

Handwriting identification using deep convolutional neural network method

Oka Sudana, I Wayan Gunaya, I Ketut Gede Darma Putra

Department of Information Technology, Udayana University, Indonesia

Article Info

Article history:

Received Dec 8, 2019

Revised Feb 28, 2020

Accepted Mar 13, 2020

Keywords:

Biometrics

Convolutional neural network

Transfer learning

Writer identification

ABSTRACT

Handwriting is a unique thing that produced differently for each person. Handwriting has a characteristic that remain the same with single writer, so a handwriting can be used as a variable in biometric systems. Each person have a different form of handwriting style but with a small possibility that same characters have something commons. We propose a handwriting identification method using sentence segmented handwriting forms. Sentence form is used to get more complete handwriting characteristics than using a single characters or words. Dataset used is divided into three categories of images, binary, grayscale, and inverted binary. All datasets have same image with different in color and consist of 100 class. Transfer learning used in this paper are pre-trained model VGG19. Training was conducted in 100 epochs. Highest result is grayscale images with genuince acceptance rate of 92.3% and equal error rate of 7.7%.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

I Wayan Gunaya,

Department of Information Technology,

Udayana University,

Raya Kampus Unud Jimbaran St., Kec. Kuta Sel., Kabupaten Badung, Bali, Indonesia.

Email: wayangunaya@student.unud.ac.id

1. INTRODUCTION

The handwriting identification is a problem that is currently still being studied. Handwriting is easy to recognize because of a number of things, like the form of letters and different styles of writing between each person. Handwriting identification can be very useful especially for security sector, banking, and forensics. The banking sector requires handwriting recognition to verify receipts or when disbursing check money. In forensic field, handwriting identification can be used as a benchmark for a criminal case related to a handwriting [1].

Handwriting identification has been done using simple and long-standing methods. Machine learning is a new field of science where this field use a collection of data and use that data to study it patterns to get a certain knowledge. When the learning process about existing patterns deepens, it will become deep learning. Deep learning is the process of learning artificial neural networks where patterns are studied more deeply to produce better results. Deep learning mimics the ability of human brain to remember things.

Implementation of deep learning will be able to learn every feature that exists if the network is made deeper. Deep learning is designed to use several layers of artificial neural network to learn about each feature. Architecture in deep learning models play an important role in the ability of network to study existing data. There are several models in deep learning, such as deep neural networks, convolutional neural network, and recurrent neural network [2].

Convolutional neural network (CNN) is a deep learning algorithm that use convolution layer for feature extraction and a fully connected layer for its classification. Training process in the convolutional neural network use backpropagation algorithm to update the weight and bias for each epoch. Each neuron found on CNN will be connected to the input image, so architectures of CNN will need several deep layers. The classification process will use fully connected layers to classify existing features with predetermined classes. One of the disadvantages of the CNN method is that the amount of data needed is quite large to get optimal results. The hardware needed must also be sufficient enough [3].

Handwriting identification by utilizing deep learning algorithms will be able to produce higher accuracy values compared to other methods. Handwriting collected from various authors that will be used as data to be studied. The use of deep learning algorithm is expected to be able to improve the results of handwriting identification so that it can be used properly [4]. There are several studies and research paper that related to this research that have been carried out. Some of them in the form of character recognition and handwriting identification using other or older method. Research about character recognition has been conducted by Dewa entitled "Convolutional Neural Networks for Handwritten Javanese Characters Recognition" [5]. This study used convolutional neural network method to recognize Javanese traditional characters in the form of handwriting. They used 2000 data which is included in 20 classes of Javanese characters. This research uses OpenCV library with a self-made CNN architecture. The results show that the accuracy rate obtained is 90%.

Research related to handwriting identification has been carried out by several previous researchers, one of them is a study conducted by Dhandra [6]. In this research, handwriting identification uses Kannada characters and use random transform method and discrete cosine transform which is combined to identify the author. The result of this study indicate that the level of accuracy reaches 100%. In addition, it was also concluded that sentences, varying structural variations, and/or a combination of two or more words had a significant impact on the identification of handwriting's author. Another research was conducted where handwriting identification is done online [7]. This study used data acquired online that will be processed directly. The method used in this research is CNN with drop segment method. This method used data augmentation by dropping each segment of the entered character. The results of this study are 98% accuracy for English handwriting and 95% for Chinese handwriting.

