Ghost-Vanilla Feature Maps: A Novel Hybrid Architecture for Efficient Fine-Grained Songket Motif Classification

Yohannes, Muhammad Ezar Al Rivan, Siska Devella, Tinaliah
Informatics, Universitas Multi Data Palembang, Palembang, Indonesia

***Abstract*—South Sumatra songket motifs present a challenging fine-grained classification task due to high inter-class similarity and substantial intra-class variability. This study proposes the Ghost-Vanilla Feature Map, a novel hybrid architecture that integrates low-cost ghost-generated features with the lightweight structural stability of VanillaNet to enhance discriminative feature learning while reducing computational burden. The proposed architecture is designed to address the inefficiency of conventional convolution-heavy networks in capturing subtle motif variations. Experimental evaluation on a dataset comprising 20 songket motif classes demonstrates that a ghost ratio 2 achieves the best trade-off, attaining an accuracy of 0.98 with more than 75% parameter reduction. Increasing the ghost ratio to 3 preserves high classification performance with an accuracy of 0.97, while ratios 4 and 5 further reduce model size at the expense of marginal accuracy degradation. Comparative results indicate that the Ghost-Vanilla Feature Map consistently outperforms lightweight CNN baselines, including MobileNetV3-Small, MobileNetV4-Conv-Small, EfficientNetV2-Small, and ShuffleNetV2. The proposed architecture substantially surpasses the Vanilla-only baseline, which achieves an accuracy of only 0.860 despite requiring 30.19 million parameters, highlighting the limitations of conventional convolution-dominant designs in fine-grained textile classification. The hybrid configuration with a ghost ratio 2 delivers superior accuracy while nearly halving the parameter count and significantly reducing computational overhead. Overall, the Ghost-Vanilla Feature Map provides an efficient and highly discriminative solution for fine-grained songket motif classification, achieving strong performance while substantially reducing model complexity through a balanced hybrid representation.**

Keywords*—Ghost Module; fine-grained classification; lightweight deep learning; songket motif classification; VanillaNet

I. Introduction

South Sumatra songket is an intangible cultural heritage with symbolic, aesthetic, and philosophical value that dates back to the Sriwijaya era and the Palembang Darussalam Sultanate. This fabric functions not only as traditional attire but also as a marker of social identity, status legitimacy, and a medium of cultural expression in ceremonial contexts. The intricate geometric patterns, floral and faunal ornaments, and symbolic details make songket an artifact of remarkable visual complexity. This complexity underscores the urgent need for accurate and efficient digital documentation and classification systems for songket motifs [1], [2].

With rapid technological advancements, deep learning serves as one of the main strategies for motif classification. Research on Lombok songket, for instance, has employed ResNet50V2 with AdamW and adaptive transfer learning [3], conventional CNNs [4], and comparative models such as AlexNet and VGG19 [5]. Additional studies have explored feature-based approaches for batik motifs [6], [7] and CNN-driven approaches for Lombok songket, further illustrating the breadth of computational techniques applied to traditional motif recognition [8], [9], [10]. While these approaches achieved promising accuracy, they remained computationally demanding and inefficient for deployment on resource-limited devices. This highlights the need for lightweight architectures that maintain performance without imposing heavy computational costs in traditional textile motif recognition.

A more targeted study on Palembang songket introduced hierarchical Ghost feature maps with a pooling strategy, successfully reducing parameter counts, model size, and computation time [11]. Another study utilized regularization and ResNet-based augmentation with dropout to mitigate overfitting [12]. Although both approaches improved performance, they were still restricted to conventional or slightly modified CNNs. As a result, their feature representations were not fully optimized to address the inherent visual complexity of songket motifs.

Songket motif recognition inherently aligns with the characteristics of fine-grained image classification, where inter-class differences are subtle and often localized within dense geometric or ornamental structures. Such tasks demand feature representations capable of capturing minute variations that distinguish one motif from another, particularly when many motif classes share similar macro-patterns. This fine-grained nature further reinforces the need for architectures that balance discriminative precision with computational efficiency [13].

Despite their promising results, existing approaches exhibit specific limitations that hinder practical deployment. ResNet50V2-based methods [3], while achieving high accuracy, suffer from excessive parameter counts (>23M parameters) and substantial computational overhead. AlexNet and VGG19 [5] also impose high memory footprints (>138M and >528M parameters, respectively). MobileNetV3 [14], though designed for mobile efficiency through neural architecture search, incorporates architectural complexity that complicates interpretability and training. EfficientNetV2 [15] requires careful scaling-coefficient tuning and remains computationally intensive during training. ShuffleNetV2 [16], while achieving efficiency through its channel-split and channel-shuffle operations, faces difficulty in capturing fine-grained textural details due to its limited representational capacity and the insufficient discriminative strength of its channel-shuffling mechanism. Meanwhile, the hierarchical Ghost approach [11] still relies on pooling strategies that may discard subtle discriminative details, and ResNet-based augmentation [12] focuses on reducing overfitting without addressing fundamental architectural inefficiencies. Crucially, none of these methods combine efficient feature generation with a minimalist backbone design, resulting in either computational redundancy or insufficient representational depth when dealing with the intricate geometric structures of songket motifs.

This issue becomes particularly relevant given that the songket dataset is relatively balanced. The main challenge is not class imbalance but rather the motifs' intrinsic complexity. Subtle geometric similarities, ornamental variations, and fine-grained textures lead to high inter-class similarity alongside significant intra-class variability. These challenges necessitate feature representations that are more discriminative, moving beyond simply deepening networks or increasing parameter counts. Therefore, new approaches are required that simultaneously balance computational efficiency and representational power.

Within the domain of lightweight architectures, two notable methods hold strong potential: the Ghost Module and VanillaNet. The Ghost Module generates additional feature maps using inexpensive convolution operations, thereby reducing computational complexity [17], [18]. Meanwhile, VanillaNet embodies a minimalist design philosophy, employing very few parameters while maintaining competitive accuracy across multiple domains [19], [20], [21]. Both approaches have shown effective performance in diverse applications, including facial recognition [18], multimodal medical image classification [22], road damage detection [20], and underwater crack classification [21].

Hybrid studies further highlight the value of integrating lightweight architectural principles. For instance, a Ghost-convolution–enlightened Transformer improves grape leaf disease diagnosis by combining Ghost efficiency with transformer-level representation capacity [23], while Van-DETR integrates VanillaNet with advanced feature fusion to enhance real-time object detection [24]. Recent advancements in re-parameterization for lightweight Vanilla-based Vision Transformers also demonstrate how combining structural simplicity with adaptive computations can improve accuracy without increasing complexity [25]. These findings suggest that combining Ghost Modules with VanillaNet has strong potential to yield architectures that are both lightweight and highly discriminative, especially for complex textile motifs.

