



Neural Style Transfer and Geometric Transformations for Data Augmentation on Balinese Carving Recognition using MobileNet

I Wayan Agus Surya Darma^{1,2} Nanik Suciati^{1*} Daniel Siahaan¹

¹Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

²STMIK STIKOM Indonesia, Denpasar, Indonesia

* Corresponding author's Email: nanik@if.its.ac.id

Abstract: The preservation of Balinese carving data is a challenge in recognition of Balinese carving. Balinese carvings are a cultural heritage found in traditional buildings in Bali. The collection of Balinese carving images from public images can be a solution for preserving cultural heritage. However, the lousy quality of taking photographs, e.g., skewed shots, can affect the recognition results. Research on the Balinese carving recognition has existed but only recognizes a predetermined image. We proposed a Neural Style Geometric Transformation (NSGT) as a data augmentation technique for Balinese carvings recognition. NSGT is combining Neural Style Transfers and Geometric Transformations for a small dataset solution. This method provides variations in color, lighting, rotation, rescale, zoom, and the size of the training dataset, to improve recognition performance. We use MobileNet as a feature extractor because it has a small number of parameters, which makes it suitable to be applied on mobile devices. Eight scenarios were tested based on image styles and geometric transformations to get the best results. Based on the results, the proposed method can improve accuracy by up to 16.2%.

Keywords: Balinese carving, Data augmentation, Neural style geometric transformations (NSGT), MobileNet, Traditional building.

1. Introduction

Bali is one of the world's best tourist destinations, which becomes the destination for both local and foreign tourists. Bali has power as a cultural, historical, and artistic destination that makes them interested in visiting Bali [1]. The concept of religion is the basis of Balinese Culture regulating human behavior. For Balinese people, any art reflects people's lives to express values that contain beauty, humanity, devotion, and harmony between soul and mind that are widely applied to traditional buildings in Bali [2]. The traditional building of a society in Bali has a *Tri Hita Karana* concept. This concept is a harmonization of life that will achieve if the relationship between humans and humans (*Pawongan*), humans with the environment (*Palemahan*), and humans with God (*Parahyangan*). Therefore, the construction of the traditional Balinese building applies *Tri Hita Karana* concept. One

example is the decoration in the form of carving patterns that adorn traditional Balinese buildings.

Several researchers have carried out several studies on object recognition on cultural objects. In Marathi Text Document recognition, Ramteke [3] proposed A Streamlined OCR System using support vector machine (SVM) and Adaptive Cuckoo Search algorithm. Our previous research on Balinese character recognition uses k-Nearest Neighbor as a classifier produces a pretty good recognition accuracy [4]. Related research on recognition of Balinese carving patterns was carried out by Mahawan [5]. This research using a Histogram of Oriented Gradient (HOG) and Principal Component Analysis (PCA) in feature extraction and Learning Vector Quantization (LVQ) in the recognition process. In the case of traditional cloth image retrieval, texture, and color are the dominant features and visual characteristics that distinguish between one fabric and another. Nasir [6] combined the Local

Binary Pattern (LBP) texture features that are invariant to the rotation with the color features of the HSV color space to reconvene the traditional fabric image. The combination of LBP texture features that are invariant to the rotation with the color features of the HSV color space produces the best 100% recall in the Batik dataset and 100% in the *Songket* dataset using the Manhattan distance. Minarno [7] has conducted research related to comparing the performance of k-Nearest Neighbor (k-NN) and Support Vector Machine (SVM) in the process of classification of Batik images. As for feature extraction using Multi Texton Histogram (MTH). This study uses 300 batik image datasets divided into 50 classes. The experimental results show that the average accuracy of the k-NN classifier is 5.09%, higher than SVM. Besides, the accuracy achieved by using SVM and k-NN is 76% and 82%, respectively. Related research has been carried out by Qin [8] on characters carved in stone by proposing an Entropy-Based Feature Extraction Algorithm (EFEA). This study proposed a combination of conventional image processing with deep learning to extract stone texts in natural scenes. Other studies on object recognition have also been carried out on fish recognition [9-10].

In object recognition, a limited number of datasets becomes a problem that often occurs, especially in recognizing unique objects. Various public datasets are available, for example, COCO and ImageNet, yet they still consist of common images. Hence, it cannot use in exceptional cases e.g., the recognition of cultural heritage objects or unique patterns that only exist in an area. It indeed still a challenge to be able to recognize objects with a limited amount of data. One of the widely used solutions is to provide variations on the original image with a pattern or style; thus, it produces a new image. Many researchers have conducted research to produce new image variations for a limited number of datasets. Related research has been carried out in the application of Neural Style Transfer to improve the classification process in breast MRI images and skin melanoma diagnosis by applied Neural Style Transfer to produce new images that combined original image content with other images [11]. Sheng [12] proposed a Convolutional Neural Network Style Transfer to automatically generates Chinese paintings towards transferring with unique ink and wash characteristics. Zhao [13] applied for Neural Style Transfer with a combination of local and global optimization to reduce artifacts in the resulting image. Related research on Neural Style Transfer has been conducted by Cheng [14] to produce more variations of image-based captcha aimed at increasing security.

This research resulted in good usability captcha based on images and fonts.

Recently, many researchers use deep learning to solve various problems related to object recognition. MobileNet is one of the widely applied models to mobile devices because of the small number of parameters compared to other models. Related research on deep learning has been applied by Fan [15] to quickly identify the various types of rocks suitable for geological survey work; thus, it can improve work efficiency and limit the cost of rock identification. Hiary [16] proposed a two-step deep learning classifier to distinguish the different flower types. The first step is to segment the flower region with a minimum bounding box. It is followed by the second stage by applying CNN in the classification process. Researchers apply deep learning in object recognition in various domains. In the Health domain, researchers have applied deep learning based on medical images [17-19]. In the security domain, deep learning has been applied to palmprint classification [20] and eye feature point detection [21]. Deep learning is also widely applied to text character recognition by combining CNN and RNNs for sequence dynamics [22], and a combination of CNN and strength-driven compression to detect bacterial species [23].

The limited number of data is still a challenge in object recognition. To overcome this problem, many researchers apply data augmentation with various techniques. Geometric Transformation was proposed to overcome insufficient training data to improve recognition performance on Indonesian *batik* using variations in rotation and scaling [24]. Related research has been carried out by applying augmentation techniques by providing rotations with various degrees to provide data variations in Pest localization and recognition [25]. A similar technique has been applied as data augmentation on Optical Remote Sensing Imagery to improve recognition performance [26].

Previous studies have not considered image conditions, e.g., color variation and skewed images that can affect recognition results. Current research uses only predetermined images, not yet considering variations in the taking of objects with diverse positions. The object's position is very influential on the results of recognition, so we need a robust method in recognizing objects with a variety of retrieval variations. Limited Balinese carving data is also a challenge in this study, so we need augmentation techniques to provide variations of data. Based on the discussion above, the availability of Balinese carving data is still a challenge in the recognition of Balinese carving. This is due to the limited building's



Figure. 1 Balinese carving: (a) *karang daun*, (b) *karang goak*, (c) *karang boma*, (d) *karang barong*, and (e) *karang gajah*

availability in Bali, which still uses Balinese carving in the middle of modern buildings. Public image datasets e.g., COCO and ImageNet, only provide general image data; hence it cannot be used in exceptional cases such as the recognition of Balinese Carvings.

In this paper, we proposed the combination of Neural Style Transfer and Geometric Transformations (NSGT) to the Balinese carving dataset to provide data variations on a small dataset. The proposed method provides a new data augmentation technique to create new data variation in rotation, rescale, lighting, and color variations to enrich the dataset so it could improve recognition performance. We used MobileNet architecture in Balinese carvings recognition. We chose the MobileNet architecture because it used a small number of parameters, which makes it suitable to be applied on mobile devices. The proposed method consists of three phases. First, creating a new dataset by combined Neural Style Transfer and Geometric Transformation (NSGT). Second, create a new model based on NSGT. Third, the recognition process uses the NSGT model on Balinese carvings.

