Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)
Vol. 7, No. 2, September 2024, pp. 266 – 274
p-ISSN: 2614-3372 | e-ISSN: 2614-6150

Data Augmentation Using Test-Time Augmentation on Convolutional Neural Network-Based Brand Logo Trademark Detection

[1]Suyahman, [2]Sunardi, [3]Murinto, [4]Arfiani Nur Khusna
[1]Master Program of Informatics, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[2]Department of Electrical Engineering, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[3,4]Department of Informatics, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
Email: [1]2307048007@webmail.uad.ac.id, [2]sunardi@mti.uad.ac.id,
[3]murintokusno@tif.uad.ac.id, [4]arfiani.khusna@tif.uad.ac.id

Article Info

Article history:
Received Nov 22th, 2023
Revised Mar 23th, 2024
Accepted Apr 25th, 2024

Keyword:
Convolutional Neural Network
Data Augmentation
Test Time Augmentation
Trademark Detection
TTA

ABSTRACT

The detection and acknowledgment of logos holds significant importance in the corporate sphere, facilitating the detection of unauthorized logo usage and ensuring trademark uniqueness within specific industry sectors. Presently, convolutional neural networks powered by deep learning are widely utilized for image recognition. However, their effectiveness is dependent on a substantial volume of training images which may not always be readily available. This study suggests employing Test Time Augmentation to address dataset constraints by expanding the original dataset, thereby enhancing classification accuracy and preventing overfitting. Test-Time Augmentation is a method used to improve the accuracy of convolutional neural networks by creating numerous augmented variations of the test images and then merging their predictions. The research findings indicate that the application of TTA has the highest performance on the VGG16 model with 98% precision, 99% recall, and 98% F1-score, and 98.87% accuracy.

Corresponding Author:
Suyahman,
Master Program of Informatics,
Universitas Ahmad Dahlan,
Yogyakarta, Indonesia, 55166
Email: 2307048007@webmail.uad.ac.id

DOI: http://dx.doi.org/10.24014/ijaidm.v7i2.28804

1. INTRODUCTION

Detecting and identifying logos is a crucial responsibility in the industrial sector as it involves identifying unauthorized usage of company logos, and ensuring that there are no similar existing logos within the same industry for trademark registration purposes [1].

Deep learning with convolutional neural networks is widely utilized for image detection [2], but it often demands an extensive collection of training image samples [3]. Access to such datasets may not always be feasible. To address this issue, data augmentation can be implemented as a regularization method to expand the original dataset, thereby improving classification accuracy and preventing overfitting [4].

Many researchers have used deep learning to solve a variety of picture identification problems [5]-[10]. Murinto et al. researched the categorization of different types of coffee beans and employed Convolutional Neural Network as well as Transfer Learning with VGG16 and MobileNetV2 Models for their study. The CNN-transfer learning model, MobileNetV2, demonstrated an impressive accuracy rate of 96% in classifying images of coffee beans [11].

Guotai Wang et al. investigated the application of data augmentation in medical image segmentation. They proposed a theoretical framework for test-time augmentation using Monte Carlo simulation to predict outcomes in deep learning for image recognition. The research focused on uncertainty estimation in segmenting fetal brain and brain tumors from 2D and 3D Magnetic Resonance Images. The findings indicated that their test-time augmentation approach outperformed a single-prediction baseline and dropout-based multiple predictions. For 3D brain tumor segmentation by 3D U-Net and V-Net, the Dice results (%) were as follows: intact tumor at 88.52±5.95, tumor core at 79.61±17.02, and enhancing core at 75.70±20.41. The study demonstrated improved uncertainty estimation while reducing overconfident incorrect predictions [12].

Nikita Moshkov et al. conducted a study on the advancement of deep learning in analyzing microscopy images of cells. They incorporated test-time augmentation into segmentation methods like U-Net for semantic segmentation and Mask R-CNN for instance segmentation, leading to a noteworthy improvement in prediction accuracy, even with simple augmentations. Their research utilized images from the Data Science Bowl 2018 nuclei segmentation competition and showed that TTA augmentation produced the highest mAP score of 0.644, highlighting its effectiveness in enhancing precision in cell analysis [13].