Ongoing developments in Ghost-based architectures, such as GhostNet [17], GhostNetV2 [26], GhostNetV3 [27], GhostFaceNets [18], GCNN [28], Ghost-YOLOv5 [29], and Ghost-YOLOv8 [30], continue to demonstrate the strength of efficient feature generation. Their applicability has also been demonstrated across a wide range of practical domains, including lung nodule detection [31], guava fruit detection in complex orchard environments [32], and optimized fruit classification using enhanced deep learning strategies [33]. In agricultural contexts, improved lightweight YOLO-based architectures have further demonstrated strong performance in real-time detection of multi-stage apple fruit in complex environments [34]. Concurrently, VanillaNet has proven adaptable across domains such as prostate zone segmentation in medical imaging [35] and hyperspectral image classification [36]. This strengthens the rationale for integrating these two concepts into a unified model that addresses limitations found in MobileNetV3 [14], EfficientNetV2 [15], ShuffleNetV2 [16], and prior Palembang songket studies [11], [12].

Complementary strategies such as transfer learning, augmentation, and resampling have also shown effectiveness [12], [37], but on a balanced dataset like songket, the most significant improvements are expected to come from architectural-level innovations rather than preprocessing techniques. Consequently, this research prioritizes designing a hybrid architecture capable of extracting highly representative fine-grained features while maintaining computational efficiency.

To address the main research problem of balancing computational efficiency and discriminative capability in fine-grained classification of South Sumatra songket motifs, this study introduces the Ghost-Vanilla Feature Map. This hybrid architecture is explicitly designed to overcome the limitations of existing approaches that either rely on deep, high-complexity networks or adopt lightweight models that fail to capture the subtle geometric and ornamental variations inherent in songket motifs. The Ghost Module is employed to generate additional feature maps at low computational cost, thereby reducing redundancy, while VanillaNet serves as a minimalist backbone that preserves representational depth without introducing structural complexity. This architectural combination provides a clear and principled justification, as it directly targets the intrinsic visual challenges of songket motifs, including high inter-class similarity and significant intra-class variability. The proposed approach offers domain-specific optimization that has not been explicitly addressed by previous Ghost-VanillaNet integrations and contributes to ongoing research on lightweight deep learning architectures for fine-grained classification tasks.

A. Problem Identification

CNN-based models are widely used in textile and motif recognition due to their ability to learn edge, texture, and fundamental shape attributes. For relatively simple patterns, these models are often sufficient. However, South Sumatra songket motifs exhibit high visual complexity, characterized by repetitive geometric structures, delicate floral-faunal ornaments, and symbolic details that produce both strong inter-class similarity and substantial intra-class variability. Although the dataset is balanced, the main challenges arise from subtle motif similarities, variations in scale and rotation, weaving irregularities, and metallic thread reflections that alter texture under different lighting conditions. Conventional CNNs rely heavily on local receptive fields, limiting their capacity to capture long-range dependencies across motif elements. Increasing model depth or parameter size to improve accuracy further exacerbates computational inefficiency, particularly on resource-constrained devices. Previous studies on Palembang Songket [11], [12] have demonstrated these limitations. A single

Ghost Module [11] reduces parameters but produces suboptimal feature representations due to an unoptimized backbone, while ResNet-based approaches [12] remain computationally heavy despite reducing overfitting. Lightweight variants reduce model size and computation time but often fail to optimize fine-grained feature representations necessary for distinguishing highly similar motifs. Therefore, a hybrid architecture is needed, one that combines low-cost feature expansion with a minimalist backbone design to balance efficiency and representational power.

B. Main Contributions

1) Hybrid Ghost-Vanilla Feature Map design: This study introduces the Ghost-Vanilla Feature Map algorithm, which integrates VanillaNet as a lightweight backbone with the Ghost Module to generate additional features via low-cost convolution operations. This integration yields more discriminative feature representations while preserving computational efficiency.

2) Optimized feature extraction for complex motifs: The proposed algorithm specifically addresses challenges such as high visual similarity and fine-grained variation in South Sumatra songket. Ghost-Vanilla Feature Maps balance representational depth and efficiency, unlike conventional CNNs that demand heavy computational resources.

3) Improved efficiency and generalization: By combining the redundancy-reduction capability of the Ghost Module with the structural simplicity of VanillaNet, the proposed model achieves an optimal trade-off between classification accuracy and complexity. This makes it suitable for resource-limited deployment while ensuring robust generalization across diverse motif datasets.

4) Empirical validation on fine-grained classification: Through comprehensive experiments, this study demonstrates that the Ghost-Vanilla integration is effective for fine-grained classification tasks characterized by high inter-class similarity, contributing to broader academic discussions on lightweight architecture design.

II. Dataset Description

A curated dataset of 20 South Sumatran songket motif classes was constructed from six regions: Palembang, Ogan Ilir, Banyuasin, Ogan Komering Ilir, Prabumulih, and PALI (see Fig. 1). Images were collected under a standardized protocol using a fixed 45 cm capture distance, 0-degree frontal angle, uniform illumination, and identical camera settings. All motif labels were verified by a songket expert to ensure authenticity and adherence to traditional weaving standards. Each motif region was cropped to 2048 × 2048 pixels at 300 dpi.

The dataset contained 2,000 images (100 per class), which were resized to 256 × 256 pixels for model input. To avoid data leakage, all images derived from the same motif source were assigned exclusively to a single train, validation, or test partition using motif-level grouped splitting. This ensures that evaluation reflects true generalization rather than memorization of repeated patterns. All images were captured using the same device and lighting configuration to maintain acquisition consistency and prevent device-dependent biases.

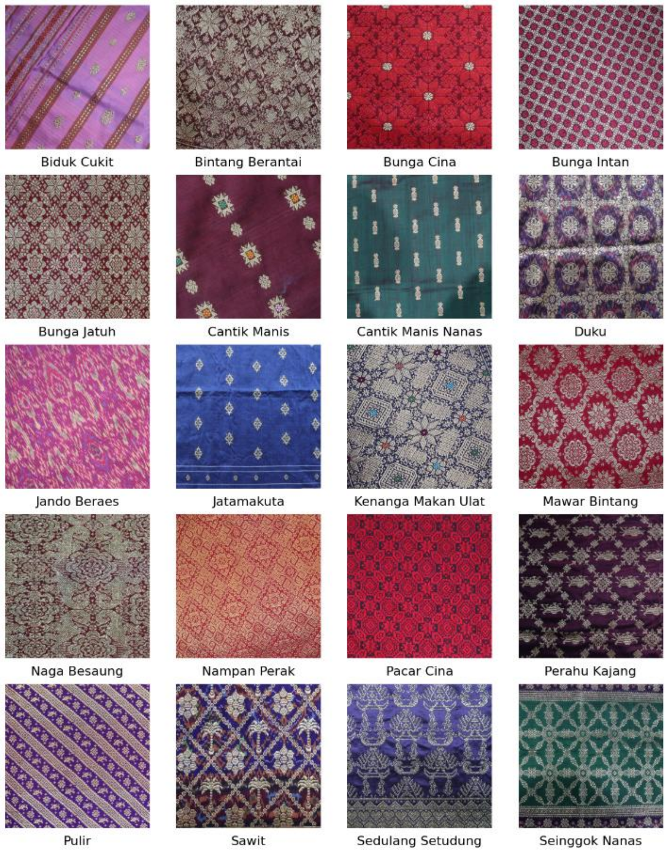


Fig. 1. South Sumatran songket motif dataset.

Although the dataset provides adequate intra-class variation, its size and regional scope remain limited. Future work will expand data collection across additional regions and devices and incorporate external textile datasets to further assess cross-domain robustness.

