IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 27 October 2022, accepted 8 December 2022, date of publication 12 December 2022,
date of current version 15 December 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3228382

RESEARCH ARTICLE

GFF-CARVING: Graph Feature Fusion for the Recognition of Highly Varying and Complex Balinese Carving Motifs

I WAYAN AGUS SURYA DARMA[1,2], (Member, IEEE), **NANIK SUCIATI**[1], (Member, IEEE), AND **DANIEL SIAHAAN**[1], (Member, IEEE)

[1]Department of Informatics, Faculty of Intelligent Electrical and Informatics Technology, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia
[2]Department of Informatics, Faculty of Technology and Informatics, Institut Bisnis dan Teknologi Indonesia, Denpasar 80225, Indonesia

Corresponding author: Nanik Suciati (nanik@if.its.ac.id)

ABSTRACT The recognition of Balinese carving motifs is challenging due to the highly varying and interrelated motifs of Balinese carvings and in addition to the scantiness of Balinese carving data. This study proposed a method named GFF-CARVING for the recognition of Balinese carving motifs. GFF-CARVING is a deep learning architecture based on the Graph Convolutional Network (GCN) and Convolutional Neural Network (CNN) to extract image and graph features. GFF-CARVING applies feature fusion to improve the discriminative ability of the model to overcome these challenges and therefore improve its recognition performance. The proposed method consists of three main modules, the image representation learning module, the graph representation learning module, and the prediction module. The image representation learning module is based on ResNet and extracts the image features using global max pooling. The graph representation learning module is based on GCN and extracts the graph features. The graph features are handcrafted features that are built based on the occurrence relationship between the constituent sub-motifs of Balinese carvings. The feature fusion generates new features that take into account the occurrence relationship between the sub-motifs. These new features are used in the prediction module to accurately recognize the Balinese carving motifs. Based on the experimental results, GFF-CARVING achieved the highest recognition accuracy of 98.93% compared to other state-of-the-art models. These results indicated that feature fusion based on the handcrafted graph features and image features improved the discriminative ability of GFF-CARVING in recognizing Balinese carving motifs.

INDEX TERMS Balinese carvings, feature fusion, graph convolutional network, graph features, image features.

I. INTRODUCTION

Balinese carvings are a work of art that is considered to be a cultural heritage in Bali. Balinese carvings found in sacred temples have unique motifs that adorn each element of the temple. Motifs of Balinese carvings are carved on a compressed sand media. Most temples in Bali that were built in the past have unique motifs. Preservation efforts have been carried out by digitally collecting and archiving various motifs of Balinese carvings, in which the first step is the automatic and accurate recognition of these motifs.

However, the recognition of Balinese carving motifs is challenging due to two reasons. Firstly, a single motif may vary in appearance when it is present in different carvings. Furthermore, Balinese carvings are comprised of sub-motifs that are mostly interrelated to one another. Secondly, there are no public Balinese carving datasets that is currently available. The complex characteristics of the motifs and also the scantiness of Balinese carving data makes the recognition process very challenging.

Several methods have been proposed to overcome the challenges of limited or scant data and high data variation [1], [2], [3]. In the context of Balinese carving, Darma et al. [4] proposed a data augmentation technique based on

The associate editor coordinating the review of this manuscript and approving it for publication was Li Zhang.

generative adversarial networks (GANs) and geometric transformation to generate synthetic data to improve recognition performance. A transfer learning approach was proposed by Darma et al. [5] to improve the performance of several pre-trained convolutional neural networks (CNNs) for the recognition of Balinese carvings. Mahawan and Harjoko [6] proposed a feature extraction method based on histogram of oriented gradient (HOG) and principal component analysis (PCA), in which the features of the training data are stored into a table using learning vector quantization (LVQ) and used for the recognition of Balinese carvings. However, these methods have yet to achieve a significant recognition performance and also have yet to fully overcome the problems faced in the recognition of Balinese carvings, namely the highly varying and interrelated sub-motifs.

Furthermore, several studies have proposed the use of graph features from images to improve recognition performance. Zhang et al. [7] proposed a modularity-based graph learning module to build the graph representation of features extracted using CNN and with the use of a graph convolutional network (GCN) module, independent CNN features and mutual GCN features are integrated to represent the retinal images and boost the recognition performance. Zhang et al. [8] proposed a structure-feature fusion adaptive GCN (SFAGCN) for skeleton-based action recognition, in which SFAGCN was shown to surpass the accuracy of state-of-the-art methods by more than 0.6% on average. Mou et al. [9] proposed a nonlocal GCN for the classification of hyperspectral images which exhibited competitive results compared to other spectral classifiers.

Based on the studies above, this study proposed a method named GFF-CARVING which is based on GCN and CNN to address the challenges faced in the recognition of Balinese carving motifs, namely the highly varying and interrelated motifs of Balinese carvings and the scantiness of Balinese carvings data. The proposed GFF-CARVING applies feature fusion to improve the discrimination ability of the model and in turn improves the recognition performance of the model. The crucial contribution of this study for the recognition of Balinese carving motifs are summarized below:

- We propose a hybrid model that combines CNN and GCN into a unified architecture to extract image features and graph features for the recognition of Balinese carving motifs.
- We built handcrafted graph features based on the occurrence relationship between the constituent sub-motifs of Balinese carvings that represent the Balinese carving images.
- We propose feature fusion of the image and graph features to improve the discriminative ability of GFF-CARVING in recognizing Balinese carving motifs.

To the best of our knowledge, there are only a few studies that conduct the recognition of Balinese carving motifs. The experimental results show that GFF-CARVING can overcome the challenges faced in the recognition of Balinese carving motifs, namely the highly varying and interrelated motifs of Balinese carvings and the scantiness of Balinese carving data.

The rest of the paper is organized as follows: Section II discusses the related works. Section III presents the proposed GFF-CARVING method for Balinese carving motif recognition. Section IV discusses the experimental results. Finally, Section V presents the conclusion and future works.

II. RELATED WORKS

Image recognition is the process of identifying images and classifying them into classes. The classification process is carried out based on features of the objects present in the images. Image feature representation is the value used to distinguish the classes of each object. Ling et al. [10] proposed a self-residual attention-based CNN for deep face recognition. This study used Resnet-50 and Resnet-101 as the backbone networks and implemented a self-residual spatial attention block and a self-residual channel attention block to decrease the redundancy between channels and to focus on the more significant parts of the face images. Wang et al. [11] proposed a method to classify pulmonary images based on Inception-v3 and transfer learning. Kui et al. [12] proposed a depthwise separable residual neural network (ResNet) for the classification of hyperspectral images that distinguishes the spectral and spatial information of the images and reduces the network size to prevent overfitting. Pal et al. [13] proposed a deep metric learning-based framework that is configured into CNNs to generate class-distinctive image feature descriptors for the classification of cervical images. Sutramiani et al. [14] proposed a data augmentation technique named MAT-AGCA to improve the performance of CNNs for the recognition of Balinese characters. MAT-AGCA addresses the challenge of limited availability of Balinese character datasets.

Transfer learning is a strategy to improve the performance of pre-trained models for the recognition of objects. Zhou et al. [15] conducted transfer learning based on the Inception-v3 and VGG19 models to differentiate benign and malignant breast tumors. This research examined various depths of transfer learning and evaluated the effects on the classification performance. Huo et al. [16] used deep transfer learning and semisynthetic training data for the classification of underwater objects in sonar images. This research applied fine-tuning and transfer learning to the VGG19 model. Sutramiani et al. [17] conducted transfer learning based on the MobileNet model for the recognition of Balinese characters. This research fine-tuned the number of trainable parameters of the pre-trained model and achieved an accuracy of 86.23%. Fan et al. [18] carried out the recognition of rock lithology using transfer learning based on the SqueezeNet and MobileNet models. The research achieved the highest recognition accuracy of 94.55% compared to other state-of-the-art methods.

Ensemble learning is a technique that combines several models or classifiers in an attempt to improve classification performance. Several studies have conducted ensemble learning in various fields. Ali et al. [19] used ensemble learning for lung nodule classification. They extracted deep features using several CNN models and these features were used to train two classifiers, namely SVM and AdaBoostM2. It was shown that their method with the used of SVM outperformed other state-of-the-art methods and achieved an accuracy of 90.46%. Kusetogullari et al. [20] proposed DIGITNET-rec, an ensemble of three CNN model to recognize digit strings based on majority voting. Nanni et al. [21] proposed an ensemble of CNNs for bioimage classification, in which the scores of the models were combined using sum rules. This study compared the performance of several models based on learning rates, batch sizes, and topologies. Banerjee et al. [22] proposed an ensemble of selected features of several CNNs based on a two-stage feature selection algorithm, namely fuzzy entropy (FE) and total contribution score (TCS) for erythrocytes detection. Liu et al. [23] proposed a deep ensemble model for facial expression recognition. A hybrid feature representation method was used to acquire high-level discriminative features and a lightweight backbone fusion based on VGG16 and ResNet was constructed to achieve low-calculation training. The model achieved an accuracy of over 94% on four benchmark datasets. Patel et al. [24] proposed feature fusion based on several modalities, features, classifier decision scores for human action recognition. Liu et al. [25] proposed a deep feature fusion ResNet for insect pest recognition. Based on these studies, feature fusion can improve recognition performance by combining several features.

Several studies have proposed several other techniques to improve recognition performance of various objects. Lee et al. [26] proposed a data augmentation method using conditional GAN (cGAN) to address the scarcity of labeled iris image data. It was shown that the method improved the iris recognition accuracy. Man et al. [27] proposed a method to classify breast histopathological images named DenseNet121-AnoGAN which utilized anomaly detection with GAN to screen mislabeled patches and DenseNet to extract multi-layered features of the discriminative patches. Liu et al. [28] proposed a method for Covid-19 diagnosis from CT images based on a two-dimensional sparse matrix profile and DenseNet. This study used the sparse matrix profile method to generate anomaly enhanced CT images which was used to train the DenseNet model. Furthermore, this study also used data augmentation techniques to achieve the best classification performance. Wang et al. [29] proposed a method to classify single chromosome images into 24 types based on extended ResNet. This study used Hausdorff distance to calculate the vector of the input image and the 24 label feature vectors. Furthermore, Lu et al. [30] proposed an efficient algorithm based on the ResNet model to predict protein-protein interactions. Other research proposed Dimension-Based Generic Convolution Block to improve the recognition accuracy and reduced the optimized the model [31].

