JCSE Journal of Computing Science and Engineering

Regular Paper
Journal of Computing Science and Engineering,
Vol. 13, No. 3, September 2019, pp. 124-130

Fish Species Recognition Using VGG16 Deep Convolutional Neural Network

Praba Hridayami*, I Ketut Gede Darma Putra, and Kadek Suar Wibawa
Department of Information Technology, Udayana University, Badung, Bali, Indonesia
prabahridayami@student.unud.ac.id, ikgdarmaputra@unud.ac.id, suar_wibawa@unud.ac.id

Abstract
Conservation and protection of fish species is very important in aquaculture and marine biology. A few studies have introduced the concept of fish recognition; however, it resulted in poor rates of error recognition and conservation of a small number of species. This study presents a fish recognition method based on deep convolutional neural networks such as VGG16, which was pre-trained on ImageNet via transfer learning method. The fish dataset in this study consists of 50 species, each covered by 15 images including 10 images for training purpose and 5 images for testing. In this study, we trained our model on four different types of dataset: RGB color space image, canny filter image, blending image, and blending image mixed with RGB image. The results showed that blending image mixed with RGB image trained model exhibited the best genuine acceptance rate (GAR) value of 96.4%, following by the RGB color space image trained model with a GAR value of 92.4%, the canny filter image trained model with a GAR value of 80.4%, and the blending image trained model showed the least GAR value of 75.6%.

Category: Smart and Intelligent Computing

Keywords: Fish recognition; Deep convolutional neural network; Transfer learning; Canny filter; VGG16

I. INTRODUCTION

The oceans constitute approximately 70% or two-thirds of the earth surface geographically. Marine ecosystems play an important role in the balance of nature, with a diverse range of fish scattered in the sea. Indeed, more than 22,000 species of fish account for more than half of all the 55,000 species of living vertebrates [1]. Studying fish species is, therefore, very important for the preservation and protection of aquaculture and marine biology.

Previously, fish species were identified manually by observation, and therefore, it was necessary to remember or study various fish characteristics to recognize fish species. Advanced technology equipped with artificial intelligence (AI) using deep learning methods facilitates the recognition of a diverse range of fish species. Convolutional neural network (CNN) is one of the most popular deep learning methods used currently. CNN is widely used as an object classification method, which is a further development of multilayer perceptron (MLP). Image classification via CNN has been successful due to the use of multiple dimensions affecting the overall scale of an object [2, 3].

All the currently available studies deal with small datasets distinguishing fewer species or show poor accuracy [4]. Our study used the transfer learning method in the absence of adequate labeled training data to train our networks. We used VGG16, which has been trained

Open Access http://dx.doi.org/10.5626/JCSE.2019.13.3.124 http://jcse.kiise.org

Received 06 July 2019; **Revised** 14 September 2019; **Accepted** 17 September 2019
*Corresponding Author

 pISSN: 1976-4677 eISSN: 2093-8020

previously in ImageNet dataset. The need for large amounts of data and high computational resources to train CNN from scratch encourage researchers to adapt pre-trained networks for a desired task domain by fine tuning the domain-specific data [5]. Training a CNN using small datasets leads to overfitting and reducing their ability to generalize unseen invariant data [6]. We used a deep convolutional neural network such as VGG16 to train 50 classes. Pretrained VGG16 is much deeper and consists of 16 weight layers.

The classification of fish species was based on features such as fish color, fish shape, and fish texture. We also performed a blending image process to accentuate the special features of fish species without losing its color. In this study, we compared the following four types of dataset for training models: RGB color space image dataset, canny filter image dataset, blending image dataset and blending image mixed with RGB image dataset.

II. RELATED WORK

AI has been growing rapidly in recent years. The number of publications in the context of AI growth has increased over the past 16 years, suggesting expanding collaboration in the field of AI and widening of the scope of research projects. CNN is one of the deep learning methods that has been successfully applied to various fields such as plants (flowering species), plant organs, and animal species. In this case, supporting studies were taken from various journals and previous studies pertinent to our investigation.