III. Proposed Method

The proposed model is structured around two core modules: one for feature learning and the other for classification, each addressing distinct aspects of the model's workflow. Fig. 2 presents a schematic overview of the complete architecture.

In the feature learning phase, Ghost-Vanilla Feature Maps are employed, implementing Ghost Feature Maps across four sequential stages, as defined in the VanillaNet-6 architecture. Stage 1 applies a Ghost Module with 1024 channels, followed by max pooling. Stages 2 and 3 expand the channels to 2048 and 4096, progressively capturing more complex textures and motif patterns. Stage 4 maintains 4096 channels to extract high-level semantic features, after which average pooling with a kernel size of 7 compresses the representation into compact feature maps. The output is subsequently flattened into a one-dimensional vector, providing a discriminative input for the classification phase.

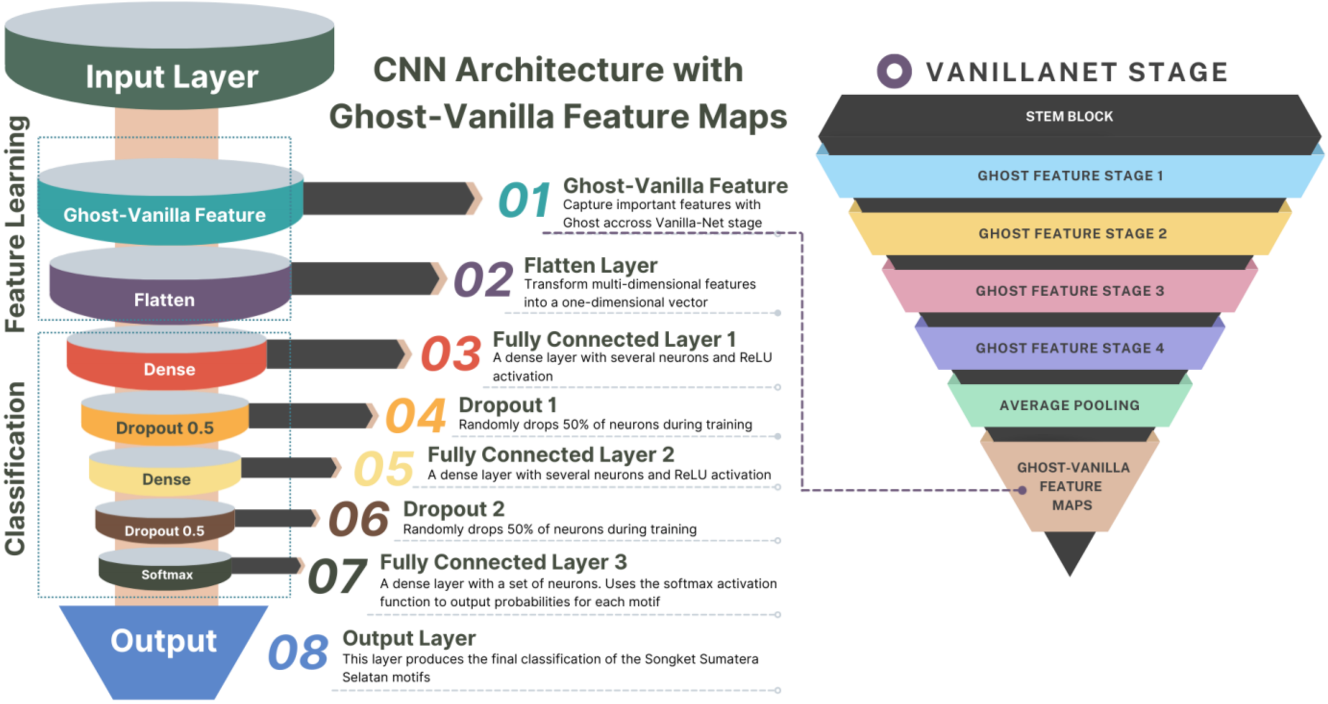

CNN Architecture with
Ghost-Vanilla Feature Maps
Input Layer
Feature Learning
Ghost-Vanilla Feature
Flatten
Classification
Dense
Dropout 0.5
Dense
Dropout 0.5
Softmax
Output
01 Ghost-Vanilla Feature
Capture important features with Ghost accross Vanilla-Net stage
02 Flatten Layer
Transform multi-dimensional features into a one-dimensional vector
03 Fully Connected Layer 1
A dense layer with several neurons and ReLU activation
04 Dropout 1
Randomly drops 50% of neurons during training
05 Fully Connected Layer 2
A dense layer with several neurons and ReLU activation
06 Dropout 2
Randomly drops 50% of neurons during training
07 Fully Connected Layer 3
A dense layer with a set of neurons. Uses the softmax activation function to output probabilities for each motif
08 Output Layer
This layer produces the final classification of the Songket Sumatera Selatan motifs
VANILLANET STAGE
STEM BLOCK
GHOST FEATURE STAGE 1
GHOST FEATURE STAGE 2
GHOST FEATURE STAGE 3
GHOST FEATURE STAGE 4
AVERAGE POOLING
GHOST-VANILLA FEATURE MAPS

Fig. 2. Architecture of the CNN model utilizing Ghost-Vanilla Feature Maps.

The classification phase begins with a dense layer of 512 units activated by ReLU, serving as the first stage of non-linear transformation to project the learned features into a more discriminative subspace. A dropout layer with a rate of 0.5 follows, acting as a regularization mechanism to reduce overfitting by randomly deactivating neurons during training. This design ensures robustness in handling the limited dataset of songket motifs while maintaining model generalization. The second dense layer, with 512 units, further refines the discriminative representations by learning deeper correlations between the extracted features. Another dropout layer with a rate of 0.5 is applied, providing additional regularization and stability during the optimization process. Finally, the output is passed through a dense softmax layer with the number of units equal to the total motif classes, producing a probability distribution across all categories.

A. *Ghost-Vanilla Feature Maps*

- The hierarchical arrangement of the network demonstrates a progressive refinement of feature representations across different stages (see Fig. 3). The initial stem layer, with a 4×4 convolution and a stride of 4, performs aggressive spatial downsampling, ensuring that redundant information is reduced while preserving the essential texture structures of songket motifs. Each subsequent stage is designed to expand the representational capacity through the Ghost Module, which generates feature maps efficiently, followed by MaxPooling layers that further condense the spatial dimensions. The kernel sizes and pooling strategies are carefully chosen to balance the preservation of discriminative motif details with the reduction of computational redundancy, thereby improving the model's efficiency and scalability.
- Furthermore, the sequence from Stage 1 to Stage 4 reflects a conceptual hierarchy of visual processing, transitioning from low-level edges and repetitive weave patterns to more abstract and semantically rich motif structures. The integration of AveragePooling and fully connected layers consolidates the extracted features into a compact and discriminative representation, making it suitable for classification. This conceptual flow aligns with established principles in deep learning architecture design, where convolutional layers combined with pooling progressively transform input images into high-level abstractions that are more separable in the classification space. In this context, the architectural design is specifically adapted to handle the repetitive, highly detailed, and structurally similar characteristics of songket motifs, which demand a balance of depth, resolution reduction, and efficient feature extraction.

