 14th century, which was compared with other artifact findings as a visualization of female figures, so that through the study of iconography, female figures can be distinguished from male figures, as well as depicting female iconography outside the palace. Apart from that, the visualization of women's iconography is also in line with studies on the iconography of female statues of Gayatri, Tribhuwanatunggadewi, and Suhita, which represent the iconography of palace women [11, 12]. These studies have explained the various iconographies of Majapahit society, especially women, however, visualization efforts are still needed to translate the female form of that period because several sources, such as terracotta artifacts, temple reliefs, and statues are distorted from the human form into idealized forms according to their creators or contemporary trends.

There have been many implications for Majapahit-era society in the modern world, such as the application of digital games set in the history of Majapahit, as well as virtual visual reconstruction efforts [13]. Research involving Majapahit in various media has been carried out, but visually, it is still based on subjective perspectives and follows current visual trends. In this research, objective visualization efforts will be made, especially on certain popular figures in the history of Majapahit, by referring to visuals in historical evidence, characters, and performance efforts precisely using AI-based technology.

Currently, AI technology is widely used to reconstruct the faces of historical figures from various civilizations. However, the use of AI Stable Diffusion, especially to produce illustrations of Majapahit figures based on archaeological evidence, is still a novelty. This technology offers the possibility to produce images that are more realistic and closely resemble to the original situation based on data and historical artifacts. For example, Zhang, J. et al (2024) used LoRA and ControlNet for the innovative revitalization of historical building facades with cultural value [14]. Changes in function, while considering cultural guidelines, are needed during a renovation or conversion of a building. Alfalasi, A et al. (2024) discuss the use of SDXL and LoRA to generate images and patterns of traditional Emirati cultural motifs from the United Arab Emirates called Al-Sadu [15]. The resulting images will be used as motifs to help relieve mental disorders and support well-being. The generative AI method is used as an effort to preserve the motifs used.

The main concern that will be examined in this research is how to implement the AI Stable Diffusion training framework to create illustrations of figures from the Majapahit Kingdom, based on limited historical evidence from statues, reliefs, and existing literature. The problem formulation includes interpreting existing data, statues, and reliefs to obtain adequate information about the physical characteristics of Majapahit figures, using the AI Stable diffusion technology to produce realistic and adequate illustrations of these figures, and evaluating whether the image results produced using this approach were accurate and authentic.

To overcome this problem, the approach that will be used is to integrate archaeological, historical, and cultural arts data with AI Stable Diffusion technology [16]. The initial step will involve an in-depth analysis of the statues and reliefs associated with Majapahit to identify the relevant physical characteristics, accessories, and clothing of the figures depicted. Once the analysis of existing sources has been gathered, a dataset must be created to train the Stable Diffusion model. After being trained, the AI Stable Diffusion technology will be applied to produce realistic illustrations of Majapahit figures. The accuracy and authenticity of the AI-generated images will be validated through a questionnaire distributed to cultural actors, archaeologists, and historical sources.

By introducing historical figures of Majapahit through generative AI, we can increase awareness and appreciation of their cultural heritage [17]. This can attract the attention of global and local audiences, as well as encourage interest in creative products related to history. Furthermore, teachers in the field of history will gain new material on the history of Majapahit, which is the center of civilization with various extraordinary works of art, literature, and architecture.

In addition, by providing a successful framework for training AI Stable Diffusion to depict historical figures from Majapahit, it will open up opportunities for developing visual representations of other historical figures, especially where there are not many graphical sources available, such as Indonesian history. Most of the evidence exists only in the form of archaeological remains and literary narratives that support the splendor of these figures. This is also true for the Majapahit Kingdom. The results of this research can hopefully be applied to various digital content formats, including games, animated films, digital books, and social media. This will certainly create new opportunities in the creative industry and expand the market for culture-based products.

II. Materials and Methods

To analyze how generative AI can help in visualizing historical Indonesian figures, especially from the Majapahit era, this research proposes a training framework as follows: data collection from literature studies, in-depth interviews with archaeologists, visual interpretation and illustration, dataset creation, generative AI training, and evaluation (Figure 1).

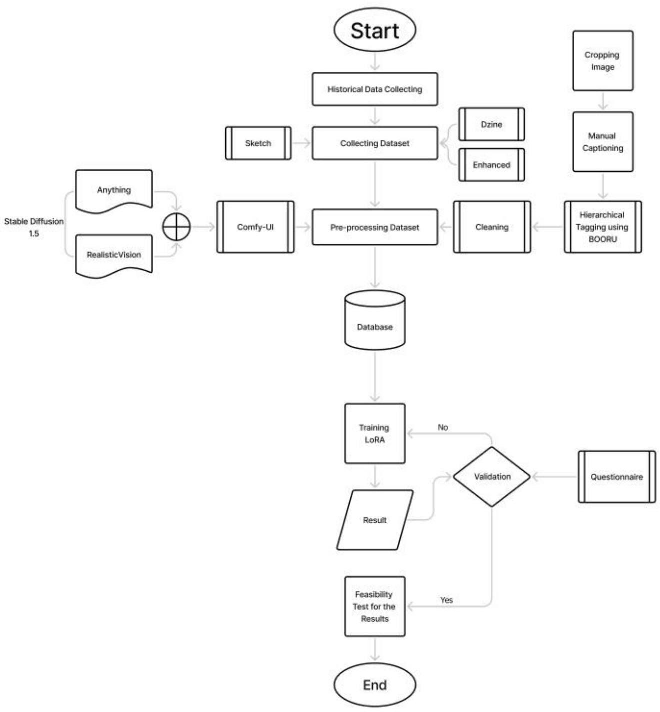

Figure 1. Generative deep learning workflow for visualization of Majapahit-era characters

A. *Literature Study*

As mentioned before, temple reliefs are often speculative and inadequate as the main source of illustration since most of the reliefs tend to portray individuals as gods or goddesses. Most historical records in Indonesia do not depict historical figures. Temple reliefs are made of andesite stone or terracotta, and they do not accurately present the colors and materials presented by the reliefs. Hence, several historical records such as Serat Pararaton, Kakawin Nagarakretagama, which contain information about figure's characters, clothes, and accessories, serve as secondary sources to understand how each historical figure should be depicted. Other sources are used, such as Serat Calon Arang, Ma Huan chronicles, Odorico da Pordenone's record, and other inscription stones.

B. *Visual Interpretation, Illustration, and Dataset Preparation*

Stable diffusion needs images to learn specific accessories, clothing, and other parts of historical figures to be able to generate such images. Given the temple reliefs, statues, and historical records, we created datasets of the king and queen in several poses by interpreting statues and temple reliefs and improved them by using written records (Figure 2.).

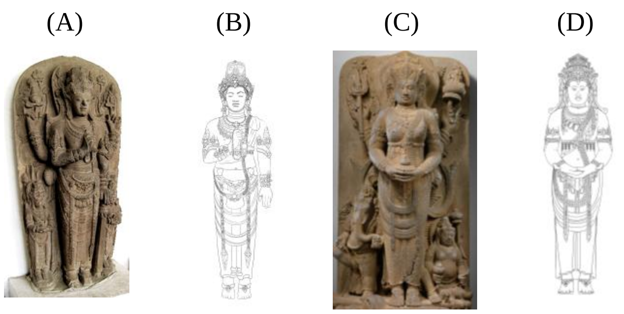

Figure 2. Process of statue interpretation for Early Era of Majapahit Kingdom, (A) Harihara (a combination of Shiva and Vishnu) statue of King Raden Wijaya (Kertarajasa) from Simping temple, Blitar, now in the collection of the Indonesia National Museum. Source: Gunawan Kartapranata CC BY-SA 3.0, (B) The illustration of Raden Wijaya based on the statue, (C) Depiction of Tribhuwana from The Metropolitan Museum of Art. Origin of Bhre Kahuripan Site, Klinterejo Village, Sooko District, Mojokerto Regency. (D) The illustration of Tribhuwana Tunggadewi based on the statue.