The study entitled "Tree Species Identification Based on Convolutional Neural Networks" authored by Zhou et al. [7] identifies tree species using CNN by analyzing leaves based on variation in color, shape and leaf streaks. This study used several preprocessing steps to clear the image without affecting the training process, which altered the image to grayscale, using the Otsu algorithm to segment the foreground and background, and changed the background value of the image to (255, 255, 255) for a clean background. Twenty-five classes of tree species were introduced and trained 200 times. The accuracy obtained for the model without rotational variation was 87.57%, and the accuracy of the model with rotational variation was 91.36%.

Guo and Gao [8] from the Beijing School of Electronic and Information Engineering Beihang University, China reported "A Multi-Organ Plant Identification Method Using Convolutional Neural Networks" involving plant organ identification using CNN as an algorithm for training classification. The dataset used was Plant Photo Bank of China from the Institute of Botany, at the Chinese Academy of Science (IB-CAS). This study used several algorithms, for example, linear weighted with an accuracy of 84.5%, SVM and RBF with 85.4% accuracy, and SVM and sigmoid with 86.6% accuracy.

The publication entitled "Flower Species Recognition System using Convolutional Neural Networks and Transfer Learning" by Gogul and Kumar [9] investigated flower species using CNN. The feature extraction approach used in this study was transfer learning. CNN combined with the transfer learning approach outperformed all manual feature extraction methods such as local binary pattern, Color Channel Statistics, Color Histograms, Haralick Texture, Hu Moments, and Zernike Moments. CNN combined with the transfer learning approach yielded an accuracy of 73.05%, 93.41%, and 90.60%, using the OverFeat, Inception-v3, and Xception architectures, respectively, as feature extractors in the FLOW-ERS102 dataset [9].

III. PROPOSED METHOD

A. Convolutional Neural Networks

The popularity of machine learning is increasing with the emergence of artificial neural network (ANN). ANN is a computational processing system that was inspired by the operation of biological nervous system (such as the human brain) [10]. CNN is one of the most impressive forms of ANN architecture.

CNNs (ConvNet) are a class of deep ANNs, which are applied to analyze visual imagery [11]. CNN consists of neurons that carry weight, bias and activation functions. Similar to other neural networks, CNN comprises input, hidden and output layers, which carry out operations that alter data in order to investigate the specific features. Three of the most common layers include convolution, activation or rectified linear unit (ReLU), and union [8].

Convolution exposes the input image through a series of convolutional filters, which activate the corresponding features of the image. ReLU enables faster and more effective training by mapping negative to zero values and maintaining positive values. ReLU is occasionally referred to as activation, because only the activated feature is carried forward to the next layer. Pooling simplifies output via non-linear down sampling, reducing the number of parameters that the network needs to learn. These three operations are repeated across more than tens or hundreds of layers, in which each learning layer is associated with different features.

The next layer is a fully connected layer that produces a dimension K vector, where K is the number of classes that the network can predict. This vector contains probabilities for each class of any classified image. The last layer of the CNN architecture uses a classification layer such as softmax to yield a classification output of more than two classes or sigmoid to generate output classifications of fewer than or equal to two classes.

VGG16 is one of the VGGnet models that uses 16

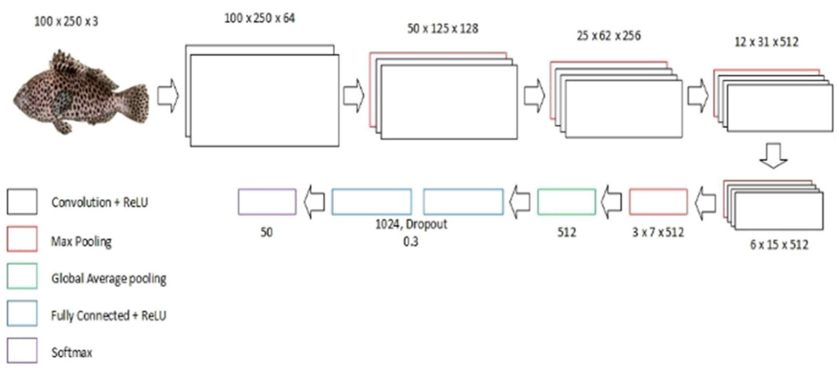

Fig. 1. VGG16 architecture.

layers as a model architecture (Fig. 1). A normal VGG16 consists of 5 convolutional blocks before connecting to a MLP classifier. The convolutional block is connected to three MLP layers comprising two hidden layers and one output layer. The output layer consists of nodes that directly represent the number of classes, and softmax activation functions (for more than two classes) or sigmoid activation functions (for classes fewer than or equal to two) [12].