This paper is organized as follows. Related work and contribution are shown in Section 1. Section 2 describes the Balinese carving, which consists of five types of *Kekarangan*. Section 3 presents our proposed methodology for the Balinese carving recognition using Neural Style Geometric Transformations (NSGT). The experiment and

results are present in Section 4. Section 5 discusses the analysis of the results. Section 6 is about the conclusion and the future work of this research.

2. Balinese carving

The carvings found on each building element, have various engraving patterns. The decorative decoration is the main element that adds to the aesthetic value of the building, which is shown through its harmonious, integrated shape, color, texture, and material. In traditional Balinese buildings, the decoration is applied in the form of carvings with specific shapes and patterns. Balinese carving art is one of carving art with various decorative patterns, as outlined in wood and concrete media. *Kekarangan* is a type of Balinese carving that is widely used in traditional buildings in Bali. The five types of *Kekarangan* that are most widely used are *Karang Goak*, *Karang Gajah*, *Karang Boma*, *Karang Daun*, and *Karang Barong*. *Kekarangan* has a unique characteristic that resembles animals, leaves, and creatures found in culture in Bali. For example, *Karang Goak* shown in Fig. 1 (b) resembles a crow's head that is decorated with leaves ornament. *Karang Gajah* combined the shape of an elephant's head with flowers or leaves ornament shown in Fig. 1 (e).

The form of Balinese carving is unique because it carved by different artist. It becomes a challenge in recognizing the Balinese carving. *Karang Barong* and *Karang Boma* have similarities that can affect the

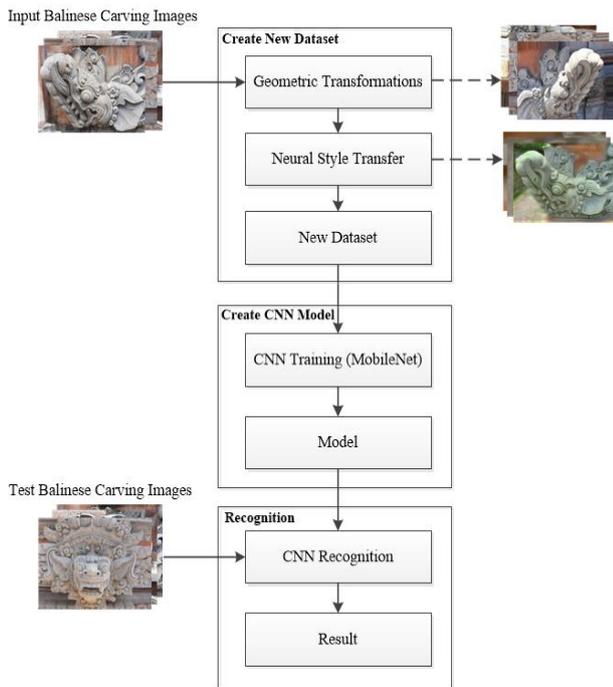


Figure. 2 Neural style geometric transformations

results of the recognition. In model training, it is necessary to vary the data to strengthen the results of the recognition. Balinese carvings also have similar motif, and this is a challenge to be able to recognize the Balinese carving precisely. In this study, we applied NST and Geometric Transformation as data augmentation to improve recognition performance. We used five Balinese image carving taken using a smartphone camera that collected from various traditional buildings in Bali. We collected 109 images of Balinese carvings consisting of five classes. Fig. 2 represents the image of Balinese carvings consisting of five classes.

3. Proposed method

In this study, we collect images of Bali carvings using general smartphone cameras, and we do not use particular types of smartphones. The challenge in this research is the limited number of Balinese carving images. For this reason, we use neural style transfers to produce variations of Balinese carving images and transfer learning to generate a new model. We proposed a Neural Style Geometric Transformations (NSGT) method combining Neural Style Transfer and Geometric Transformations for data augmentation. This method generates a variation for a small dataset to produce a higher recognition performance. The addition of this variation is also to overcome the Balinese carving image's poor quality in terms of tilt angle and image position. Fig. 2 shows the proposed recognition method.

Table 1. MobileNet network architecture

Type/Stride	Filter Shape	Input Size	
Conv/s2	3x3x3x32	224x224x3	
Conv dw/s1	3x3x32 dw	112x112x32	
Conv/s1	1x1x32x64	112x112x32	
Conv dw/s2	3x3x64 dw	112x112x64	
Conv/s1	1x1x64x128	56x56x64	
Conv dw/s1	3x3x128 dw	56x56x128	
Conv/s1	1x1x128x128	56x56x128	
Conv dw/s2	3x3x128 dw	56x56x128	
Conv/s1	1x1x128x256	28x28x128	
Conv dw/s1	3x3x256 dw	28x28x256	
Conv/s1	1x1x256x256	28x28x256	
Conv dw/s2	3x3x256 dw	28x28x256	
Conv/s1	1x1x256x512	14x14x256	
5x	Conv dw/s1	3x3x512 dw	14x14x512
	Conv/s1	1x1x512x512	14x14x512
Conv dw/s2	3x3x512 dw	14x14x512	
Conv/s1	1x1x512x1024	7x7x512	
Conv dw/s2	3x3x1024 dw	7x7x1024	
Conv/s1	1x1x1024x1024	7x7x1024	
Avg Pool/s1	Pool 7x7	7x7x1024	
FC/s1	1024x1000	1x1x1024	
Softmax/s1	Classifier	1x1x1000	

Previous research on the Balinese carvings recognition conducted by Mahawan [5], has not considered the various angles of the image capture position. The image used must be adjusted to a good position before being recognized. So that the results of the recognition only use predetermined Balinese carving images. In this study, we proposed a combination of Neural Style Transfer and Geometric Transformation as data augmentation. Neural Style Transfer (NST) can provide color variations by transferring styles from the style image to the content image to produce a new image [12]. NST can handle color variations and lighting on Balinese carving objects that will be recognized. This variation can produce a new image that can increase the number of datasets. Neural Style Transfer cannot represent objects that vary in rotation and rescale, so it is not effective to recognize objects with these variations.

Geometric Transformation is a data augmentation technique for generating new images with geometric variations, e.g., rotation and rescale [25]. This technique provides variations on the dataset to represent various geometric conditions on the object. Fig. 3 shows the results of applying Geometric Transformation to Balinese carving images. Geometric transformations cannot represent variations in color and lighting on a recognized object.