This paper aims to analyze handwriting identification by convolutional neural network method. Dataset in this study is a collection of handwritten images from IAM handwriting database [8]. We will create three separated datasets with different image colors such as, grayscale, binary, and inverted binary. The handwritten features that will be a differentiator is the thickness, slope, and distance between the letters. The features, images, and methods used in this study will be a differentiator compared to previous research. The method used in this paper is convolutional neural network. Convolutional neural network (CNN) is included in the type of deep neural network, because of the network depth and applied mainly to imagery. CNN method proved successfully in surpassing other machine learning methods, such as SVM. Each neuron in CNN is presented in two-dimensional form.

Transfer learning is one of the methods that being used for convolutional neural network. Transfer learning on convolutional neural network mean the re-use of model that has been trained using a huge amount of data and already performed well in a certain task [9-11]. This pre-trained model is a result from a competition to create a new algorithm for objects detection and classification [12, 13]. Result of using pre-trained model depend on which architecture that will be used. There are several architectures like DenseNet, Resnet, and VGG [14]. Each of this architecture will have a different model and the result will have different accuracy value. Concept of transfer learning is to utilize knowledge acquired for one task to solve another related task. There are several benefits of using transfer learning, such as higher start, higher slope, and higher asymptote [11, 15].

2. RESEARCH METHOD

This research will mainly use CNN without any addition in features extraction method or addition in classification method. We would like to use CNN only with pre-trained model to show the performance in accuracy and processing time that we will get. We used VGG or Visual Geometry Group pre-trained model that have 19 deep layers, or VGG19 [16]. This architecture shows that a deep layer in CNN is an important factor to create a classification system that have high result accuracy. VGG19 has been trained using ImageNet datasets which have 1000 classes of images and can help overcome the problem of limited number of images in the dataset that being used in training process [10]. We used VGG19 as pre-trained model and replace the top layer with our own top layer with total number of our classes. We used 100 classes in this research with dataset of sentences image with dimension of 100x1200 pixel.

Figure 1 shows the architecture of CNN that being used in this research. We use VGG19 base layer and freeze it to keep its value in first training step. VGG19 used 5 blocks of convolution layer, with first two blocks have two convolution layer and last three blocks have four layers of convolution. We freeze these blocks to keep its data from previous training. After the base layer produce a feature map, we apply global average pooling to get a single feature map. Top layers of VGG19 replaced by modified top layer with 100 classes and two fully connected layer.