1) Stem block: The architecture begins with a Conv2D layer consisting of 512 filters with a 4×4 kernel and stride 4. This configuration simultaneously reduces the spatial dimension of the input and generates initial low-level feature representations related to edges and basic textures. The use of a relatively large stride at this stage accelerates computation while maintaining essential information from high-resolution images.

2) Ghost feature stage 1: Feature Expansion. Stage 1 is composed of a GhostModule with 1024 output channels, kernel size 1×1, and ratio 2, followed by a MaxPooling layer of size

2×2 with stride 2. The GhostModule refines the early representation by applying a channel transformation through the 1×1 kernel, which allows efficient linear projection across channels without increasing spatial complexity. The subsequent pooling operation reduces the spatial resolution and enhances translational invariance, thereby mitigating the sensitivity of the network to motif position variations within the image.

3) *Ghost feature stage 2*: Deep Feature Extraction. Stage 2 employs a GhostModule with 2048 output channels and a 1×1 kernel, followed by a MaxPooling operation of size 2×2 with a stride of 2. Increasing the number of channels at this stage enables the network to capture more complex mid-level features, which are particularly relevant for distinguishing motif classes with subtle structural similarities. Pooling at this stage contributes to the reduction of spatial dimensions while reinforcing the ability of the network to preserve dominant features in a compact and discriminative representation.

4) *Ghost feature stage 3*: High-Dimensional Encoding. Stage 3 integrates a GhostModule with 4096 channels and a 1×1 kernel, followed by a 2×2 pooling operation with a stride of 2. The significant increase in the number of channels allows the construction of higher-level feature abstractions, where the representation is a nonlinear composition of multiple mid-level features extracted earlier. Spatial downsampling ensures compactness of the representation while retaining global contextual information. This aligns with the hierarchical representation theory in CNNs, where deeper layers capture increasingly complex semantic concepts built upon simpler features from earlier stages.

5) *Ghost feature stage 4*: Final Abstraction. Stage 4 consists solely of a GhostModule with 4096 channels. The absence of pooling in this stage allows the model to preserve the full resolution of channel-level features, ensuring that high-dimensional abstractions are maintained before transitioning into the classification process. This design emphasizes the semantic correlation between channels, which forms the final abstraction layer of the feature hierarchy.

6) *Feature maps and global representation*: Following the stacked stages, an AveragePooling layer with kernel size 7×7 aggregates spatial information into a global vector representation. This operation summarizes consistent features across the entire image and minimizes dependence on specific spatial locations. The subsequent Flatten layer transforms the pooled feature map into a one-dimensional vector suitable for processing by fully connected layers.

B. *Parameter Distribution of Feature Maps*

The comparison between Fig. 4 and Fig. 5 highlights the impact of architectural design on model complexity and efficiency. Fig. 4 illustrates the Vanilla Feature Maps, which are adopted from the VanillaNet-6 structure [19] and rely entirely on Conv2D layers. This design yields approximately 27.8 million parameters, with the majority concentrated in the deeper convolutional stages. While this configuration provides very high representational capacity, it also introduces significant computational overhead.

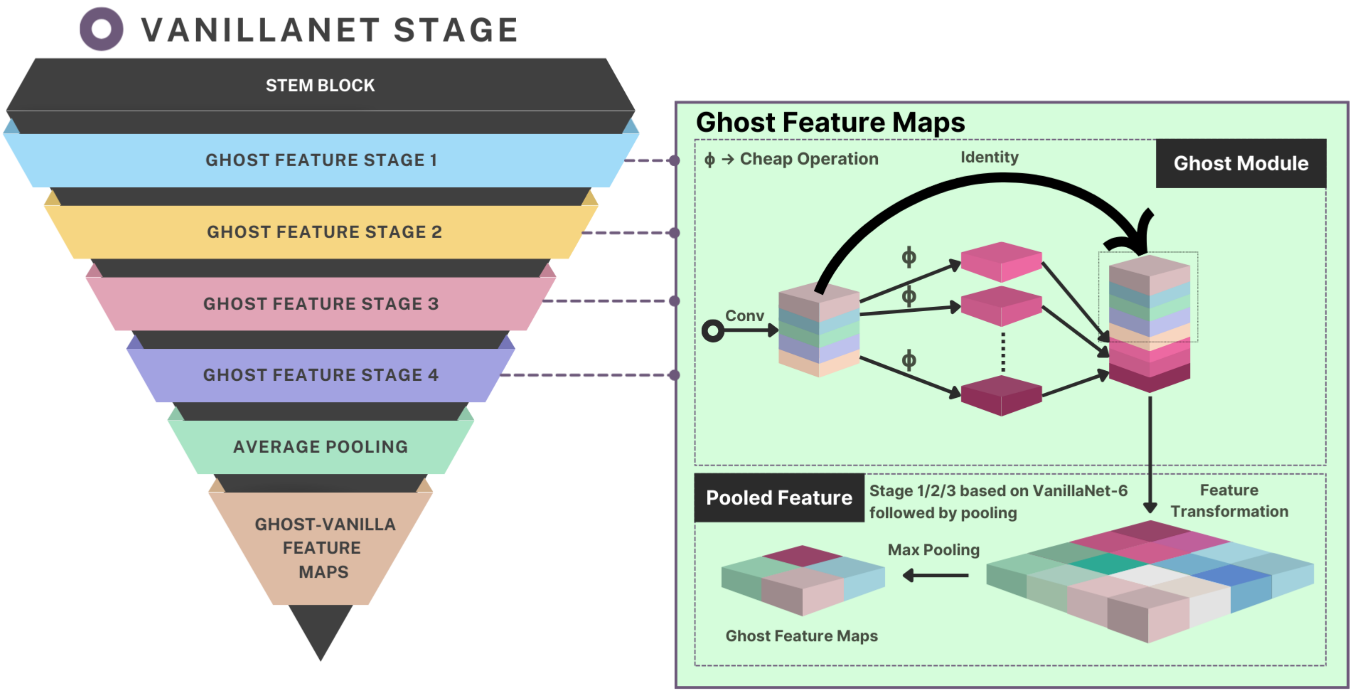
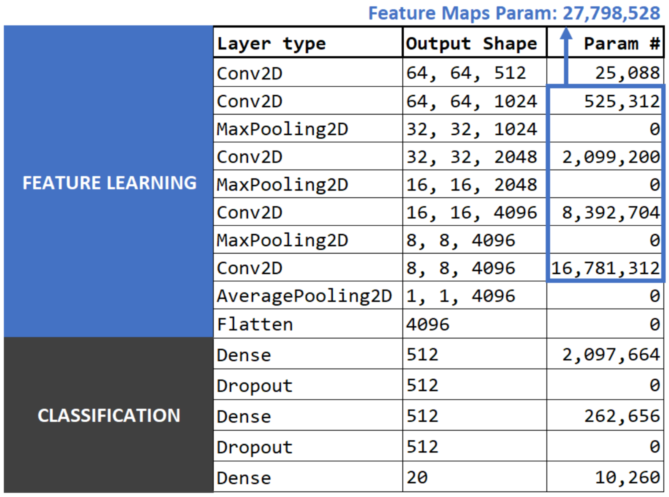


Fig. 3. Architecture design of Ghost-Vanilla Feature Maps.