The proposed method combined Neural Style Transfer and Geometric Transformations as a data augmentation in a Balinese carving dataset. We



Figure. 3 Geometric transformations on balinese carving images

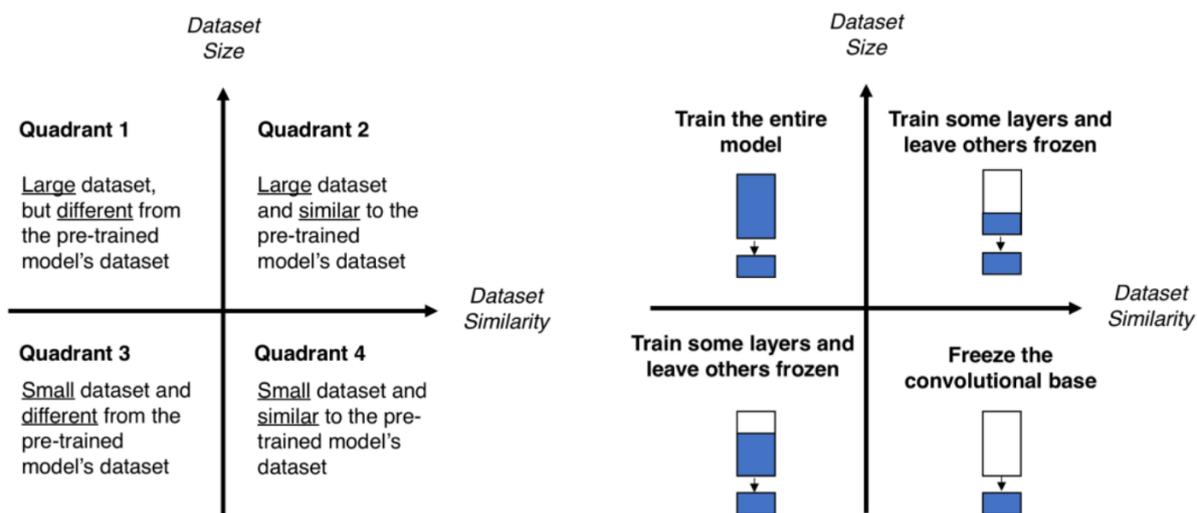


Figure. 4 Dataset size-similarity matrix

combined these two techniques to produce an augmentation technique representing various conditions in color, lighting, rotation, and rescale. NSGT aims to produce new images to increase the amount of data on a small dataset. The proposed method consists of three phases. First, creating a new dataset by combined Neural Style Transfer and Geometric Transformation (NSGT). Second, create a new model based on NSGT. Third, the recognition process uses the NSGT model on Balinese carvings. We chose MobileNet as a feature extractor because it uses very few parameters, so it is suitable for mobile devices. The following are the steps in the proposed method.

3.1 Transfer learning

CNN training requires a long time to be able to form a new model. To speed up convergence, we

initializing the weights using a trained/convergent model. This technic is called transfer learning. In this study, we use the MobileNet architecture, which uses 3,228,864 parameters; hence it is suitable for use on a mobile platform. Compared to other models such as VGG16 with 138,357,544 parameters [27] and Resnet-50 with parameters reaching 23 million, MobileNet is a model with much fewer parameters.

MobileNet is built based on depthwise separable convolutions, which are a form of standardized convolution factorization into depthwise and pointwise convolution. The depthwise convolution applies a single filter to each input channel in MobileNet Architecture. Pointwise convolution then applies 1×1 convolution to combined output with depthwise convolution. The depthwise can be written as:

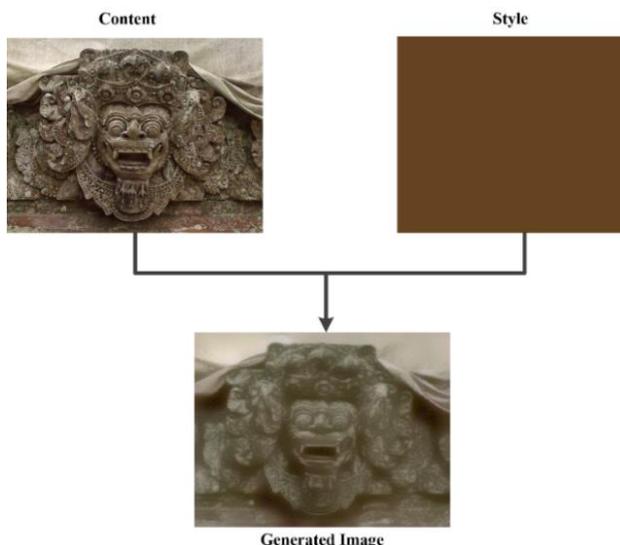


Figure. 5 Neural style transfer

Table 2. Variation of style images in neural style transfer experiment

Number of Styles	Iterations	Generated Images
1 style image	30	1395
2 style images	30	2745
3 style images	30	4095

$$\hat{G}_{k,l,m} = \sum_{i,j} \hat{K}_{i,j,m} \cdot F_{k+i-1,l+j-1,m} \quad (1)$$

Where \hat{K} is kernel depthwise convolution with size $D_K \times D_K \times M$ applied to the m_{th} channel on F to produce m_{th} channel from an output feature map \hat{G} . D_K is the spatial dimension of the kernel assumed to be square and M is number of input channels. Table 1 shows the MobileNet network architecture that is built based on depthwise separable convolutions. Only the first layer uses full convolution. Each layer on the MobileNet network is followed by a batchnorm and nonlinear ReLU except at the final fully connected layer. The Filter Shape column represents the filter size at each convolution layer. The input size column represents the size of the input image processed at each convolution layer on MobileNet. The default input image is 224x224, with three channels. MobileNet Network Architecture is shown in Table 1.

We apply the transfer learning to speed up the process of model training. There are four quadrants based on the size and similarity of the dataset to the pre-trained model [28]. In this study, the data used are Balinese carving images that have no resemblance to the dataset used in the pre-trained model. The number of datasets that we use also affects the quadrant that corresponds to our case. We use small dataset

because of the limited amount of Balinese carving data. Based on the size and similarity of the dataset, we use quadrant 3 in the transfer learning process. Fig. 4 is used to classify the problem based on the Dataset Size-Similarity Matrix. In this study, we applied quadrant 3 to the transfer learning process based on Dataset Size-Similarity Matrix.

The model we use is MobileNet, which has 3,228,864 parameters. MobileNet uses the fewest parameters compared to other models, so it is suitable for mobile devices. In the transfer learning process, we conducted training on 2,145,088 parameters on MobileNet using the dataset that we have. The number of parameters that we did not train was 1,083,776. The training process is based on the Dataset Size-Similarity Matrix, shown in Fig. 4.

3.2 Neural style transfer

The limited number of Balinese carving data is one of the challenges in this study. To overcome this problem, we use neural style transfer to generate new images by combining style images with original images. The image of Balinese carvings we used in this study consisted of 5 classes. Fig. 5 show the image of Balinese carvings combined with styles image to produce new images using neural style transfers.

We use Neural Style Transfer to add variations to the training images that we have collected. We use three image styles: solid yellow, solid brown, and solid blue. This color selection provides color variations based on lighting conditions and Balinese carving media found in traditional buildings. Fig. 6 shows the image of the Neural Style Transfer. Style images provide color variations that are based on lighting conditions and media used in traditional buildings. Color variations in the image of the NST can provide variations on the dataset to improve recognition performance.

In this study, we applied a Neural Style Transfer using VGG19 as a pre-trained model. VGG19 uses three images as input: style image, target image, and placeholder that will contain the generated image. In the application of NST, three-loss functions need attention, content loss function, style loss function, and final loss. The Content loss is used to make the composite image approximate the content image as regards content features by minimizing the difference between the activations of higher layers. The content loss is defined as,

