A Novel Optimization Strategy for CNN Models in Palembang Songket Motif Recognition

Yohannes, Muhammad Ezar Al Rivan, Siska Devella, Tinaliah
Informatics, Faculty of Computer Science and Engineering, Universitas Multi Data Palembang,
Palembang, Indonesia

Abstract*—Palembang Songket is an essential part of Indonesian cultural heritage, and its introduction and preservation present challenges, particularly in recognizing various motifs. This research introduces a novel strategy to optimize the performance of Convolutional Neural Networks (CNNs) by presenting a hierarchical integration of Ghost Module operations and Max Pooling, referred as Ghost Feature Maps. While the Ghost Module is effective in reducing parameters and enhancing feature extraction, it has limitations in filtering irrelevant information. To address this shortcoming, we propose a hierarchy in which Max Pooling works in conjunction with the Ghost Module, strengthening its performance by not only extracting dominant features but also eliminating excess, non-essential information. This hierarchical design enables more efficient feature extraction, thus enhancing the model's recognition accuracy. By combining Ghost Modules and Max Pooling in a structured manner, this approach advances established methodologies and offers a new perspective on feature representation within CNN architectures. Utilizing a dataset of 10 augmented classes of Palembang Songket motifs totaling 1000 images, we conducted experiments using varying ratios of Ghost Feature Maps. The results indicate that a ratio of 2 achieves an impressive accuracy of 0.98 with minimal parameter reduction. Additionally, a ratio of 3 results in a 34% decrease in parameters while maintaining a competitive accuracy of 0.95. Ratios of 4 and 5 continue to demonstrate robust performance, achieving accuracy levels of 0.93 while delivering over 60% reductions in model size and parameters. This research not only contributes to the optimization of CNN architectures but also supports the preservation of cultural heritage by improving the recognition capabilities of Palembang Songket motifs.

Keywords*—*Convolutional neural network; ghost module; palembang songket motif; recognition

I. Introduction

One of the artistically significant pieces of Indonesian cultural heritage is Songket. The term "sungkit" which describes the process of embroidering gold and silver threads, is the source of the word "songket" [1]. Ten connected steps are involved in the production of Songket fabric, including thread dyeing, klose processing, lungsin coating, thread type selection, and weaving designs with lidi. Songket fabric has different characteristics and philosophies depending on its region of origin [2]. One of the Songket fabrics registered as an Indonesian Intangible Cultural Heritage is Palembang Songket [3]. Palembang Songket has various types of motifs. The motifs of Palembang Songket reflect its beauty, uniqueness, and traditional values. Palembang Songket faces challenges in its preservation because the motif recognition process still relies on manual or semi-automatic approaches that are vulnerable to human error and limitations. Successfully recognizing and understanding Palembang Songket motifs is essential for preserving local art and culture and has significant economic impacts through promoting Songket products in the global market.

The problem of identifying Songket motifs has been addressed in the past by a number of feature extraction techniques, including Felzenzswalb segmentation [4], and Gray Level Co-occurrence Matrix (GLCM) [5]. These techniques are then combined with a variety of classifier algorithms, including Naive Bayes [6], Decision Tree [6], and Support Vector Machine (SVM) [5]. Even while these techniques have produced acceptable outcomes in certain situations, there are still a number of important drawbacks. The primary drawback of these methods is their dependence on manual feature extraction, which frequently results in challenges in capturing the crucial elements of the complex Songket motifs.

Enhancement of motif identification performance is a promising area in deep learning. Convolutional Neural Network (CNN) models have demonstrated their capacity to recognize patterns in a variety of domains, including the identification of images. CNN has a lot of potential to improve Songket motif recognition. Applying CNN to motif recognition has the benefit of minimizing reliance on manual feature extraction by automatically extracting pertinent characteristics from data. Although there has been some prior research, not much has been accomplished in terms of training CNN to identify Songket motifs [7], [8]. The primary constraint is to the model's capability to manage intricate motif modifications, such as rotation, scaling, and distortion. The intricacy of Palembang Songket patterns is too great for a conventional CNN training model to handle.

Due to their numerous convolution layers, CNNs have substantial computational costs. In each convolution layer, mathematical operations are conducted to each input in order to extract important properties from the input data. The number of convolution layers that are conducted therefore increases the amount of computing operations needed, which could result in significant computational expenses in terms of processing time and energy consumption [9], [10], [11]. Thus, the convolution operations of CNN can be assumed by the Ghost Module. By using a method called the Ghost Module, which involves joining convolution filters with smaller "ghost filters", the CNN model is able to retain a significant amount of representation information while requiring fewer calculations and parameters [12]. In order to improve the performance of Palembang

Songket motif recognition, this research suggests a novel method that incorporates the Ghost Module into the CNN architecture.

The main contribution in this research is the introduction of Ghost Feature Maps through the integration of Ghost Module as a substitute for conventional convolutional layers, with Max Pooling applied hierarchically afterward. Ghost Feature Maps function to facilitate feature learning in CNN models, with the aim of increasing the efficiency of Palembang Songket motif recognition. In this hierarchical approach, Ghost Module is applied first to improve feature extraction efficiency by reducing computational complexity. Afterward, Max Pooling is used to further reduce spatial dimensions, enhancing computational efficiency and focusing on dominant features while suppressing irrelevant information. This combination reduces the number of parameters and model size while increasing accuracy performance. In addition, this study offers a solution to the challenge of recognizing complex Songket motif variations, such as rotation, scale, and deformation. This approach provides an innovative solution in addressing problems that have not been fully resolved in Palembang Songket motif recognition.

The remaining content of the paper is organized as follows: Section II discusses the related works in the field of Palembang Songket Motif, CNN and Ghost Module. Section III presents a detailed explanation of the proposed method. Section IV presents results and discussion, and performance. The last Section V summarizes the overall conclusion of the paper.

II. Literature Review

This section discusses the literature review of the Indonesian traditional fabrics, CNN, and Ghost Module research.

A. *Image-based Recognition of Indonesian Traditional Fabrics*

Sriani, Hasibuan, and Ananda [5] used SVM to classify Batu Bara Songket motifs, which are characterized by distinctive patterns such as Bunga Tanjung, Pucuk Betikam, Pucuk Cempaka, Pucuk Pandan, Tampuk Manggis, and Tolab Berantai. Gray-level texture features were extracted using the Co-Occurrence Matrix method, considering parameters like Contrast, Correlation, Energy, and Homogeneity. These extracted features were then processed as input for classification using the SVM. Despite the challenges, the study achieved a classification accuracy of 57% with 60 training data and 30 test data.

Aprianti et al. [6] classified Lombok songket fabric motifs, which are characterized by geometric patterns, varying density, color, and motif positioning. The algorithms used included Naive Bayes and Decision Tree, tested with different pixel sizes to compare accuracy levels. The results showed that the Naive Bayes algorithm achieved the highest accuracy of 90% at a 100×100 pixel size, while the Decision Tree algorithm was optimal at 400×400 pixels with the same accuracy. This approach demonstrated that combining algorithms with pixel size adjustments could significantly enhance motif recognition accuracy.

Ariessaputra et al. [7] classified Lombok songket motifs using a Convolutional Neural Network (CNN) algorithm, demonstrating the potential of image processing for traditional fabric pattern recognition. The dataset consisted of 20 Lombok Songket images with identical motifs and colors, 14 with the same but different colors, and 10 with various motifs and colors. In the preprocessing phase, each image underwent resizing, followed by CNN layers for convolution, pooling, and fully connected operations. Data augmentation through 150-degree rotations was applied to enhance model robustness. The results indicated that motif classification with consistent colors achieved an accuracy of 84%, highlighting the effectiveness of CNNs for identifying and distinguishing Lombok Songket motifs across varying visual parameters.