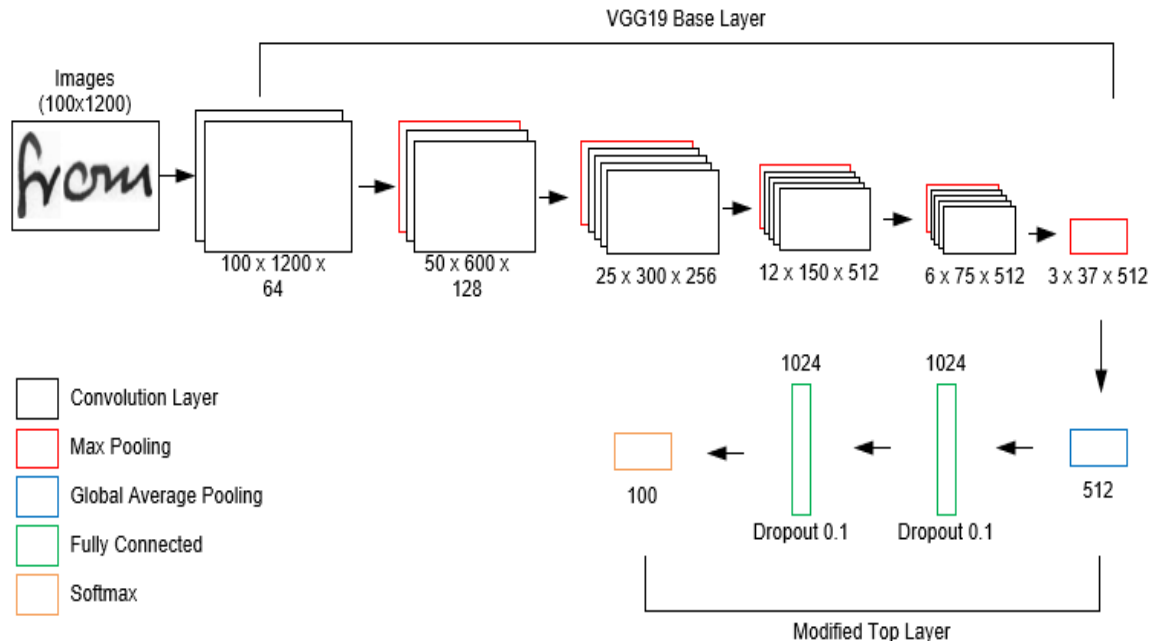


Figure 1. VGG19 with Modified top layer architecture

Top layer that used in this architecture is obtained by tuning its parameter. Parameter tuning is important because it can affect the results of classification process [17]. Tuning parameter in this research is done manually by trial and error process, until we get the best parameter to train our data. First, we replace flatten with global average pooling or GAP to generate one feature map for each corresponding output from block convolution layer [18]. The use of global average pooling is to reduce the overfitting that prone when we use fully connected. Furthermore, global average pooling takes the sums out value of spatial information, so it will speed up the training process [19].

We combine global average pooling with fully connected layer to obtained better result. After global average pooling, we use two fully connected layer as hidden layer with a dropout. Fully connected layer has a function to take the feature map and use them to classify the image into corresponding label, but fully connected layer is prone to overfitting especially when large neural nets is trained on relatively small datasets. Because of this, we use dropout to overcome this problem [20, 21]. Dropout have a function to randomly setting the outgoing edges of hidden layer units to 0 at each update of training phase. Because of this, we set a certain amount of dropout in the fully connected layer to avoid overfitting.

Optimizer have a role to update the weight parameters to minimize the loss function. In this research we use SGD or stochastic gradient descent as optimizer because SGD perform computations not on the whole dataset, but compute on small subset or even random selection of data examples, so it will make training process more efficient [22, 23]. Also, SGD can produce the same performance as regular gradient descent when the learning rate are quite low. The summarize of CNN architecture can be seen in Table 1. Table 1 shows the summarize of CNN architecture that used in this research. This architecture than trained using Google Colaboratory because its compatibility and the resource available on its cloud service is faster than physical resource and Colaboratory provide a free-of-charge resource that is enough to solve demanding real-world problems [24].

Table 1. Summarize of CNN architecture

	Layer Type	Output Shape		Layer Type	Output Shape
Input		100, 1200, 3			
Block 1	Convolution 1	100, 1200, 64	Block 4	Convolution 1	12, 150, 512
	Convolution 2	100, 1200, 64		Convolution 2	12, 150, 512
	Max Pooling	50, 600, 64		Convolution 3	12, 150, 512
Block 2	Convolution 1	50, 600, 128	Block 5	Convolution 4	12, 150, 512
	Convolution 2	50, 600, 128		Max Pooling	6, 75, 512
	Max Pooling	25, 300, 128		Convolution 1	6, 75, 512
Block 3	Convolution 1	25, 300, 256	Convolution 2	6, 75, 512	
	Convolution 2	25, 300, 256	Convolution 3	6, 75, 512	
	Convolution 3	25, 300, 256	Convolution 4	6, 75, 512	
	Convolution 4	25, 300, 256	Max Pooling	3, 37, 512	
	Max Pooling	12, 150, 256	Global Average Pooling	512	
			Dense + Dropout	1024	
			Dense + Dropout	1024	
			Softmax	100	

3. RESULTS AND ANALYSIS

In this section, we show results that we get based on method that we use. First, we give brief explanation about the image dataset used in this research. Then, training and testing results will be evaluated with comparison to overlook the effect of different preprocessing image method and use of pre-trained model.

3.1. Dataset

This research is done using dataset from IAM Handwriting Database [8]. The image used in this paper is an image of handwriting from various writer. The image from IAM Handwriting Database is already segmented in sentences form, and in grayscale. This research use 5467 images with 100 classes of different writers and 40-100 images in each class for training process, and five images for testing process. Writer then identified with id, like 000, 052, and 670, that we get from IAM handwriting dataset.

This image separated into three categories, grayscale images, binary images, and inverted binary images. Binary image is created using grayscale image with Otsu's thresholding [25], and inverted image created using binary image. Figure 2 show example of three categories image that being use in this research. These three categories have the same image with different color, and trained using the same CNN Architecture.