$$L_{content} = \frac{1}{2} \sum_{i,j} (A_{ij}^l(g) - A_{ij}^l(c))^2 \quad (2)$$

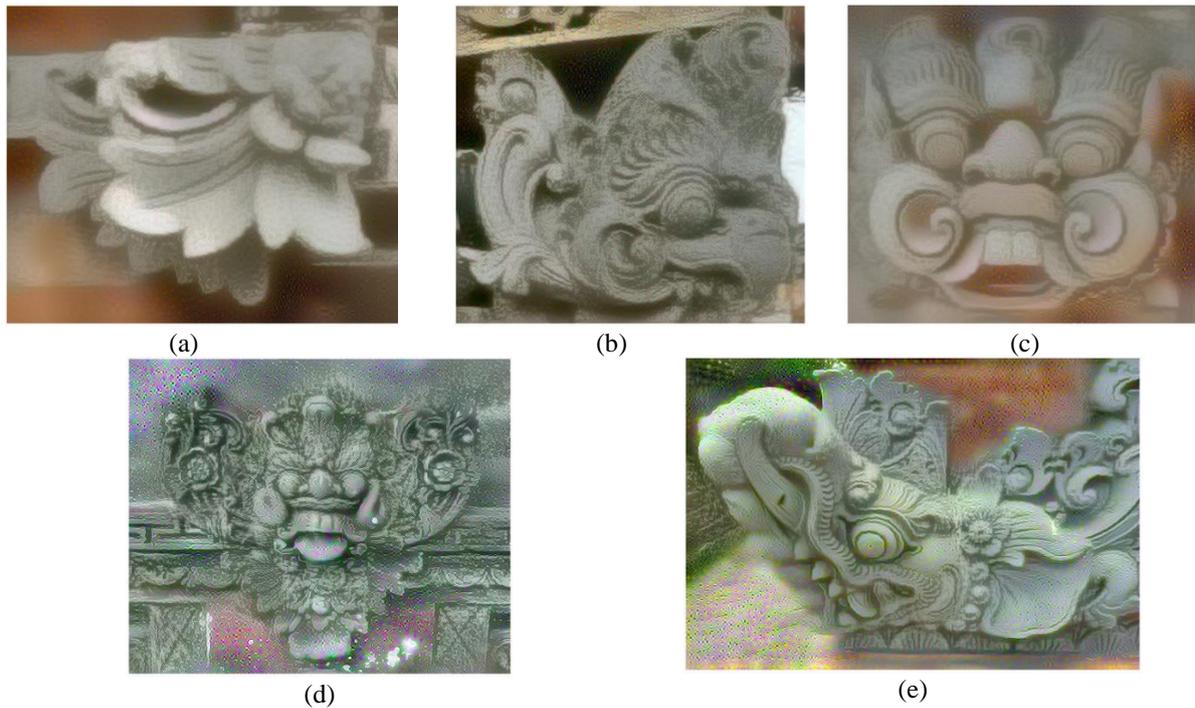


Figure. 6 Neural style transfer result on balinese carving: (a) *karang daun*, (b) *karang goak*, (c) *karang boma*, (d) *karang barong*, and (e) *karang gajah*

Table 3. Scenario details

Scenario	Amount of Data	Augmentation	Neural Style Transfer
1	45	-	-
2	405	7 type of Geometric Transformations	-
3	1.395	-	1 Style Image
4	2.745	-	2 Style Image
5	4.095	-	3 Style Image
6	9.765	7 type of Geometric Transformations	1 Style Image
7	19.215	7 type of Geometric Transformations	2 Style Image
8	28.665	7 type of Geometric Transformations	3 Style Image

Essentially $L_{content}$ captures the root mean squared error between the activations produced by the content image and the generated image. In Eq (2), $A_{ij}^l(I)$ be the activation of the l_{th} layer, i_{th} feature map and j_{th} position obtained using the image I . In higher layers, different feature maps are activated in the presence of different objects, so if two images with the same content will have similar activations in the higher layers.

There are several steps to define style loss. This is using all layers to extract style information from the VGG19 network. The amount of correlation present between feature maps in each layer is used to measure the style information [14]. Style loss is the difference in correlation present between feature maps, which are calculated by the resulting image and the style image. Style loss is defined as,

$$L_{Style} = \sum_l w^l L_{style}^l \text{ where,} \quad (3)$$

$$L_{style}^l = \frac{1}{M^l} \sum_{ij} (G_{ij}^l(s) - G_{ij}^l(g))^2 \text{ where,} \quad (4)$$

$$G_{ij}^l(I) = \sum_k A_{ik}^l(I) A_{jk}^l(I). \quad (5)$$

In Eq. (3), w^l is a weight given to each layer during loss computation. In Eq. (4), M^l is an hyperparameter that depends on the size of the l_{th} layer. Style matrix for the style image and the generated image is computed by minimizing the mean-squared distance between the entries of the gram matrix from the style image and the generated image. The gram matrix is computed using Eq. (5), where $A_{ik}^l(I)$ be the activation of the l_{th} layer, i_{th} feature map and k_{th} position obtained using the image I and $A_{jk}^l(I)$ be the activation of the l_{th} layer, j_{th} feature map and k_{th} position obtained using the image I .

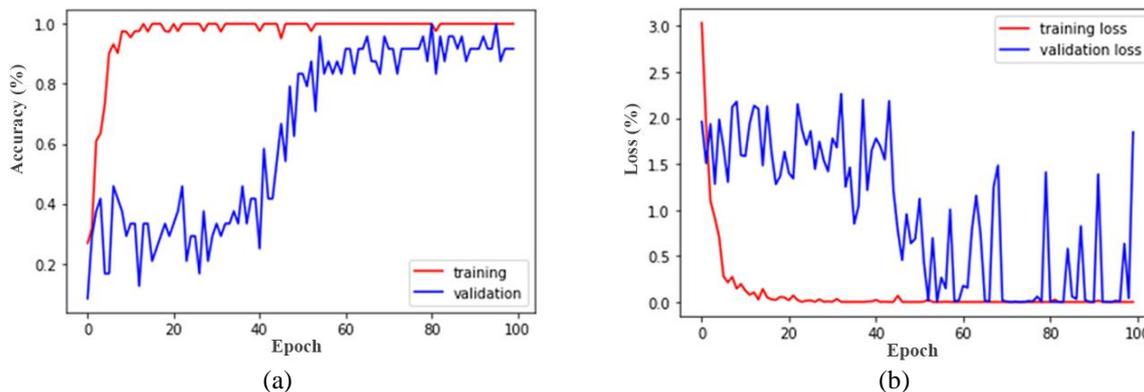


Figure. 7 Graph of models training on original dataset: (a) model accuracy and (b) model loss

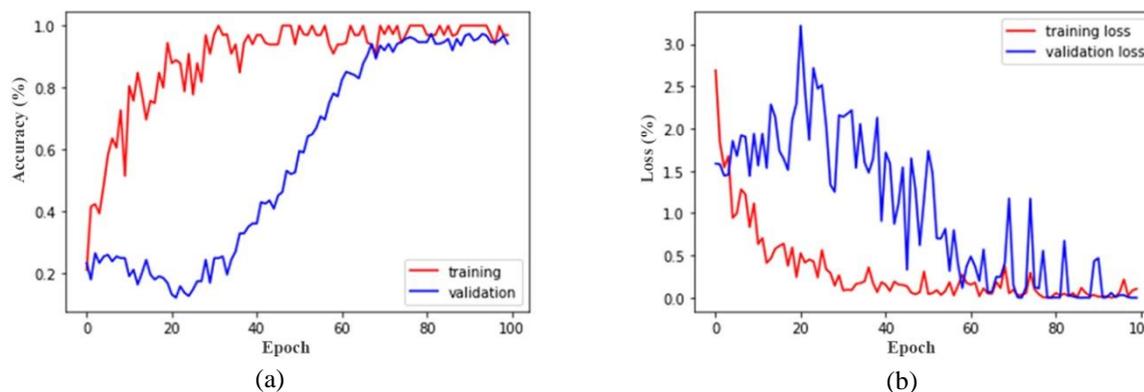


Figure. 8 Graph of models training on dataset with geometric transformations: (a) model accuracy and (b) model loss

To generate images from a combination of image content and style image, the final loss must be minimized by determining α and β to control the amount of content and style added into the generated image. Final loss is defined as,

$$L_{final} = \alpha L_{content} + \beta L_{style} \quad (6)$$

To generate the new image that combined the content and style images in Fig. 2, we use the lost function in Eq. (6) to minimize the noise from the content and style images. Where α is the weighting factor for content reconstruction and β is the weighting factor for style reconstruction.