The dataset must contain two things. First, the image, and second, the prompt or caption for each image. LoRA needs images to learn specific accessories, clothing, and other parts of historical figures in order to generate such images. Based on statues and historical records, we created several images of kings and queens in several poses by interpreting from the written records and temple reliefs. In the preprocessing, the sketches that originally imitated statues were transformed and developed into several poses such as close up, side view walking, front view walking, sitting cross-legged, sitting on a throne, and lying down (Figure 3. and Figure 4.). The illustrations with different poses were used to enrich the dataset for generative AI machine learning based on the Stable Diffusion 1.5 model with visual style support from LoRA.

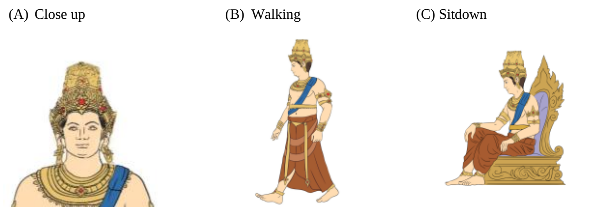

Figure 3. The illustration with different poses to enrich the King dataset

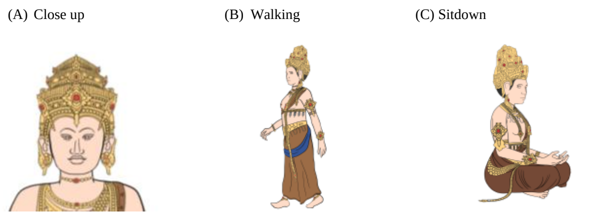

Figure 4. The illustration with different poses to enrich the Queen dataset

In this research, kings and queens will be the main focus because of their distinct clothes and accessories. Furthermore, each illustration of the kings and queens will be drawn with several possible poses to enrich the dataset. To create a realistic depiction of kings and queens' figures, each vector will be fed into image-to-image AI using Dzine.ai to generate a realistic representation of historical Majapahit figures (Figure 5. and Figure 6.). These images are then curated to ensure the accuracy of each figure's attributes and overall appearance (Figure 7.).

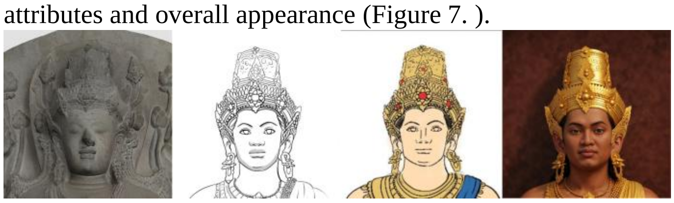

Figure 5. Dataset Preprocessing Process for King Raden Wijaya

Figure 6. Dataset Preprocessing Process for Queen Tribhuwana Tunggadewi

(A) King (B) Queen

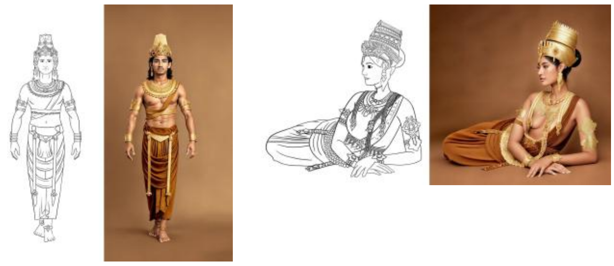

Figure 7. Image-to-image AI using Dzine.ai to generate a realistic representation of the King and Queen Majapahit figures

For the text prompt generation process, we used automatic tagging with WD 1.4 ConvNext Tagger V2, which is implemented in Kohya-ss [18]. The resulting tags were then curated manually using the Booru Dataset Tag Manager. Booru was chosen because it has the advantage of hierarchical prompting, which helps machine learning better understand the structure of image composition (Figure 8.)

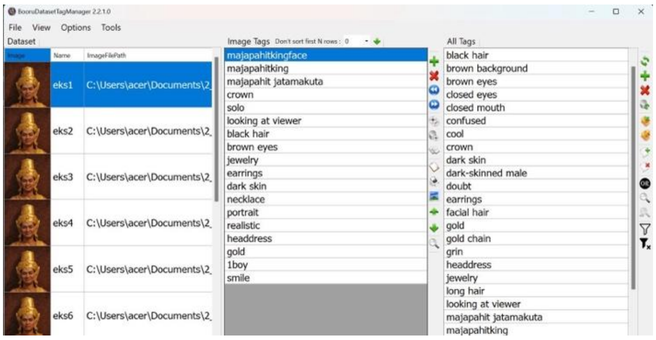

Figure 8. The preview of Booru tag manager used to help organizing caption based on visual hierarchy

C. *Training of Generative AI Model*

To train the Generative AI model, several steps are involved. The training model is conducted using a computer with the following specifications:

1. OS: Linux Mint 22
2. Processor: Intel i5
3. Ram: 16GB
4. VGA: NVidia RTX 3060 Lite Hash VRAM 12GB
5. Library: Kohya-ss v24.1.4
6. Python: 3.10.14

The steps are outlined below:

1) Dataset configuration

The previously created dataset is divided into four sections that are king pose, queen pose, king close-up, and queen close-up. Each image in every section is accompanied by a text prompt in the form of a .txt file, which will be used as a caption.

2) Model training

After the dataset is ready to be used, the next process is to select the base checkpoint as the base model in generative AI. At this stage, only base checkpoints that are of producing faces with the Austronesian race with attributes similar to those used in the Majapahit era are selected. This step is important to increase the likelihood of producing accurate Majapahit character data. The results identified two base checkpoints, namely RealisticVisionV60B1_v51VAE and Anithing_V30Pruned, which were downloaded on an open platform (https://civitai.com/) on July 19, 2024.

The training process is carried out with two pre-selected base checkpoints using the LoRA technique. LoRA is known to have a small file size, which makes it easier to store. In training with LoRA, several hyperparameters are used:

a. Image repetition, which is used to repeat the data multiple times to improve the AI's understanding to detect the prompts provided:
   1. king pose 2
   2. queen pose 2
   3. king close-up 7
   4. queen close-up 7

   For the pose dataset, only two repetitions are used to minimize overfitting as the pose dataset contains the complete attributes of a king and queen. On the other hand, in the close-up dataset, seven repetitions were performed to improve the AI's ability to reconstruct the face in detail.