In this study, we adopted several of the approaches above to address the challenges faced in the recognition of Balinese carving motifs, namely the highly varying and interrelated motifs of Balinese carvings and the scantiness of Balinese carving data to improve the recognition of Balinese carving motifs. The transfer learning approach can be used to improve the recognition performance but does not overcome the challenges encountered in the recognition of Balinese carving motifs. Furthermore, ensemble learning by combining several models and features can also improve recognition performance. However, in the context of Balinese carving motifs, a more suitable approach that takes into consideration the complex characteristics of Balinese carving motifs is needed. Therefore, we built handcrafted features that exploit the occurrence relationship between the sub-motifs as the graph features. Based on these handcrafted features, we applied feature fusion to enrich the image features. Therefore, improve recognition performance. This study proposes a method to recognize Balinese carving motifs named GFF-CARVING which applies feature fusion on the handcrafted graph features extracted using GCN and the image features extracted using ResNet. This feature fusion approach improves the discriminative ability of the model to recognize Balinese carving motifs.

III. METHODOLOGY

Fig. 1 shows the different variants of Balinese carving motifs and the interrelated unique sub-motifs. The Balinese carving is consisted of seven sub-motifs i.e., *Barong, Gajah, Karang Goak, Karang Daun, Patra Punggel, Patra Cina*, and *Keketusan Kakul kakulan.* A single sub-motif may vary in appearance when it is present in different carvings because Balinese carvings are carved by different craftsmen and not printed, as shown in Fig. 1a. The scantiness of Balinese carving data is also a challenge in the recognition task. To our knowledge, there is currently no publicly available Balinese carving dataset. In addition, Balinese carvings have unique sub-motifs that are interrelated to one another, as shown in Fig. 1b. Each Balinese carving is constituted of several interrelated sub-motifs. For example, *Karang Barong* is composed of three sub-motifs i.e., *Barong*, *Patra Cina*, and *Keketusan Kakul kakulan. Barong* and *Keketusan Kakul kakulan* sub-motifs also appear in *Karang Barong 2*. Hence these sub-motifs are interrelated. The combination of these constituent sub-motifs is described in more detail in Table 1. Based on the unique characteristics of Balinese carving sub motifs, we exploit these characteristics to construct graph features. This graph feature is combined with image features to produce new features to improve the model's discriminative ability in classifying Balinese carving motifs on highly varying and limited dataset.

Fig. 2 shows the proposed Balinese carving recognition method that consists of four steps. The first step is to construct a directed graph based on a weighted adjacency matrix to represent the occurrence relationship between the submotifs. The second step is dataset vectorization. In this step,

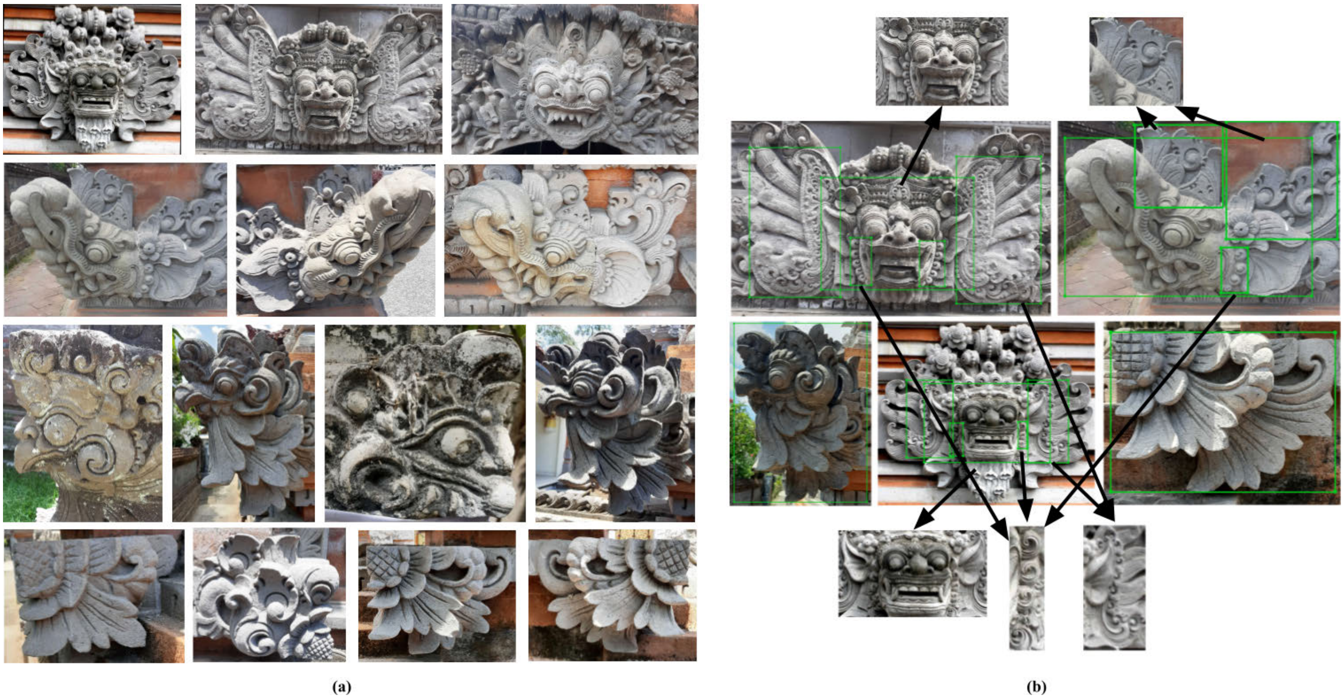


FIGURE 1. Images of Balinese carvings that are highly varying and interrelated: (a) Balinese carving motifs that are highly varying (b) a variety of interrelated constituent sub-motifs.

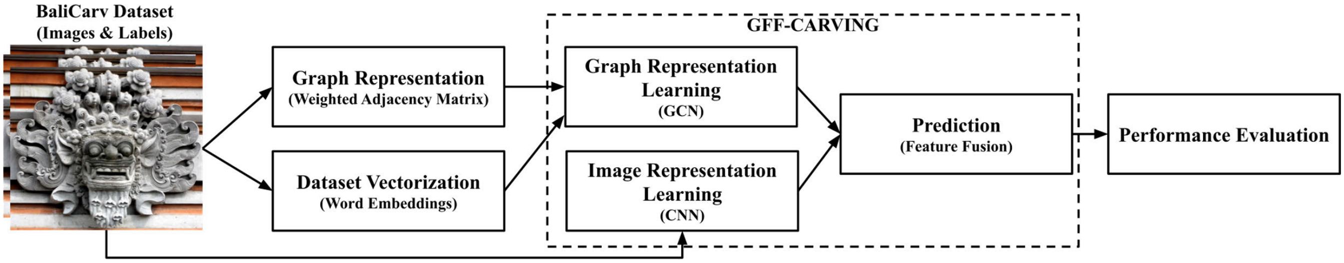


FIGURE 2. The proposed Balinese carving motifs recognition based on graph feature fusion. The Balinese carving images are used as the input of the image representation learning module and the carving labels as the input of graph representation and dataset vectorization process.

we applied word embedding to the BaliCarv dataset based on the identified sub-motif labels. We utilized FastText to generate a vectorized form of the BaliCarv dataset. The third step is the proposed GFF-CARVING method, which consists of an image representation learning module, a graph representation learning module, and a prediction module. In this step, we applied ResNet to extract image features and GCN to capture graph features then predict the labels of the motifs based on feature fusion. We combined the image and graph features to improve the discriminative ability of the model in recognizing Balinese carving motifs. The last step is performance evaluation. We evaluated the performance of the proposed GFF-CARVING with other state-of-the-art CNN models on the BaliCarv dataset.

A. WEIGHTED ADJACENCY MATRIX FOR GRAPH REPRESENTATION

The occurrence relationships between the sub-motif labels are represented by a weighted adjacency matrix and shown as a graph. A graph is a structure that encodes object connections. Objects in a graph are represented by nodes, while edges that connect nodes reflect the relationship between nodes. Weights can be assigned to edges to indicate the strength of the link between nodes. A weighted directed graph is used to represent the graph in this scenario. Each Balinese carving contains sub-motifs that constitute the carving as a whole. The weighted graph is based on the occurrence of two different sub-motifs in one Balinese carving.

We constructed handcrafted graph features based on the occurrence relationships between the sub-motif labels. Each sub-motif label is represented as a node in the graph and each directed edge between the nodes has a weight that represents the probability of occurrence of other sub-motif labels when a particular sub-motif label is present. Furthermore, each sub-motif label has a probability of occurrence which is calculated by the number of sample images that contain this particular sub-motif divided by the total number of sample images. A conditional probability can be used to represent

TABLE 1. Balinese carving motifs combination with its constituent sub-motifs, the number of motif images, and the number of sub-motif images.