In this study, we applied for neural style transfer with variations in the number of different style images. The third experiment was conducted to determine the effect of the number of style images used on the model's performance. We applied neural style transfer to 45 training data with 30 iterations in each experiment. In this experiment, we succeeded in producing 1395 training data, which were used as the first NST model, 2745, and 4095 training data in the other two experiments. Table 2 shows the experiments we carried out by applying NST with variations in the number of style images.

3.3 Scenario details

In this paper, we conducted eight test scenarios to produce the best model with the highest recognition accuracy. Refers to Table 3, eight experiments were performed based on the type of augmentation and the number of image styles. The first scenario was conducted using a training dataset without adding augmentation to the dataset. In the first scenario, a limited number of datasets are used to determine the accuracy of the resulting recognition. The second scenario used a dataset that has been added with augmentation, such as rotation, rescale and other augmentation based on Geometric Transformations. In the third, fourth, and fifth scenarios, we add variations to the training dataset using Neural Style Transfer with the different number of style images. In the sixth, seventh and eighth scenarios, we combined the Neural Style Transfer result dataset with a different number of style image and Geometric Transformations. We use an RMSprop optimizer with a learning rate of 1×10^{-5} and 100 epochs for all scenarios.

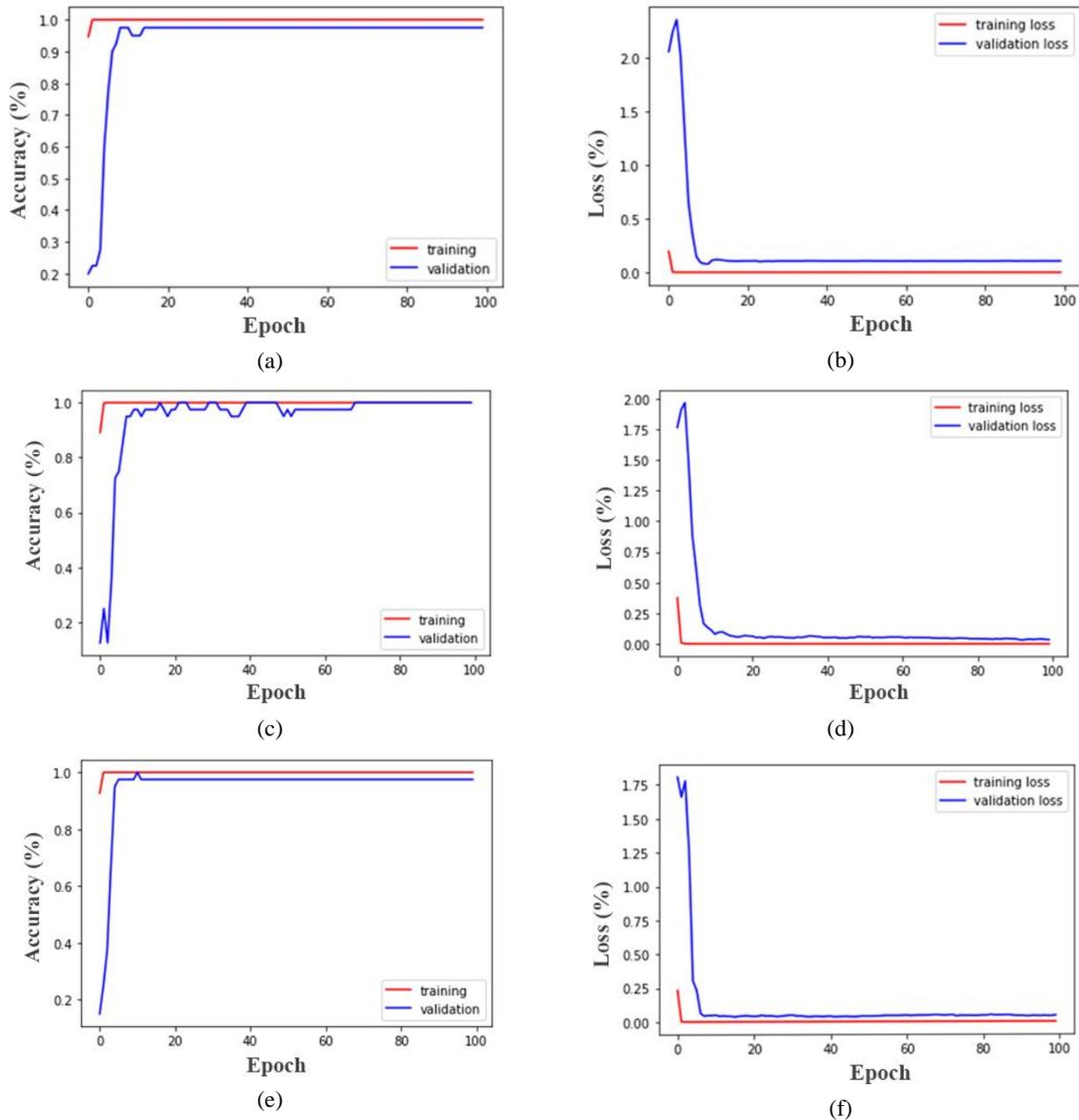


Figure. 9 Graph of models training on NST dataset: (a) 1 style NST model accuracy, (b) 1 style NST model loss, (c) 2 style NST model accuracy, (d) 2 style NST model loss, (e) 3 style NST model accuracy, and (f) 3 style NST model loss

4. Experiment and result discussion

In this section, we conducted eight experiments using MobileNet architecture. The first experiment is using the original dataset with 45 images without augmentation processes. The second experiment is using a dataset that generated using Geometric Transformations. This augmentation consists of seven types of augmentation, which are applied to 45 original data to produce 405 images. The third, fourth, and fifth experiments are using a dataset generated by applying Neural Style Transfer using several numbers of image styles. The sixth, seventh, and eighth experiments using datasets that produced with

the application of Geometric Transformations to the Neural Style Transfer datasets. These experiments use eight types of datasets, as shown in Table 3 Each experiment carried out is as follows.

4.1 Original dataset

In the first scenario, we use the original Balinese carving dataset taken using a smartphone camera. This dataset consists of five classes without the addition of data augmentation with a total of 45 image training data and 40 image validation data. We conducted training on train and validation data of 100 epochs to produce the first model. This model shows

the results of the training accuracy of 94.7% and validation accuracy of 79.1%. Fig. 7 (a) shows a graph of accuracy and validation in the first model training conducted with 100 epochs. The results of the first scenario show the accuracy and loss of validation cannot follow the accuracy and loss of training. Validation accuracy increased at epoch 42, while loss validation shows unstable results and tends to produce high validation loss values. The convergence of the model takes a long time until the 80th epoch. This model produces training loss of 5.89×10^{-5} , and validation loss of 12.5% is shown in Fig. 7 (b).

To determine the performance of the model, we conducted a test using 191 Balinese carving images that have gone through an augmentation process. This augmentation process uses seven types of Geometric Transformations, which represent various conditions for taking Balinese carving images. Based on the results of testing on this model, the use of the original dataset without Geometric Transformations is only able to produce an accuracy of recognition of 75.4%.

4.2 Original dataset with geometric transformations

In the second scenario, we use an original Balinese carving dataset taken using a smartphone camera. This dataset consists of 5 classes with the addition of data augmentation using Geometric Transformation on training and validation data. The distribution of datasets used in the second training model is 360 training images and 320 validation images. We use Geometric Transformations such as rescale, rotation, width and height shifting, zoom, and horizontal flip on this data augmentation. The selection of this augmentation type is adjusted to the variations of Balinese carving image taking using a smartphone camera. Fig. 7 shows the Balinese carving images that have gone through the data augmentation process using Geometric Transformations.