B. Transfer Learning

Once the CNN architecture is designed, the next phase is to train it with a huge amount of data. The re-use of a model that has performed well in a certain task to solve a different but related task is known as transfer learning (TL) [13]. TL aims to transfer knowledge from a large dataset known as source domain to a smaller dataset, which is the target domain. In this study, we used VGG16 which was trained previously in ImageNet dataset as the model. ImageNet is an image dataset organized according to the WordNet hierarchy (currently only the nouns), in which each node of the hierarchy is depicted by hundreds and thousands of images. Currently, each node carries an average of over 500 images.

C. Introduction to Dataset

The dataset used was derived from QUT FISH at Queensland University of Technology Robotics and includes fish species images and labels. QUT FISH provides 750 cropped images from 50 species [14]. The dataset image sizes are not similar to one another. Therefore, we resized the images to a uniform size of 100×250 pixels.

D. Image Enhancement

The dataset assigned during the training process is one of the factors that determines the performance of neural network. If the datasets assigned are few in number, the accuracy of the neural network is worse than that of the neural network trained using large numbers of datasets. To overcome this problem, data augmentation is needed to add variety to the existing dataset. Data augmentation refers to any method that artificially inflates the original training set with label-preserving transformations [6]. Data augmentation can be represented as shown below:

$$Ø : S \rightarrow T, \tag{1}$$

where S is the original training set and T denotes the augmented set of S. The artificially inflated training set is represented as follows:

$$S' = S \cup T, \tag{2}$$

where S' contains the original training set and the respective transformations defined by Ø.

Data augmentation can be divided into two methods: geometric and photometric methods. Transformation is carried out in geometric methods such as zoom, flipping, rotation and cropping schemes, followed by transformation using photometric methods such as color jittering, edge enhancement and fancy PCA [6].

In this study, we used three kinds of data augmentation: zoom range by 0.1, rotation range by 10, and horizontal and vertical flip. Sample data augmentation results in this study are presented in Fig. 2.

To improve the features used as traits in the fish image, it is necessary to perform edge detection in order to mark detailed portions on images and enhance blurry images. Edge detection is one of the most important processes in image processing, and the detection results directly affect image analysis [15]. A point (x, y) represents the edge of the image if that point exhibits a high-intensity difference compared with its neighbor.

The aim of Canny edge detection algorithm is to ensure adequate detection (minimum number of false edges), appropriate localization (proximity of the real edge and the detected edge) and minimal response (one edge should be detected only once) [16]. Canny edge detection can facilitate the segmentation process to yield the fish shape by forming contour lines as shown in Fig. 3(b).

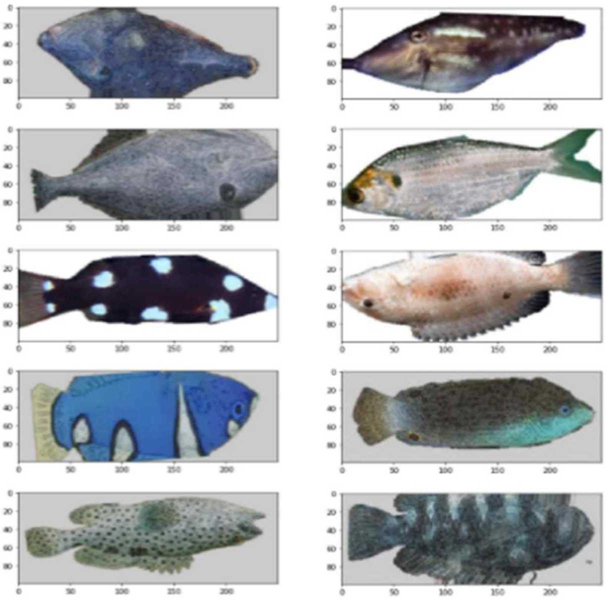

Fig. 2. Data augmentation.