b. Batch-size: 2

c. Learning-rate: 0.0001

d. LR warmup: 10%

e. Max resolution: 512x512

f. Epoch: 10

g. Optimizer: AdamW8bit

h. Augmentation: None

3) Image test generation

After the training process is complete, each LoRA model checkpoint is saved and tested using the following parameters for preliminary test:

a. LoRA weight: [0.5,0.6,0.7,0.8,0.9,1.0]

b. Seed: 0

c. Clip-skip:
   1. *RealisticVisionV60B1_v51VAE: -1*
   2. *Anithing_V30Pruned: -2*

d. Sampler and scheduler:
   1. *RealisticVisionV60B1_v51VAE: DPM++ SDE, karras*
   2. *Anithing_V30Pruned: Euler, normal*

e. Steps and CFG:
   1. *RealisticVisionV60B1_v51VAE: 25 and 7.5*
   2. *Anithing_V30Pruned: 30 and 4*

f. Prompt:
   1. Queen: majapahitqueenface, 1woman, chubby, javanese, bare shoulder, topless, majapahit royal attire, warm lighting, cosy atmosphere, Instagram style, upper body shot, (cinematic, black and red:0.85), (sunset beautiful background:1.3), sharp, dim colors, (colorful photo with vibrant colors:0.7), (high resolution:1.3), selective focus,

(small:1.3), (poor quality photo with low key lighting:1.5)

2. King: majapahitkingface, 1man, chubby, javanese, bare shoulder, topless, majapahit royal attire, warm lighting, cosy atmosphere, Instagram style, upper body shot, (cinematic, black and red:0.85), (sunset beautiful background:1.3), sharp, dim colors, (colorful photo with vibrant colors:0.7), (high resolution:1.3), selective focus, (small:1.3), (poor quality photo with low key lighting:1.5)

After the training, the LoRA checkpoint is used to generate an image based on prompts with varying weights and epochs (Figure 9. – Figure 12.). The most suitable weight and epoch will serve as the basis for future evaluations.

Figure 9. Kings Anithing_V30Pruned

Figure 10. Queen Anithing_V30Pruned

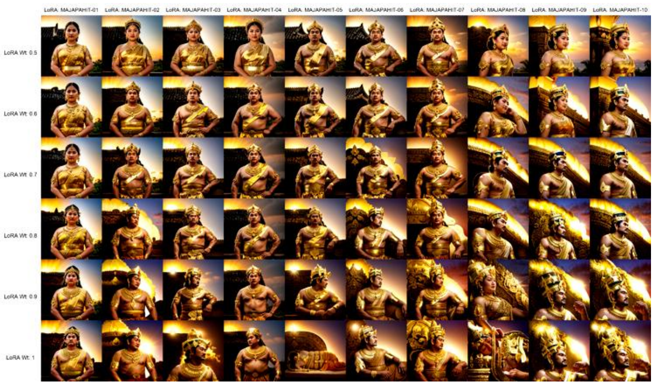

Figure 11. King RealisticVisionV60B1_v51VAE

Figure 12. Queen RealisticVisionV60B1_v51VAE; some images are censored due to showing bare female chest based on visualization of the statue data (see figure 2C)

The results are curated based on the similarity of facial features, body build, and clothing attributes. The results of the preliminary test are shown in Table 1.

Table 1. LoRA Checkpoint dan LoRA Weight Selection

Base Checkpoint	LoRA Checkpoint	LoRA Weight
RealisticVisionV60B1_v51VAE	7	0.7
Anithing_V30Pruned	7	1

The selected LoRA checkpoint and LoRA weight will be used to generate additional images for human evaluation, applying the following parameters:

a. Seed: [1,2,3,4]

b. Clip-skip:
   1. *RealisticVisionV60B1_v51VAE*: -1
   2. *Anithing_V30Pruned*: -2

c. Sampler dan scheduler:
   1. *RealisticVisionV60B1_v51VAE*: DPM++ SDE, karras
   2. *Anithing_V30Pruned*: Euler, normal

d. Steps and CFG:
   1. *RealisticVisionV60B1_v51VAE*: 25 dan 7.5
   2. *Anithing_V30Pruned*: 30 and 4

e. King Prompt:
   1. MajapahitKingFace, 1man, chubby,Javanese, double chin, round face, upper body, majapahit royal attire, sunset in background, in the forest, moody, gloomy, dark picture, stylized, linquivera, liiv1
   2. MajapahitKingFace, 1 man, chubby, Javanese, double chin, round face,majapahit jatamakuta, confused, solo, looking at viewer, upper body, black hair, 1man, brown eyes, jewelry, male focus, earrings, dark skin, necklace, crown, portrait, realistic, gold, smile, 40mm, professional photo, 28mm, realistic
   3. majapahitking Walking in a beach, chubby, MajapahitKingFace, Majapahit, Javanese, double chin, round face, 1man, sunset, full body, barefoot, king solo, long hair, black hair, hat, jewelry, full body, male focus, necklace, bracelet, muscular, facial hair, crown, beard, armlet, topless male, MajapahitKingFace
   4. photo of an MajapahitKing, MajapahitKingFace,

chubby, full body, 1 man, Javanese, double chin, round face, Indonesian, king solo, black hair, jewelry, male focus, barefoot, bracelet, facial hair, crown, armlet, realistic, anklet, sitting on the throne

f. Queen Prompt:

1. majapahitqueenface, 1woman, chubby, Javanese, double chin, round face, bare shoulder, topless, majapahit royal attire, warm lighting, cosy atmosphere, Instagram style, upper body shot, (cinematic, black and red:0.85), (sunset beautiful background:1.3), sharp, dim colors, (colorful photo with vibrant colors:0.7), (high resolution:1.3), selective focus, (small:1.3), (poor quality photo with low key lighting:1.5)
2. majapahitqueenface, 1woman, chubby, Javanese, double chin, round face, dress, bare shoulders, topless, jewelry, confused, solo, looking at viewer, black hair, 1woman, brown eyes, jewelry, earrings, dark skin, necklace, crown, portrait, realistic, gold, smile, deep shadow, dark theme, 40mm, professional photo, 28mm, analog, (National Geographic),
3. majapahitqueen Walking in a beach, chubby, head out of frame, full body, 1 woman, Javanese, double chin, round face, sunset, barefoot, queen solo, long hair, black hair, hat, bare shoulders, topless, jewelry, necklace, bracelet, muscular, facial hair, crown, armlet, topless female, majapahitqueenface
4. photo of an Majapahitqueen, chubby, head out of frame, full body, 1 woman, Javanese, Indonesian, queen solo, black hair, hat, jewelry, barefoot, bracelet, facial hair, crown, armlet, realistic, anklet, sitting on the throne

4) *Evaluation by a human evaluator*

The 16 images of kings and 16 images of queens from each base checkpoint are used for collecting survey responses from 180 students majoring in Indonesian history and culture. The respondent is tasked with selecting from two images from each base checkpoint that have having higher similarity to the sketch (Figure 13.). These results are used to figure out which base checkpoint will produce the highest similarity to the sketch.

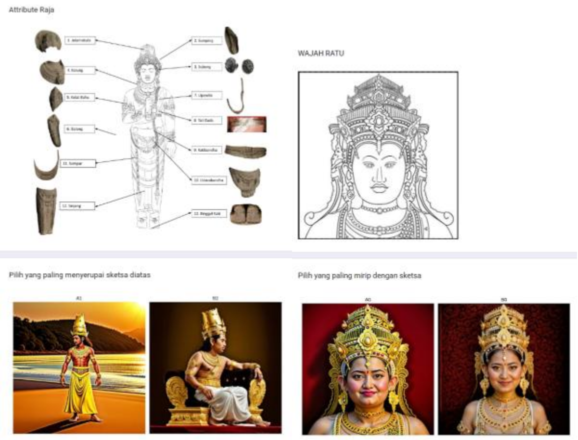

Figure 13. Example of image comparison for human-based evaluation

5) *Evaluation by CLIPScore*

CLIPScore is a widely used evaluation method for Generative AI that evaluates an image's similarity to its accompanying prompt or caption [19]. One of the advantages of CLIPScore is that it can be used reference-free, meaning no candidate caption is required. In the field of Generative AI, this method can determine whether the generated image is similar enough to the provided prompt or caption. The CLIP model calculates the cosine similarity between an image embedding *c* and text embedding *v*. The cosine similarity is then scaled by using a learned parameter τ (often called temperature). Once trained, the parameter τ = 94.80423736572266

$$CLIPScore(c, v) = \frac{\cos(c, v)}{\tau} \tag{1}$$

To evaluate the generated image, especially for Majapahit figures, we must first train the CLIP model so that it understands the nuances of Majapahit figures and what might be present in a sample image.