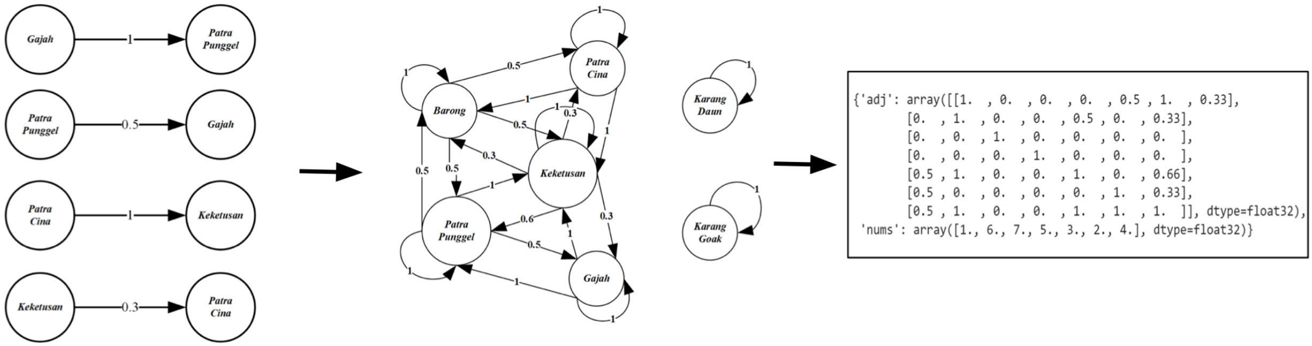
Balinese Carving Motif	Sub-Motif 1	Sub-Motif 2	Sub-Motif 3	#Motif Images	#Sub-Motif Images
Karang Barong 1	*Barong*	*Patra Cina*	*Keketusan Kakul-kakulan*	218	1,526
Karang Barong 2	*Barong*	*Patra Punggel*	*Keketusan Kakul-kakulan*	851	5,643
Karang Gajah	*Gajah*	*Patra Punggel*	*Keketusan Kakul-kakulan*	376	1,238
Karang Daun	*Daun*	-	-	221	221
Karang Goak	*Goak*	-	-	698	698
TOTAL				2,364	9,326

TABLE 2. Co-occurrence matrix of sub-motif pairs ($A \in R^{C \times C}$) and Number of the sub-motif occurrences in the Balinese carving motif combination (N).

	L_j							
L_i	*Barong*	*Gajah*	*Karang Goak*	*Karang Daun*	*Patra Punggel*	*Patra Cina*	*Keketusan Kakul-kakulan*	*N*
Barong	0	0	0	0	1	1	1	2
Gajah	0	0	0	0	1	0	1	1
Karang Goak	0	0	0	0	0	0	0	1
Karang Daun	0	0	0	0	0	0	0	1
Patra Punggel	1	1	0	0	0	0	2	2
Patra Cina	1	0	0	0	0	0	1	1
Keketusan Kakul-kakulan	1	1	0	0	2	1	0	3

TABLE 3. Conditional probabilities of the sub-motif pairs.

L_j L_i	*Barong*	*Gajah*	*Karang Goak*	*Karang Daun*	*Patra Punggel*	*Patra Cina*	*Keketusan Kakul-kakulan*
Barong	0	0	0	0	0.5	1	0.3
Gajah	0	0	0	0	0.5	0	0.3
Karang Goak	0	0	0	0	0	0	0
Karang Daun	0	0	0	0	0	0	0
Patra Punggel	0.5	1	0	0	0	0	0.6
Patra Cina	0.5	0	0	0	0	0	0.3
Keketusan Kakul-kakulan	0.5	1	0	0	1	1	0



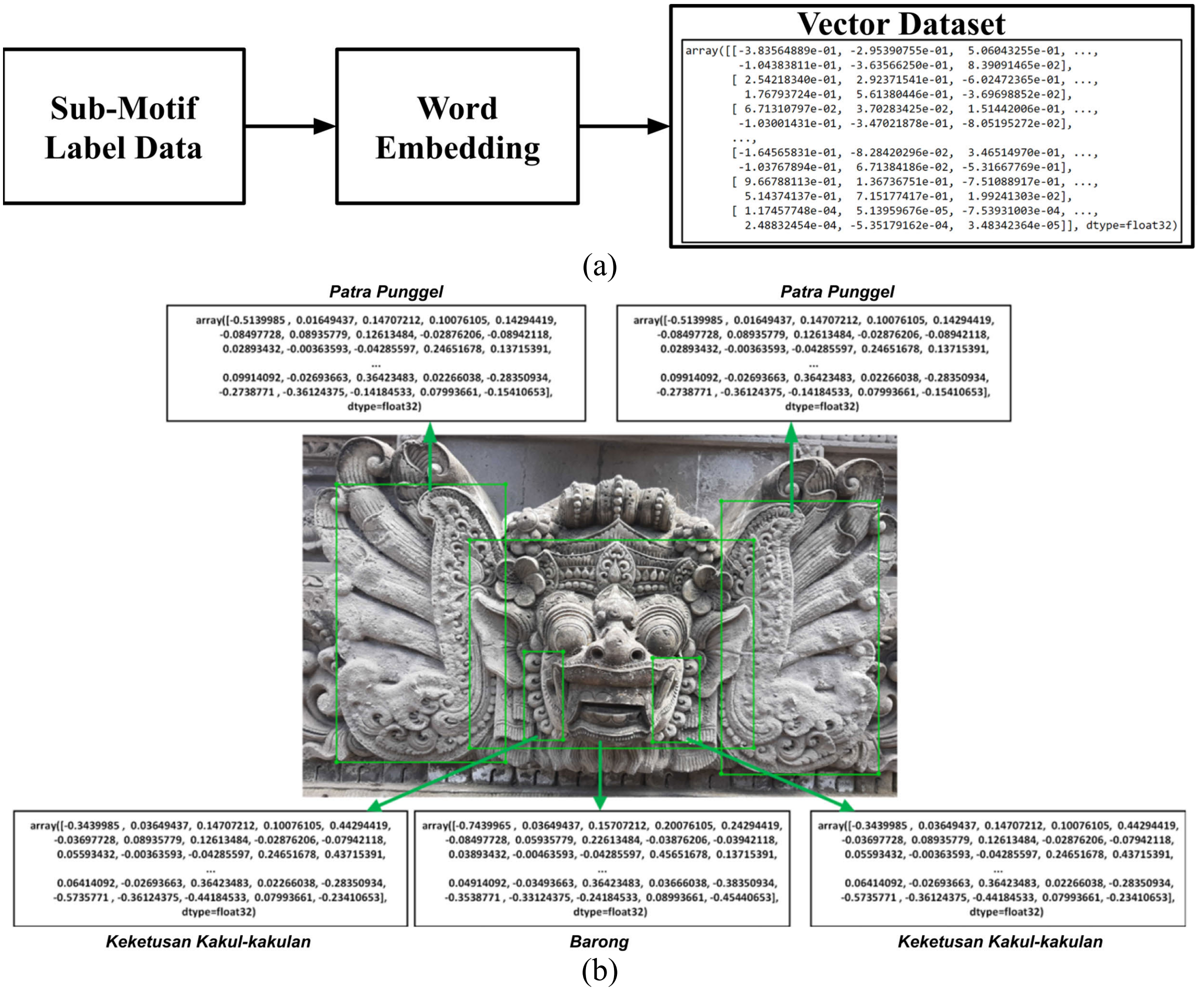
```
{'adj': array([[1.  , 0.  , 0.  , 0.  , 0.5 , 1.  , 0.33],
       [0.  , 1.  , 0.  , 0.  , 0.5 , 0.  , 0.33],
       [0.  , 0.  , 1.  , 0.  , 0.  , 0.  , 0.  ],
       [0.  , 0.  , 0.  , 1.  , 0.  , 0.  , 0.  ],
       [0.5 , 1.  , 0.  , 0.  , 1.  , 0.  , 0.66],
       [0.5 , 0.  , 0.  , 0.  , 0.  , 1.  , 0.33],
       [0.5 , 1.  , 0.  , 0.  , 1.  , 1.  , 1.  ]], dtype=float32),
 'nums': array([1., 6., 7., 5., 3., 2., 4.], dtype=float32)}
```

(a) (b) (c)

FIGURE 3. (a) The illustration of the conditional probabilities between sub-motifs, (b) The directed graph relationship of carving sub-motifs and the corresponding conditional probabilities, and (c) The adjacency matrix that represents the directed graph relationship of the carving sub-motifs.

the occurrence relationship between sub-motif labels, namely $P\left(L_j \mid L_i\right)$, to indicate the probability that the L_j label appears when the L_i label appears.

The graph and the corresponding weighted adjacency matrix can be constructed based on the Balinese carving motifs combination. First, we identified the different motifs present in the Balinese carving images, which are shown in Table 1. Each Balinese carving motif is composed of sub-motifs. The combination of this sub-motifs is spread over 2,364 images in the BaliCarv dataset, consisting of 9,326 sub-motifs. Afterward, we counted the number of occurrences of sub-motif pairs in the Balinese carving motifs combination to obtain the co-occurrence matrix of sub-motif pairs $A \in R^{C \times C}$, where C is the number of labels is shown in Table 2. Then, we counted the number of occurrences of each sub-motif in the Balinese carving combination (N). Finally, the weighted adjacency matrix of the occurrence relationship between sub-motif pairs based on the conditional



```
array([[-3.83564889e-01, -2.95390755e-01,  5.06043255e-01, ...,
        -1.04383811e-01, -3.63566250e-01,  8.39091465e-02],
       [ 2.54218340e-01,  2.92371541e-01, -6.02472365e-01, ...,
         1.76793724e-01,  5.61380446e-01, -3.69698852e-02],
       [ 6.71310797e-02,  3.70283425e-02,  1.51442006e-01, ...,
        -1.03001431e-01, -3.47021878e-01, -8.05195272e-02],
       ...,
       [-1.64565831e-01, -8.28420296e-02,  3.46514970e-01, ...,
        -1.03767894e-01,  6.71384186e-02, -5.31667769e-01],
       [ 9.66788113e-01,  1.36736751e-01, -7.51088917e-01, ...,
         5.14374137e-01,  7.15177417e-01,  1.99241303e-02],
       [ 1.17457748e-04,  5.13959676e-05, -7.53931003e-04, ...,
         2.48832454e-04, -5.35179162e-04,  3.48342364e-05]], dtype=float32)
```

Patra Punggel

```
array([-0.5139985 , 0.01649437, 0.14707212, 0.10076105, 0.14294419,
 -0.08497728, 0.08935779, 0.12613484, -0.02876206, -0.08942118,
 0.02893432, -0.00363593, -0.04285597, 0.24651678, 0.13715391,
 ...
 0.09914092, -0.02693663, 0.36423483, 0.02266038, -0.28350934,
 -0.2738771 , -0.36124375, -0.14184533, 0.07993661, -0.15410653],
 dtype=float32)
```