Hambali, Mahayadi, and Imran [8] applied CNN to classify Lombok Songket motifs, focusing on Songket from two prominent Lombok regions, Sade and Pringgasela. The study utilized a dataset of 64 images, comprising 40 samples from Sade and 24 from Pringgasela. The model's testing results showed an accuracy of 86%, with 87% precision and 86% recall. These results demonstrated CNN's effectiveness in differentiating textures within traditional Songket fabrics, offering valuable insights for preserving and recognizing regional textile characteristics.

Andrian et al. [13] used CNN architectures, including AlexNet, EfficientNet, LeNet, and MobileNet, to classify Lampung Batik motifs. The study utilized a dataset of 1000 images representing ten distinct motifs, enhanced through preprocessing techniques like rotation, shifting, brightness adjustment, and zooming. The results showed that LeNet achieved the highest accuracy of 99.33%, highlighting its suitability for small datasets, while other architectures also demonstrated strong performance despite occasional classification errors due to motif similarities.

Elvitaria et al. [14] proposed an ensemble deep learning method for batik image classification that combines texture feature extraction using Gray Level Co-occurrence Matrix (GLCM) with the Residual Neural Network (ResNet) classification model. By extracting texture features such as contrast, dissimilarity, and entropy using GLCM and combining them with ResNet, the proposed ensemble method achieved high performance, with accuracy, precision, recall, and F1-score all above 90%. The study demonstrated that the ensemble deep learning approach, particularly with the standard deviation feature, improved classification accuracy and can be applied to preserve batik culture digitally.

Muliono, Iranita, and Syah [15] proposed a deep learning model for classifying traditional Batak Ulos fabrics, utilizing CNN to recognize and classify different Ulos motifs. The study employed the Modular Neural Network (MNN) to simplify complex computations, achieving an accuracy of 97.83% with a loss value of 0.0793 during training. The validation results showed a validation loss of 2.1885 and a validation accuracy of 74.29%, demonstrating the model's strong performance while indicating areas for potential improvement in generalization.

Overall, the studies reviewed highlight significant advancements in image-based recognition techniques for recognizing Indonesian traditional fabrics, such as Songket and

Batik. Various machine learning models, including SVM, Naive Bayes, Decision Trees, CNN, and ensemble deep learning methods, have been applied to address the challenges of recognizing intricate fabric patterns characterized by texture, color, and motif positioning. These findings underscore the potential of image-based recognition systems in advancing the digital preservation and recognition of traditional Indonesian fabrics, offering valuable contributions to the cultural heritage field.

B. Ghost Module and CNN

Han et al. [12] proposed the Ghost Module in their research to produce feature maps with low computational costs, thereby simplifying the architecture of conventional CNN. GhostNet, built using the Ghost Module, demonstrated better recognition performance than MobileNetV3, with an accuracy of 75.7% on the ImageNet ILSVRC-2012 dataset. According to these results, the Ghost Module has the potential to be an efficient solution that can be added to current CNN networks.

Wang and Li [16] discussed the creation of a CNN model to enhance recognition systems' efficiency. This study used the GhostNet and Convolutional Block Attention Module (CBAM) methods. GhostNet and Ghost Bottleneck improved the model's capacity to extract significant image characteristics. Furthermore, employing GhostNet resulted in fewer parameters while retaining good accuracy.

Zhao and Cheng [17] proposed a more efficient approach that minimizes computation compared to traditional CNN by merging the Yolov5 model and GhostNet. This method increases processing speed and accuracy. The proposed approach's performance test showed that it achieved high detection accuracy while needing less memory and compute. Overall, this study provides a solution for image identification for human security screening purposes. This framework is also done for vehicle detection [18].

Huangfu, Li, and Yan [19] proposes the Ghost-YOLO v8 algorithm to improve the detection of surface floating litter in artificial lakes. This efficient and lightweight algorithm includes an SE mechanism for better feature extraction, a small-target detection layer to reduce semantic loss, and a GhostConv module to decrease computational demands.

Fang, Chen, and He [20] suggested an efficient CNN solution for facial expression recognition called Ghost-based Convolutional Neural Network (GCNN). This approach seeks to overcome CNN-related overfitting concerns. The Ghost Module architecture is less computationally expensive since it may minimize the number of parameters while producing more feature maps than CNN approaches. Based on this research, GCNN can efficiently extract and classify face expression features.

Alansari et al. [21] introduced Lightweight Face Recognition using GhostFaceNets, which requires less processing than normal CNN models. GhostFaceNets is a very accurate and efficient face recognition system. The proposed method was tested using a variety of datasets, including LFW, AgeDB-30, IJB-B, IJB-C, and MegaFace. GhostFaceNets uses the Ghost Module method to execute linear modifications on feature maps, resulting in better and more thorough feature extraction.

Luan, Mu, and Yuan [22] addresses challenges in Online Signature Verification (OSV), by proposing the one-dimensional Ghost-ACmix Residual Network (1D-ACGRNet). The network is designed to combine convolution with a self-attention mechanism to effectively capture both global and local signature features. Simplification of operations is achieved through the Ghost-based Convolution and Self-Attention (ACG) block, which reduces computational load. Significant accuracy improvements are shown in experiments on the MCYT-100 and SVC-2004 Task2 datasets, with equal error rates reaching as low as 0.91% for genuine and forged signatures.

Paoletti et al. [23] conducted Hyperspectral Image Classification (HSI) which is one of the remote sensing techniques used in Earth observation for health, robotic vision, and quality control. The challenge in HSI is that each HSI image has hundreds of spectral bands that produce large amounts of data, requiring high computation. This study introduces an approach by combining ghost-module architecture and CNN. Test results show that using Ghost module can reduce HSI's cost and computation time.

Tang et al. [24] proposed an efficient mechanism called DFC attention is proposed, using GhostNetV2. GhostNetV2 can overcome limitations in conventional CNN methods. In testing, GhostNetV2 showed better performance than CNN architecture, achieving an accuracy of 75.3% on the ImageNet dataset, with efficient FLOPs. Therefore, GhostNetV2 is the right choice for mobile applications requiring efficiency and high performance.

Liu et al. [25] investigated effective training strategies for compact neural networks to address performance gaps and proposed GhostNetV3. The strategy focused on essential methods, including re-parameterization to improve efficiency, knowledge distillation to enhance smaller model performance through learning from larger models, and optimized learning schedules and data augmentation to increase training data diversity. As a result, GhostNetV3 achieves an optimal balance between accuracy and inference costs.

He et al. [26] proposed the Ghost module-based convolution network approach for superresolution (SR) in satellite video in this research. This approach is called Ghost module-based video SR (GVSR) and consists of two main modules: the preliminary image generation module and the SR results' reconstruction module. Experimental results on Jilin-1 and OVS-1 videos show that this method is superior in quality and quantity to other Deep Learning methods.

Liu et al. [27] introduced a new approach for hyperspectral image classification called Ghost module extended morphological profile (GhostEMP), which employs Ghost Module and extended morphological profile (EMP) features. This method can reduce model complexity and the amount of calculations, hence increasing operational efficiency. The experimental results suggest that this strategy effectively preserves model performance by maximizing hyperspectral data features. Not only does it apply to hyperspectral data, but it also utilizes the Ghost Features Network (GFN) for super-resolution by cascading residual-in-residual ghost blocks [28].

The studies emphasize the effectiveness of the Ghost Module in enhancing the efficiency of deep learning models for diverse applications, including image recognition and hyperspectral image processing. GhostNet and its variants, such as GhostNetV2 and GhostNetV3, have significantly improved accuracy while reducing computational complexity. These advancements make Ghost Module in GhostNet architecture a valuable solution for high-performing applications with minimal resource usage.