As with any machine learning model, it requires datasets for training. In this section, we use the dataset we created earlier by interpreting statutes and reliefs, along with their prompts or captions. The CLIP model was then trained using PyTorch Lightning [20]. To train the CLIP model, the hyperparameters are as follows:

a. Base model = openai/clip-vit-base-patch32 [21]
b. Learning rate = 0.0001
c. Maximum caption length = 77
d. Optimizer = AdamW
e. Epoch = 50
f. Batch size = 2
g. Shuffle = True

After the training was completed, we tested the CLIP model to calculate the CLIPScore using the training data. As a result, the CLIPScore for the training data achieved an average of 22.08, as depicted in Figure 14. . This score will serve as the baseline for benchmarking the generated images produced by the Stable Diffusion models.

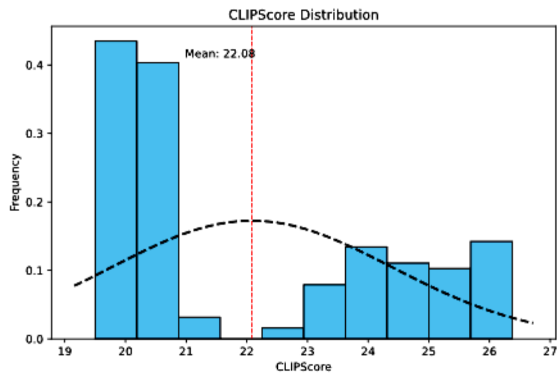

Figure 14. CLIPScore of finetuning the CLIP model using training data

6) *Evaluation by VQAScore*

The VQA Score is a method used to evaluate the quality of a generated image by AI by assessing the content of the image about a given prompt [22]. This method evaluates the result by using a Visual Question Answering (VQA) model, which processes an image and answers the question "Does this figure show '{prompt}' ? Please answer yes or no". The output of this model is a probability that the VQA model answers yes, as described in equation (2).

$$P(\text{"yes"} \mid image, question) \tag{2}$$

The VQA Score model itself is based on the CLIP-FlanT5xl variant [22]. CLIP is used for tokenizing the image input, which will be processed together with the question tokens. Once the tokenization process is complete, all tokens will be processed with the FlanT5 model [23]. The complete diagram on how the VQA Score is calculated is depicted in Figure 15.

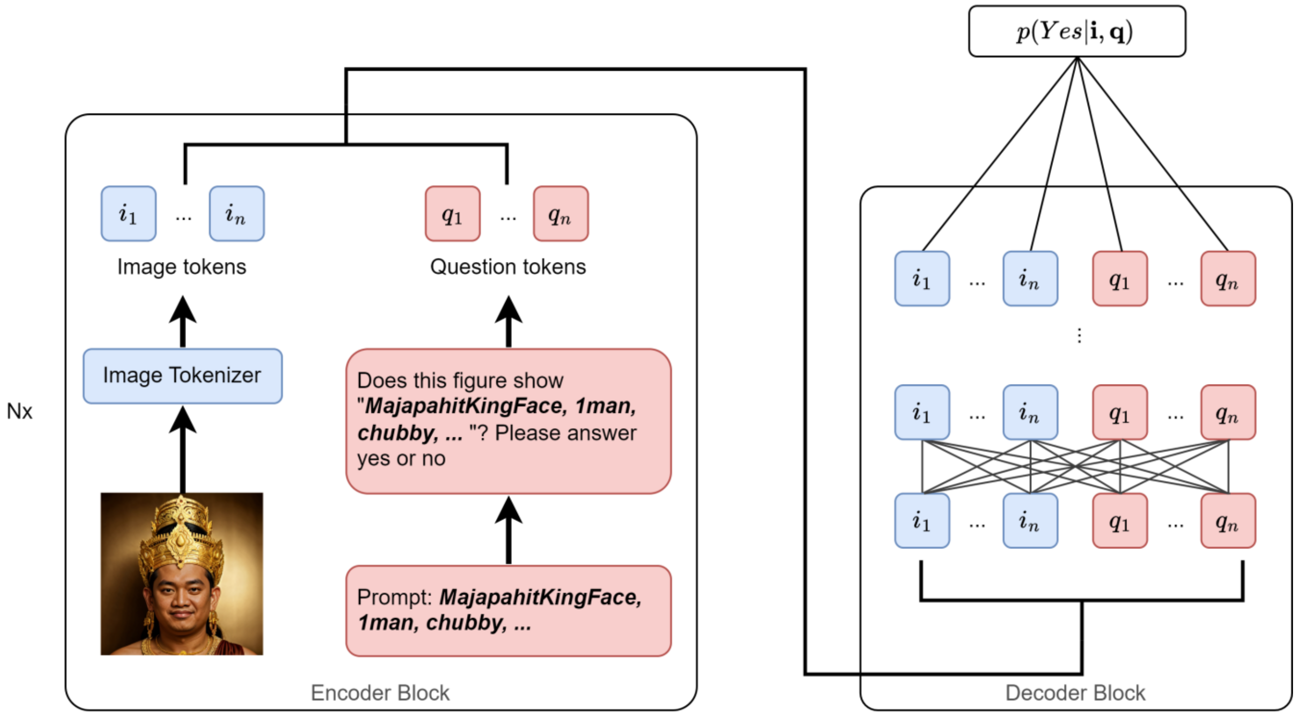

Figure 15. CLIP-FlanT5xl model for calculating VQA Score. The model consists of an encoder block, which takes in an image and a question, and a bidirectional decoder block, which will calculate the probability.

III. Results and Discussion

A. Generated Majapahit Figures by Human Evaluator

The comparison results from 180 respondents regarding the facial similarities of the king and queen, as well as their attributes, are presented in the form of heatmaps. Figure 16. visualizes the facial comparison results, and Figure 17. Illustrates the clothing attributes of the king and queen. Each row in the heatmap represents a compared image, and each column represents a comparison image. The color scale on the heatmap is used to illustrate the intensity of respondents' preferences, where higher values indicate that more respondents believe the comparison image (in the column) is more similar to the original sketch than the image being compared (in the row). In other words, the brighter or more intense the color on the scale, the greater the agreement between the comparison image and the sketch, according to the respondent's assessment.

The comparison data are normalised by dividing the sum of the entire matrix and summing each row. The resulting numbers represent the probability of each image in the row being selected as the one with higher similarity to the sketch compared to the image generated by the other base checkpoint. The results for each image are shown in Figure 18. and Figure 19. To understand the comparison results more deeply, the data obtained from respondents were normalized by dividing each matrix element by the total number of all elements in the matrix. This normalization process is important to remove the absolute scale of respondents' preferences, allowing us to focus on relative comparisons between images objectively. After normalization, the results are accumulated for each row in the matrix, which represents each image tested.