Patra Punggel

```
array([-0.5139985 , 0.01649437, 0.14707212, 0.10076105, 0.14294419,
 -0.08497728, 0.08935779, 0.12613484, -0.02876206, -0.08942118,
 0.02893432, -0.00363593, -0.04285597, 0.24651678, 0.13715391,
 ...
 0.09914092, -0.02693663, 0.36423483, 0.02266038, -0.28350934,
 -0.2738771 , -0.36124375, -0.14184533, 0.07993661, -0.15410653],
 dtype=float32)
```

Keketusan Kakul-kakulan

```
array([-0.3439985 , 0.03649437, 0.14707212, 0.10076105, 0.44294419,
 -0.03697728, 0.08935779, 0.12613484, -0.02876206, -0.07942118,
 0.05593432, -0.00363593, -0.04285597, 0.24651678, 0.43715391,
 ...
 0.06414092, -0.02693663, 0.36423483, 0.02266038, -0.28350934,
 -0.5735771 , -0.36124375, -0.44184533, 0.07993661, -0.23410653],
 dtype=float32)
```

Barong

```
array([-0.7439965 , 0.03649437, 0.15707212, 0.20076105, 0.24294419,
 -0.08497728, 0.05935779, 0.22613484, -0.03876206, -0.03942118,
 0.03893432, -0.00463593, -0.04285597, 0.45651678, 0.13715391,
 ...
 0.04914092, -0.03493663, 0.36423483, 0.03666038, -0.38350934,
 -0.3538771 , -0.33124375, -0.24184533, 0.08993661, -0.45440653],
 dtype=float32)
```

Keketusan Kakul-kakulan

```
array([-0.3439985 , 0.03649437, 0.14707212, 0.10076105, 0.44294419,
 -0.03697728, 0.08935779, 0.12613484, -0.02876206, -0.07942118,
 0.05593432, -0.00363593, -0.04285597, 0.24651678, 0.43715391,
 ...
 0.06414092, -0.02693663, 0.36423483, 0.02266038, -0.28350934,
 -0.5735771 , -0.36124375, -0.44184533, 0.07993661, -0.23410653],
 dtype=float32)
```

FIGURE 4. Carving sub-motif vectorization process: (a) The sub-motif label vectorization through word embedding process. (b)The illustration of vector data that represents each carving sub-motif. Each carving sub-motif is represented as a vector data to be processed in the graph representation learning module.

probabilities of sub-motif pairs is constructed by dividing each row of the co-occurrence matrix A by the number of occurrences of the corresponding labels (N). The conditional probability $P\left(L_j \mid L_i\right)$ for the pair of sub-motif labels L_j and L_i is calculated using the following formula:

$$P\left(L_j \mid L_i\right) = \frac{A_{ij}}{N_j} \tag{1}$$

where A_{ij} is the number of occurrences of the sub-motif pairs L_j and L_i, and N_j is the number of occurrences of L_j. We calculated the conditional probabilities for each pair of sub-motif labels using Eq. 1. The conditional probabilities for each pair of sub-motif labels are shown in Table 3 and the final weighted adjacency matrix is shown in Figure 3c. The conditional probability that the L_i label appears when the label itself appears is represented with a weight of 1.

TABLE 4. The BaliCarv dataset sub-motif labels. Each Balinese carving image is labeled with the interrelated constituent carving sub-motifs.

Sub-motif classes	#sub-motif labels
Karang Daun	221
Gajah	376
Karang Goak	698
Barong	1,069
Patra Cina	872
Keketusan Kakul-kakulan	2,387
Patra Punggel	3,703
TOTAL	9,326

Fig. 3a depicts the nodes of sub-motif labels, the corresponding directed edge that connects the nodes, and the weight that is given to each edge based on the conditional probability between the sub-motifs. It can be seen that when the gajah label appears, the probability that the *Patra Punggel* label appears is represented with a weight of 1. However, as the edges are directed, this does not apply the other way around. When the *Patra Punggel* label appears, the probability that the gajah label appears is represented with a weight of 0.5. Another example is the relationship between the *Patra Cina* label and the *Keketusan* label. When the *Patra Cina* label is present, the probability of occurrence of the *Keketusan* label is represented with a weight of 1, but when the *Keketusan* label is present, the probability of occurrence of the *Patra Cina* label is represented by a weight of 0.3.

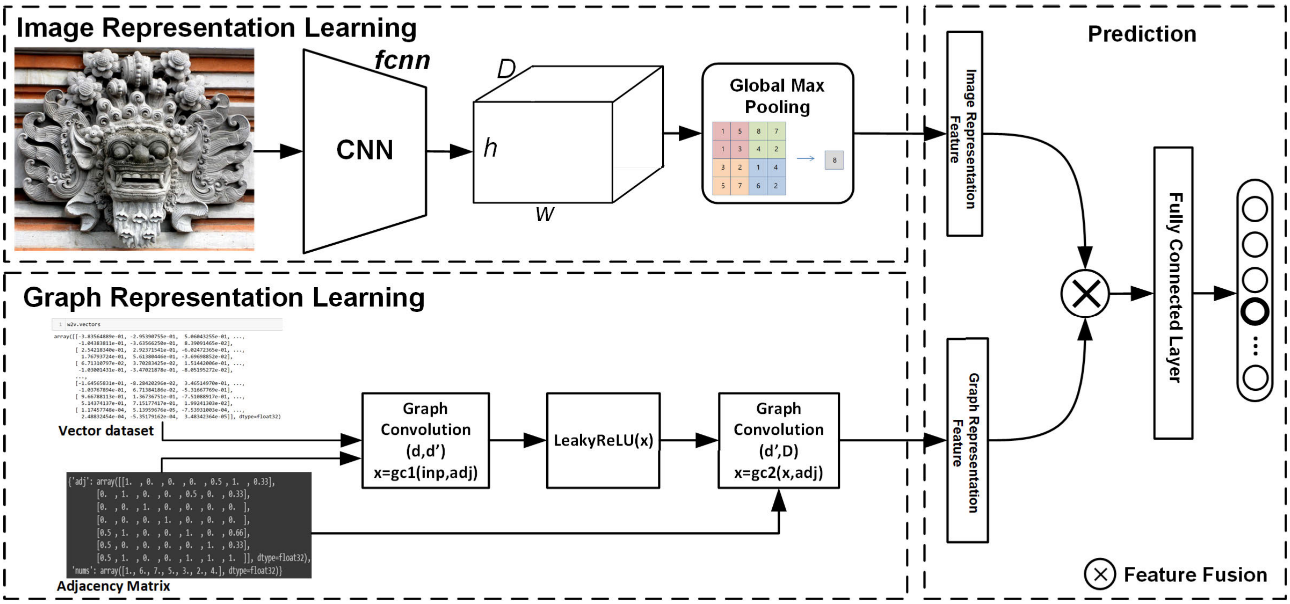


Image Representation Learning
fcnn
CNN
D
h
w
Global Max Pooling
1 5 6 7
1 3 4 2
3 2 1 4
5 7 6 2
8
Prediction
Image Representation Feature
Fully Connected Layer
Graph Representation Learning
Graph Convolution (d,d') x=gc1(inp,adj)
LeakyReLU(x)
Graph Convolution (d',D) x=gc2(x,adj)
Graph Representation Feature
Feature Fusion
Vector dataset
Adjacency Matrix

```
1  w2v.vectors
```

```
array([[-3.83564889e-01, -2.95390755e-01,  5.06043255e-01, ...,
        -1.04383811e-01, -3.63566250e-01,  8.39091465e-02],
       [ 2.54218340e-01,  2.92371541e-01, -6.02472365e-01, ...,
         1.76793724e-01,  5.61380446e-01, -3.69698852e-02],
       [ 6.71310797e-02,  3.70283425e-02,  1.51442006e-01, ...,
        -1.03001431e-01, -3.47021878e-01, -8.05195272e-02],
       ...,
       [-1.64565831e-01, -8.28420296e-02,  3.46514970e-01, ...,
        -1.03767894e-01,  6.71384186e-02, -5.31667769e-01],
       [ 9.66788113e-01,  1.36736751e-01, -7.51088917e-01, ...,
         5.14374137e-01,  7.15177417e-01,  1.99241303e-02],
       [ 1.17457748e-04,  5.13959676e-05, -7.53931003e-04, ...,
         2.48832454e-04, -5.35179162e-04,  3.48342364e-05]], dtype=float32)
```

```
{'adj': array([[1.  , 0.  , 0.  , 0.  , 0.5 , 1.  , 0.33],
        [0.  , 1.  , 0.  , 0.  , 0.5 , 0.  , 0.33],
        [0.  , 0.  , 1.  , 0.  , 0.  , 0.  , 0.  ],
        [0.  , 0.  , 0.  , 1.  , 0.  , 0.  , 0.  ],
        [0.5 , 1.  , 0.  , 0.  , 1.  , 0.  , 0.66],
        [0.5 , 0.  , 0.  , 0.  , 0.  , 1.  , 0.33],
        [0.5 , 1.  , 0.  , 0.  , 1.  , 1.  , 1.  ]], dtype=float32),
 'nums': array([1., 6., 7., 5., 3., 2., 4.], dtype=float32)}
```

FIGURE 5. The GFF-CARVING model. The model consists of three main modules, namely the image representation learning module to extract the image features, the graph representation learning module to capture the graph features, and the prediction module that utilizes feature fusion to predict the motif labels.

Fig. 3b shows the illustration of weighted directed graph of the occurrence relationship between sub-motif labels. Fig. 3c is the final weighted adjacency matrix that represents the weighted directed graph. This graph will be used in the graph representation learning module.

B. SUB-MOTIF LABEL VECTORIZATION

We applied sub-motifs label vectorization to enrich the motif carving features so as to improve the discriminative ability of the model. Fig. 4 shows the sub-motif label vectorization through word embedding process and the illustration of the vector values of the carving sub-motif labels. The vectorized form of the BaliCarv dataset was generated by applying word embedding based on the sub-motif labels. Word embedding maps each label of the BaliCarv dataset into a dense vector. The dense vector is a numerical representation of the semantic meaning of each sub-motif label. Each sub-motif label in the Balinese carving dataset is encoded into a very dense and high-dimensional vector. The abstract meaning and relationship of each label are coded numerically.