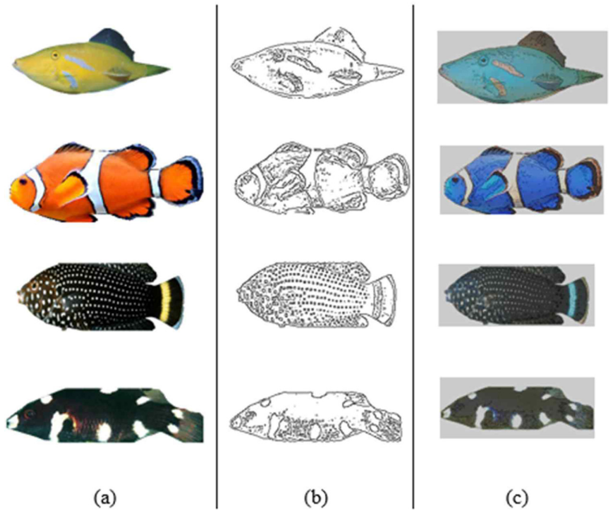

Fig. 3. (a) RGB images, (b) canny filter images, and (c) blending images.

Canny uses Gaussian derivative kernel to filter noise from the image to obtain a smooth edge followed by image gradient to calculate edge strength and direction. Furthermore, the non-maximum suppression stage is carried out in order to accurately determine the position of the edges and ensure that each edge has a width of one pixel. Finally, the thresholding is determined to classify each pixel whether or not it is included in the edge category.

Images are built from a collection of pixels consisting of thousands or millions of pixels with a uniform arrangement. Each pixel has its own intensity or brightness value ranging from the darkest and the lightest of three different colors: Red, Green, and Blue [17]. The various combinations of these color intensities produce a color image.

Blending image combines two or more images simultaneously, which allows the perception of concurrent images, although with different emphasis [18]. In this study, we blended RGB color space and canny filter images, as shown in Fig. 3(a). The results of canny filter on RGB fish images are shown in Fig. 3(b). After the image has been changed into the RGB color space and using the canny filter, image blending was performed to yield the image as shown in Fig. 3(c).

The main purpose of this study is to compare four types of dataset including RGB color space image dataset, canny filter image dataset, blending image dataset, and blending image mixed with RGB image dataset.

Fig. 3(a) shows the result of changing the image color space into RGB color space. Each original image with a different color space before was changed into RGB color space to equalize them. We used RGB because RGB displays match our perception system.

Fig. 3(b) presents canny filter images showing the necessary features of each fish species that is distinct. Canny filter is a segmentation that can be used under different situations. Canny edge detection technique is used for object recognition and pattern matching, where it is necessary to retain the features even in case of noisy images [19].

Fig. 3(c) shows blending images, which strengthen the features of the fish species. In this study, the image blending was carried out by combining the RGB color space fish image and the image with canny filter.

IV. EXPERIMENTAL RESULTS AND ANALYSES

Our study sought to obtain the lowest equal error rate (EER) in fish species recognition. We derived the EER value from false acceptance rate (FAR) and false rejection rate (FRR). FAR occurs when a fish species is identified with another fish species. This type of issue is also referred to as a false positive event. FRR involves missed detection and failure to identify a legitimate fish. This type of issue is commonly known as a false negative event. If we plot both the FAR and FRR on a graph, the EER is the point where the two lines intersect [20]. EER is sometimes used as a measure of the accuracy of biometric systems. Genuine acceptance rate (GAR) is defined as a percentage of genuine users accepted by the system.

$$\mathrm{FAR} = \frac{\textit{Total number of fish species identified with another fish species}}{\textit{Total number of Tests Performed}}$$

$$\mathrm{FRR} = \frac{\textit{Total number genuine of fish species rejected}}{\textit{Total number of Tests Performed}}$$

$$\mathrm{GAR} = 1 - FRR \tag{3}$$

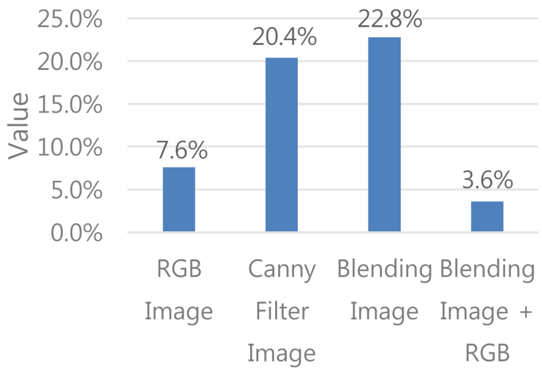

Fig. 4. FAR chart.