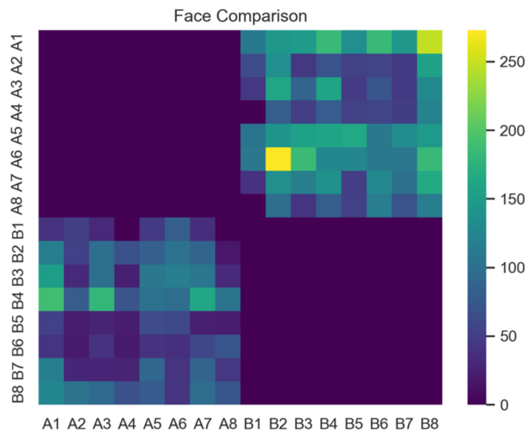

Figure 16. Comparison of face images. A(i) generated results from Anithing_V30Pruned, and B(i) generated results from RealisticVisionV60B1_v51VAE. This chart shows the probability that the Y-axis will be chosen as more similar compared to the X-axis.

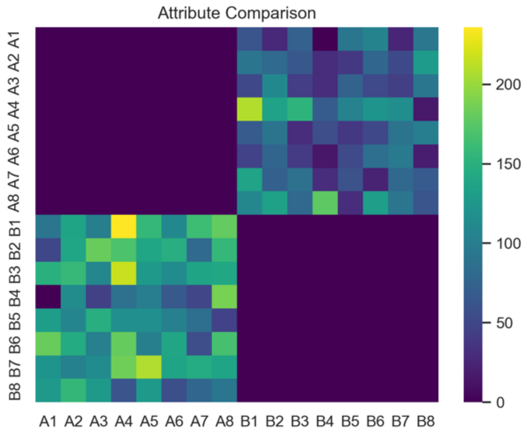

Figure 17. Comparison of attribute images. A(i) generated results from Anithing_V30Pruned, and B(i) generated results from RealisticVisionV60B1_v51VAE. This chart shows the probability that the Y-axis will be chosen as more similar compared to the X-axis.

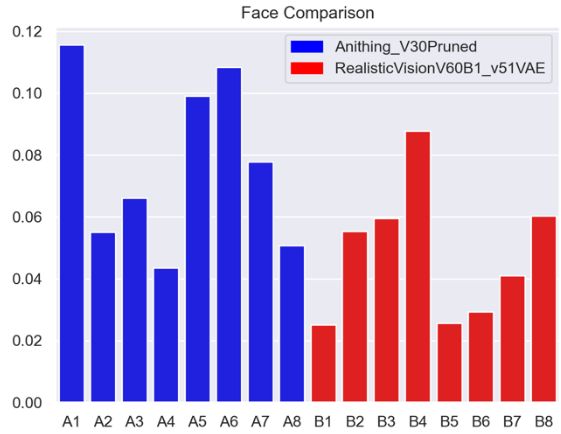

Figure 18. Probability of facial image similarity with the provided sketch

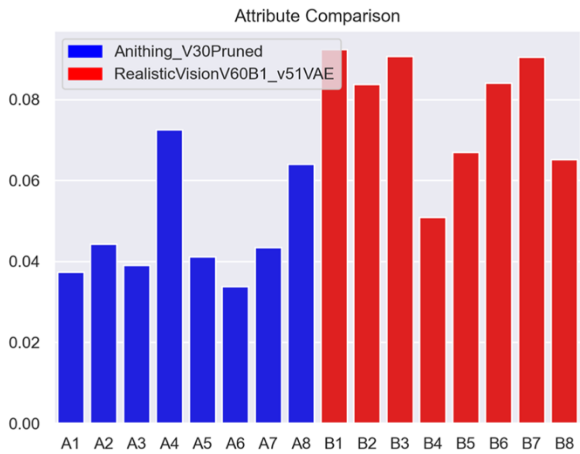

Figure 19. Probability of attribute image similarity with the provided sketch

The result of this process is the probability that an image in a particular row has a higher level of similarity to the sketch compared to images produced by other base model checkpoints (those in columns). This probability describes how big a chance it is that the image in the row is considered more appropriate to the sketch by the respondents compared to the image in the column. In the context of visualizing Majapahit figures, these probability results provide valuable insights into how effective the model is at producing images that align with respondents' expectations based on the reference sketches. Figure 18. illustrates the probabilities for images of the faces of kings and queens, while Figure 19. Presents the probabilities related to the clothing attributes worn by these figures.

This statistical approach enables a more objective analysis of the comparison results between images. The resulting probabilities not only indicate which images respondents select more frequently but also provide in-depth insights into the distribution of preferences. These insights help identify which checkpoints perform better at generating certain elements, such as facial features or clothing attributes, based on qualitative evaluations from respondents. Using this approach, the effectiveness of various generative models, such as Anithing_V30Pruned and RealisticVisionB1_v51VAE, can be quantitatively evaluated, allowing for a systematic comparison of their performance in visually reconstructing historical characters based on calibrated validation data.

Based on the results presented in Figure 18 and Figure 19, Anithing_V30Pruned has 61.62% probability of generating facial images of kings and queens that are closer to the sketch. On the other hand, RealisticVisionB1_v51VAE has a 62.44% probability of generating the correct clothing attributes. According to Figure 18, the results reflect that the model has been able to represent the historical background, texture, and visual features relevant to faces, although some degree of inaccuracy remains. These results show that in the domain of facial recognition, Majapahit era characters, Anithing_V30Pruned succeeded in capturing several key features such as facial shape, bone structure, and general expression.

In contrast, the RealisticVisionB1_v51VAE model shows a slightly higher probability (62.44%) of generating correct clothing attributes. These results indicate that this model has a stronger representation of the visual features associated with Majapahit traditional clothing, which may be more complex and richer in detail. The VAE (Variational Autoencoder) technique used in this model allows the model to explore variations in clothing attributes, such as texture, pattern, and embellishment, more flexibly compared to conventional models [24]. This performance also highlights the importance of explicit modelling of clothing attributes, as the attire of kings and queens in the Majapahit era had very specific characteristics, including ornamentation and symbolism tied to social status. This shows that the RealisticVisionB1_v51VAE model is more adept to the complexity of historical and symbolic visual attributes.

Furthermore, the model's ability to generate similar faces and attributes is strongly influenced by the training data used and the generalization capabilities of the applied LoRA technique [25]. LoRA enables fine-tuning without requiring major modifications to the neural network [26]. In this case, LoRA appears to have successfully improved key facial features and attributes based on the limited training data. This approach has demonstrated its capacity to improve model generalization without requiring very large computational resources. However, achieving greater precision in generative design will require further testing, especially to enhance the resolution and accuracy of specific attributes related to culture and history.

B. Generated Majapahit Figures Performance by CLIPScore

After completing the training steps for the CLIP model described above, the trained CLIP model can now be used to calculate the embedding of the generated images and their corresponding prompts. Using both embeddings, the CLIPScore is calculated using equation (1) for the results produced by Stable Diffusion with the base models Anithing and RealisticVision. The distribution of CLIPScore for the base models Anithing and RealisticVision, in generating close-up facial images with the given prompts, is shown in Figure 20. and Figure 21. respectively.