Fig. 4a shows the sub-motif label vectorization through word embedding process. The BaliCarv dataset consists of seven sub-motif classes. The vectorization process is implemented for all the seven motif labels of the BaliCarv dataset that consists of 2,364 images which contains 9,326 sub-motifs labels. Table 4 shows the sub-motif labels of the BaliCarv dataset along with the number of occurrences of each sub-motif label within the images. The FastText model was applied to map each sub-motif label into a vector to generate the vectorized form of the BaliCarv dataset. The training process using FastText was carried out for 10,000 iterations using a vector dimension of 300. This vectorized form of the BaliCarv dataset and the weighted adjacency matrix are used as input data to the graph representation learning process. As an illustration, Fig. 4b shows the vector values that represent each carving sub-motif.

C. GRAPH FEATURE FUSION CARVING

Fig. 5 shows in detail the GFF-CARVING process, which consists of three main modules. The first module is the image representation learning module, which is responsible for extracting image features. An image feature is part of the pattern of an image object to recognize or differentiate from other objects. The second module is the graph representation learning module, which is responsible for extracting graph features based on the interrelationships between the sub-motif labels. The third module is the prediction module, which is responsible for feature fusion of the extracted image and graph features, and also predicting the Balinese carving motifs. The purpose of extracting image and graph features is to produce new features of Balinese carving by combining image features and graph features, thereby enriching the features, and increasing the model's discriminatory ability in classifying Balinese carving motifs.

1) IMAGE REPRESENTATION LEARNING

The first module aims to extract features from the images. In our experiment, we used ResNet as the base model. The resolution of the input image is 500 × 500 pixels. To obtain the features from the images, we used global max pooling with the following formula:

$$x = f_{GMP}\left(f_{cnn}\left(I;\theta_{cnn}\right)\right) \in R^{D} \tag{2}$$

where x is the features extracted from the image, I is the input image, θ_{cnn} is the model parameter, and D is the dimension of the feature map where $D = 2048$.

To optimize the neural network, we used SGD as the optimizer with a momentum = 0.9, L2 regularization with a weight decay = 1×10^{-4}, and an initial learning rate = 1×10^{-2} which decays every 40 epochs by a factor of 10. The image representation learning module generates a 2048-dimensional feature map.

2) GRAPH REPRESENTATION LEARNING

The second module aims to extract graph features. This module implemented a GCN network consisting of two layers. Each GCN layer received a node representation from the previous layer and outputs a new node representation. In the first layer, the weighted adjacency matrix is used as input for the label-level word embedding, where d is the vector dimension of the word embedding and $d = 300$. The output from the first layer goes through the LeakyReLU activation function; hence the model can learn the relationship between complex labels by stacking multiple GCN layers. The results of the GCN convolution in the first layer produces an output with a dimension $d' = 1024$. The output of the first layer and the weighted adjacency matrix is used as the input for the second layer to produce a feature vector with a dimension of 2048. The final output of the GCN module is a 2048-dimensional feature map that is based on the graph of interrelated sub-motif labels.

The weighted adjacency matrix and the vectorized form of the BaliCarv dataset are used as the input for the graph representation learning module. The weight adjacency matrix represents the occurrence relationship between the sub-motifs that constitute the Balinese carvings. The vectorized form of the BaliCarv dataset is built through the word embedding process.

3) PREDICTION BASED ON FEATURE FUSION

The prediction module combines the extracted image features and graph features. Feature fusion is implemented on both features by utilizing the matrix multiplication function. We used the *torch.matmul* function to combine image features and graph features instead of *torch.mm* and *torch.bmm*. *Torch.mm* and *torch.bmm* function does not broadcast matrix product, so it cannot treat arrays with different shapes during arithmetic operations. On the other hand, *torch.matmul* functions treat arrays of different shapes during arithmetic operations. The smaller array is broadcast across the larger array to have a compatible shape. In addition, *torch.matmul* can perform tensor multiplication with high-dimensional input. The feature fusion is calculated with the following formula:

$$(x_1 x_2 \ldots x_n) \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = x_1 \times y_1 x_2 \times y_2 \ldots x_n \times y_n \qquad (3)$$

In Eq. 3, $x_1 x_2 \ldots x_n$ is the first vector and $y_1 y_2 \ldots y_n$ is the second vector. The *torch.matmul* function calculates both vectors to perform feature fusion to generate a new vector value. After applying feature fusion, the results are fed through a fully connected layer to predict the labels of the Balinese carving motifs. In addition, the feature fusion generates features that takes into account the occurrence relationship between sub-motifs and can improve the discriminative ability of the model; therefore, improving the performance of the model in recognizing Balinese carving motifs.

IV. EXPERIMENT RESULTS AND DISCUSSION

A. DATASET

We used the BaliCarv dataset, composed of 2,364 images containing 9,326 sub-motif labels and seven sub-motif classes. The BaliCarv dataset was built through a data generation process using neural style transfer and geometric transformation described in [4]. We applied K-Fold cross-validation to the model training process. First, we split the dataset into 5-folds, thus dividing the dataset into five types of data trained on each model. Then, each training process was conducted on 1,892 train and 472 test data.

B. PERFORMANCE EVALUATION

We evaluated the performance using precision, recall, F1, and accuracy on a RTX3060 GPU. We evaluated the performance of GFF-CARVING by comparing the performance with other benchmark CNN models. These models were trained using the BaliCarv dataset that consists of 2,364 Balinese carving images. The performance is evaluated with the following formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (4)$$

$$Precision = \frac{TP}{TP + FP} \qquad (5)$$

$$Recall/True\ Positive\ Rate(TPR) = \frac{TP}{TP + FN} \qquad (6)$$

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \qquad (7)$$

$$False\ Positive\ Rate(FPR) = \frac{FP}{TN + TP} \qquad (8)$$

$$AUC = \int_0^1 TPRd(FPR) \qquad (9)$$

where TP is the number of correctly predicted positive labels, TN is the number of correctly predicted negative labels, FP is the number of incorrectly predicted positive labels, and FN is the number of incorrectly predicted negative labels. Accuracy is the ratio of true predictions to the overall data. Precision is the ratio between TP to the total number of positive predictions. Recall/TPR is the ratio between TP to the total number of positive data. The F1 score is the harmonic mean of the precision and recall. In addition, we applied K-Fold cross validation, receiver operating characteristic (ROC), and area under the curve (AUC) scores to evaluate each model.

TABLE 5. Detailed view of the proposed model summary. We experimented with three variants of ResNet as the GFF-CARVING backbone.

Stage	Output	GFF-CARVING-ResNet-50	GFF-CARVING-ResNet-101	GFF-CARVING-ResNeXt-50
conv1	112×112	7×7, 64, stride 2	7×7, 64, stride 2	7×7, 64, stride 2
		3×3, max pool, stride 2	3×3, max pool, stride 2	3×3, max pool, stride 2
conv2_x	56×56	$\begin{bmatrix} 1\times 1, & 64 \\ 3\times 3, & 64\ C=32 \\ 1\times 1, & 128 \end{bmatrix}\times 4$	$\begin{bmatrix} 1\times 1, & 64 \\ 3\times 3, & 64\ C=32 \\ 1\times 1, & 256 \end{bmatrix}\times 4$	$\begin{bmatrix} 1\times 1, & 128 \\ 3\times 3, & 128\ C=32 \\ 1\times 1, & 256 \end{bmatrix}\times 4$
conv3_x	28×28	$\begin{bmatrix} 1\times 1, & 128 \\ 3\times 3, & 128\ C=32 \\ 1\times 1, & 256 \end{bmatrix}\times 4$	$\begin{bmatrix} 1\times 1, & 128 \\ 3\times 3, & 128\ C=32 \\ 1\times 1, & 512 \end{bmatrix}\times 4$	$\begin{bmatrix} 1\times 1, & 256 \\ 3\times 3, & 256\ C=32 \\ 1\times 1, & 512 \end{bmatrix}\times 4$
conv4_x	14×14	$\begin{bmatrix} 1\times 1, & 256 \\ 3\times 3, & 256\ C=32 \\ 1\times 1, & 1024 \end{bmatrix}\times 6$	$\begin{bmatrix} 1\times 1, & 256 \\ 3\times 3, & 256\ C=32 \\ 1\times 1, & 1024 \end{bmatrix}23$	$\begin{bmatrix} 1\times 1, & 512 \\ 3\times 3, & 512\ C=32 \\ 1\times 1, & 1024 \end{bmatrix}\times 6$
conv5_x	14×14	$\begin{bmatrix} 1\times 1, & 512 \\ 3\times 3, & 512\ C=32 \\ 1\times 1, & 2048 \end{bmatrix}\times 3$	$\begin{bmatrix} 1\times 1, & 512 \\ 3\times 3, & 512\ C=32 \\ 1\times 1, & 2048 \end{bmatrix}\times 3$	$\begin{bmatrix} 1\times 1, & 1024 \\ 3\times 3, & 1024\ C=32 \\ 1\times 1, & 2048 \end{bmatrix}\times 3$
	1×1	MaxPool2d	MaxPool2d	MaxPool2d
gc1		300→1024	300→1024	300→1024
gc2		1024→2048	1024→2048	1024→2048
#Parameters		25.5×10^6	44.5×10^6	25×10^6
FLOPs		10.4G FLOPs	20.1G FLOPs	10.8G FLOPs

TABLE 6. Experimental results of GFF-CARVING using several variations of ResNet using K-Fold cross validation.