III. Methodology

A. Dataset

The dataset used in this research consists of 10 classes of Palembang Songket motifs, namely Bintang Berantai, Bunga Cina, Bunga Jatuh, Cantik Manis, Jando Beraes, Kenanga Makan Ulat, Naga Besaung, Nampan Perak, Pacar Cina, and Pulir. The motif images were captured using a Canon 7D DSLR camera equipped with a Canon EF 70-200mm F2.8 lens, Canon Speedlite 600 Mark II, and tripod, with a consistent portrait distance of 45 cm from the object and a front-facing angle of 0 degrees. A total of 50 Songket fabrics were photographed and thoroughly validated by a Songket motif expert. The expert ensured that each image was following the traditions and authenticity of Palembang Songket culture.

Fig. 1. Dataset of palembang songket motif.

The Palembang Songket motif photos were cropped to 2048 x 2048 pixel size with 300 dpi resolution, and then augmented. Augmentation techniques applied include rotation, scaling, and horizontal and vertical flipping, as these techniques are more suitable for the complex patterns of Songket and do not alter the basic motif design. This augmentation technique produces a total of 1000 images, with each motif containing 100 images that correspond to the Palembang Songket motif collection with geometric patterns, as shown in Fig. 1. Afterwards, the images were resized to 256 × 256 for use as input in the CNN architecture with Ghost Feature Maps. As part of the preprocessing stage, each Songket motif image was resized to 256 x 256 pixels to ensure uniform image sizes.

B. Proposed Method

The proposed model architecture can be divided into two major components, namely feature learning and classification, each responsible for different aspects of its functionality. Fig. 2 illustrates the architecture of the proposed model.

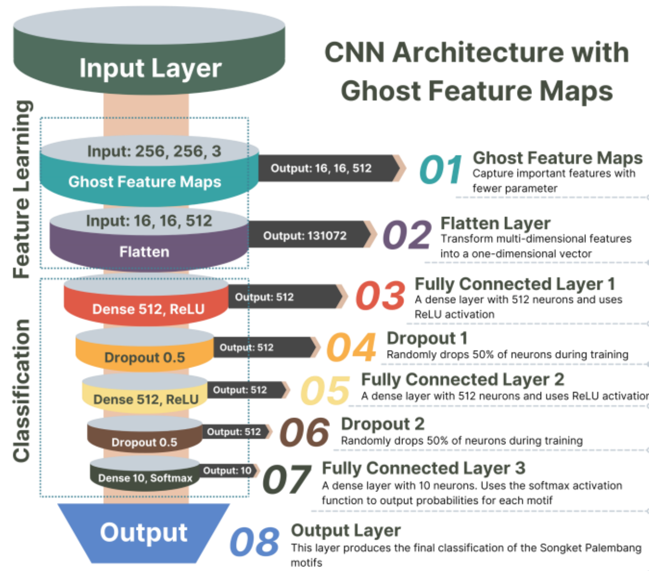

Fig. 2. CNN architecture with ghost feature maps.

1) Feature Learning: The process starts with applying Ghost Feature Maps. This layer is designed to extract feature maps efficiently by generating primary feature maps and applying simple transformations to produce additional maps. The objective here is to capture essential patterns in the data while minimizing computational resources and parameters, making the approach both efficient and effective. Following this, the extracted feature maps are processed by a Flatten layer. The Flatten layer transforms these high-dimensional feature maps into a one-dimensional vector, which is necessary for subsequent classification layers. The Flatten layer ensures that all the spatial information captured during the feature learning process is retained but represented in a format compatible with fully connected (dense) layers.

2) Classification: The classification phase includes three Fully Connected (Dense) layers. The first two Dense layers each consist of 512 neurons and use the ReLU activation function to introduce non-linearity, enabling the model to learn more complex relationships in the data. These layers take the flattened feature vector from the feature learning phase and process it further to identify the patterns necessary for classification. To address the risk of overfitting, the model incorporates two Dropout layers after the first and second Dense layers. Dropout randomly turns off 50% of the neurons during each training iteration, encouraging the model to learn more robust and generalized features by not overly relying on specific neurons. The final step involves a third Dense layer, the output layer. This layer contains ten neurons, corresponding to the 10 Palembang Songket motifs being classified. It uses the softmax activation function to generate a probability distribution across the ten classes, with the class having the highest probability selected as the predicted motif.

C. *Ghost Module*

The GhostNet architecture, which features a layer known as the Ghost bottleneck, was the primary inspiration for this work. The Ghost bottleneck combines the Ghost module, a batch normalization, two or three Ghost Modules in a row, and then interleaved depthwise convolution, batch normalization, and ReLU activation [12]. However, in this research, only the Ghost module is employed without the full implementation of the Ghost bottleneck.

The dataset is a Palembang Songket pattern and it is unique and thus very rare, so it is hard to get a lot of data. Batch normalization was excluded from the Ghost Module. Instead, the Ghost Module was used alone, assuming that its simplicity would adequately convey the distinctive qualities of the Palembang Songket motifs without the added complication of the Ghost bottleneck.

The Ghost module consists of multiple phases, including primary convolution, cheap convolution, feature concatenation, and channel trimming. Each of these levels contributes significantly to the module's efficiency by reducing the amount of parameters and computational complexity while maintaining overall network performance. The Ghost module is a way to create more feature maps while doing fewer computations than usual, which is very helpful in deep learning models. Fig. 3 shows how the Ghost module works.

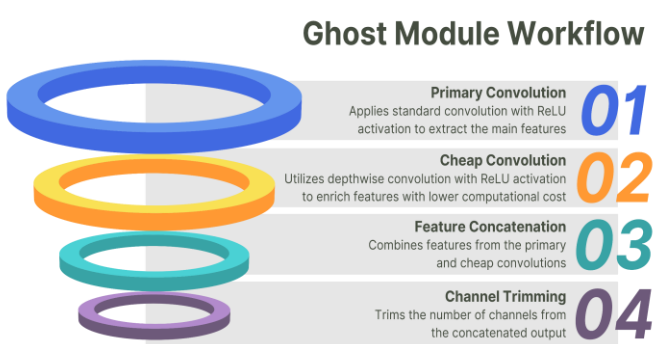

Fig. 3. Ghost module workflow.

1) Primary convolution: In the initial stage, the Ghost module conducts feature extraction through a primary convolution operation. This convolutional process decreases the number of output channels by a defined reduction ratio, denoted as r, which indicates the degree of parameter reduction in comparison to a typical convolution operation. If the desired number of output channels is C_{out}, the output channels from the primary convolution, C_{cheap} are calculated in Eq. (1).

$$C_{cheap} = \frac{C_{out}}{r} \tag{1}$$

This convolution employs a kernel size of $k \times k$, where k represents the kernel dimension. It is applied with the same padding to maintain the input's spatial dimensions. A ReLU activation function is applied after the convolution to introduce non-linearity, enabling the model to capture more complex patterns. The mathematical expression for this step is calculated in Eq. (2).

$$\boldsymbol{P} = \sigma(\boldsymbol{X} * \boldsymbol{W_1} + \boldsymbol{b}) \tag{2}$$

Where, $\boldsymbol{X}$ is the input, $\boldsymbol{W_1}$ is the convolution filter of size k, $\boldsymbol{b}$ is the bias term, and σ represents the ReLU activation function. The result of this process, $\boldsymbol{P}$, contains features with C_{cheap} channels, representing a portion of the total desired channels, C_{out}.

2) Cheap convolution: In the next stage, the features produced by the primary convolution undergo a second convolution process using a depthwise convolution. This operation processes each feature channel independently, meaning each channel is convolved with a separate filter, which reduces computational cost while generating additional feature details. The depthwise convolution is applied with a kernel size of $k \times k$, where k is the size of the depthwise kernel. The output of this process can be calculated in Eq. (3).

$$\boldsymbol{C} = \sigma(\boldsymbol{P} * \boldsymbol{W_2}) \tag{3}$$

Where $\boldsymbol{W_2}$ is the depthwise convolution filter applied to each feature channel in $\boldsymbol{P}$, and $\boldsymbol{C}$ represents the output of this operation. A ReLU activation function (σ) is also applied to maintain non-linearity in the feature representation. The Ghost module enriches the features with minimal computational overhead compared to full convolutions by using depthwise convolution, contributing to a more efficient model.