The second scenario in Fig. 8 shows the accuracy and loss of validation that is still not able to follow the accuracy and loss of training. The accuracy graph is shown in Fig. (8) shows a large gap between accuracy and validation. The convergence of the model still takes longer until the 65th epoch. Likewise, with the loss graph shown in Fig. 8 (b), the value of training loss and validation loss are also far adrift. Validation loss follows the value of training loss when entering the 65th epoch. This model produces 96.9% training accuracy with a loss of 10% and validation accuracy of 94.1% with a loss of 3.7%. We tested the model using 191 Balinese carving

images. This model can produce a recognition accuracy of 83.7%. These results indicate that Geometric Transformations as data augmentation on the dataset can increase recognition accuracy by 8.3% compared to the first model.

4.3 Original dataset with neural style transfer

In this experiment, we conducted three scenarios that apply the different number of style images on Neural Style Transfer. These three experiments conducted to determine the effect of the number of style images on recognition performance. These experiments were carried out by applying Neural Style Transfer to produce three models with several different style images. Table 2 shows three experiments conducted in this scenario. The first experiment applies for Neural Style Transfer by generating a new image with 30 iterations using one style image. We produced 1395 images that were trained to form new models. To determine the effect of the number of style images on the resulting model, we tried again by applying two image styles and three image styles to produce two new models, detail of the images generated with NST can be seen in Table 3. Fig. 9 shows a graph of accuracy and loss in the training model produced by applying for Neural Style Transfer with several style images.

In the first model, we applied NST with one style image to produce 1395 images. This model produces 100% training accuracy with 9.8×10^{-10} training loss and 81.8% validation accuracy with 2.2×10^{-2} loss validation. The model accuracy graph is shown in Fig. 9 (a) shows an improvement in performance compared to the model that only uses geometric transformation shown in Fig. 8 (a). This model also produces a small training and validation loss, showing that this model's performance is quite good. To find out the effect of the number of style images used on NST, we add more style images on NST to create new model. Details on the number of images generated by the NST can be seen in Table 2. Fig. 9 (c) and 9 (d) show an improvement in model performance by producing good accuracy and loss values on models with two style images.

In model with three style images, the performance of the validation accuracy decreases but still with a high accuracy. Three experiments in this scenario showed an improvement in the model's performance using NST as data augmentation. Based on the experiments, the number of style images used in NST has a significant effect on improving model performance.

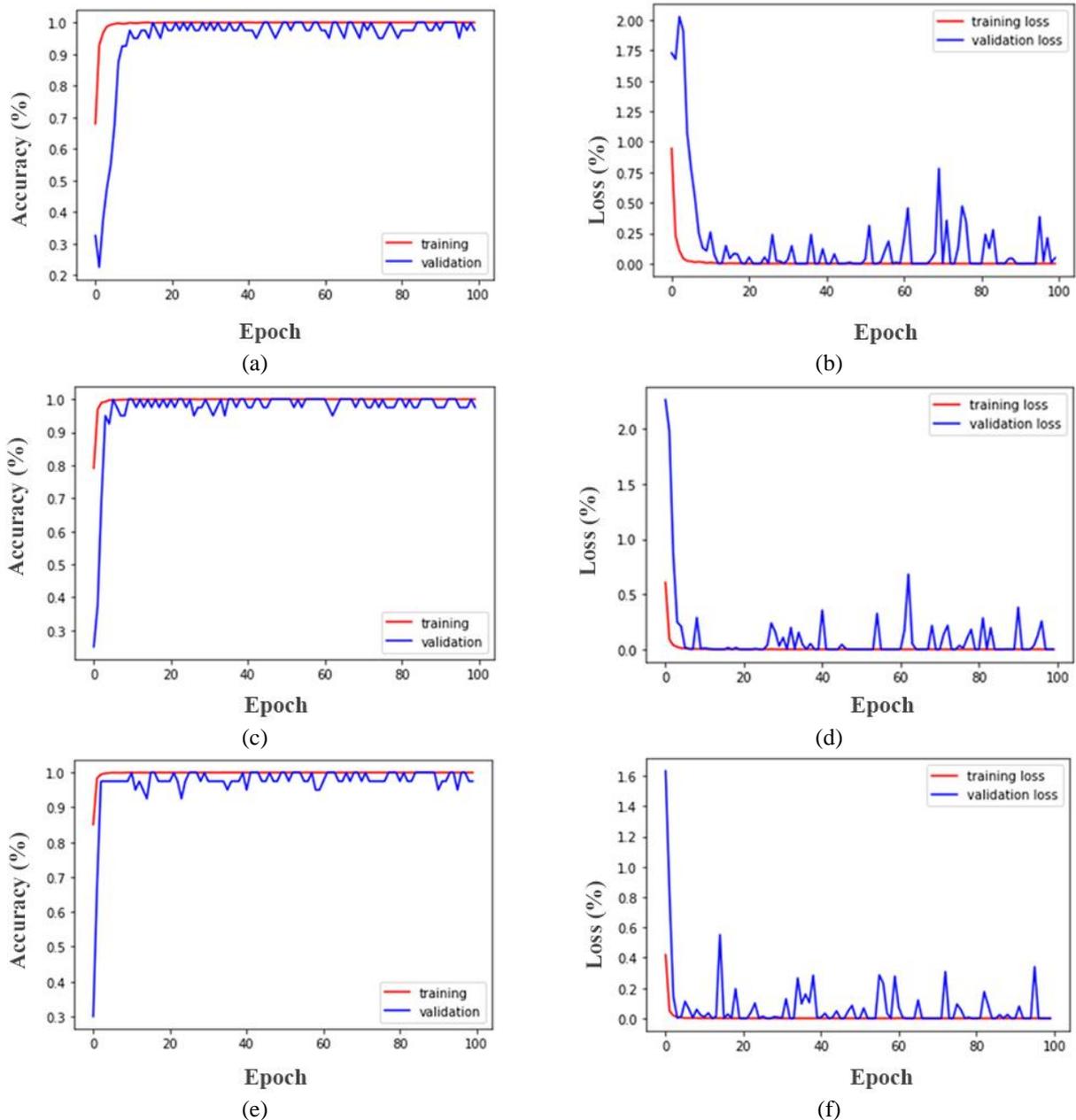


Figure. 10 Graph of models training on NSGT: (a) 1 Style NST model accuracy, (b) 1 style NST model loss, (c) 2 style NST model accuracy, (d) 2 style NST model loss, (e) 3 Style NST model accuracy, and (f) 3 style NST model loss

4.4 Original dataset with neural style transfer and geometric transformations

We conducted the next experiment by applying Geometric Transformations to the NST dataset. In the Balinese carvings recognition, the recognition results are greatly influenced by way of taking photographs such as the tilt angle, zoom, and position of the Balinese carvings. The application of Geometric Transformations on the dataset can provide variations of the image based on the tilt angle, zoom, and position of the images, thus increasing the accuracy of the recognition. In this experiment, we conducted three scenarios. These scenarios are carried out by

applying seven Geometric Transformations to three types of NST datasets. Detailed testing scenarios and the number of generated images are shown in Table 3.