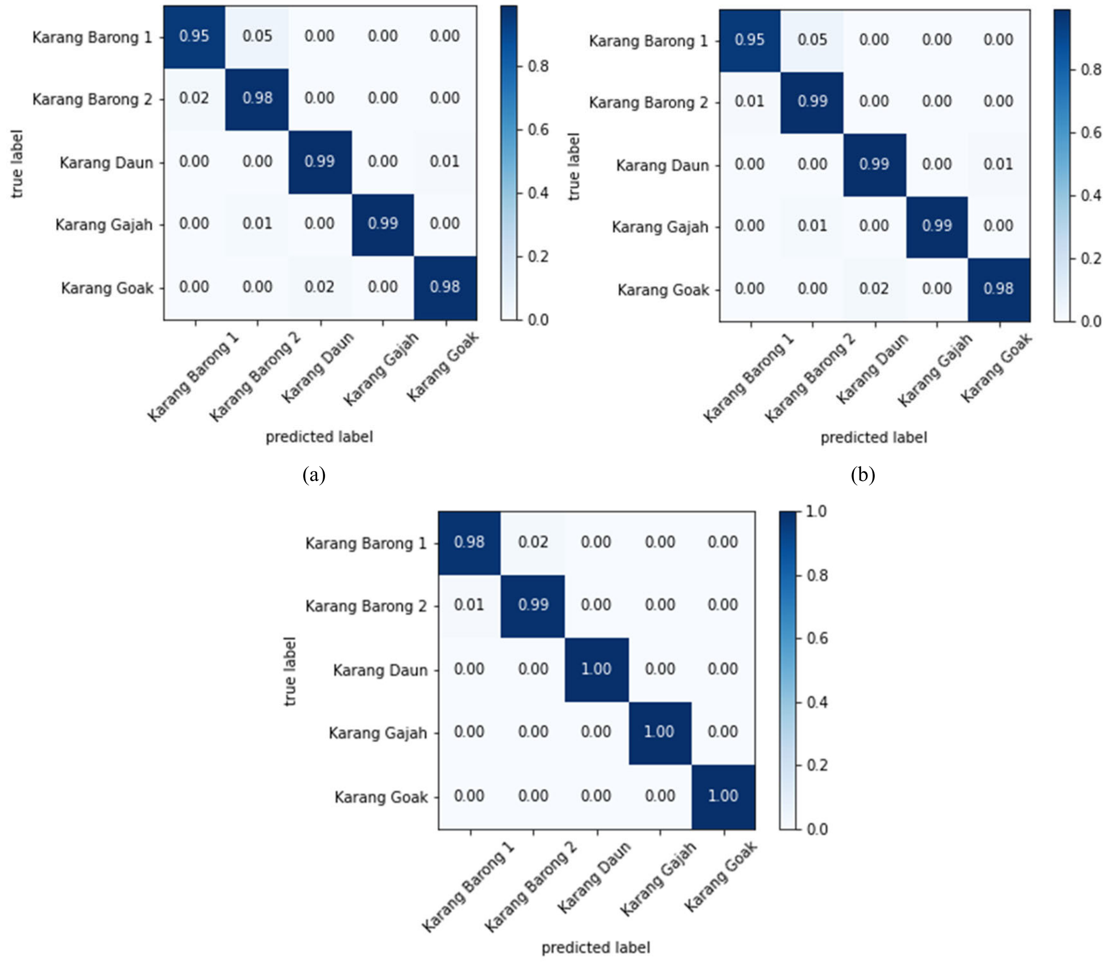
Model	K-Fold	Precision (%)	Recall (%)	F1 (%)	Testing Acc. (%)	Average acc. (%)
GFF-CARVING-ResNet-50	1	87,91	91,57	89,7	97,31	97,88
	2	95,11	96,55	96,83	98,50	
	3	**93,85**	**98,81**	**96,27**	**98,78**	
	4	96,04	90,79	93,34	97,44	
	5	89,21	93,11	91,12	97,39	
GFF-CARVING-ResNet-101	1	91,22	94,67	92,91	98,20	98,06
	2	94,20	90,23	92,17	97,86	
	3	**94,50**	**97,26**	**95,86**	**98,89**	
	4	95,76	90,98	93,31	97,07	
	5	93,61	95,34	94,47	98,30	
GFF-CARVING-ResNeXt-50	1	90,49	92,15	91,31	98,55	98,93
	2	95,54	97,32	96,62	98,89	
	3	**92,04**	**98,18**	**95,01**	**99,38**	
	4	95,57	91,36	93,61	98,48	
	5	94,17	95,63	94,89	99,35	

*Best model on each ResNet variant highlighted with bold font

C. BALINESE CARVING RECOGNITION

We conducted the recognition of Balinese carving motifs by applying our proposed method, namely GFF-CARVING, on the BaliCarv dataset. The Balinese carving recognition process was carried out by first extracting the image features and the graph features. Subsequently, feature fusion was applied to combine the image and graph features to generate features that takes into account the occurrence relationship between the sub-motifs. Finally, these features are fed through a fully connected layer for label prediction.

We used ResNet as the backbone model for the image representation learning module. We trained the model for 150 epochs. In the graph representation learning module, we used the final weighted adjacency matrix as the graph data and the vectorized form of the BaliCarv dataset to train the GCN model. The graph representation learning module is built based on GCN with two layers. The first layer accepts input data with a vector dimension = 300. Convolution in the first layer produces a 1024-dimensional feature vector based on the graph data and vectorized form of the BaliCarv dataset.



(a) (b) (c)

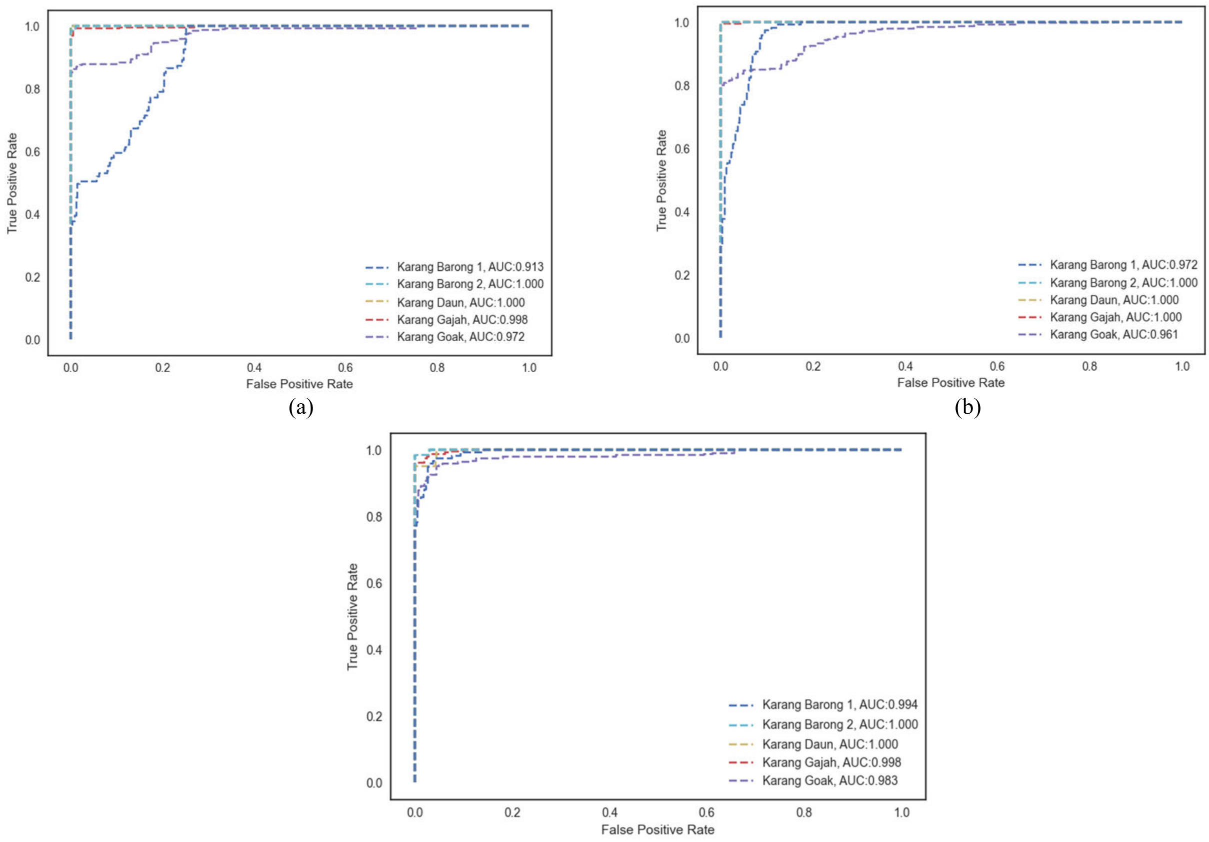
FIGURE 6. Confusion matrix comparison on three ResNet variant based on the model. (a) GFF-CARVING-ResNet-50, (b) GFF-CARVING-ResNet-101, (c) GFF-CARVING-ResNeXt-50.

We used LeakyReLU to learn and model the interrelated nodes in the graph convolution process in the first layer. In the second layer, the graph convolution process generates a 2048-dimensional feature vector. The features generated in the graph representation learning module contain information on the occurrence relationship between the sub-motif labels.

We experimented with three variants of ResNet as the GFF-CARVING backbone. Table 5 shows a detailed view of the proposed model summary. Each model has a different number of parameters, especially in Resnet-101, which has almost twice the parameter size. Table 6 shows the experimental results of the GFF-CARVING method using different variations of ResNet. We conducted experiments using three variants of ResNet as the backbone network. We performed K-Fold cross-validation to evaluate the model's performance. We split the dataset on each training into 5-folds. We used ResNet-50 as the backbone network in the first scenario to learn the image features. The experimental results in the first scenario show that the model can recognize Balinese carvings with an accuracy of 97.88%. Furthermore, we also evaluated the performance of GFF-CARVING with the use of ResNet-50 as the backbone network using precision, recall, and F1, in which it achieved the best results with 93.85%, 98.81%, and 96.27%, respectively.

In the second scenario, we applied ResNet-101 as the backbone network to extract image features. ResNet-101 has more parameters than ResNet-50. This second scenario aims to determine the number of network parameters in the feature learning process. The experimental results show that the used of ResNet-101 produced better recognition performance compared to the used of ResNet-50 as the backbone network. The GFF-CARVING method with the use of ResNet-101 achieved the best precision, recall, and F1 of 94.50%, 97.26%, and 95.86%, respectively. The 5-fold cross-validation yielded an average accuracy of 98.06%.