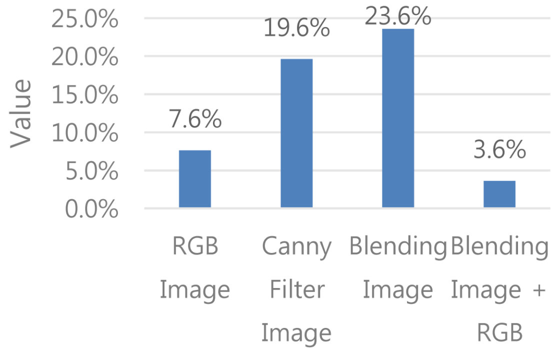

Fig. 6. FRR chart.

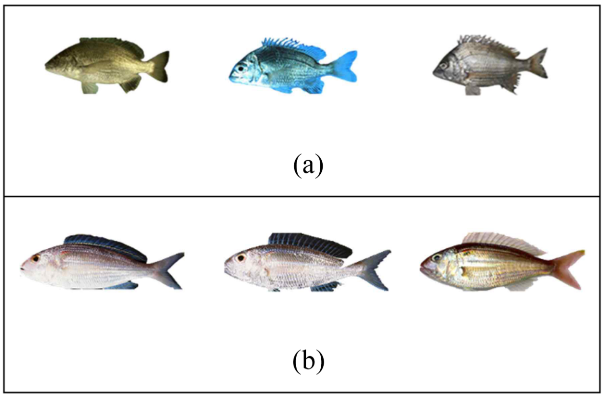

Fig. 5. (a) *Acanthopagrus berda* species and (b) *Nemipterus hexodon* species.

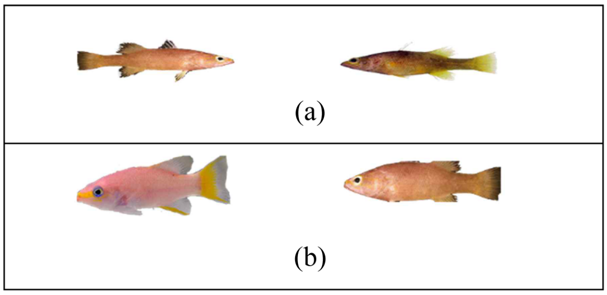

Fig. 7. (a) *Liopropoma mitratum* unaccepted and (b) *Liopropoma* mitratum accepted.

We made several attempts to test our trained model using RGB image dataset to obtain the best result using different image datasets. The final result showed that the blending image mixed with RGB image trained model yielded the best performance among all the experiments. We visualized the three values using a bar chart, and compared all four experiments based on the specific values.

Fig. 4 shows that the blending image mixed with RGB image trained model yielded the lowest false acceptance rate suggesting that from 250 test images only 9 images were accepted in the wrong classes. We found that the 9 images that were included in the wrong classes showed almost similar features to the classes they were accepted under.

Fig. 5(a) illustrates the test images with false-positive results. *Acanthopagrus berda* test images were mistaken as *Nemipterus* hexodon species. We also suspect that the blending image mixed with RGB trained model showed the best training result because it included additional images for training compared with the other models. In addition, we included the original RGB images in the training dataset so the test process not only tested the test data in the blending image, which strengthens the features, but also tested it with the original RGB image training result.

Fig. 6 also shows that the blending image mixed with RGB image trained model exhibited the lowest false rejection rate. Based on the result obtained, the 9 images were not recognized in any classes.