While data augmentation has been widely used during the training phase, its application and effect at test-time requires more exploration, especially across different CNN architectures. This paper presents the use of TTA [14] for enhancing the precision of CNN. The effectiveness of the suggested model will be evaluated by comparing it with earlier studies on CNN architectures like VGG16 [15], VGG19 [16], and ResNet50 [17]. The main objective of this research is to fill this gap by evaluating the impact of applying test-time augmentation on the performance of VGG16, VGG19, and ResNet50 in the context of image processing tasks.

This research adds to our knowledge about brand logo classification using TTA in combination with CNN. The primary aim of this study is to enhance the accuracy of the model through data preparation, model testing and validation, as well as performance assessment. This research extends the understanding of the application of data augmentation, not only during training but also in the test-time phase, offering a new perspective on CNN model optimization. The research also compares and contrasts the effects of various test-time augmentation techniques on different CNN architectures, providing practical recommendations for practitioners and researchers in choosing the most effective augmentation strategy.

To support this research, we have reviewed various related literature. For example, Takahashi highlighted the importance of data augmentation during training to improve model generalization [18]. Meanwhile, Tursun began to explore the use of data augmentation during the test phase as a strategy to improve model performance on datasets not seen during training [19]. However, both studies have not specifically targeted the effect of test-time augmentation on different CNN architectures, signaling the significance of our study in the existing literature.

2. RESEARCH METHOD

The study gathered 490 images from the Brand Logo Trademark dataset [20], encompassing 10 separate classes. These images had a resolution of 224 x 224. To facilitate testing, the dataset will be divided into training and testing data at a ratio of 70:30, resulting in 343 training samples and 147 testing samples. The details of the dataset are presented in Table 1 and Figure 1 displaying image samples.

Table 1. Class and number brand logo trademark images in the dataset

Classes	Images
Ajaib	21
Akulaku	50
Alfamart	50
Alodokter	50
Apple	20
Adidas	50
Aqua	50
BCA	50
Miniso	49
Uniqlo	50
Total	490

Figure 1. Sample images on the dataset

2.1. Convolutional Neural Network

A convolutional neural network is a form of neural network utilized for the categorization of images [21]. It has multiple layers that perform specific operations such as convolution, pooling, and loss estimation. A typical CNN design begins with an initial layer that has a number of neurons equal to the pixels in the input image. The following layers are composed of convolutional layers, which act as feature extractors by convolving the input data with different filters termed kernels. The size of the kernel affects how sensitive the layer is to input features, as each neuron responds only to a specific area of the previous layer known as its receptive field [22]. The result from these layers creates an activation map that illustrates how applying a specific filter impacts the input data. An activation layer is usually placed after the convolutional layers, introducing non-linearity to activation maps, sometimes using a pooling layer to decrease dimensionality, and ultimately utilizing fully connected layers to extract high-level features [23]. For further insight into this process, refer to Figure 2 depicting an overview of inner workings within a CNN's layer.

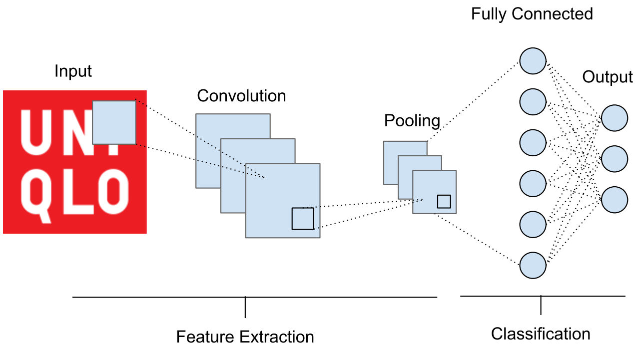

Figure 2. CNN architecture for image classification

The CNN architectures that will be used to test the impact of using TTA are VGG16, VGG19, and ResNet50. The VGG16 and VGG19 models are complex convolutional neural networks created by the Visual Graphics Group at Oxford University. [24]. The VGG16 model comprises 16 layers, whereas the VGG19 model consists of 19 layers. Both designs include a sequence of convolutional layers, followed by max-pooling and fully connected layers. These models are efficient for tasks involving the classification of images, especially with small datasets [25]. On the other hand, ResNet50 is a deeper architecture from the Residual Network family that addresses vanishing gradients through shortcut connections and consists of 50 layers incorporating residual blocks for training significantly deeper networks compared to traditional architectures [26]. These three architectures can be observed in Figure 3.