In the third scenario, we carried out the recognition of Balinese carving motifs using ResNeXt-50 as the backbone network. ResNeXt-50 is a variant of ResNet that applies a repeated building block that aggregates a set of transformations with the same topology [32]. The results of the third



(a)

(b)

(c)

FIGURE 7. **ROC curves and AUC scores comparison on three ResNet variant. (a) GFF-CARVING-ResNet-50, (b) GFF-CARVING-ResNet-101, (c) GFF-CARVING-ResNeXt-50.**

scenario were better than the previous two scenarios. The 5-fold cross-validation yielded an average accuracy of 98.93%. The GFF-CARVING method with the use of ResNetXt-50 achieved a precision, recall, and F1 of 92.04%, 98.18%, and 95.01%, respectively.

Fig. 6 shows the confusion matrix of the three ResNet variants used as the GFF-CARVING backbone. The three variants of ResNet showed almost the same performance, with the performance of recognizing Balinese carving motifs in each class with the lowest error of 5%. *Karang Barong 1* is the motif with the lowest recognition results per class, with 95% in the GFF-CARVING-ResNet-50 and GFF-CARVING-Resnet-101 models. In GFF-CARVING-ResNeXt-50, *Karang Barong 1's* recognition performance increased to 98%. Generally, the experimental results showed that a high recognition performance reached 98.93% using different variants of ResNet. Based on the experimental results, the model with the highest performance yielded by GFF-CARVING-ResNeXt-50 model. However, the GFF-CARVING-ResNet-101 outperforms GFF-CARVING-ResNeXt-50 based on the precision and F1, but this model has more parameters and higher FLOPs than the other ResNet variants. Fig. 7 shows the ROC curves and AUC scores comparison on three ResNet variants. The ROC curve on GFF-CARVING-ResNet-50 shows *Karang Barong 1* has the lowest AUC score of 0.913. On the other hand, GFF-CARVING-ResNet-101 and GFF-CARVING ResNeXt-50 achieved higher AUC scores in Karang Barong 1 were 0.972 and 0.994, respectively. Based on AUC scores, GFF-CARVING-ResneXt-50 produced the best performance with a higher AUC score than the ResNet-50 and Resnet-101 variants.

Therefore, based on evaluation metrics and parameter size, ResNeXt-50 as the backbone produced the highest performance. The higher FLOPs on GFF-ResNet-101 are because the ResNet-101 variant used in the image representation learning module has a larger parameter. Therefore, the ResNeXt-50 model was chosen as the primary model in GFF-CARVING. In the next section, we extend our experiment to compare the performance of the GFF-CARVING model with benchmark CNN models.

D. RECOGNITION PERFORMANCE COMPARISON

Table 7 shows the comparison of recognition performance of benchmark CNN models and the GFF-CARVING model. We evaluated the performance of GFF-CARVING by comparing it to benchmark CNN models in previous research that applied transfer learning and data augmentation strategies to overcome limited or scant data and high variation. Based on the previous method, we applied a fine-tuning strategy

TABLE 7. Comparison of recognition performance.

Model	Params. (M)	Precision (%)	Recall (%)	F1 (%)	Testing Acc. (%)	FLOPs (G)
MobileNet + TLS [33][5]	4.3	87.00	78.00	79.00	81.70	1.15
VGG16 + TLS [5]	138	85.00	84.00	80.00	83.60	31
VGG19 + TLS [5]	143	86.00	92.00	89.00	84.52	39.3
MobileNetV2 + TLS [34]	3.5	91.00	84.00	85.00	85.46	0.61
DenseNet169 + NST + GT [14]	14.3	89.00	86.00	84.00	88.76	6.76
InceptionResNetV2 + NST + GT [14]	55.9	90.00	86.00	85.00	87.16	26.4
Xception + NST + GT [35]	22.9	92.00	89.00	89.00	89.31	16.8
EfficientNetB4 + NST + GT [36]	19.5	91.00	87.00	87.00	88.21	8.98
ResNet-50 + NST + GT [37]	25.5	87.00	81.00	80.00	84.57	7.76
ResNet-101 + NST + GT [37]	44.5	90.00	94.00	91.00	88.76	15.2
ResNeXt-50 + NST + GT [38]	25	91.00	94.00	92.00	90.46	10.8
MobileNet + NST + GT [4]	5.2	93.00	92.00	91.00	91.60	1.15
GFF-CARVING (Proposed)	25	93.56	94.93	94.28	**98.93**	10.8

*TLS=Transfer Learning Strategies, NST =Neural Style Transfer, GT=Geometric Transformation
*Best model highlighted with bold font

and data augmentation method to each benchmark model to improve recognition performance on limited and high data variation data. In addition, we also compared the performance of GFF-CARVING with our previous method [4], that applied neural style transfer and geometric transformation as data augmentation method. Furthermore, we conducted further experiments by conducting an ablation study on the effect of the handcrafted graph features on the final prediction. Each benchmark CNN model was also trained using the BaliCarv dataset.

The MobileNet model achieved the lowest performance with an accuracy of 81.70%. MobileNet is a model with small parameters. MobileNet achieved precision, recall, and F1 of 87%, 78%, and 79%, respectively. The VGG16 and VGG19 models that applied transfer learning strategies with larger parameter size and higher model complexity exhibited better performance than the MobileNet model, in which the models achieved an accuracy of 85% and 84.52%, respectively. Furthermore, the ResNet-101 model achieved a higher recognition performance of 88.37%, while the MobileNetV2 achieved the performance with an accuracy of 85.46% and a precision, recall, and F1 of 91%, 84%, and 85%, respectively. Furthermore, we performed more experiments using several different architectures i.e., DenseNet169, InceptionResNetV2, Xception, and EfficientNetB4 that yielded accuracy of 88.76%, 87.16%, 89.31%, and 88.21%, respectively.

The highest accuracy achieved by the benchmark models was only 89.31%. Our previous study proposed a data augmentation technique based on neural style transfer and geometric transformation to address the scantiness of Balinese carving data [4], which increased the recognition accuracy to 91.60%. These results indicated that the previous studies that applied transfer learning strategy and data augmentation method could not fully overcome the challenges faced in recognition of Balinese carving motifs. The GFF-CARVING model proposed in this study was designed to overcome challenges faced in the recognition of Balinese carving motifs. The proposed graph-based fusion feature increased the recognition accuracy of Balinese carving motifs, which reached 98.93%. Furthermore, compared to our previous data augmentation technique, GFF-CARVING enhanced the recognition accuracy, which reached 7.33%.

We conducted further experiments by conducting an ablation study on the effect of the handcrafted graph features on the final prediction. We compared our GFF-CARVING models with the baseline ResNet-50, ResNet-101, and ResNeXt-50 that only used imaging features to determine the performance improvement of the GFF-CARVING model. Table 7 shows that the GFF-CARVING outperforms the baseline ResNet variants. Baseline ResNet variants can only achieve accuracy that reaches 90.46%. These results indicated that the handcrafted features built by utilizing the occurrence relationship between the sub-motifs label enhanced the recognition accuracy of Balinese carving motifs. Furthermore, feature fusion of the image and graph features increased the model's discriminative ability to recognize Balinese carving motifs. However, our proposed method has a larger FLOPs size than the MobileNet variants in terms of model complexity due to the used of ResNet variants in the image representation module. Hence, it requires higher computing resources.

Based on these exhaustive experiments, GFF-Carving outperformed other state-of-the art models in experiments. Handcrafted features that are built based on the characteristics of Balinese carving motifs can enrich features, thereby increasing model recognition performance. In other research domains, a similar strategy by exploiting data characteristics to build new graph features can be a strategy to improve recognition performance.

V. CONCLUSION

In this study, we proposed GFF-CARVING, a Graph Feature Fusion method for the recognition of Balinese carving motifs that addresses the challenges faced in the recognition of Balinese carving motifs, namely the highly varying and interrelated sub-motifs of Balinese carvings and the scantiness of Balinese carving data. The proposed GFF-CARVING consists of three modules, namely the image representation learning module, graph representation learning module, and prediction module. GFF-CARVING combines CNN and GCN into a unified architecture to extract image and graph features for Balinese carving recognition. We built handcrafted graph features based on the occurrence relationship between the constituent sub-motifs of Balinese carvings and extracted image features using ResNet from images of Balinese carvings. Then, feature fusion was implemented on the image and graph features to improve the discriminative ability of GFF-CARVING in recognizing Balinese carving motifs. Based on the experimental results, the proposed GFF-CARVING outperforms benchmark CNN models and achieved an accuracy of 98.93%. The experimental results indicate that the handcrafted graph features can significantly enhance the recognition of Balinese carving motifs. In addition, the proposed feature fusion of the image and graph features can generate enriched features that can improve the discriminative ability of GFF-CARVING; therefore, overcoming the challenges faced in the recognition of Balinese carving motifs. However, the proposed model requires a higher computational cost than the MobileNet variant. Therefore, it cannot be applied to mobile devices.

In future work, there is still a need for improvement in terms of model complexity in the hybrid deep learning approach. Therefore, the hybrid model can be applied to mobile devices with limited computing resources. In addition, we will apply GFF-CARVING for image retrieval of Balinese carvings to digitally archive Balinese carvings in various temples to preserve cultural heritage. This study is a significant breakthrough in the conservation of Balinese carvings. In addition, further research can be done by trying to apply GFF-Carving to other domains, by exploiting the characteristics of the data.

REFERENCES

[1] F. J. Moreno-Barea, J. M. Jerez, and L. Franco, "Improving classification accuracy using data augmentation on small data sets," *Exp. Syst. Appl.*, vol. 161, Dec. 2020, Art. no. 113696, doi: 10.1016/j.eswa.2020.113696.

[2] R. Li, X. Jia, M. Hu, M. Zhou, D. Li, W. Liu, R. Wang, J. Zhang, C. Xie, L. Liu, F. Wang, H. Chen, T. Chen, and H. Hu, "An effective data augmentation strategy for CNN-based pest localization and recognition in the field," *IEEE Access*, vol. 7, pp. 160274–160283, 2019, doi: 10.1109/ACCESS.2019.2949852.

[3] C. Dewi, R.-C. Chen, Y.-T. Liu, X. Jiang, and K. D. Hartomo, "Yolo V4 for advanced traffic sign recognition with synthetic training data generated by various GAN," *IEEE Access*, vol. 9, pp. 97228–97242, 2021, doi: 10.1109/ACCESS.2021.3094201.