In the first experiment, we applied seven Geometric Transformations to the NST 1 style image dataset. This scenario generates 9,765 images. The graph of model accuracy and loss in this scenario is shown in Fig. 10 (a), and Fig. 10 (b). The result of training accuracy is 100%, with a training loss of 6.5×10^{-8} . This model also produces a validation accuracy of 100% with a validation loss of 4.7×10^{-2} . The validation accuracy can well follow the training accuracy of the 100 epochs. This shows that the

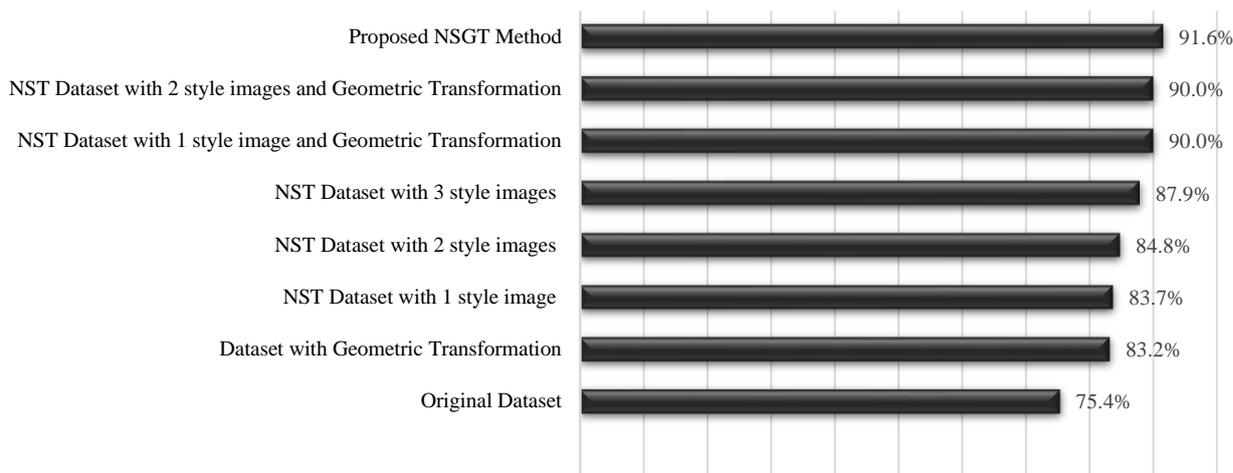


Figure. 11 Comparison of recognition accuracy on eight scenarios

Table 4. Performance comparison of NSGT with other augmentation methods on balinese carving recognition

Augmentation Method	Datasets	Training Accuracy	Validation Accuracy	Recognition Accuracy
Without Augmentation	45 images	94,7%	79,1	75,4%
Geometric Transformation [24][25][26]	405 images	96,9%	94,10%	83,2%
Neural Style Transfer [11][12][13][14]	4095 images	100%	90,40%	87,9%
Proposed Method (NSGT)	28665 images	100%	93,4%	91,6%

model has good performance. The second experiment was carried out by applying Geometric Transformations to the two style NST datasets. This model also shows good performance, which can be seen in Fig. 10 (c) and 10 (d) with lower loss validation compared to previous experiments. The application of Geometric Transformations on the three styles NST datasets also shows an improvement in the model performance, which is quite good with a lower validation loss compared to the two styles NST dataset.

5. Analysis and discussion

We conducted a test using 191 Balinese carving images to determine the model’s performance that produced in these eight scenarios. The results show the application of NSGT to the dataset can improve the model’s performance. The proposed method can provide higher recognition accuracy compared to NST or Geometric Transformation. Fig. 11 shows the comparison of the recognition accuracy of the eight models.

The proposed method improved the recognition performance by providing color variations, lighting, variations in geometric transformations e.g., rotation and rescale. Increasing the data variety in the dataset

can enrich the features, thereby increasing recognition performance. We analyzed the performance between NSGT and other augmentation methods that only use Neural Style Transfer [11-14]. Neural Style Transfer can provide variations in the color of the augmented image. However, it is not effective if used to recognize objects with variations in positions such as rotation and zoom. Our proposed method can provide better recognition performance because it provides color variations, lighting, variations in geometrical transformations e.g., rotation and rescales. Our proposed method can increase recognition performance by up to 3.7% compared to only using Neural Style Transfer as data augmentation.

Then we compare NSGT with Geometric Transformation, which has been used as data augmentation [24-26] Geometric Transformation provides geometric variations, e.g., rotation and rescale, to increase the dataset’s size by enriching data variations. This method can improve recognition performance by 7.8% compared to the original dataset without data augmentation techniques. However, Geometric Transformation cannot provide variations in color and lighting, so the recognition performance is not optimal in data with color and lighting variations. In this study, we combined color

variations, lighting, rotation, and rescale to enrich data variations and increase the data in the dataset. NSGT can provide data variations that can improve the model's performance in recognizing Balinese carving objects with various conditions. Table 4 shows a comparison of NSGT performance with other methods that only use Geometric Transformation and Neural Style Transfer. Our proposed method can improve recognition performance by up to 16.2%. These results indicate that a good model needs to be trained using data that can represent test data. By providing data variations with color, lighting, rotation, and rescale variations, NSGT can produce significant performance improvements. NSGT can increase the amount of data in a small dataset and improved the model's performance.

6. Conclusion

In this paper, we proposed a data augmentation technique representing images that have variations in color, lighting, rotation, and rescale, taken using a smartphone. We proposed a Neural Style Geometric Transformations (NSGT) method combining Neural Style Transfer and Geometric Transformations for data augmentation. NSGT aims to provide data variations on the small dataset. The limited number of Balinese carving images as a dataset is the main problem in the Balinese carving recognition. NSGT provides a new data augmentation technique to create new data variation in rotation, rescale, lighting and color variations to enrich the dataset to improve recognition performance. These variations on the dataset to overcome the Balinese carving image's poor quality in terms of tilt angle and image position. In the training process, the NSGT model can reduce training loss and validation loss compared to conventional techniques, e. g., variations in the image's geometric position.

We conducted eight experiments to determine the performance of NST and Geometric Transformations on the Balinese carving recognition. Based on the test results, the addition of Geometric Transformations can improve recognition accuracy by up to 7.8% from the original dataset. The application of NST with variations in the number of image styles can improve recognition accuracy by up to 12.5%. To get the best results, we combined NST and Geometric Transformations to improve recognition accuracy by up to 16.2%. Based on this result, this proposed method can be used as data augmentation to improve the accuracy of recognition in the case of a limited Balinese carving dataset.

As future work, we would like to develop a new architecture that can recognize Balinese carvings that have a variety of patterns at once. The challenge in recognizing Balinese carvings is the similar patterns in each carving, i.e., *Karang Barong* and *Karang Boma*, so special methods are needed to distinguish these patterns. Towards this, we might develop this research to inventory the Balinese carvings scattered in Bali digitally.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization, I Wayan Agus Surya Darma, Nanik Suciati and Daniel Siahaan; methodology, I Wayan Agus Surya Darma, Nanik Suciati and Daniel Siahaan; software, I Wayan Agus Surya Darma; validation, I Wayan Agus Surya Darma, Nanik Suciati and Daniel Siahaan; formal analysis, I Wayan Agus Surya Darma, Nanik Suciati and Daniel Siahaan; investigation, I Wayan Agus Surya Darma, Nanik Suciati and Daniel Siahaan; resources, I Wayan Agus Surya Darma, Nanik Suciati and Daniel Siahaan; data curation, I Wayan Agus Surya Darma, Nanik Suciati and Daniel Siahaan; writing—original draft preparation, I Wayan Agus Surya Darma; writing—review and editing, Nanik Suciati and Daniel Siahaan; visualization, Nanik Suciati and Daniel Siahaan; supervision, Nanik Suciati and Daniel Siahaan.