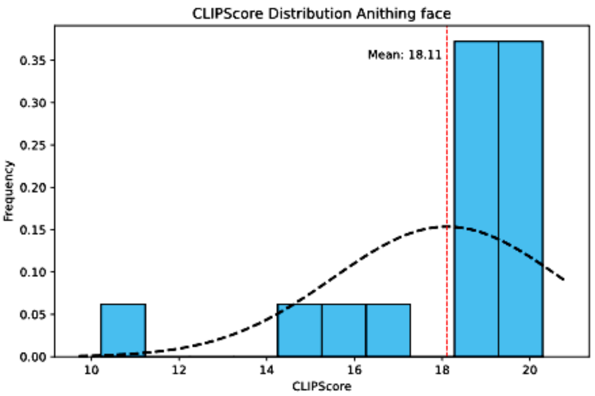

Figure 20. CLIPScore distribution for Anithing_V30Pruned in generating facial images

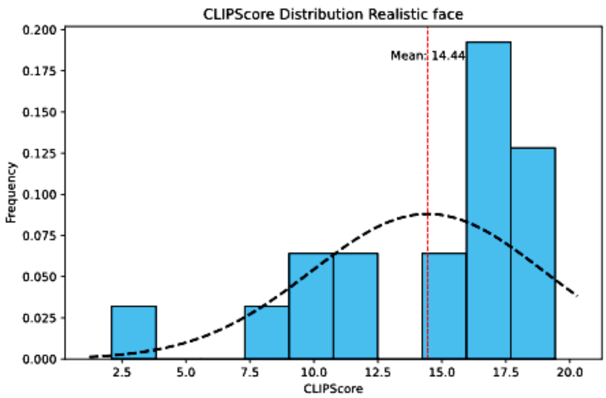

Figure 21. CLIPScore distribution for RealisticVisionB1_v51VAE in generating facial images

As shown above, Anithing_V30Pruned can achieve CLIPScore with an average of 18.11, whereas RealisticVisionB1_v51VAE achieves an average of 14.44. Furthermore, the lowest CLIPScore for Anithing_V30Pruned is 10.220, while the lowest CLIPScore for RealisticVisionB1_v51VAE is 2.099. This clearly indicates that Anithing_V30Pruned model surpasses RealisticVisionB1_v51VAE in its ability to generate images that are closer to the given prompt, specifically for facial images.

In the generation of clothing attributes, similar trends in the results can be observed. Anithing_V30Pruned achieves a CLIPScore with an average of 19.45 and a lowest CLIPScore of 15.120, whereas RealisticVisionB1_v51VAE achieves a CLIPScore with an average of 15.67 and a lowest CLIPScore of 1.563. The distribution of CLIPscore for both models is depicted in Figure 22. and Figure 23. respectively.

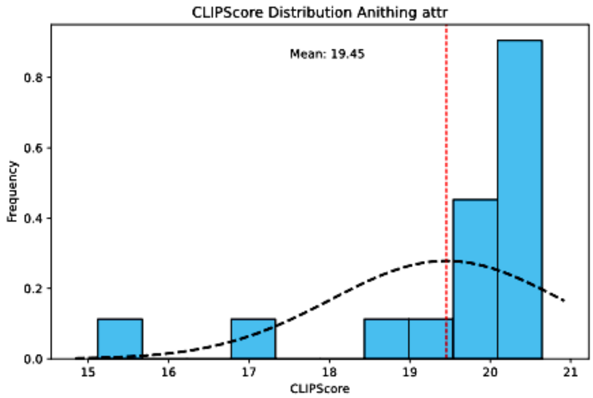

Figure 22. CLIPScore distribution for Anithing_V30Pruned in generating clothing attributes

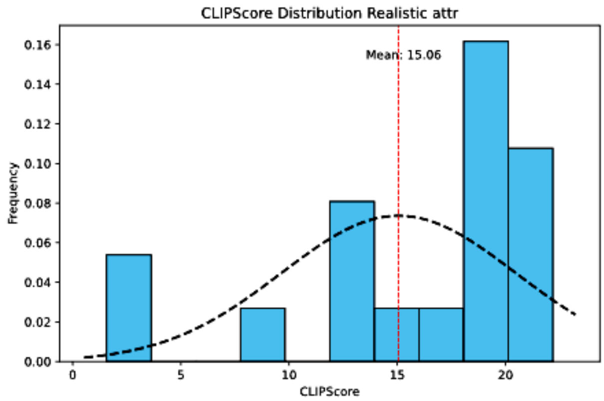

Figure 23. CLIPScore distribution for RealisticVisionB1_v51VAE in generating clothing attributes

C. Analysis of Human Evaluator with CLIPScore

As shown above, the human evaluator has a similar conclusion for facial images, with Anithing_V30Pruned outperforming the RealisticVisionB1_v51VAE model. In contrast, for clothing attributes, the human evaluator preferred RealisticVisionB1_v51VAE over Anithing_V30Pruned, while CLIPScore results indicated the opposite. It is suspected that it is likely due to the quality of the generated images, as illustrated in Figure 24. .

The human evaluator prefers higher-quality images over lower-quality ones even though the validity of the clothing attributes is better demonstrated in Anithing_V30Pruned results. Furthermore, since CLIPScore evaluates on the entire image, the generated environment is also taken into consideration. On the other hand, most of the facial images only use plain backgrounds, as shown in Figure 25. meaning that only the figure itself is evaluated by the CLIPScore.

Figure 24. The example of comparison between the models in showing the clothing attribute. (Left: RealisticVisionB1_v51VAE. Right: Anithing_V30Pruned)

Figure 25. The example of comparison between the models in showing facial. (Left: Anithing_V30Pruned. Right: RealisticVisionB1_v51VAE)

D. *Generated Majapahit Figures Performance by VQAScore*

Using the same images and prompt for CLIPScore, we also calculate the VQAScore for each prompt and figure. The comparison is shown in Figure 26. By calculating the average score, we conclude that RealisticVisionB1_v51VAE is more favorable for images depicting clothing attributes, whereas Anithing_V30Pruned is more favorable for portraying close-up facial images, as shown in Table 2.

Table 2. VQA Score for every attribute and facial images based on the figures and model

Visualization	*Anithing_V30Pruned*	*RealisticVisionB1*
ATTRIBUTES	**0.657507597**	**0.732824335**
King	0.718964413	0.78050645
Queen	0.59605078	0.673221692
FACE	**0.606097266**	**0.547666937**
King	0.630986571	0.550080481
Queen	0.581207961	0.544650007

E. *Analysis of Human Evaluator with VQA Score*

As shown above, the human evaluator reached a similar conclusion for facial images, with Anithing_V30Pruned performing better than the RealisticVisionB1_v51VAE model, and RealisticVisionB1_v51VAE performs better than Anithing_V30Pruned in portraying clothing attributes. Although the FlanT5_xl model is not fine-tuned to understand Majapahit clothing attributes and facial features, other tokens might have been depicted more accurately by each model, respectively. As a result, the scores achieved are only 60%-70% similar to the given prompt.