[4] W. A. S. Darma, N. Suciati, and D. Siahaan, "Neural style transfer and geometric transformations for data augmentation on balinese carving recognition using MobileNet," *Int. J. Intell. Eng. Syst.*, vol. 13, no. 6, pp. 349–363, Dec. 2020, doi: 10.22266/ijies2020.1231.31.

[5] I. W. A. S. Darma, N. Suciati, and D. Siahaan, "Balinese carving recognition using pre-trained convolutional neural network," in *Proc. 4th Int. Conf. Informat. Comput. Sci. (ICICoS)*, Nov. 2020, pp. 1–5, doi: 10.1109/ICICoS51170.2020.9299021.

[6] I. M. A. Mahawan and A. Harjoko, "Pattern recognition of balinese carving motif using learning vector quantization (LVQ)," in *Proc. Int. Conf. Soft Comput. Data Sci.*, vol. 788, 2017, pp. 43–55, doi: 10.1007/978-981-10-7242-0_4.

[7] G. Zhang, J. Pan, Z. Zhang, H. Zhang, C. Xing, B. Sun, and M. Li, "Hybrid graph convolutional network for semi-supervised retinal image classification," *IEEE Access*, vol. 9, pp. 35778–35789, 2021, doi: 10.1109/ACCESS.2021.3061690.

[8] Z. Zhang, Z. Wang, S. Zhuang, and F. Huang, "Structure-feature fusion adaptive graph convolutional networks for skeleton-based action recognition," *IEEE Access*, vol. 8, pp. 228108–228117, 2020, doi: 10.1109/ACCESS.2020.3046142.

[9] L. Mou, X. Lu, X. Li, and X. X. Zhu, "Nonlocal graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8246–8257, Dec. 2020, doi: 10.1109/TGRS.2020.2973363.

[10] H. Ling, J. Wu, L. Wu, J. Huang, J. Chen, and P. Li, "Self residual attention network for deep face recognition," *IEEE Access*, vol. 7, pp. 55159–55168, 2019, doi: 10.1109/ACCESS.2019.2913205.

[11] C. Wang, D. Chen, L. Hao, X. Liu, Y. Zeng, J. Chen, and G. Zhang, "Pulmonary image classification based on inception-V3 transfer learning model," *IEEE Access*, vol. 7, pp. 146533–146541, 2019, doi: 10.1109/ACCESS.2019.2946000.

[12] K. Li, Z. Ma, L. Xu, Y. Chen, Y. Ma, W. Wu, F. Wang, and Z. Liu, "Depthwise separable ResNet in the MAP framework for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2020.3033149.

[13] A. Pal, Z. Xue, B. Befano, A. C. Rodriguez, L. R. Long, M. Schiffman, and S. Antani, "Deep metric learning for cervical image classification," *IEEE Access*, vol. 9, pp. 53266–53275, 2021, doi: 10.1109/ACCESS.2021.3069346.

[14] N. P. Sutramiani, N. Suciati, and D. Siahaan, "MAT-AGCA: Multi augmentation technique on small dataset for balinese character recognition using convolutional neural network," *ICT Exp.*, vol. 7, no. 4, pp. 521–529, Dec. 2021, doi: 10.1016/J.ICTE.2021.04.005.

[15] L. Zhou, Z. Zhang, X. Yin, H.-B. Jiang, J. Wang, G. Gui, Y.-C. Chen, and J.-X. Zheng, "Transfer learning-based DCE-MRI method for identifying differentiation between benign and malignant breast tumors," *IEEE Access*, vol. 8, pp. 17527–17534, 2020, doi: 10.1109/ACCESS.2020.2967820.

[16] G. Huo, Z. Wu, and J. Li, "Underwater object classification in sidescan sonar images using deep transfer learning and semisynthetic training data," *IEEE Access*, vol. 8, pp. 47407–47418, 2020, doi: 10.1109/ACCESS.2020.2978880.

[17] N. P. Sutramiani, N. Suciati, and D. Siahaan, "Transfer learning on balinese character recognition of lontar manuscript using MobileNet," in *Proc. 4th Int. Conf. Informat. Comput. Sci. (ICICoS)*, Nov. 2020, pp. 1–5, doi: 10.1109/ICICoS51170.2020.9299030.

[18] G. Fan, F. Chen, D. Chen, and Y. Dong, "Recognizing multiple types of rocks quickly and accurately based on lightweight CNNs model," *IEEE Access*, vol. 8, pp. 55269–55278, 2020, doi: 10.1109/ACCESS.2020.2982017.

[19] I. Ali, M. Muzammil, I. U. Haq, A. A. Khaliq, and S. Abdullah, "Deep feature selection and decision level fusion for lungs nodule classification," *IEEE Access*, vol. 9, pp. 18962–18973, 2021, doi: 10.1109/ACCESS.2021.3054735.

[20] H. Kusetogullari, A. Yavariabdi, J. Hall, and N. Lavesson, "DIGITNET: A deep handwritten digit detection and recognition method using a new historical handwritten digit dataset," *Big Data Res.*, vol. 23, Feb. 2021, Art. no. 100182, doi: 10.1016/j.bdr.2020.100182.

[21] L. Nanni, S. Ghidoni, and S. Brahnam, "Ensemble of convolutional neural networks for bioimage classification," *Appl. Comput. Informat.*, vol. 17, no. 1, pp. 19–35, Jan. 2021, doi: 10.1016/j.aci.2018.06.002.

[22] S. Banerjee and S. S. Chaudhuri, "Total contribution score and fuzzy entropy based two-stage selection of FC, ReLU and inverseReLU features of multiple convolution neural networks for erythrocytes detection," *IET Comput. Vis.*, vol. 13, no. 7, pp. 640–650, Oct. 2019, doi: 10.1049/iet-cvi.2018.5545.

[23] J. U. N. Liu, Y. Feng, and H. Wang, "Facial expression recognition using pose-guided face alignment and discriminative features based on deep learning," *IEEE Access*, vol. 9, pp. 69267–69277, 2021, doi: 10.1109/ACCESS.2021.3078258.

[24] C. I. Patel, S. Garg, T. Zaveri, A. Banerjee, and R. Patel, "Human action recognition using fusion of features for unconstrained video sequences," *Comput. Electr. Eng.*, vol. 70, pp. 284–301, Aug. 2018, doi: 10.1016/j.compeleceng.2016.06.004.

[25] W. Liu, G. Wu, F. Ren, and X. Kang, "DFF-ResNet: An insect pest recognition model based on residual networks," *Big Data Mining Anal.*, vol. 3, no. 4, pp. 300–310, Dec. 2020, doi: 10.26599/BDMA.2020.9020021.

[26] M. B. Lee, Y. H. Kim, and K. R. Park, "Conditional generative adversarial network-based data augmentation for enhancement of iris recognition accuracy," *IEEE Access*, vol. 7, pp. 122134–122152, 2019, doi: 10.1109/ACCESS.2019.2937809.

[27] R. Man, P. Yang, and B. Xu, "Classification of breast cancer histopathological images using discriminative patches screened by generative adversarial networks," *IEEE Access*, vol. 8, pp. 155362–155377, 2020, doi: 10.1109/ACCESS.2020.3019327.

[28] Q. Liu, C. K. Leung, and P. Hu, "A two-dimensional sparse matrix profile DenseNet for COVID-19 diagnosis using chest CT images," *IEEE Access*, vol. 8, pp. 213718–213728, 2020, doi: 10.1109/ACCESS.2020.3040245.

[29] C. Wang, L. Yu, X. Zhu, J. Su, and F. Ma, "Extended ResNet and label feature vector based chromosome classification," *IEEE Access*, vol. 8, pp. 201098–201108, 2020, doi: 10.1109/ACCESS.2020.3034684.

[30] S. Lu, Q. Hong, B. Wang, and H. Wang, "Efficient ResNet model to predict protein-protein interactions with GPU computing," *IEEE Access*, vol. 8, pp. 127834–127844, 2020, doi: 10.1109/ACCESS.2020.3005444.

[31] C. Patel, D. Bhatt, U. Sharma, R. Patel, S. Pandya, K. Modi, N. Cholli, A. Patel, U. Bhatt, M. A. Khan, S. Majumdar, M. Zuhair, K. Patel, S. A. Shah, and H. Ghayvat, "DBGC: Dimension-based generic convolution block for object recognition," *Sensors*, vol. 22, no. 5, p. 1780, Feb. 2022, doi: 10.3390/s22051780.

[32] S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He, and U. S. Diego, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1492–1500, [Online]. Available: https://github.com/facebookresearch/ResNeXt

[33] A. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.

[34] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4510–4520.

[35] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1800–1807, doi: 10.1109/CVPR.2017.195.

[36] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Mach. Learn. ICML* May 2019, pp. 10691–10700.

[37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[38] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5987–5995, doi: 10.1109/CVPR.2017.634.

I WAYAN AGUS SURYA DARMA (Member, IEEE) received the master's degree in information systems and computer management from Udayana University, in 2015, and the Ph.D. degree in computer science from the Department of Informatics, Faculty of Intelligent Electrical and Informatics Technology, Institut Teknologi Sepuluh Nopember, in 2022. His research interests include computer vision, image processing, and artificial intelligence.

NANIK SUCIATI (Member, IEEE) received the master's degree in computer science from the University of Indonesia, in 1998, and the Dr.-Eng. degree in information engineering from the University of Hiroshima, in 2010. She is currently an Associate Professor with the Department of Informatics, Faculty of Intelligent Electrical and Informatics Technology, Institut Teknologi Sepuluh Nopember. She has published more than 50 journal articles and conference papers related to computer science. Her research interests include computer vision, computer graphics, and artificial intelligence.

DANIEL SIAHAAN (Member, IEEE) received the master's degree in software engineering from Technische Universiteit Delft, in 2002, and the P.D. (Eng.) degree in software engineering from Technische Universiteit Eindhoven, in 2004. He is currently an Associate Professor with the Department of Informatics, Faculty of Intelligent Electrical and Informatics Technology, Institut Teknologi Sepuluh Nopember. He has published more than 50 journal articles and conference papers related to software engineering. His research interests include software engineering, requirements engineering, and natural language processing.