3) Feature Concatenation: Following the two convolution processes, the outputs from the primary and depthwise convolutions are aggregated. Feature aggregation combines these outputs to create a richer feature representation, merging primary and additional features into a single output set. If the output of the primary convolution is denoted as P and the output of the depthwise convolution as C, the aggregation process can be expressed mathematically in Eq. (4).

$$\boldsymbol{O} = [\boldsymbol{P}, \boldsymbol{C}] \tag{4}$$

Where $\boldsymbol{O}$ represents the aggregation output, and the notation $[.,.]$ signifies concatenation along the channel dimension. This aggregation ensures that the final feature set captures both the main and additional details from the input, leading to a more informative feature representation without significantly increasing computational complexity.

4) Channel Trimming: The final stage is channel trimming to ensure that the number of output channels corresponds to the target, C_{out}. After feature aggregation, the number of channels may surpass the target number. Channel trimming is done to keep only the necessary C_{out} channels. If the aggregated feature map $\boldsymbol{O}$ has C_{concat} channels, where $C_{concat} > C_{out}$, only retain the first C_{out} channels that can be represented as in Eq. (5).

$$\boldsymbol{Y}(i,j) = \boldsymbol{O}(i,j,1{:}C_{out}) \tag{5}$$

Where, $\boldsymbol{O}(i,j,1{:}C_{out})$ denotes selecting the first C_{out} channels from the aggregated feature map $\boldsymbol{O}$, where i and j are the spatial indices of the feature map. The operation trims the excess channels, ensuring the correct output dimensionality, thus providing an output feature map $\boldsymbol{Y}$ with dimensions $H \times W \times N$, where H and W are the height and width of the spatial dimensions, and N is the number of required channels. This trimming guarantees that the final output size is consistent with the architecture's design without unnecessarily increasing computational costs.

D. Ghost Feature Maps

The proposed Ghost feature maps, illustrated in Fig. 4, integrate Ghost modules with Max Pooling. In CNNs, multiple filters are applied within each convolutional layer to extract various features [29]. These convolutional and pooling layers are arranged sequentially, forming a hierarchical structure that progressively captures and reduces feature dimensions [30]. In this research, traditional convolutional layers are replaced by Ghost modules, which provide a more efficient alternative while maintaining the standard CNN architecture. Utilizing a ratio (r) parameter can generate more features from input with significantly fewer parameters. This research uses a kernel size k of 3 for both the primary [12] and depthwise convolution in the Ghost module, with the bias $\boldsymbol{b}$ set to 0. For instance, in the first layer, Ghost Module (32, r = 2) produces 32 features using only half the parameters conventional convolution requires. This not only accelerates training but also reduces the risk of overfitting. Additionally, as the number of features increases in subsequent Ghost Modules (32, 64, 128, 512), the model captures more complex variations in the input data. Increasing the feature channels as the network deepens enriches representation and allows the model to learn more abstract features.

Max pooling then is strategically applied after each Ghost module to reduce the dimensionality of the features significantly extracted. It also greatly reduces computational complexity in later layers, allowing the model to focus on the most dominant features, significantly improving its generalization ability. Max pooling's function in promoting invariance makes the model much more robust to small rotational or translational changes in the input data, which is very important in pattern recognition problems, where such variations are the norm. This makes the audience feel more confident that the model is not only efficient but strong as well.

Fig. 4. Ghost feature maps workflow.

Doing that four times, Ghost module and max pooling, gives even more advantages. Every pair builds a pyramid of feature representation. As the layers get deeper, the learned feature hierarchies become more and more complex, starting from simple features in the lower layers to more abstract representations in the higher layers. The architecture actually exploits this by using a Ghost module and then max pooling right after to ensure maximum feature extraction, but without losing any computational efficiency.

E. Parameter Distribution of Proposed Model Architecture

The distribution of parameters in the proposed CNN model's layer structure is presented in Fig. 5. Fig. 5 illustrates a comparative analysis of the total number of parameters between standard 2D convolutional (Conv2D) and Ghost module layers with varying expansion ratios (2 to 5). The Conv2D layer has the highest parameter count at 683,584, indicating its substantial computational complexity. In contrast, the Ghost module layers demonstrate a progressive reduction in the number of parameters as the expansion ratio increases, with 344,736 parameters at a ratio of 2; 150,633 at a ratio of 3; 87,120 at a ratio of 4, and 54,603 at a ratio of 5. This trend highlights the efficiency of Ghost module in significantly reducing parameter count while maintaining performance. The results suggest that higher Ghost module ratios can drastically minimize the model's computational load, offering a more resource-efficient alternative to traditional Conv2D layers. This parameter reduction is advantageous for applications with limited computational resources without compromising the model's representative capacity.

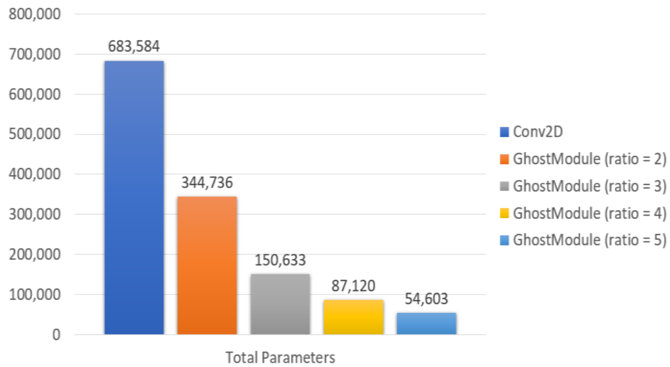

Fig. 5. Comparison of total parameters by layer type.

F. Experimental Setup

The experimental setup is to test the Ghost Feature Maps with ratios of 2, 3, 4, and 5 [12] versus Conv2D layers for image classification. The research will utilize a dataset of Palembang Songket motif images organized into ten distinct classes, with the data split structured as 80% for training, 10% for validation, and 10% for testing purposes. A single learning rate of 0.001 and a batch size of 32 will be employed, with the Adam optimizer selected for model training over 50 epochs.

The model architectures will encompass Ghost Modules for each specified ratio, followed by a Flatten layer, three Dense layers, and a Dropout for regularization. In parallel, a standard 2D convolutional model will be constructed with a similar architectural framework to facilitate direct comparison.

The evaluation will follow some standards: accuracy, precision, recall, and F1-score. The experimental procedure will also preprocess the data so the dataset is normalized and a consistent input is entered into the model. After training, however, all models will be tested thoroughly on the test set, and a complete analysis will be done comparing Ghost Feature Maps to normal convolutional layers and examining how variations in hyperparameters affect the model's overall performance.

G. Classification Performance

Many different performance measures are used to assess the classification models effectiveness. Some more common ones are accuracy, precision, recall, F1-score, and overall accuracy. They all express the model's capability to classify the Palembang Songket patterns differently.