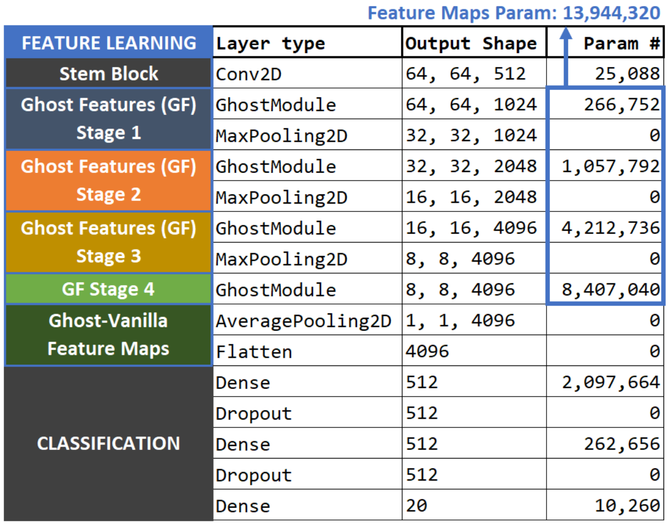
Feature Maps Param: 27,798,528

	Layer type	Output Shape	Param #
FEATURE LEARNING	Conv2D	64, 64, 512	25,088
	Conv2D	64, 64, 1024	525,312
	MaxPooling2D	32, 32, 1024	0
	Conv2D	32, 32, 2048	2,099,200
	MaxPooling2D	16, 16, 2048	0
	Conv2D	16, 16, 4096	8,392,704
	MaxPooling2D	8, 8, 4096	0
	Conv2D	8, 8, 4096	16,781,312
	AveragePooling2D	1, 1, 4096	0
	Flatten	4096	0
CLASSIFICATION	Dense	512	2,097,664
	Dropout	512	0
	Dense	512	262,656
	Dropout	512	0
	Dense	20	10,260

Fig. 4. Layer structure of vanilla feature maps.

In contrast, Fig. 5 presents the Ghost-Vanilla Feature Maps, which are derived from the VanillaNet-6 structure by replacing the Conv2D layers with Ghost Modules. This modification reduces the total parameters to approximately 13.9 million, nearly half of the parameters in the Vanilla configuration. The reduction stems from the efficiency of Ghost Modules, which generate additional feature maps through cheaper linear operations while maintaining the same hierarchical depth of up to 4096 channels.

Overall, the comparison shows that Ghost-Vanilla Feature Maps achieve a better balance between representation power and computational efficiency. By significantly reducing the parameter count without compromising hierarchical feature extraction, the architecture in Fig. 5 offers a more scalable and resource-efficient alternative for complex image recognition tasks compared to the Vanilla Feature Maps shown in Fig. 4.



Feature Maps Param: 13,944,320

FEATURE LEARNING	Layer type	Output Shape	Param #
Stem Block	Conv2D	64, 64, 512	25,088
Ghost Features (GF) Stage 1	GhostModule	64, 64, 1024	266,752
	MaxPooling2D	32, 32, 1024	0
Ghost Features (GF) Stage 2	GhostModule	32, 32, 2048	1,057,792
	MaxPooling2D	16, 16, 2048	0
Ghost Features (GF) Stage 3	GhostModule	16, 16, 4096	4,212,736
	MaxPooling2D	8, 8, 4096	0
GF Stage 4	GhostModule	8, 8, 4096	8,407,040
Ghost-Vanilla Feature Maps	AveragePooling2D	1, 1, 4096	0
	Flatten	4096	0
CLASSIFICATION	Dense	512	2,097,664
	Dropout	512	0
	Dense	512	262,656
	Dropout	512	0
	Dense	20	10,260

Fig. 5. Layer structure of Ghost-Vanilla Feature Maps.

C. Experimental Setup

The experiment assesses Ghost-Vanilla Feature Maps implemented at ratios of 2, 3, 4, and 5 against a standard VanillaNet architecture composed of Conv2D layers. The dataset contains South Sumatran songket motif images grouped into twenty classes, partitioned into training (80%), validation (10%), and testing (10%). Training was carried out using the Adam optimizer, a learning rate of 0.001, a batch size of 32, across 50 epochs.

In Ghost-Vanilla configurations, Ghost Modules are applied sequentially according to the defined ratios and followed by dense layers with Dropout. In the modified VanillaNet, Conv2D layers are replaced with Ghost Modules to assess their impact on feature extraction efficiency.

Model performance is evaluated using accuracy, precision, recall, and F1-score metrics. Input images are preprocessed to normalize pixel values, ensuring consistent data representation across all models. After training, models are thoroughly tested on the reserved test set to examine the comparative benefits of Ghost-Vanilla Feature Maps over standard convolutional architectures, as well as the influence of ratio variations on overall classification performance.

IV. Results

The comparative evaluation between Vanilla and Ghost-Vanilla Feature Maps demonstrates that the integration of Ghost modules significantly enhances classification performance across the majority of songket motifs, as shown in Table I. While the Vanilla model exhibits relatively high accuracy, its performance is inconsistent across specific motifs, particularly in terms of recall. For instance, motifs such as Biduk Cukit, Nampan Perak, and Naga Besaung exhibit recall values as low as 0.40–0.50 under the Vanilla configuration, suggesting that the model frequently fails to identify actual instances of these motifs. In contrast, the Ghost-Vanilla model significantly improves these metrics, achieving recall values of up to 1.00 in most cases, thereby providing a more reliable and sensitive recognition process.

A closer examination of the precision and F1-scores further highlights the superiority of Ghost-Vanilla Feature Maps. In motifs with complex visual structures, such as Bintang Berantai and Jatamakuta, the Vanilla model records relatively modest precision scores (0.56–0.88), which may lead to misclassifications. However, by leveraging Ghost modules, the Ghost-Vanilla configuration consistently elevates precision values to 1.00 and F1-scores to the maximum threshold. This indicates that the hybrid approach not only reduces false positives but also achieves a more balanced trade-off between precision and recall, which is critical in motif recognition where inter-class similarities are common.

Another important observation is the stability of performance across motifs that are inherently easier to classify. Motifs such as Cantik Manis, Kenanga Makan Ulat, Mawar Bintang, and Sedulang Setudung already achieve perfect scores under the Vanilla model. Interestingly, the Ghost-Vanilla configuration preserves this level of performance without any degradation, suggesting that the hybrid approach is robust and does not compromise accuracy on simpler motifs. This stability further supports the generalizability of the proposed method across varying motif complexities.