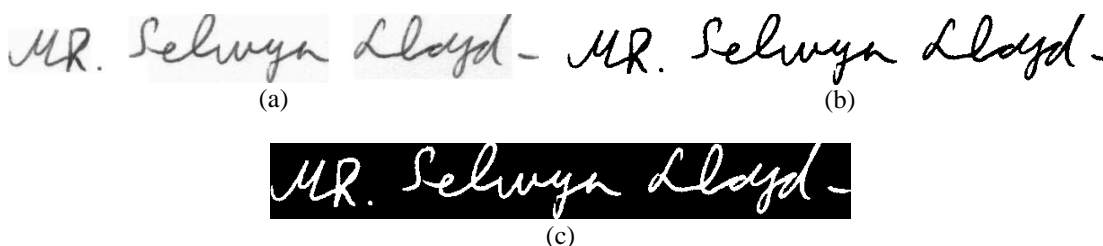


Figure 1. Image dataset used with format: (a) grayscale, (b) binary, and (c) inverse binary

3.2. Training result

Training process done by using Google Colaboratory [24] with number of epoch used in this training is 100 epoch. Training process also use 20% of training data as validation data to validate training after one epoch. It take almost 10 hours to complete the training process. Figure 3 show comparison between training accuracy from the uses of binary image, grayscale image, and inverted binary image as training dataset. Based on this figure, training accuracy from inverted binary image have steeper curve than the other two images dataset, and have a more stable learning curve. Inverted binary image and grayscale image have a training accuracy that reach 99%, grayscale image have training accuracy of 97% and binary image have training accuracy of 97%. Figure 4 show validation accuracy between the three datasets. Validation process is used to evaluate the model after one epoch is done and is used to give an unbiased estimate of the trained model [26]. Validation dataset uses in this research is 20% from training dataset, with total of 1056 images. Based on this figure, validation of grayscale image dataset is a more stable, with 90% of accuracy compared to the other two image datasets.

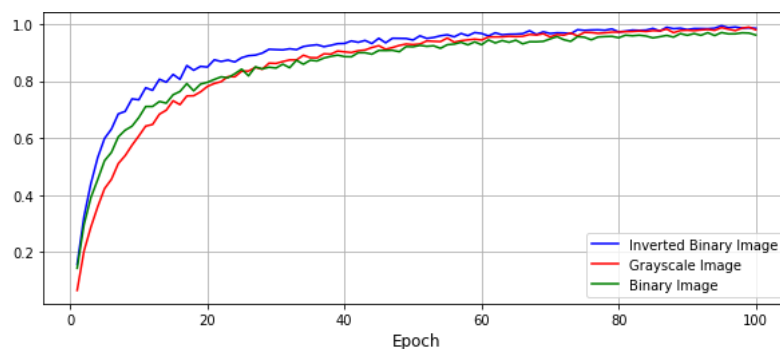


Figure 3. Comparison of training accuracy

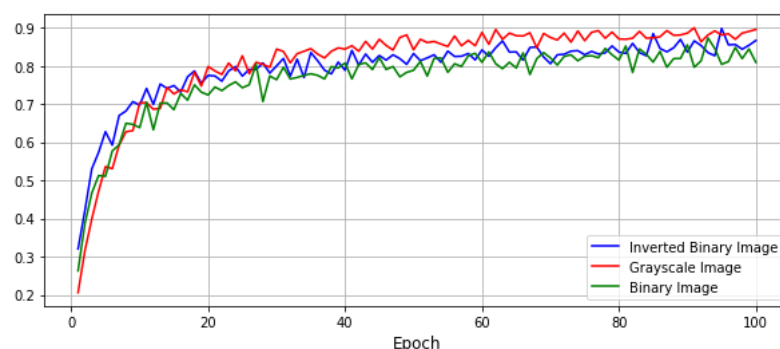


Figure 4. Comparison of validation accuracy

3.3. Testing result and analysis

This research tests the method used with the value of EER (equal error rate). We sought to obtain the lowest equal error rate between the three datasets, and then compared it with other method. To find equal error rate, first we find false acceptance rate (FAR) and false rejected rate (FRR). FAR occurs when the test identified handwriting owner to another writer, or false positive. FRR occurs when handwriting is rejected because of a failure to identify, or false negative [27]. To find EER, FAR and FRR is plotted together in a graph. EER is the point where FAR (1) and FRR (2) meet [10]. EER is often used to measure biometric systems. Genuine acceptance rate (GAR) (3) is defined as percentage of genuine writer accepted by the system.