1) Accuracy: simply the ratio of correct predictions to the total number of instances [31]. The accuracy value can be calculated using Eq. (6).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{6}$$

2) Precision: a measure used to determine how accurate the model is in its predictions. It is calculated by dividing the number of true positives by the number of predicted positives. Precision is the number of true positives divided by the total number of positives the model predicted [31]. The precision value can be calculated using Eq. (7).

$$Precision = \frac{TP}{TP+FP} \tag{7}$$

3) Recall: is the model's ability to find all the true positives. It is the percentage of positive examples that the model correctly labels [31]. The recall value can be calculated using Eq. (8).

$$Recall = \frac{TP}{TP+FN} \tag{8}$$

4) F1-Score: is the harmonic mean of precision and recall, used as a single value to balance precision and recall [31]. It can be calculated using Eq. (9).

$$F1-Score = \frac{2 \times Precision \times Recall}{Precision+Recall} \tag{9}$$

5) Overall accuracy: is simply the number of correct predictions divided by the total number of samples over all of the classes. It is especially useful when performing multiclass classification [31]. The overall accuracy value can be calculated using Eq. (10).

$$Overall\ Accuracy = \frac{\sum TP+\sum TN}{\sum TP+\sum TN+\sum FP+\sum FN} \tag{10}$$

IV. Results and Discussion

A. Results

A comparison of two Convolutional Neural Network models using Conv2D layers optimized with Ghost Module produces significant differences. This difference can be seen from the pattern of accuracy or loss that is difficult to stabilize and fluctuations increase towards several epochs when considering training and validation data. The Conv2D model (Fig. 6) in the accuracy and loss graph has four sharp fluctuations in several epochs. The increasing loss and accuracy fluctuations indicate that the model chooses unstable predictions under certain conditions. This pattern indicates the difficulty of the model in identifying patterns in data that are consistent for each epoch in training and validation data. This condition affects the process to remain stable so that the model cannot generalize to new data.

When given the Ghost Module (Fig. 7), some fluctuations in the learning process appear better. The number of sharp fluctuations decreases from four to two, and all remaining conditions are flatter. This effect indicates that the Ghost Module can make the model stable for each epoch so that the trend of decreasing loss becomes more stable and accuracy increases continuously. The remaining training and validation data provide smoother and less extreme patterns like the Conv2D model so that it can detect better patterns each epoch. In addition, the validation data also looks better by decreasing the trend in several epochs. Accuracy increases more stably, and the loss decreases, making the learning model more effective on data without overfitting training and validation data. Overall, Ghost Module can provide a stable model and reduce sharp fluctuations as evidenced by more stable accuracy and decreased loss in training and validation datasets.

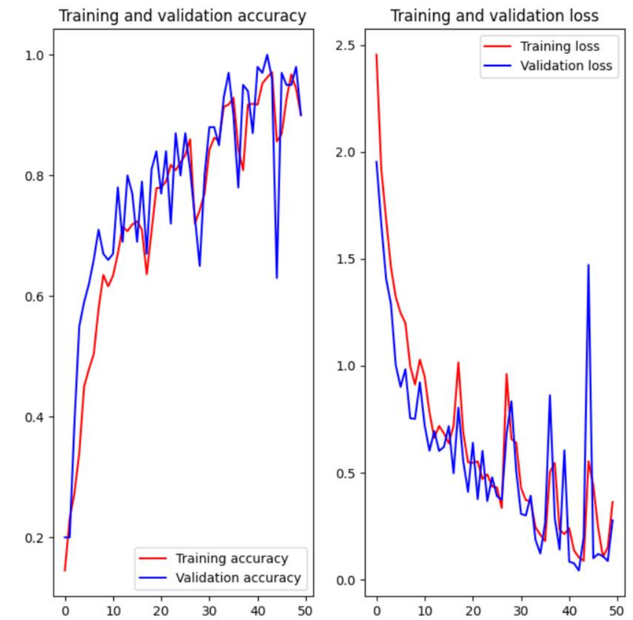

Fig. 6. Training and validation performance, Conv2D.

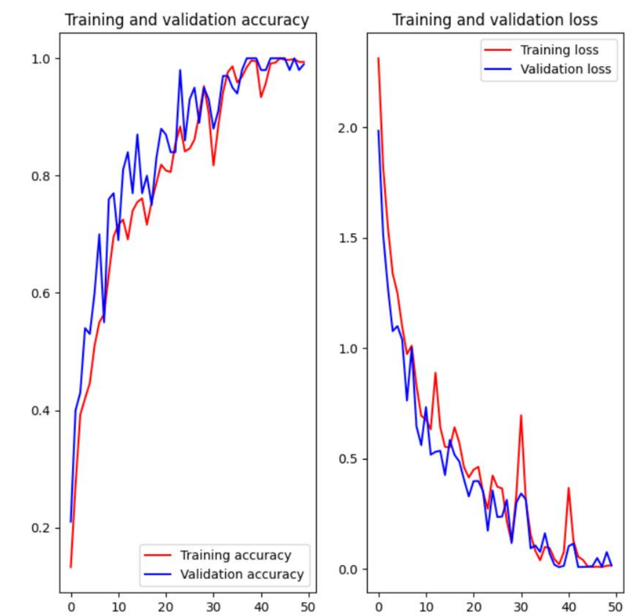

Fig. 7. Training and validation performance, Ghost Feature Maps (*r* = 2).

Furthermore, based on Fig. 8, the accuracy and loss results for the training and validation process of the Ghost Module model with a ratio of 3 are similar to the Ghost Module model with a ratio of 2. However, at ratio 3, the graph shows slightly inconsistent accuracy and loss fluctuation in training and validation. This occurs at epochs 40 to 50, where a gap begins to widen slightly in the accuracy and loss values for the training and validation processes. From the training model graph results, it can be seen that this model can still learn the Palembang Songket motif data well, where there are still similarities in the fluctuations in accuracy and loss values with the Ghost module ratio 2.

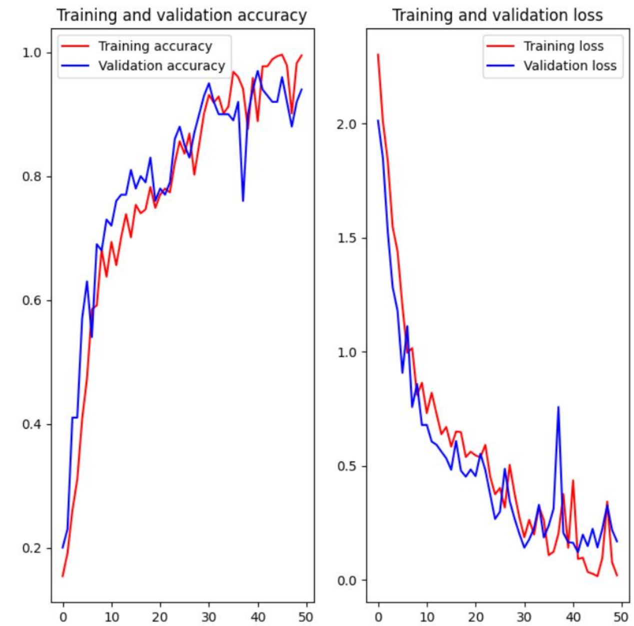

Fig. 8. Training and validation performance, Ghost Feature Maps (*r* = 3).

For the performance results of the Ghost module with a ratio of 4 in Fig. 9, the fluctuations in loss values for the training and validation processes begin to increase. At least starting from epochs 10 to 20, there have been quite large fluctuations, but for the following epochs up to 50, the loss value begins to decrease. Compared to ratio 3, the Ghost module for ratio 4 tends to be less stable at epochs 30 to 50. In this epoch range, a gap begins to move away from the accuracy and loss values. The results of ratio four show that the training model began to experience a decrease in performance for the classification of Palembang Songket motifs compared to ratio 3.