TABLE I. MODEL PERFORMANCE EVALUATION ON SOUTH SUMATRA SONGKET MOTIF CLASSIFICATION

Motif Class	Accuracy Score		Precision Score		Recall Score		F1-score	
	Vanilla	Ghost-Vanilla	Vanilla	Ghost-Vanilla	Vanilla	Ghost-Vanilla	Vanilla	Ghost-Vanilla
Biduk Cukit	0.96	1.00	0.63	1.00	0.50	1.00	0.56	1.00
Bintang Berantai	0.96	0.98	0.56	0.71	1.00	1.00	0.71	0.83
Bunga Cina	0.99	1.00	1.00	1.00	0.70	1.00	0.82	1.00
Bunga Intan	1.00	1.00	1.00	1.00	0.90	1.00	0.95	1.00
Bunga Jatuh	0.99	1.00	0.89	1.00	0.80	1.00	0.84	1.00
Cantik Manis	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Cantik Manis Nanas	0.98	1.00	0.80	1.00	0.80	1.00	0.80	1.00
Duku	0.99	1.00	0.90	1.00	0.90	1.00	0.90	1.00
Jando Beraes	0.99	1.00	0.83	1.00	1.00	1.00	0.91	1.00
Jatamakuta	0.98	1.00	0.88	1.00	0.70	1.00	0.78	1.00
Kenanga Makan Ulat	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Mawar Bintang	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Naga Besaung	0.96	0.98	0.63	1.00	0.50	0.60	0.56	0.75
Nampan Perak	0.97	1.00	1.00	1.00	0.40	1.00	0.57	1.00
Pacar Cina	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Perahu Kajang	0.98	1.00	0.71	1.00	1.00	1.00	0.83	1.00
Pulir	0.99	1.00	0.77	1.00	1.00	1.00	0.87	1.00
Sawit	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Sedulang Setudung	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Seinggok Nanas	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

TABLE II. EVALUATION SUMMARY OF THE PROPOSED MODEL

Model	Overall Accuracy	Total Parameters	FLOPs (B)
CNN with Vanilla Feature Maps	0.860	30,194,196	15.2557
CNN with Ghost-Vanilla Feature Maps (ratio = 2)	0.980	16,339,988	7.8003
CNN with Ghost-Vanilla Feature Maps (ratio = 3)	0.970	7,958,079	4.0705
CNN with Ghost-Vanilla Feature Maps (ratio = 4)	0.925	4,911,380	2.6600
CNN with Ghost-Vanilla Feature Maps (ratio = 5)	0.905	3,440,215	1.9481

Overall, the findings highlight the effectiveness of Ghost-Vanilla Feature Maps in addressing the limitations of conventional Vanilla architectures. The hierarchical incorporation of Ghost modules enhances feature representation by capturing both dominant and subtle discriminative patterns while suppressing irrelevant information. As a result, the model achieves near-perfect classification across all songket motifs, confirming its potential as a robust and efficient framework for fine-grained cultural pattern recognition.

Evaluation outcomes (see Table II) demonstrate that the proposed Ghost-Vanilla Feature architecture substantially outperforms the Vanilla-only baseline while simultaneously reducing computational complexity. The Vanilla Feature Maps model achieves an accuracy of only 0.860 despite requiring 30.19 million parameters and 15.2557B FLOPs, indicating that conventional convolution-heavy designs impose significant computational overhead without delivering commensurate discriminative benefits for fine-grained textile classification. In contrast, the hybrid configuration with ghost ratio 2 attains the highest accuracy of 0.980 while reducing parameters to nearly half of the Vanilla baseline and cutting FLOPs by almost 50%. This finding highlights the architectural advantage of integrating low-cost ghost-generated features with structurally stable VanillaNet representations, enabling the model to extract more diverse and discriminative feature patterns under a significantly more efficient computational cost.

A progressive reduction in parameters and FLOPs across higher ghost ratios further illustrates the flexibility of the hybrid architecture, though with diminishing returns beyond an optimal threshold (see Table II). The ghost ratio 3 and 4 models maintain strong accuracies of 0.970 and 0.925, respectively, while achieving substantial reductions in model size, demonstrating the design's ability to balance expressivity and efficiency. However, the model with ghost ratio 5 exhibits a noticeable drop in accuracy (0.905), revealing that excessive reliance on ghost-generated features limits the capture of high-frequency visual cues essential for fine-grained motif discrimination. Taken together, the results establish ghost ratio 2 as the most effective configuration, offering an optimal synergy between computational parsimony and discriminative power.

V. Discussion

The improvements achieved by the Ghost-Vanilla Feature Maps highlight the effectiveness of integrating Ghost modules into the conventional Vanilla architecture. By producing both intrinsic and inexpensive feature maps, the model can capture essential discriminative information while filtering out redundant patterns. This hierarchical representation leads to more consistent recognition outcomes across motifs, particularly in reducing classification variability between complex and simple categories. Importantly, the hybrid approach achieves this enhancement without increasing computational cost excessively, which underscores its efficiency as a feature extraction strategy.

Beyond performance gains, the robustness of the Ghost-Vanilla model demonstrates its potential for fine-grained cultural motif recognition. The ability to significantly improve classes with lower baseline performance while maintaining perfect scores for easier categories highlights its balanced generalization capacity. This indicates that the model not only addresses the limitations of standard convolutional layers but also establishes a scalable solution that can be extended to other domains requiring high precision and reliability in visual pattern analysis.

TABLE III. Performance Evaluation of the Proposed Feature Maps

Ghost Ratio	Feature Maps	Total Parameters	Overall Accuracy
ratio 2	Ghost	67,727,028	0.965
	Ghost-Vanilla	**16,339,988**	**0.980**
ratio 3	Ghost	44,988,541	0.960
	Ghost-Vanilla	7,958,079	0.970
ratio 4	Ghost	33,914,980	0.920
	Ghost-Vanilla	4,911,380	0.925
ratio 5	Ghost	27,066,179	0.855
	Ghost-Vanilla	3,440,215	0.905

The Ghost feature map, which has been employed in prior studies [11], is re-evaluated in this work as a comparative baseline using a different and more challenging dataset, where the number of motif classes is expanded from 10 to 20, in order to assess its robustness under increased fine-grained classification complexity.

The experimental findings reveal that the Ghost-Vanilla Feature Map consistently enhances classification performance across all ghost ratio configurations compared to the use of Ghost Modules alone, as summarized in Table III. The highest accuracy is achieved at a ghost ratio 2, where the Ghost-Vanilla model attains 0.980, surpassing the standard Ghost model at 0.965, while simultaneously reducing the parameter count from 67.7 million to 16.3 million, a reduction of more than 75%. Similar improvements are observed at ghost ratios 3 and 4, where Ghost-Vanilla yields modest yet consistent accuracy gains while maintaining significantly fewer parameters. These results indicate that the integration of VanillaNet strengthens the representational capacity of the model by providing structurally efficient yet highly discriminative features, making the hybrid design particularly suitable for fine-grained classification tasks such as textile motifs.

At higher ghost ratios, particularly ratio 5, both architectures exhibit a decline in accuracy due to the excessive reliance on cheap-operation feature maps, which limits the expressive capability typically preserved by standard convolutions. Nevertheless, the Ghost-Vanilla variant maintains a notable performance advantage over the pure Ghost model, demonstrating its stabilizing effect even under extreme reductions in convolutional complexity. These findings collectively suggest that ghost ratio 2 represents the optimal configuration for balancing accuracy, parameter efficiency, and deployability. Overall, the results confirm that the proposed Ghost-Vanilla architecture effectively addresses the dual challenge of computational efficiency and representational richness, providing a compelling solution for resource-constrained environments and fine-grained recognition problems.