Fig. 7(a) represents the test images showing false-negative results, which were unaccepted in any class with a 0% recognition result. Fig. 7(b) represents the test images that were accepted for the *Liopropoma mitratum* species class. The worst performance of the trained model involved testing of the blending image. We suspect that the blending image only trained with model color was not compatible with the test dataset, which contains the original RGB images. The canny filter-trained model exhibited similar limitations.

Fig. 8 shows the GAR chart presenting all the data tests, which were accepted by the recognition system. The highest value of GAR was 96.4% comprising blending images mixed with RGB trained model.

The test results are shown in Table 1, which lists the test values of CNNs using VGG16 pre-trained on ImageNet using a combined model of data enhancement. The results suggest that the VGG16 model with the best performance had the lowest EER of 3.6% that crossed at

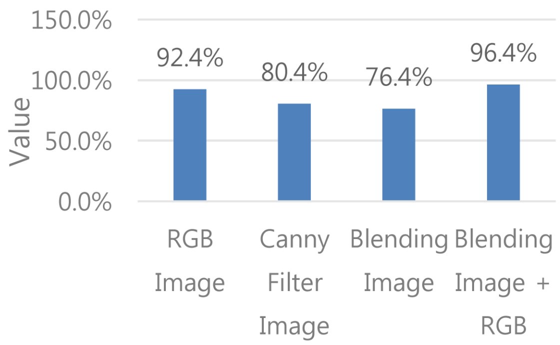

Fig. 8. GAR chart.

Table 1. The result of four attempts

	FAR (%)	FRR (%)	Threshold
RGBI	7.6	7.6	62
CFI	20.4	19.6	79
BI	22.8	24.4	85
BI + RGBI	3.6	3.6	72

RGBI: red green blue image, CFI: canny filter image, BI: blending image.

threshold 72 in the blending image model mixed with trained RGB image. The poorest performance of VGG16 model, which had the highest EER value is represented by the blending image-only trained model scoring 24.4% that crossed at threshold 85.

V. CONCLUSION

We used TL to identify fish species based on the CNN architecture of VGG16 model. We executed the model using the QUT FISH dataset comprising 50 species of fish, with 15 images each. We also performed an augmentation process to decrease the risk of overfitting in numerous image variations. The augmentation techniques were based on image transformations such as zoom, rotation, and flipping. Based on the experiment results and analysis, the pre-trained VGG16 yielded the best performance by blending image mixed with RGB image dataset. Our finding indicates that pre-training models on ImageNet and image enhancement can overcome the limited number of images in the dataset and decrease the error rate of the identification result.

ACKNOWLEDGMENTS

This study has been supported by Department of Information Technology, Faculty of Engineering, Udayana University, Bali, Indonesia. We would also like to thank the QUT (Queensland University of Technology) FISH for the fish species dataset.

REFERENCES

1. G. Helfman, B. B. Collette, D. E. Facey, and B. W. Bowen, *The Diversity of Fishes: Biology, Evolution, and Ecology.* Chichester: Wiley-Blackwell, 2009.
2. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, 2016, pp. 770-778.
3. V. Allken, N. O. Handegard, S. Rosen, T. Schreyeck, T. Mahiout, and K. Malde, “Fish species identification using a convolutional neural network trained on synthetic data,” *ICES Journal of Marine Science*, vol. 76, no. 1, pp. 342-349, 2018.
4. D. Rathi, S. Jain, and S. Indu, “Underwater fish species classification using convolutional neural network and deep learning,” in *Proceedings of 2017 9th International Conference on Advances in Pattern Recognition (ICAPR)*, Bangalore, India, 2017, pp. 1-6.
5. M. M. Ghazi, B. Yanikoglu, and E. Aptoula, “Plant identification using deep neural networks via optimization of transfer learning parameters,” *Neurocomputing*, vol. 235, pp. 228-235, 2017.
6. L. Taylor and G. Nitschke, “Improving deep learning with generic data augmentation,” in *Proceedings of 2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, Bangalore, India, 2018, pp. 1542-1547.
7. H. Zhou, C. Yan, and H. Huang, “Tree species identification based on convolutional neural networks,” in *Proceedings of 2016 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, Hangzhou, China, 2016, pp. 103-106.
8. P. Guo and Q. Gao, “A multi-organ plant identification method using convolutional neural networks,” in *Proceedings of 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, 2017, pp. 371-376.
9. I. Gogul and V. S. Kumar, “Flower species recognition system using convolution neural networks and transfer learning,” in *Proceedings of 2017 4th International Conference on Signal Processing, Communication and Networking (ICSCN)*, Chennai, India, 2017, pp. 1-6.
10. K. O’Shea and R. Nash, “An introduction to convolutional neural networks,” 2015; https://arxiv.org/abs/1511.08458.
11. Standford University, “Introduction to convolutional neural networks,” 2018; https://web.stanford.edu/class/cs231a/lectures/intro_cnn.pdf.
12. M. A. H. Abas, N. Ismail, A. I. M. Yassin, and M. N. Taib, “VGG16 for plant image classification with transfer learning and data augmentation,” *International Journal of Engineering & Technology*, vol. 7, no. 4.11, pp. 90-94, 2018.
13. S. K. Manikonda and D. N. Gaonkar, “A novel islanding detection method based on transfer learning technique using