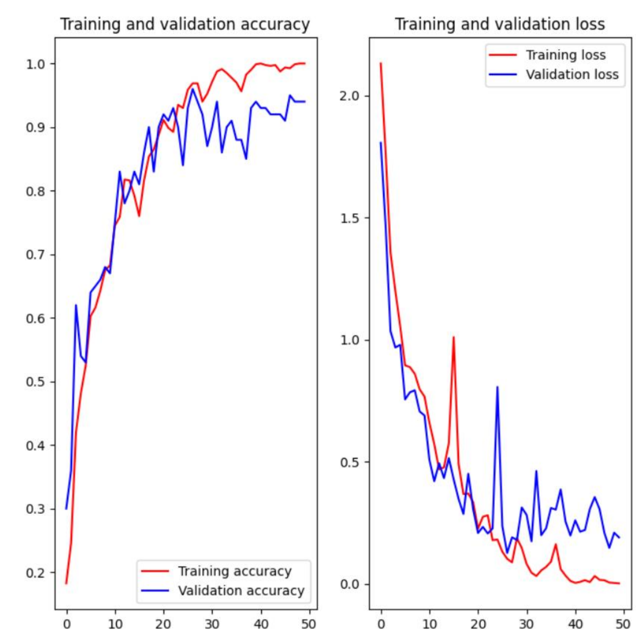

Fig. 9. Training and validation performance, Ghost Feature Maps (*r* = 4).

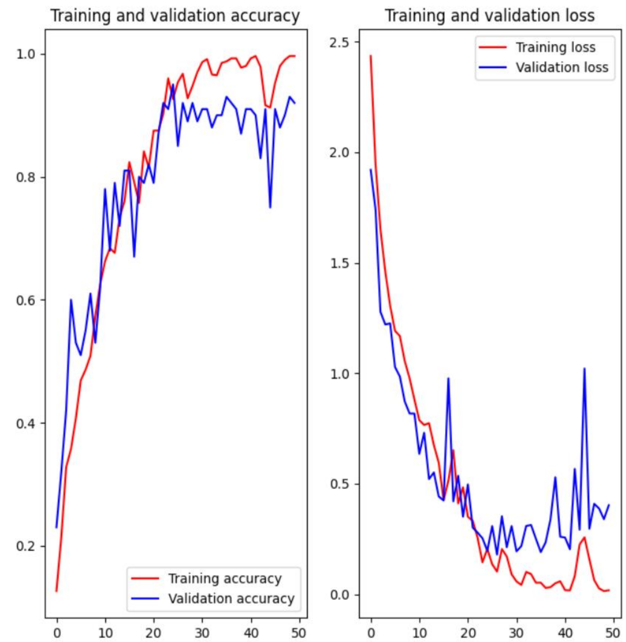

Fig. 10. Training and validation performance, Ghost Feature Maps (r = 5).

At ratio five shown in Fig. 10, the performance of the Ghost Module shows a wider gap for accuracy and loss values. This is almost the same as ratio four, where the condition of the training model experienced quite large fluctuations from epoch 30 to 50. However, more significant fluctuations occurred in epoch 40 to 50 compared to ratios 2, 3, and 4. This is due to a significant reduction in the number of parameters in the Ghost module, so it cannot capture the pattern of the Palembang Songket motif features.

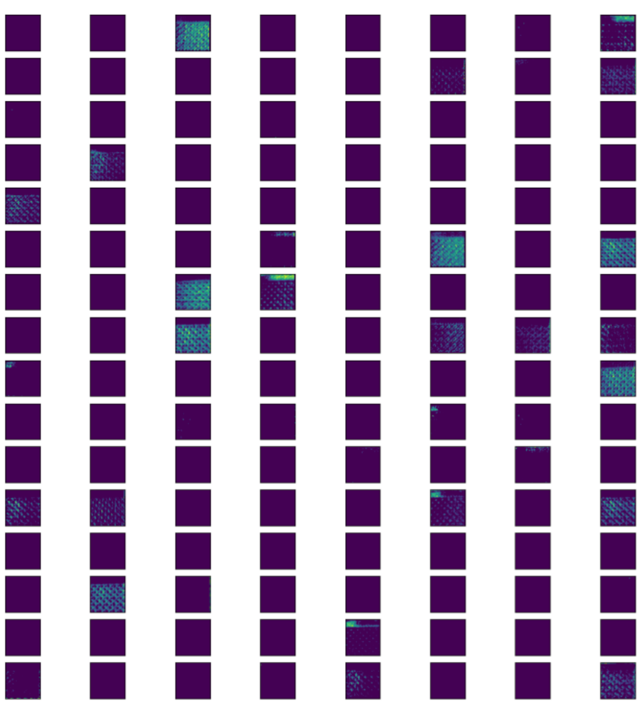

Fig. 11. Visualization of Conv2D feature maps.

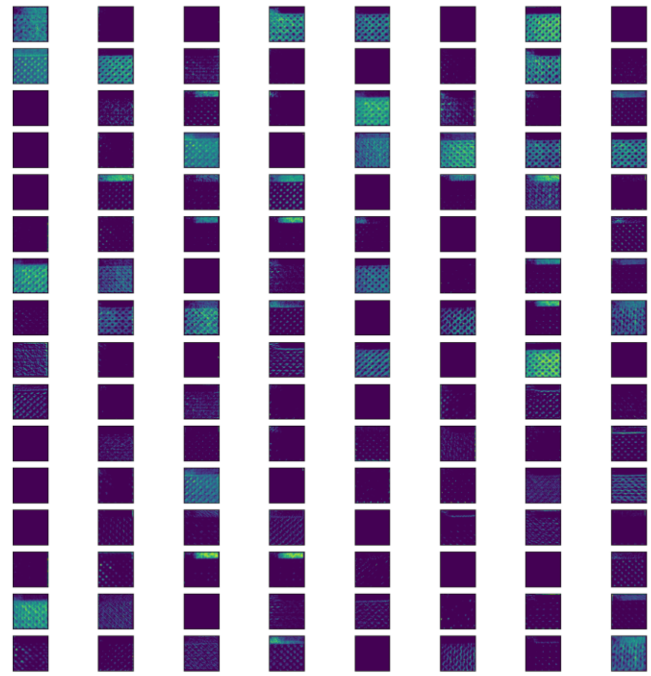

Fig. 12. Visualization of Ghost Feature Maps.

Another way to test the model is to visualize the feature maps and see what kind of features the model has learned from the Palembang Songket motif image. The visualization results are then compared with the CNN model with standard Conv2D and Ghost Feature Maps. As shown in Fig. 11, the visualization results of the feature maps that the CNN model with Conv2D produced have a lot of dark spots and only a few obvious motif feature patterns using this trained model. It turns out that even with the trained model it is still hard to get the Palembang Songket motif pattern.

As seen in Fig. 12, unlike the regular CNN model with Conv2D, the CNN model with Ghost Feature Maps can actually yield clearer feature maps than the regular CNN model. As evidence by the feature map of the Ghost feature map, it performs better than the normal Conv2D. The CNN model with ghost really brings out the Songket motif pattern in many places. The shape and edge features are more distinct than in the regular CNN model. This is good because it will allow the model to learn to recognize the intricate Palembang Songket pattern much more accurately than the old Conv2D model.

Table I shows the results of the comparison of the classification performance of the CNN model with Ghost Feature for 10 classes of Palembang Songket motifs. Compared to CNN with Conv2D, CNN with Ghost feature (ratio 2) is able to provide better classification performance. This is proven by the increase in accuracy, precision, recall, and f1-score for all classes of Palembang Songket motifs. For the Songket Bintang Berantai motif, there was an increase in accuracy from 0.94 to 0.98, precision from 0.63 to 0.83, recall remained at 1.00, and f1-score from 0.77 to 0.91. Furthermore, the Bunga Jatuh motif increased with accuracy from 0.99 to 1.00, precision remained at 1.00, recall from 0.90 to 1.00, and f1-score from 0.95 to 1.00. Then, the Kenanga Makan Ulat motif increased with accuracy from 0.98 to 1.00; precision remained at 1.00, recall from 0.80

to 1.00, and f1-score from 0.89 to 1.00. After that, for the Naga Besaung motif, accuracy increased from 0.95 to 0.98, precision from 0.86 to 1.00, recall from 0.60 to 0.80, and f1-score from 0.71 to 0.89. Then, for the Nampan Perak motif, there was an increase in accuracy from 0.98 to 1.00, precision from 0.90 to 1.00, recall from 0.90 to 1.00, and f1-score from 0.90 to 1.00. The rest, the Bunga Cina, Cantik Manis, Jando Beraes, Pacar Cina, and Pulir motifs, did not increase because they were already very well recognized.