$$FAR = \frac{\text{Total number of writer identified wrongly}}{\text{Total number of tests performed}} \quad (1)$$

$$FRR = \frac{\text{Total number of writer rejected}}{\text{Total number of tests performed}} \quad (2)$$

$$GAR = 1 - FRR \quad (3)$$

We test our trained model using 500 images (five images for each class). The results show that using grayscale images dataset we obtained that grayscale images dataset have the lowest FAR and FRR, and the highest GAR value. The result then followed by inverted binary images dataset and binary images dataset. Figure 5 shows that grayscale image yielded lowest false acceptance rate compare to the other two. From 500 images we tested, only 38 images were identified as wrong writer. We found that some of the 38 images have a similarity with the classes the were identified as, with example can be seen in Figure 6. Figure 6 illustrates the similarity between the two writer classes. The similarity can be seen in some way alphabets and/or words is written. Based on test result, image (a) identified as class 387 with 89% accuracy. It shows that similarity in the way alphabets is written, even only some of them, have a great impact in identification process. Figure 5 also shows that grayscale dataset has lowest false rejected rate. Based on the tests, only 39 images rejected. Some of it got identified as the right class but with low accuracy and some got identified as

wrong class but with low accuracy. Figure 5 also shows genuine acceptance rate of all three categories dataset. The highest value of GAR was grayscale images dataset with value of 92.3%.

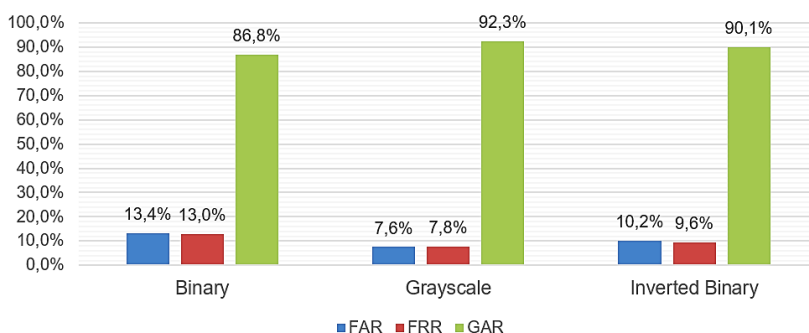


Figure 5. Chart of FAR, FRR, and GAR value

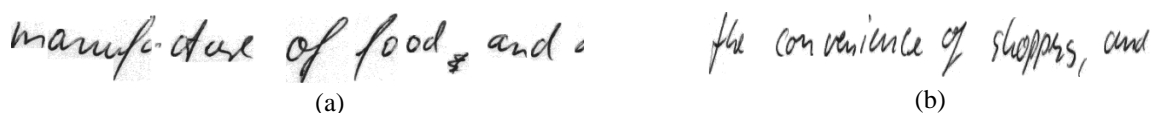


Figure 6. Comparison handwriting of different writer: (a) writer with id 385 and, (b) writer with id 387

Figure 7 shows the ROC or receiver operation characteristics of grayscale images dataset. This graph shows where FAR and FRR value meet. Based on this figure, equal error rate can be found on threshold 57. The complete result can be seen in Table 2. Based on this result, grayscale image dataset has the best result compare with the other datasets. Grayscale dataset get equal error rate (EER) of 7.7% that crossed at threshold 51. It shows that pre-trained model VGG19 have a good compatibility with grayscale images. The poorest performance is with binary dataset with highest EER value of 13.2% that crossed at threshold 52. Inverted binary has a good performance with EER value of 9.9% that crossed at threshold 60.