The experimental results (see Table IV) reveal a fundamental limitation of conventional lightweight CNN architectures, such as MobileNetV3-Small, MobileNetV4-Conv-Small, EfficientNetV2-Small, and ShuffleNetV2 1.0×, when applied to fine-grained textile classification. Although these architectures offer low computational overhead and small parameter sizes, their accuracies remain modest, ranging from 0.425 to 0.615. This performance gap underscores that the inherent structural compression of lightweight models is insufficient to capture the subtle intra-class variations and high inter-class similarities typical of Songket motifs. Notably, EfficientNetV2-Small illustrates a critical observation: increasing FLOPs or depth alone (7.5516B FLOPs) does not guarantee improved discriminative capability, indicating that representational quality in this domain is governed by the specificity of feature construction rather than mere architectural scale.

A deeper examination of the two baseline feature construction strategies provides additional insight into the nature of this limitation. The model employing Ghost Feature Maps achieves high accuracy (0.965) with minimal FLOPs, yet its parameter count swells to 67 million, revealing substantial structural redundancy despite its efficient convolutional operations. In contrast, the Vanilla Feature Maps model exhibits the opposite behavior: it consumes extremely high computational resources (15.2557B FLOPs), but yields an accuracy of only 0.860. This discrepancy highlights a core challenge in feature engineering for fine-grained tasks: reducing redundancy alone (as in Ghost) or increasing representational depth alone (as in Vanilla) is insufficient. Neither approach, when used in isolation, is capable of producing compact yet semantically rich feature embeddings required for resolving fine-grained motif distinctions.

TABLE IV. PERFORMANCE EVALUATION OF THE PROPOSED FEATURE MAPS

Approach	Params	FLOPs (B)	Acc
MobileNetV3-Small	1,538,356	0.1558	0.615
MobileNetV4-Conv-Small	4,158,324	0.4944	0.500
EfficientNetV2-Small	21,842,788	7.5516	0.425
ShuffleNetV2 1.0x	2,323,704	0.3975	0.580
CNN with Ghost Feature Maps	67,727,028	1.4413	0.965
CNN with Vanilla Feature Maps	30,194,196	15.2557	0.860
CNN with Ghost-Vanilla Feature Maps (Ours)	**16,339,988**	**7.8003**	**0.980**

The proposed Ghost-Vanilla Feature architecture demonstrates a decisive improvement by synergistically combining the strengths of both approaches. Achieving the highest accuracy of 0.980 with a substantially reduced parameter count (16.3M) and moderate FLOPs (7.8003B), the model exemplifies an optimal balance between computational efficiency and discriminative power. This performance gain suggests that the complementary interaction between Ghost-generated low-cost feature enrichments and the structural regularity of VanillaNet effectively mitigates redundancy while preserving essential high-frequency visual cues. The hybrid design not only enhances feature robustness but also provides evidence that fine-grained textile classification benefits from architectures that integrate lightweight generative feature expansion with stabilized backbone representations. These findings emphasize the architectural significance of the proposed approach and position it as a strong candidate for deployment in both high-performance and resource-constrained environments.

Further, comparative analysis in Table IV shows that the proposed Ghost-Vanilla Feature Maps address limitations not explicitly handled by state-of-the-art lightweight CNNs. While architectures such as MobileNet, ShuffleNet, and EfficientNet mainly reduce complexity through depthwise separable convolutions or compound scaling, they lack explicit mechanisms for enriching fine-grained local representations required for discriminating highly similar textile motifs. In contrast, the Ghost-Vanilla architecture employs a fundamentally different feature construction strategy that balances controlled feature expansion with structural regularity, resulting in more compact, yet semantically expressive representations.

This study is limited to the architectural evaluation of the proposed Ghost-Vanilla Feature Map within a controlled experimental environment. The dataset employed has been curated and standardized, and the research does not extend to deployment-oriented aspects such as user interface development, real-world field testing under non-standardized imaging conditions, or long-term model adaptation mechanisms. These practical considerations fall outside the present scope and are recommended for investigation in future work.

VI. CONCLUSION

The findings of this study demonstrate that the proposed Ghost-Vanilla Feature Map provides an effective and computationally efficient solution for fine-grained textile motif classification. By integrating low-cost ghost-generated features with the structurally stable representations of VanillaNet, the hybrid architecture achieves the highest accuracy of 0.980 at a ghost ratio 2 while reducing parameters by more than 75% compared to the pure Ghost model. These improvements significantly outperform existing lightweight CNN architectures such as MobileNetV3-Small, MobileNetV4-Conv-Small, EfficientNetV2-Small, and ShuffleNetV2 1.0×, whose accuracies range only from 0.425 to 0.615 despite their compact computational footprints. These results highlight that fine-grained motif recognition demands not only architectural compactness but also a carefully engineered feature extraction strategy capable of capturing subtle intra-class textures and high inter-class similarities, underscoring the architectural robustness and efficiency of the Ghost-Vanilla design for both high-performance and resource-limited environments.

Future work can explore the integration of reparameterization techniques from RepGhost into the Ghost-Vanilla Feature Maps architecture to further improve model efficiency. By employing this strategy, convolutional kernel transformations can markedly reduce parameter count without affecting the Ghost Module's fundamental feature extraction capabilities. The integration of reparameterization allows the model to maintain discriminative feature representations for complex and multi-scale songket motifs while simultaneously reducing overall model size. This approach can be evaluated across different Ghost ratios to identify the optimal balance between efficiency and classification accuracy and to test its applicability on devices with limited computational resources.

Although songket motifs serve as an appropriate benchmark due to their high inter-class similarity and substantial intra-class variability, the primary contribution of this study is methodological rather than domain-specific. The core architectural principles, leveraging low-cost feature expansion in conjunction with a lightweight backbone, hold potential applicability across a wide range of fine-grained classification scenarios, including medical imaging, agricultural disease detection, industrial quality inspection, and remote sensing. Future research is necessary to evaluate the generalizability of the proposed method across these domains and to examine its progression from a research prototype to a production-ready system.

ACKNOWLEDGMENT

This research was supported by the Ministry of Higher Education, Science, and Technology of Indonesia through the research grant for the fundamental research scheme in 2025 with contract number 123/C3/DT.05.00/PL/2025.

REFERENCES

[1] D. S. Nindiati, Andayani, and H. Purwanta, "Acculturation of Palembang Songket Cloth Culture," in The 3rd International Conference on Humanities Education, Law and Social Sciences, 2024, vol. 9, no. 2, pp. 1003–1017. doi: https://doi.org/10.18502/kss.v9i2.14918.

[2] D. Djumrianti, A. F. Abdillah, and L. Lisnini, "The Influence of Cultural Acculturation (Social, Customs and Arts) in the past on Palembang Songket Motifs," Atlantis Press SARL, 2024, pp. 25–38. doi: 10.2991/978-2-38476-220-0_4.

[3] E. Wahyudi, B. Imran, Zaeniah, S. Erniwati, M. N. Karim, and Z. Muahidin, "Fine-Tuning Resnet50V2 With Adamw and Adaptive Transfer Learning for Songket Classification in Lombok," Pilar Nusa Mandiri J. Comput. Inf. Syst., vol. 21, no. 1, pp. 82–91, Mar. 2025, doi: 10.33480/pilar.v21i1.6485.