Table II shows the results of the overall comparison of the performance of the CNN model with Ghost Feature. The comparison consists of total parameters, overall accuracy, and model size. The total parameters referred to are the overall parameters used from the input layer to the output layer, namely ghost feature maps, flatten, up to the fully connected layer. Based on the results in Table II, it was found that the CNN model with Ghost Feature provided an overall accuracy of 0.98 with fewer total parameters compared to the CNN model with Conv2D which had an overall accuracy of 0.92. In addition, the CNN model with Ghost Feature ratios of 4 and 5 was able to provide the same overall accuracy of 0.93 with much more reduced parameters with a smaller model size.

TABLE I. COMPARISON OF MODEL PERFORMANCE BASED ON PALEMBANG SONGKET MOTIF CLASS

Songket Motif	Accuracy		Precision		Recall		F1 – Score	
	Conv2D	*Ghost Feature*	*Conv2D*	*Ghost Feature*	*Conv2D*	*Ghost Feature*	*Conv2D*	*Ghost Feature*
Bintang Berantai	0.94	0.98	0.63	0.83	1.00	1.00	0.77	0.91
Bunga Cina	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Bunga Jatuh	0.99	1.00	1.00	1.00	0.90	1.00	0.95	1.00
Cantik Manis	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Jando Beraes	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Kenanga Makan Ulat	0.98	1.00	1.00	1.00	0.80	1.00	0.89	1.00
Naga Besaung	0.95	0.98	0.86	1.00	0.60	0.80	0.71	0.89
Nampan Perak	0.98	1.00	0.90	1.00	0.90	1.00	0.90	1.00
Pacar Cina	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Pulir	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

TABLE II. COMPARATIVE RESULTS OF THE PROPOSED METHOD

Model	Total Parameters	Overall Accuracy	Model Size (MB)
CNN with Conv2D	68,060,746	0.92	259.63
CNN with Ghost Feature ($r = 2$)	67,721,898	**0.98**	258.34
CNN with Ghost Feature ($r = 3$)	44,983,411	0.95	171.60
CNN with Ghost Feature ($r = 4$)	33,909,850	0.93	129.36
CNN with Ghost Feature ($r = 5$)	27,061,049	0.93	103.23

TABLE III. METHOD COMPARISON

Authors	Methods	Dataset	Overall Accuracy
Ariessaputra et al. [7]	CNN (Conv2D, MaxPooling, Fully Connected)	Lombok Songket Motifs	0.84
Hambali et al. [8]	CNN (Conv2D, MaxPooling, Dropout, Fully Connected)	Lombok Songket Motifs	0.86
Ours	**CNN with Ghost Feature ($r = 2$)**	**Songket Palembang Motifs**	**0.98**

B. Discussion

A comparison was also conducted with two previous studies on Songket motif classification using CNN-based methods, as summarized in Table III. Ariessaputra et al. [7] utilized a CNN architecture comprising Conv2D, Max Pooling, and Fully Connected layers, while Hambali et al. [8] implemented a CNN model with additional Dropout layers to enhance generalization. Both approaches leveraged CNN's feature extraction and classification capability for traditional fabric patterns. These studies highlight the relevance of CNN architectures for motif recognition, providing a contextual foundation for evaluating the proposed method's design and performance.

Each method has its strengths and limitations. However, the proposed approach, which retains the basic CNN architecture with Dropout but replaces Conv2D and Max Pooling with Ghost Feature maps involving the Ghost Module and Max Pooling, shows improved performance over the previous studies. The hierarchical combination of Ghost Module and Max Pooling in the proposed method leads to better classification results compared to the state-of-the-art methods from Ariessaputra et al. [7] and Hambali et al. [8]. This improvement highlights the effectiveness of the proposed method in enhancing motif recognition accuracy.

In addition to key parameters such as the learning rate and batch size, the experiment also involved setting a Dropout rate of 0.5 in the dense layer. This configuration aimed to improve the model's generalization in recognizing complex and diverse Songket motifs. During training, Dropout randomly deactivates units, helping the model learn feature representations that are more adaptive to the distinctive patterns of Songket motifs, such as geometric curves and overlapping color variations.

The proposed model addresses the limitations of previous approaches by integrating the Ghost Module and a hierarchical combination of Max Pooling, which significantly enhances the efficiency of dominant feature extraction without losing the primary characteristics of motif patterns. Unlike conventional convolutional layers that rely on a large number of parameters, this model leverages the Ghost Module for a more lightweight feature generation mechanism, while the hierarchical integration of Max Pooling reduces redundancy and ensures a more focused feature extraction process. This approach not only improves efficiency in filtering less relevant features but also strengthens the model's ability to capture intricate patterns, especially in motifs with significant visual similarities and minor variations. With this structure, the model demonstrates enhanced efficiency and effectiveness in recognizing Songket motifs.

V. Conclusion

The use of Ghost feature maps, which involve the Ghost module, in the CNN model leads to a significant reduction in the number of parameters and the model size compared to traditional CNNs utilizing Conv2D. This efficiency is highlighted by the model achieving an impressive accuracy of 0.98 at a ratio of 2, with only a minor parameter reduction of about 0.5% and a slightly smaller model size. However, as the ratio of Ghost feature maps increases, a further decrease in parameters and model size occurs, accompanied by a decline in accuracy. Specifically, a ratio of 3 results in a 34% reduction in parameters but lowers accuracy to 0.95. Ratios 4 and 5 stabilize accuracy at 0.93 while achieving over 60% reductions in model size and parameters compared to the Conv2D model. Thus, a trade-off between accuracy and model size becomes evident, particularly at a ratio of 3, where significant size reductions are achieved with only a slight impact on accuracy.

The proposed Ghost Feature maps in this model are constructed hierarchically through a combination of Ghost Modules and Max Pooling, applied four times. Each pair forms a pyramid-like feature representation, allowing the model to learn increasingly complex feature hierarchies as the depth of the layers increases. However, the optimal number of feature repetition levels required for achieving the best performance remains to be explored. Future research should investigate whether adding deeper hierarchical layers could reduce performance or significantly improve recognition accuracy. Therefore, further development of deeper architectures and evaluation at various layer depths is necessary to determine whether this approach can significantly improve Songket motif recognition.

Acknowledgment

This research was supported by Ministry of Education, Culture, Research, and Technology of Indonesia through the research grant for the fundamental research scheme in 2024 with contract number 104/E5/PG.02.00.PL/2024.

References

[1] A. Suzianti, R. D. Amaradhanny, and S. N. Fathia, "Fashion heritage future: Factors influencing Indonesian millenials and generation Z's interest in using traditional fabrics," J. Open Innov. Technol. Mark. Complex., vol. 9, no. 4, p. 100141, Dec. 2023, doi: 10.1016/j.joitmc.2023.100141.

[2] K. Sedyastuti, E. Suwarni, D. R. Rahadi, and M. A. Handayani, "Human Resources Competency at Micro, Small and Medium Enterprises in Palembang Songket Industry," in Proceedings of the 2nd Annual Conference on Social Science and Humanities (ANCOSH 2020), 2021, pp. 248–251. doi: 10.2991/assehr.k.210413.057.

[3] "Songket Palembang," Warisan Budaya Takbenda Indonesia, 2013. https://budaya-data.kemdikbud.go.id/wbtb/objek/AA000222 (accessed Oct. 10, 2024).