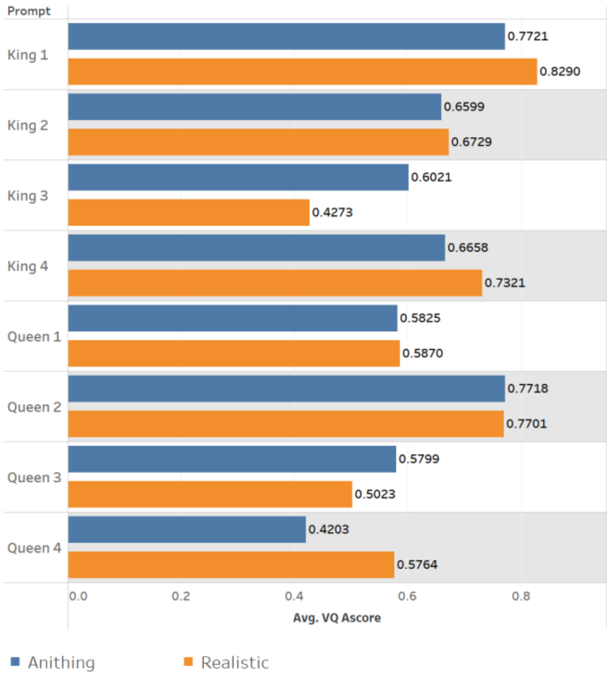

Figure 26. VQAScore for every figure and prompt averaged over 5 generated images

F. *Majapahit Figures Characterization*

The attributes of the visualized figure represent the visual embodiment of the dharma of a King or Queen, where dharma refers to the visual of a particular deity according to belief, whether Vaishnava or Chaiva [27]. Beyond that, based on the visual culture of the time, these attributes not only reflect divine characteristics but also represent the attire worn by the King or Queen during their lifetime, or at least their physical appearance. The explanation of the attributes and character traits of King and Queen figures approximated by generative AI results is based on the accuracy of the representation of historical iconography and artifacts. The model successfully reconstructed clothing details and social attributes, showing that AI can learn visual patterns from documented Majapahit artifacts. However, these results require further validation by historians to ensure their accuracy. LoRA and other AI models provide great opportunities for precise digital reconstruction but still require in-depth verification to be truly historically accurate.

G. *Potential Sector of Developing Historical Figures*

Research on the figure visualization of the Majapahit era using generative AI would contribute to various sectors, including the economy, education, and socio-culture, which strengthens appreciation of cultural heritage and opens up opportunities for digital innovation. AI-based visualization results can strengthen historical tourism by providing interactive experiences at Majapahit sites, increasing the number of visitors, and creating job opportunities [28]. In addition to that, such visualization can be integrated into creative products such as fashion and digital art, increasing sales value and introducing Majapahit culture to a global audience through digital platforms [29].

Furthermore, this research can enrich the history curriculum by presenting more accurate and interesting visual reconstructions. These visual reconstructions also support archaeological research, helping in the digital restoration of sites and artifacts to provide deeper insight into Majapahit life. Additionally, the visualization of these historical figures strengthens national identity by reintroducing the glory of Majapahit. The results of the research can be the basis for historical site conservation policies, supporting education and cultural appreciation in society [30]. Moreover, these discoveries contribute to the reconstruction of the daily life and social structure of the Majapahit era. Anthropological research can compare Majapahit culture with both contemporary societies and societies outside Indonesia, providing a broader perspective on social and cultural developments [31]. Understanding and appreciating the history of Majapahit can strengthen a sense of national pride and identity as Indonesians. As an integral part of Indonesia's cultural heritage, learning about Majapahit's achievements and contributions helps Indonesians feel more connected to their history and fosters pride in their national identity.

IV. Conclusion

This study focuses on modeling a generative artificial intelligence system to visualize figures from the Majapahit era using the Stable Diffusion approach with the LoRA technique. The aims are to reconstruct the facial images and attributes of kings and queens based on historical references. Based on the evaluation results from 180 respondents, it was found that the Anithing_V30Pruned model has a probability of 61.62% to produce the faces of the king and queen that are closer to the original sketch, while the RealisticVisionB1_v51VAE model shows a probability of 62.44% to reconstruct clothing attributes with better accuracy. Evaluation with CLIPScore, Anithing_V30Pruned reached 18.78 and RealisticVisionB1_v51VAE reached 15.06. The overall result, CLIPScore, is contrary to human evaluation. The difference here is caused by the quality of the resulting image being higher than the truth of the clothing attributes. In contrast to the VQA results, in the evaluation with the VQA approach, the results obtained are in line with the human evaluation test. The VQA test on the visualization of the king and queen's face got the highest score with 0.606 in the Anithing_V30Pruned model and 0.7328 for the visualization of the king and queen's attributes in the RealisticVisionB1_v51VAE model. Ultimately, this framework shows that the Stable Diffusion technique with LoRA can provide fairly consistent results in representing important elements of historical figures, both in terms of physical and cultural attributes. The generative model is not only able to produce aesthetically relevant visual images, but also historically appropriate to visual references taken from artifacts and iconographic evidence. In addition, normalization and probability analysis provide a more objective quantitative framework for evaluating model performance in history-based image reconstruction. This study shows that the application of AI in historical visual reconstruction has great potential to help enrich cultural preservation and historical research. The techniques used in this study can be further developed to produce more accurate images and can be used in various other applications, such as education, tourism, and culture-based creative industries.

Acknowledgment

The authors wish to express gratitude to Raka Surya Kusuma, Kelly Yumico Pie, Jocelin Anabel Forisa, Rachel Valentina Selima, Vancouver Gabe Simatupang, Enrique Valentino Akira, and Gde Bagus Andhika Wicaksana for their assistance.

References

[1] B. H. M. Vlekke, *Nusantara: Sejarah Indonesia*. Kepustakaan Populer Gramedia, 2008.

[2] D. Maulana Prasetya, S. Ladiqi, and M. A. Salleh, "Back to the past: The roots of Indonesian middlepowermanship," *Asian Journal of Comparative Politics*, p. 20578911241236518, 2024.

[3] A. A. Munandar, "Majapahit and The Contemporary Kingdoms: Interactions and Views: Majapahit dan Negeri-Negeri Sezaman: Interaksi Dan Pandangan," *Berkala Arkeologi*, vol. 40, no. 1, pp. 1-22, 2020.

[4] M. Wood, "Heroes and Golden Ages—The Hindu-Buddhist, Islamic and Colonial Past Remembered," in *Official History in Modern Indonesia*: Brill, 2005, pp. 33-82.

[5] J. Ras, "Sacral Kingship in Java," in *Fruits of Inspiration*: Brill, 2001, pp. 373-388.

[6] A. Akbar, "Cultural resource management for a Majapahit Kingdom site in Trowulan, East Java, Indonesia," *Conservation and Management of Archaeological Sites*, vol. 16, no. 4, pp. 297-307, 2014.

[7] S. O. Robson, "Java at the crossroads: aspects of Javanese cultural history in the 14th and 15th centuries," *Bijdragen tot de taal-, land-en volkenkunde*, no. 2/3de Afl, pp. 259-292, 1981.

[8] D. Fatma, "Tinjauan visual pada terakota koleksi museum majapahit, Trowulan, Kabupaten Mojokerto," State University of Surabaya, 2014.

[9] W. Muthi'ah, A. Sachari, P. Setiawan, and A. H. Destiarman, "Komparasi Busana Arca Harihara Era Majapahit: Koleksi Museum Nasional Jakarta, Pusat Informasi Majapahit Trowulan, dan Museum Anjuk Ladang Nganjuk," *HASTAGINA: JURNAL KRIYA DAN INDUSTRI KREATIF*, vol. 1, no. 01, pp. 18-25, 2021.

[10] W. Muthiah and A. Sachari, "Motif Kalamakara Pada Temuan Perhiasan Emas Era Majapahit Koleksi Museum Nasional Jakarta," *PROSIDING: SENI, TEKNOLOGI, DAN MASYARAKAT*, vol. 3, pp. 36-41, 2020.