VGG16 network," in *Proceedings of 2019 IEEE International Conference on Sustainable Energy Technologies (ICSET)*, Bhubaneswar, India, 2019, pp. 109-114.
14. K. Anantharajah, Z. Ge, C. McCool, S. Denman, C. Fookes, P. Corke, D. Tjondronegoro, and S. Sridharan, "Local inter-session variability modelling for object classification," in *Proceedings of IEEE Winter Conference on Applications of Computer Vision*, Steamboat Springs, CO, 2014, pp. 309-316.
15. W. Rong, Z. Li, W. Zhang, and L. Sun, "An improved CANNY edge detection algorithm," in *Proceedings of 2014 IEEE International Conference on Mechatronics and Automation*, Tianjin, China, 2014, pp. 577-582.
16. M. H. Shokhan, "An efficient approach for improving canny edge detection algorithm," *International Journal of Advances in Engineering & Technology*, vol. 7, no. 1, pp. 59-64, 2014.
17. E. J. Hong, A. Saudi, and J. Sulaiman, "Application of SOR iteration for Poisson image blending," in *Proceedings of the International Conference on High Performance Compilation, Computing and Communications*, Kuala Lumpur, Malaysia, 2017, pp. 60-64.
18. P. Kovesi, E. J. Holden, and J. Wong, "Interactive multi-image blending for visualization and interpretation," *Computers & Geosciences*, vol. 72, no. 147-155, 2014.
19. Kirti and A. Bhatnagar, "Image segmentation using canny edge detection technique," *International Journal of Techno-Management Research*, vol. 4, no. 4, pp. 8-14, 2017.
20. J. Andress, *The Basics of Information Security: Understanding the Fundamentals of InfoSec in Theory and Practice*. San Diego, CA: Elsevier, 2014.

Praba Hridayami

Praba Hridayami was born on December 19, 1997 in Mengwitani, Bali, Indonesia. She is a student in Department of Information Technology, Udayana University, Bali, Indonesia. Her research interests are in image processing, pattern recognition, computer vision, and machine learning.

I Ketut Gede Darma Putra

I Ketut Gede Darma Putra was born in Mengwitani, 24 April 1974. A lecturer in Department of Electrical Engineering and Information Technology, Udayana University Bali, Indonesia. He received his S.Kom degree in Informatics Engineering from Institute of Sepuluh November Technology Surabaya, Indonesia in 1997. He received his master degree on Informatics and Computer Engineering from Electrical Engineering Department, Gadjah Mada University, Indonesia in 2000 and achieved his doctorate degree on Informatics and Computer Engineering from Electrical Engineering Department, Gadjah Mada University, Indonesia in 2006. His research interests are in biometrics, image processing, data mining, and soft computing.

Kadek Suar Wibawa

Kadek Suar Wibawa was born in Sangsit on August 16, 1983. A lecturer in Department of Information Technology, Udayana University, Bali, Indonesia. He received his master degree from Department of Informatics Technology, Bandung Institute of Technology, Indonesia, in 2013. His research interests are in Internet of Things, computer network, and Android developer.