[4] Y. Yullyana, D. Irmayani, and M. N. S. Hasibuan, "Content-Based Image Retrieval for Songket Motifs using Graph Matching," Sinkron, vol. 7, no. 2, pp. 714–719, May 2022, doi: 10.33395/sinkron.v7i2.11411.

[5] S. Sriani, M. S. Hasibuan, and R. Ananda, "Classification of Batu Bara Songket Using Gray-Level Co-Occurrence Matrix and Support Vector Machine," J. Ris. Inform., vol. 5, no. 1, pp. 481–490, Dec. 2022, doi: 10.34288/jri.v5i1.469.

[6] R. Aprianti, K. Evandari, R. A. Pramunendar, and M. Soeleman, "Comparison Of Classification Method On Lombok Songket Woven Fabric Based On Histogram Feature," in 2021 International Seminar on Application for Technology of Information and Communication (iSemantic), Sep. 2021, pp. 196–200. doi: 10.1109/iSemantic52711.2021.9573223.

[7] S. Ariessaputra, V. H. Vidiasari, S. M. Al Sasongko, B. Darmawan, and S. Nababan, "Classification of Lombok Songket and Sasambo Batik Motifs Using the Convolution Neural Network (CNN) Algorithm," JOIV Int. J. Informatics Vis., vol. 8, no. 1, pp. 38–44, Mar. 2024, doi: 10.62527/joiv.8.1.1386.

[8] H. Hambali, M. Mahayadi, and B. Imran, "Classification of Lombok Songket Cloth Image Using Convolution Neural Network Method (CNN)," J. Pilar Nusa Mandiri, vol. 17, no. 2, pp. 149–156, 2021, doi: https://doi.org/10.33480/pilar.v17i2.2705.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2016. doi: 10.1109/CVPR.2016.90.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Advances in Neural Information Processing Systems, 2012, vol. 25. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf

[11] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," 3rd Int. Conf. Learn. Represent. ICLR 2015, pp. 1–14, 2015.

[12] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "GhostNet: More Features From Cheap Operations," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2020, pp. 1577–1586. doi: 10.1109/CVPR42600.2020.00165.

[13] R. Andrian, R. Taufik, D. Kurniawan, A. S. Nahri, and H. C. Herwanto, "Lampung Batik Classification Using AlexNet, EfficientNet, LeNet and MobileNet Architecture," Int. J. Adv. Comput. Sci. Appl., vol. 15, no. 11, 2024, doi: 10.14569/IJACSA.2024.0151191.

[14] L. Elvitaria, E. F. A. Shaubari, N. A. Samsudin, S. K. A. Khalid, S. -, and Z. Indra, "A Proposed Batik Automatic Classification System Based on Ensemble Deep Learning and GLCM Feature Extraction Method," Int. J. Adv. Comput. Sci. Appl., vol. 15, no. 10, 2024, doi: 10.14569/IJACSA.2024.0151058.

[15] R. Muliono, M. S. Iranita, and R. B. Syah, "An Effectivity Deep Learning Optimization Model to Traditional Batak Culture Ulos Classification," Int. J. Adv. Comput. Sci. Appl., vol. 14, no. 4, pp. 634–638, 2023, doi: 10.14569/IJACSA.2023.0140469.

[16] Z. Wang and T. Li, "A Lightweight CNN Model Based on GhostNet," Comput. Intell. Neurosci., vol. 2022, pp. 1–12, Jul. 2022, doi: 10.1155/2022/8396550.

[17] Q. Zhao and H. Cheng, "An Efficient Approach to Human Security Screening Image Recognition Through a Lightweight CNN Utilizing Yolov5s and GhostNet," Trait. du Signal, vol. 40, no. 4, pp. 1653–1660, Aug. 2023, doi: 10.18280/ts.400433.

[18] W. Chen, Y. Zhang, X. Chen, and W. Li, "Research On Vehicle Detection based on Ghost Net and Se Attention Mechanism," in 2023 11th International Conference on Information Technology: IoT and Smart City (ITIoTSC), Aug. 2023, pp. 268–271. doi: 10.1109/ITIoTSC60379.2023.00055.

[19] Z. Huangfu, S. Li, and L. Yan, "Ghost-YOLO v8: An Attention-Guided Enhanced Small Target Detection Algorithm for Floating Litter on Water Surfaces," Comput. Mater. Contin., vol. 80, no. 3, pp. 3713–3731, 2024, doi: 10.32604/cmc.2024.054188.

[20] B. Fang, G. Chen, and J. He, "Ghost-based Convolutional Neural Network for Effective Facial Expression Recognition," in 2022 International Conference on Machine Learning and Knowledge Engineering (MLKE), Feb. 2022, pp. 121–124. doi: 10.1109/MLKE55170.2022.00029.

[21] M. Alansari, O. A. Hay, S. Javed, A. Shoufan, Y. Zweiri, and N. Werghi, "GhostFaceNets: Lightweight Face Recognition Model From Cheap Operations," IEEE Access, vol. 11, pp. 35429–35446, 2023, doi: 10.1109/ACCESS.2023.3266068.

[22] F. Luan, X. Mu, and S. Yuan, "Ghost Module Based Residual Mixture of Self-Attention and Convolution for Online Signature Verification," Comput. Mater. Contin., vol. 79, no. 1, pp. 695–712, 2024, doi: 10.32604/cmc.2024.048502.

[23] M. E. Paoletti, J. M. Haut, N. S. Pereira, J. Plaza, and A. Plaza, "Ghostnet for Hyperspectral Image Classification," IEEE Trans. Geosci. Remote Sens., vol. 59, no. 12, pp. 10378–10393, Dec. 2021, doi: 10.1109/TGRS.2021.3050257.

[24] Y. Tang, K. Han, J. Guo, C. Xu, C. Xu, and Y. Wang, "GhostNetV2: Enhance Cheap Operation with Long-Range Attention," in Advances in Neural Information Processing Systems, Nov. 2022, pp. 9969–9982. [Online]. Available: http://arxiv.org/abs/2211.12905

[25] Z. Liu, Z. Hao, K. Han, Y. Tang, and Y. Wang, "GhostNetV3: Exploring the Training Strategies for Compact Models," 2024, [Online]. Available: http://arxiv.org/abs/2404.11202

[26] Z. He, D. He, X. Li, and R. Qu, "Blind Superresolution of Satellite Videos by Ghost Module-Based Convolutional Networks," IEEE Trans. Geosci. Remote Sens., vol. 61, pp. 1–19, 2023, doi: 10.1109/TGRS.2022.3233099.

[27] S. Liu, B. Ding, J. Bai, and Z. Xiao, "Hyperspectral Image Classification Based on Extended Morphological Profile Features and Ghost Module," in 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Jul. 2021, pp. 3617–3620. doi: 10.1109/IGARSS47720.2021.9554092.

[28] Y. Huang, Y. Zhou, J. Lan, Y. Deng, Q. Gao, and T. Tong, "Ghost Feature Network for Super-Resolution," in 2020 Cross Strait Radio Science & Wireless Technology Conference (CSRSWTC), Dec. 2020, pp. 1–3. doi: 10.1109/CSRSWTC50769.2020.9372549.

[29] A. Ghosh, A. Sufian, F. Sultana, A. Chakrabarti, and D. De, "Fundamental Concepts of Convolutional Neural Network," in Recent Trends and Advances in Artificial Intelligence and Internet of Things, V. E. Balas, R. Kumar, and R. Srivastava, Eds. Cham: Springer International Publishing, 2020, pp. 519–567. doi: 10.1007/978-3-030-32644-9_36.

[30] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," J. Big Data, vol. 8, no. 1, p. 53, 2021, doi: 10.1186/s40537-021-00444-8.

[31] P. Wlodarczak, Machine Learning and its Applications. University of Southern Queensland, Toowoomba, Queensland, Australia: CRC Press, 2020.