[11] A. Winaya and A. A. Munandar, "Ancient Javanese Women during the Majapahit period (14th–15th centuries CE): An Iconographic Study based on the Temple Reliefs| Perempuan Jawa Kuno periode Majapahit (Abad ke-14–15 Masehi): Suatu tinjauan Ikonografi terhadap Relief Candi," *SPAFA Journal*, vol. 5, 2021.

[12] N. M. Permata, A. A. Munandar, and E. G. Harazaki, "Makna Atribut Istimewa pada Arca Perwujudan Tokoh-tokoh Perempuan Majapahit: Kajian Arkeologi Religi Hindu-Buddha," *Heritage of Nusantara: International Journal of Religious Literature and Heritage*, vol. 11, no. 2, pp. 136-170, 2022.

[13] S. Rostianingsih, "Designing Computer Simulation Game for Learning the History of the Kingdom of Majapahit," Petra Christian University, 2014.

[14] J. Zhang, Y. Huang, Z. Li, Y. Li, Z. Yu, and M. Li, "Development of a method for commercial style transfer of historical architectural façades based on stable diffusion models," *Journal of Imaging*, vol. 10, no. 7, p. 165, 2024.

[15] A. Alfalasi, E. Khan, M. Alhashmi, R. Aldweik, and D. Svetinovic, "SDXL Finetuned with LoRA for Coloring Therapy: Generating Graphic Templates Inspired by United Arab Emirates Culture," *arXiv preprint arXiv:2403.05562*, 2024.

[16] M. Poleski, "Generative models based on diffusion," 2024.

[17] G. Pisoni, N. Díaz-Rodríguez, H. Gijlers, and L. Tonolli, "Human-centered artificial intelligence for designing accessible cultural heritage," *Applied Sciences*, vol. 11, no. 2, p. 870, 2021.

[18] S. Woo *et al.*, "Convnext v2: Co-designing and scaling convnets with masked autoencoders," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 16133-16142.

[19] J. Hessel, A. Holtzman, M. Forbes, R. Le Bras, and Y. Choi, "CLIPScore: A Reference-free Evaluation Metric for Image Captioning," in *EMNLP 2021*, M.-F. Moens, X. Huang, L. Specia, and S. W.-t. Yih, Eds., 2021/11// 2021, Online and

Punta Cana, Dominican Republic: Association for Computational Linguistics, pp. 7514-7528, doi: 10.18653/v1/2021.emnlp-main.595. [Online]. Available: https://aclanthology.org/2021.emnlp-main.595/

[20] *PyTorchLightning/pytorch-lightning: 0.7.6 release*. (2020). Zenodo. Accessed: 2025/02/09/04:50:48. [Online]. Available: https://zenodo.org/records/3828935

[21] A. Radford *et al.*, "Learning Transferable Visual Models From Natural Language Supervision," ed: arXiv, 2021.

[22] Z. Lin *et al.*, "Evaluating Text-to-Visual Generation with Image-to-Text Generation," *arXiv [cs.CV],* 2024 2024. [Online]. Available: http://arxiv.org/abs/2404.01291.

[23] H. W. Chung *et al.*, "Scaling Instruction-Finetuned Language Models," ed: arXiv, 2022.

[24] L. Pinheiro Cinelli, M. Araújo Marins, E. A. Barros da Silva, and S. Lima Netto, "Variational autoencoder," in *Variational Methods for Machine Learning with Applications to Deep Networks*: Springer, 2021, pp. 111-149.

[25] E. J. Hu *et al.*, "Lora: Low-rank adaptation of large language models," *arXiv preprint arXiv:2106.09685,* 2021.

[26] X. Wang, L. Aitchison, and M. Rudolph, "LoRA ensembles for large language model fine-tuning," *arXiv preprint arXiv:2310.00035,* 2023.

[27] K. Hall, "Ritual networks and royal power in Majapahit Java," *Archipel,* vol. 52, no. 1, pp. 95-118, 1996.

[28] N. B. Dhital, L.-C. Wang, H.-H. Yang, C.-H. Lee, W.-H. Shih, and C.-S. Wu, "Effects of the COVID-19 pandemic on public bus occupancy and real-world tailpipe emissions of gaseous pollutants per passenger kilometer traveled," *Sustainable Environment Research,* vol. 32, no. 1, p. 35, 2022.

[29] A. Chollisni, S. Syahrani, A. Shandy, and M. Anas, "The concept of creative economy development-strengthening post COVID-19 pandemic in Indonesia," *Linguistics and Culture Review,* vol. 6, pp. 413-426, 2022.

[30] A. Perkasa and R. Padawangi, "Living Heritage of Ruins? Contesting the Paradox in Trowulan's Majapahit Heritage," *Citizens, Civil Society and Heritage-making in Asia,* p. 61, 2017.

[31] J. Fazlagić and R. Skikiewicz, "Measuring sustainable development-the creative economy perspective," *International Journal of Sustainable Development & World Ecology,* vol. 26, no. 7, pp. 635-645, 2019.

Peter Ardhianto, a member of IAENG in 2024, graduated with a Doctor's degree from Asia University, Taiwan in 2022. Currently, He is a visual communication design lecturer, researcher in generative AI design for local onomatopoeia, historical figures, traditional motifs design, and bio-design. Regarding his expertise, Peter has published his research in several reputable journals.

Yonathan Purbo Santosa, a member of IAENG in 2024, graduated in 2020 with a Master's in Computer Science degree from Bonn University, Germany. At the moment, he is a lecturer and a researcher in the Informatics Engineering Program at Soegijapranata Catholic University. Recently, he has been actively publishing in several reputable journals in generative AI for traditional motif design, poster design, and sustainable architecture design.

Gustav Anandhita is a member of the IAENG in 2024. He completed his Master's degree in Architecture from the Bandung Institute of Technology (ITB) in 2015. Currently, Gustav is actively teaching courses on generative design and computational approaches in architectural design at Soegijapranata Catholic University. Additionally, he is developing a visual media platform based on artificial intelligence on social media, @ainusantara.

Sayid Mataram graduated from the cultural studies doctoral programme from Universitas Sebelas Maret in 2023 with a major in visual art and design focusing on Indonesian comics. Sayid is currently a lecturer, researcher, and practitioner in the field of visual communication design, especially in visual narrative, local onomatopoeia, comic media, art history, and Javanese culture. Regarding his expertise, Sayid has published in several reputable journals, books, and other publications.

Christian Moniaga graduated with a Master of Architecture degree from Soegijapranata Catholic University, Semarang, Indonesia, in 2015. He is a dedicated architecture lecturer, researcher, and practitioner passionate about sustainable design and wellbeing. Christian is expanding his research background in technology through AI and has published several articles.

Gita Claudia completed her Master of Accounting at Soegijapranata Catholic University and Master of Business Administration at Providence University in 2020. Currently, she is a lecturer at the Accounting Department, Soegijapranata Catholic University with a focus on teaching, research, and publication in the field of organizational audit. She also holds a certified public accountant (CPA).

Stefani Lily Indarto, a lecturer, researcher and practitioner in the field of Audit and Assurance who has several certifications including ASEAN CPA, graduated with a Master of Management from Atmajaya University Yogyakarta. Her research focuses on risk-based Audit, including risk assessment related to sustainability issues, such as governance management, as well as socio-cultural, environmental, and regulatory risk impacts. Several reputable research and journal publications, as well as books have been produced.