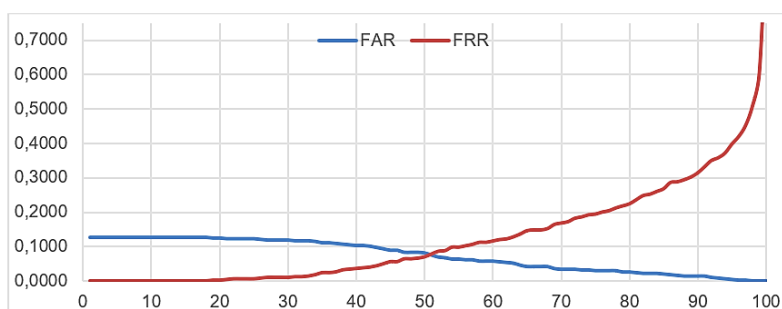


Figure 7. Receiver operation characteristics graph of grayscale dataset

Table 2. Complete result of the three datasets

Dataset	FAR (%)	FRR (%)	Threshold	GAR (%)
Binary	13.4%	13.0%	52	86.8%
Grayscale	7.6%	7.8%	51	92.3%
Inverted Binary	10.2%	9.6%	52	90.1%

The worst result we get is when using binary image as dataset. This result may be caused by the noise and data loss in the binary images that make it difficult for the CNN to learn. When binary images generated from grayscale images, its loss several features from original images. Inverted binary dataset have a high result of GAR caused by its black background recognized as a features by CNN. If we compare this

method with other research [6, 7], we can say that this method have several advantage. First, the method that we used doesn't need much of preprocessing method and another segmentation method. This make time that needed for preparing the dataset faster than the other method. Second, we use handwriting in form of sentences, so it has more characteristics and features that can be extracted and learn by CNN. But we can't compare time needed for training because the other research doesn't specify it.

4. CONCLUSION

We used transfer learning to identified writer based on their handwriting using pre-trained CNN. We trained model using IAM handwriting dataset comprising of 100 classes of writer. Based on the result of training and testing, pre-trained model VGG19 has the best performance when using grayscale images than using binary or inverted binary images. Our finding also indicates that handwriting image in sentences form can be used directly without another segmentation or feature extraction method needed. The disadvantage of this research is the time it took to finish training process. Training process took almost 10 hours to complete 100 epochs.

REFERENCES

- [1] K. Chaudhari and A. Thakkar, "Survey on handwriting-based personality trait identification," *Expert Systems with Applications*, vol. 124, pp. 282–308, Jun 2019. doi: 10.1016/j.eswa.2019.01.028.
- [2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications ACM*, vol. 60, no. 6, pp. 84–90, May 2017. doi: 10.1145/3065386.
- [3] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep Convolutional Neural Network for Inverse Problems in Imaging," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4509–4522, Sep 2017. doi: 10.1109/TIP.2017.2713099.
- [4] J. Schmidhuber, "Deep Learning in Neural Networks: An Overview," *Neural Networks*, vol. 61, pp. 85–117, Jan 2015. doi: 10.1016/j.neunet.2014.09.003.
- [5] C. K. Dewa, A. L. Fadhilah, and A. Afiahayati, "Convolutional Neural Networks for Handwritten Javanese Character Recognition," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 12, no. 1, pp. 83–94, Jan 2018. doi: 10.22146/ijccs.31144.
- [6] B. V. Dhandra and M. B. Vijayalaxmi, "A Novel Approach to Text Dependent Writer Identification of Kannada Handwriting," *Procedia Computer Science*, vol. 49, pp. 33–41, 2015. doi: 10.1016/j.procs.2015.04.224.
- [7] L. Xing and Y. Qiao, "DeepWriter: A Multi-Stream Deep CNN for Text-independent Writer Identification," arXiv:1606.06472, Jun 2016.
- [8] U. V. Marti and H. Bunke, "The IAM-database: an English sentence database for offline handwriting recognition," *International Journal on Document Analysis and Recognition*, vol. 5, no. 1, pp. 39–46, Nov 2002. doi: 10.1007/s100320200071.
- [9] A. Rehman, S. Naz, M. I. Razzak, and I. A. Hameed, "Automatic Visual Features for Writer Identification: A Deep Learning Approach," *IEEE Access*, vol. 7, pp. 17149–17157, 2019. doi: 10.1109/ACCESS.2018.2890810.
- [10] P. Hridayami, "Fish Species Recognition Using VGG16 Deep Convolutional Neural Network," *Journal of Computing Science and Engineering*, vol. 13, no. 3, pp. 124–130, 2019.
- [11] D. Han, Q. Liu, and W. Fan, "A new image classification method using CNN transfer learning and web data augmentation," *Expert Systems with Applications*, vol. 95, pp. 43–56, Apr 2018. doi: 10.1016/j.eswa.2017.11.028.
- [12] O. Russakovsky et al., "ImageNet Large Scale Visual Recognition Challenge," arXiv:1409.0575, Sep. 2014.
- [13] Y. You, Z. Zhang, C.-J. Hsieh, J. Demmel, and K. Keutzer, "ImageNet Training in Minutes," arXiv:1709.05011, Jan 2018.
- [14] S. H. S. Basha, S. R. Dubey, V. Pulabaigari, and S. Mukherjee, "Impact of Fully Connected Layers on Performance of Convolutional Neural Networks for Image Classification," *Neurocomputing*, vol. 378, p. 112–119, Oct 2019. doi: 10.1016/j.neucom.2019.10.008.
- [15] S. Giri and B. Joshi, "Transfer Learning Based Image Visualization Using CNN," *International Journal of Artificial Intelligence and Application*, vol. 10, no. 4, pp. 47–55, Jul 2019. doi: 10.5121/ijaia.2019.10404.
- [16] H. Ye, H. Han, L. Zhu, and Q. Duan, "Vegetable Pest Image Recognition Method Based on Improved VGG Convolution Neural Network," *Journal of Physics Conference Series*, vol. 1237, Jun 2019. doi: 10.1088/1742-6596/1237/3/032018.
- [17] T. Sinha, B. Verma, and A. Haidar, "Optimization of convolutional neural network parameters for image classification," *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, Honolulu, HI, pp. 1–7, 2017. doi: 10.1109/SSCI.2017.8285338.
- [18] M. Lin, Q. Chen, and S. Yan, "Network in Network," arXiv:1312.4400, Mar 2014.
- [19] L. Alzubaidi, H. A. Alwazay, and Z. M. Arkah, "Convolutional Neural Network with Global Average Pooling for Image Classification," *International Conference on Electrical, Communication, Electronics, Instrumentation and Computing (ICECEIC)*, 2020.
- [20] Jing Yang and Guanci Yang, "Modified Convolutional Neural Network Based on Dropout and the Stochastic Gradient Descent Optimizer," *Algorithms*, vol. 11, no. 3, Mar 2018. doi: 10.3390/a11030028.