References

- [1] I. G. B. R. Utama, "Keunikan Budaya Dan Keindahan Alam Sebagai Citra Destinasi Bali Menurut Wisatawan Australia Lanjut Usia", *Jurnal Kajian Bali (Journal of Bali Studies)*, Vol. 6, No. 1, pp. 149–172, 2016.
- [2] G. Hartanti and A. Nediari, "Pendokumentasian Aplikasi Ragam Hias Budaya Bali, Sebagai Upaya Konservasi Budaya Bangsa Khususnya Pada Perancangan Interior", *HUMANIORA*, Vol. 5, No. 1, pp. 521–540, 2014.
- [3] S. P. Ramteke, A. A. Gurjar, and D. S. Deshmukh, "A streamlined OCR System for Handwritten MARATHI Text Document Classification and Recognition using SVM-ACS Algorithm", *International Journal of Intelligent Engineering and Systems*, Vol. 11, No. 3, pp. 186–195, 2018.
- [4] I. W. A. S. Darma, "Implementation of Zoning and K-Nearest Neighbors in Character Recognition of Wrésastra Script", *Lontar*

Komputer: Jurnal Ilmiah Teknologi Informasi, Vol. 10, No. 1, pp. 9–18, 2019.

- [5] I. M. A. Mahawan and A. Harjoko, “Pattern Recognition of Balinese Carving Motif Using Learning Vector Quantization (LVQ)”, *Communications in Computer and Information Science*, Vol. 788, pp. 43–55, 2017.
- [6] M. Nasir, N. Suciati, and A. Y. Wijaya, “Kombinasi Fitur Tekstur Local Binary Pattern yang Invariant Terhadap Rotasi dengan Fitur Warna Berbasis Ruang Warna HSV untuk Temu Kembali Citra Kain Tradisional”, *Jurnal Inspiration*, Vol. 7, No. 1, pp. 42–51, 2017.
- [7] A. E. Minarno, A. S. Maulani, A. Kurniawardhani, F. Bimantoro, and N. Suciati, “Comparison of Methods for Batik Classification using Multi Texton Histogram”, *Telecommunication Computing Electronics and Control (Telkomnika)*, Vol. 16, No. 3, pp. 1358–1366, 2018.
- [8] X. Qin, X. Chu, C. Yuan, and R. Wang, “Entropy-Based Feature Extraction Algorithm for Stone Carving Character Detection”, *The Journal of Engineering*, Vol. 2018, No. 16, pp. 1719-1723.
- [9] E. Prasetyo, R. Purbaningtyas, and R. D. Adityo, “Cosine K-Nearest Neighbor in Milkfish Eye Classification”, *International Journal of Intelligent Engineering and Systems*, Vol. 13, No. 3, pp. 11–25, 2020.
- [10] R. A. Pramunendar, S. Wibirama, P. I. Santosa, P. N. Andono, and M. A. Soeleman, “A Robust Image Enhancement Techniques for Underwater Fish Classification in Marine Environment”, *International Journal of Intelligent Engineering and Systems*, Vol. 12, No. 5, pp. 116–129, 2019.
- [11] A. Mikołajczyk and M. Grochowski, “Data Augmentation for Improving Deep Learning in Image Classification Problem”, In: *2018 International Interdisciplinary PhD Workshop, (IIPhDW)*, Swinoujście, Poland, pp. 117–122, 2018.
- [12] J. Sheng, C. Song, J. Wang, and Y. Han, “Convolutional Neural Network Style Transfer towards Chinese Paintings”, *IEEE Access*, Vol. 7, pp. 163719–163728, 2019.
- [13] H. H. Zhao, P. L. Rosin, Y. K. Lai, M. G. Lin, and Q. Y. Liu, “Image Neural Style Transfer with Global and Local Optimization Fusion”, *IEEE Access*, Vol. 7, pp. 85573–85580, 2019.
- [14] Z. Cheng, H. Gao, Z. Liu, H. Wu, Y. Zi, and G. Pei, “Image-Based CAPTCHAs Based on Neural Style Transfer”, *IET Information Security*, Vol. 13, No. 6, pp. 519–529, 2019.
- [15] G. Fan, F. Chen, D. Chen, and Y. Dong, “Recognizing Multiple Types of Rocks Quickly and Accurately Based on Lightweight CNNs Model”, *IEEE Access*, Vol. 8, pp. 55269–55278, 2020.
- [16] H. Hiary, H. Saadeh, M. Saadeh, and M. Yaqub, “Flower Classification Using Deep Convolutional Neural Networks,” *IET Computer Vision*, Vol. 12, No. 6, pp. 855–862, 2018.
- [17] L. G. Falconi, M. Perez, and W. G. Aguilar, “Transfer Learning in Breast Mammogram Abnormalities Classification with Mobilenet and Nasnet”, In: *Proc. of 2019 International Conf. on Systems, Signals and Image Processing, Osijek, Croatia*, pp. 109–114, 2019.
- [18] S. Banerjee and S. S. Chaudhuri, “Total Contribution Score and Fuzzy Entropy Based Two-Stage Selection of FC, ReLU and InverseReLU Features of Multiple Convolution Neural Networks for Erythrocytes Detection”, *IET Computer Vision*, Vol. 13, No. 7, pp. 640–650, 2019.
- [19] S. Duraisamy and S. Emperumal, “Computer-aided Mammogram Diagnosis System Using Deep Learning Convolutional Fully Complex-Valued Relaxation Neural Network Classifier”, *IET Computer Vision*, Vol. 11, No. 8, pp. 656–662, 2017.
- [20] G. Li and J. Kim, “Palmprint Recognition with Local Micro-structure Tetra Pattern”, *Pattern Recognition*, Vol. 61, No. 1, pp. 29–46, 2017.
- [21] X. Zhao, C. Meng, M. Feng, S. Chang, and Q. Zeng, “Eye Feature Point Detection Based on Single Convolutional Neural Network”, *IET Computer Vision*, Vol. 12, No. 4, pp. 453–457, 2018.
- [22] F. Wang, Q. Guo, J. Lei, and J. Zhang, “Convolutional Recurrent Neural Networks with Hidden Markov Model Bootstrap for Scene Text Recognition”, *IET Computer Vision*, Vol. 11, No. 6, pp. 497–504, 2017.
- [23] A. Signoroni, M. Savardi, M. Pezzoni, F. Guerrini, S. Arrigoni, and G. Turra, “Combining the Use of CNN Classification and Strength-Driven Compression for the Robust Identification of Bacterial Species on Hyperspectral Culture Plate Images”, *IET Computer Vision*, Vol. 12, No. 7, pp. 941–949, 2018.
- [24] I. M. A. Agastya and A. Setyanto, “Classification of Indonesian Batik Using Deep Learning Techniques and Data Augmentation”, In: *Proc. of 2018 3rd International Conf. Information Technology, Information Systems and Electrical Engineering*, Yogyakarta,

- Indonesia, pp. 27–31, 2018.
- [25] R. Li, X. Jia, M. Hu, M. Zhou, D. Li, W. Liu, R. Wang, J. Zhang, C. Xie, L. Liu, F. Wang, H. Chen, T. Chen, and H. Hu, “An Effective Data Augmentation Strategy for CNN-Based Pest Localization and Recognition in the Field”, *IEEE Access*, Vol. 7, pp. 160274–160283, 2019.
- [26] R. Dong, D. Xu, J. Zhao, L. Jiao, and J. An, “Sig-NMS-Based Faster R-CNN Combining Transfer Learning for Small Target Detection in VHR Optical Remote Sensing Imagery”, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 57, No. 11, pp. 8534 - 8545, 2019.
- [27] D. I. Swasono, H. Tjandrasa, and C. Fathicah, “Classification of Tobacco Leaf Pests Using VGG16 Transfer Learning”, In: *Proc. of 12th International Conf. on Information and Communication Technology and Systems*, Surabaya, Indonesia, pp. 176–181, 2019.
- [28] P. Marcelino, “Transfer learning from pre-trained models”, *towardsdatascience*, 2018. [Online]. Available: <https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f124751>. [Accessed: 26- May-2020].