[4] M. D. P. Fiki, F. Syuhada, and Y. Sa'adati, "Classification of Central Lombok Songket Motifs Using Convolutional Neural Network," BINARY, vol. 1, no. 1, pp. 13–20, 2025.

[5] S. Lestari and N. Apipah, "Comparison of Classification of Songket Fabric Types Using AlexNet and VGG19 (Visual Geometry Group) Method," Int. J. Softw. Eng. Comput. Sci., vol. 5, no. 1, pp. 386–395, Apr. 2025, doi: 10.35870/ijsecs.v5i1.3815.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

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

[11] Yohannes, M. E. Al Rivan, S. Devella, and Tinaliah, "A Novel Optimization Strategy for CNN Models in Palembang Songket Motif Recognition," Int. J. Adv. Comput. Sci. Appl., vol. 16, no. 1, pp. 830–841, 2025, doi: 10.14569/IJACSA.2025.0160180.

[12] E. Ermatita, H. Noprisson, and A. Abdiansah, "Palembang songket fabric motif image detection with data augmentation based on ResNet using dropout," Bull. Electr. Eng. Informatics, vol. 13, no. 3, pp. 1991–1999, Jun. 2024, doi: 10.11591/eei.v13i3.6883.

[13] Y. Xie, Q. Gong, X. Luan, J. Yan, and J. Zhang, "A survey of fine-grained visual categorization based on deep learning," J. Syst. Eng. Electron., vol. 35, no. 6, pp. 1337–1356, 2024, doi: https://doi.org/10.23919/JSEE.2022.000155.

[14] A. Howard, M. Sandler, B. Chen, W. Wang, L.-C. Chen, M. Tan, G. Chu, V. Vasudevan, Y. Zhu, R. Pang, H. Adam, and Q. Le, "Searching for MobileNetV3," in 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 1314–1324. doi: 10.1109/ICCV.2019.00140.

[15] M. Tan and Q. Le, "EfficientNetV2: Smaller Models and Faster Training," in Proceedings of the 38th International Conference on Machine Learning, 2021, vol. 139, pp. 10096–10106. [Online]. Available: https://proceedings.mlr.press/v139/tan21a.html

[16] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design," in Computer Vision – ECCV 2018, 2018, pp. 122–138. doi: https://doi.org/10.1007/978-3-030-01264-9_8.

[17] 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.

[18] 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.

[19] H. Chen, Y. Wang, J. Guo, and D. Tao, "VanillaNet: the Power of Minimalism in Deep Learning," in Advances in Neural Information Processing Systems, 2023, vol. 36, pp. 7050–7064. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2023/file/16336d94a5ffca8de019087ab7fe403f-Paper-Conference.pdf

[20] H. Wang, M. Li, X. Gu, X. Liu, B. Yang, and K. Chen, "Research on Road Damage Recognition Based on VanillaNet Neural Network," J. Comput. Sci. Artif. Intell., vol. 3, no. 1, pp. 35–38, Apr. 2025, doi: 10.54097/j2ex8q47.

[21] S. Zhu, X. Li, G. Wan, H. Wang, S. Shao, and P. Shi, "Underwater Dam Crack Image Classification Algorithm Based on Improved VanillaNet," Symmetry (Basel)., vol. 16, no. 7, p. 845, Jul. 2024, doi: 10.3390/sym16070845.

[22] Y. Zhang, Y. Zhang, T. Li, M. Shi, and Y. Huang, "Analysis of a Multimodal Ocular Ultrasound Image Classification Algorithm Based on YOLO and VanillaNet," IEEE Access, vol. 13, no. May, pp. 148374–148383, 2025, doi: 10.1109/ACCESS.2025.3598747.

[23] X. Lu, R. Yang, J. Zhou, J. Jiao, F. Liu, Y. Liu, B. Su, and P. Gu, "A hybrid model of ghost-convolution enlightened transformer for effective diagnosis of grape leaf disease and pest," J. King Saud Univ. - Comput. Inf. Sci., vol. 34, no. 5, pp. 1755–1767, May 2022, doi: 10.1016/j.jksuci.2022.03.006.

[24] X. Lu, G. Zhan, W. Wu, W. Zhang, X. Wu, and C. Han, "Van-DETR: enhanced real-time object detection with vanillanet and advanced feature fusion," Vis. Comput., vol. 41, no. 6, pp. 4221–4238, Apr. 2025, doi: 10.1007/s00371-024-03656-0.

[25] Z. Tan, X. Li, Y. Wu, Q. Chu, L. Lu, N. Yu, and J. Ye, "Boosting Vanilla Lightweight Vision Transformers Via Re-Parameterization," 12th Int. Conf. Learn. Represent. ICLR 2024, no. 2023, pp. 1–20, 2024.

[26] 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

[27] 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

[28] 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.

[29] 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.

[30] 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.

[31] L. Yang, H. Cai, X. Luo, J. Wu, R. Tang, Y. Chen, and W. Li, "A lightweight neural network for lung nodule detection based on improved ghost module," Quant. Imaging Med. Surg., vol. 13, no. 7, pp. 4205–4221, Jul. 2023, doi: 10.21037/qims-21-1182.

[32] Z. Liu, J. Xiong, M. Cai, X. Li, and X. Tan, "V-YOLO: A Lightweight and Efficient Detection Model for Guava in Complex Orchard Environments," Agronomy, vol. 14, no. 9, p. 1988, Sep. 2024, doi: 10.3390/agronomy14091988.

[33] J. P. Shermila, V. Seethalakshmi, A. Ahilan, and M. Devaki, "SOOTY FRUIT: Fruits Item Classification Using Sooty Tern Optimized Deep Learning Network," Int. J. Intell. Eng. Syst., vol. 17, no. 4, pp. 289–298, Aug. 2024, doi: 10.22266/IJIES2024.0831.22.

[34] B. Ma, Z. Hua, Y. Wen, H. Deng, Y. Zhao, L. Pu, and H. Song, "Using an improved lightweight YOLOv8 model for real-time detection of multi-stage apple fruit in complex orchard environments," Artif. Intell. Agric., vol. 11, pp. 70–82, Mar. 2024, doi: 10.1016/j.aiia.2024.02.001.

[35] S. Fouladi, L. Di Palma, F. Darvizeh, D. Fazzini, A. Maiocchi, S. Papa, G. Gianini, and M. Alì, "Neural Network Models for Prostate Zones

Segmentation in Magnetic Resonance Imaging," Information, vol. 16, no. 3, p. 186, Feb. 2025, doi: 10.3390/info16030186.

[36] Q. Ren, B. Tu, S. Liao, and S. Chen, "Hyperspectral Image Classification with IFormer Network Feature Extraction," Remote Sens., vol. 14, no. 19, p. 4866, Sep. 2022, doi: 10.3390/rs14194866.

[37] X. Wang and J. Liu, "Vegetable disease detection using an improved YOLOv8 algorithm in the greenhouse plant environment," Sci. Rep., vol. 14, no. 1, p. 4261, Feb. 2024, doi: 10.1038/s41598-024-54540-9.