- [21] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *The Journal of Machine Learning Research (JMLR)*, vol. 15, no. 56, pp. 1929–1958, 2014.
- [22] N. S. Keskar and R. Socher, "Improving Generalization Performance by Switching from Adam to SGD," arXiv:1712.07628, Dec 2017.
- [23] A. C. Wilson, R. Roelofs, M. Stern, N. Srebro, and B. Recht, "The Marginal Value of Adaptive Gradient Methods in Machine Learning," arXiv:1705.08292, May 2018.
- [24] T. Carneiro, R. V. Medeiros Da Nobrega, T. Nepomuceno, G. B. Bian, V. H. C. De Albuquerque, and P. P. R. Filho, "Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications," *IEEE Access*, vol. 6, pp. 61677–61685, 2018. doi: 10.1109/ACCESS.2018.2874767.
- [25] C. Lv, T. Zhang, C. Liu, and School of Automation, Beijing Institute of Technology, "An Improved Otsu's Thresholding Algorithm on Gesture Segmentation," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 21, no. 2, pp. 247–250, Mar 2017. doi: 10.20965/jaciii.2017.p0247.
- [26] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights Imaging*, vol. 9, no. 4, pp. 611–629, Aug 2018. doi: 10.1007/s13244-018-0639-9.
- [27] NKA Wirdiani, T. Lattifia, I. K. Supadma, B. J. K. Mahar, D. A. N. Taradhita, "Real-Time Face Recognition with Eigenface Method," *International Journal of Image, Graphics and Signal Processing(IJIGSP)*, vol. 11, no. 11, pp. 1–9, Nov 2019. doi: 10.5815/ijigsp.2019.11.01.

BIOGRAPHIES OF AUTHORS



A. A. Kompiang Oka Sudana, currently a Lecturer in the Information Technology Study Program, Udayana University, Bali. Areas of interest: Image Processing, Information Systems, IT Implementation in Culture and Religion. S1 in Informatics Department of ITS Surabaya, Masters in Electrical Engineering UGM Yogyakarta, currently in the process of studying S3 at Udayana University.



I Wayan Gunaya was born on September 3, 1998 in Bangli, Bali, Indonesia. He is a student in Department of Information Technology, Udayana University, Bali, Indonesia. His research interests are in image processing, computer vision, data science, and machine learning.



I Ketut Gede Darma Putra currently 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.