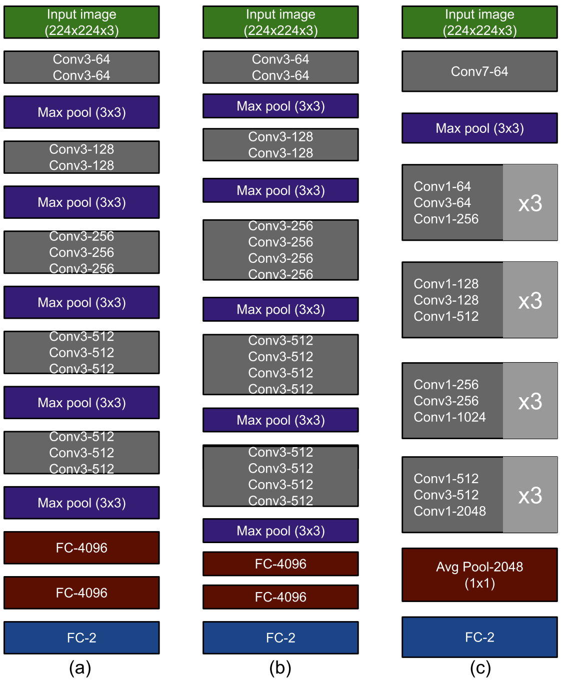

Figure 3. CNN architectures of (a) VGG-16, (b) VGG-19, and (c) ResNet-50

2.2. Test-Time Augmentation

TTA refers to a procedure that includes implementing data augmentation on input images during inference and generating multiple duplicates of the image [27]. The model then makes predictions on each image and returns an average of the results. Essentially, this provides the model with an improved opportunity to accurately classify a given image by presenting it with a slightly altered version of the original input [28]. Table 2 outlines the specific augmentation technique utilized in this process.

Table 2. Data augmentation techniques

Technique	Range
Resize	64, 352
Rotate	-180,162
Shear	-0.3,0.3
Zoom	0,1
Horizontal Flip	True
Fill	Reflect
Shift	0,1
Contour	Laplacian
Negative	-
Emboss	-
Sharpen	-
Blur	-

TTA has been consistently observed that this method results in better segmentation accuracy than predicting based solely on the original images [13]. Figure 4 further illustrates the application of TTA.

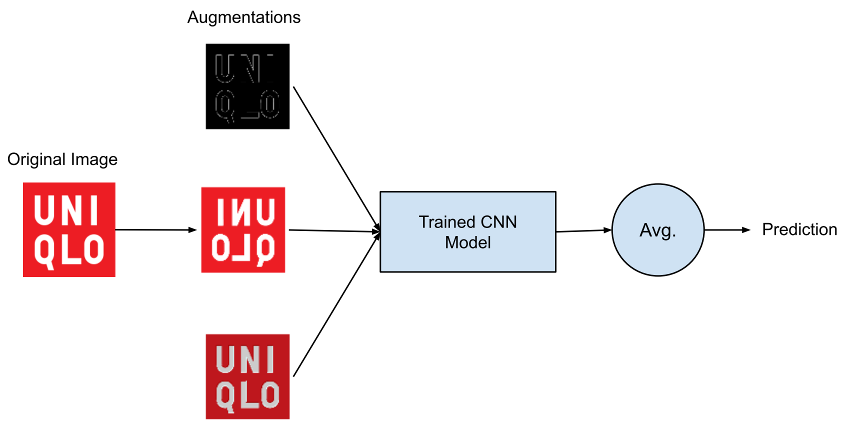

Figure 4. Test Time Augmentation

2.3. Method

The primary aim of this research is to develop a logo detection model utilizing the CNN method and the framework depicted in Figure 3. The different stages involved in building a neural network classification model are presented in Figure 5, starting from the input dataset and progressing through preprocessing, splitting the dataset into training, validation, and test sets [29]. The training set will be used for model training using the CNN model, while data augmentation with TTA will be performed for testing purposes. Subsequently, detailed implementation procedures of the model will be elaborated upon, Figure 1 displaying of research methodology.

2.4. Experimental Setup

The CNN Architecture incorporated support for GPU. The experiments were conducted on a Jupyter notebook available through Google Colaboratory. The CPU of the Jupyter notebook system runs at 2.20 GHz with an Intel(R) Xeon(R) processor and has 13 GB of RAM, while the GPU is an NVIDIA Tesla T4 with a memory capacity of 15 GB. The implementation utilized the Keras 2.12.0 framework, an open-source Python library for deep neural networks.

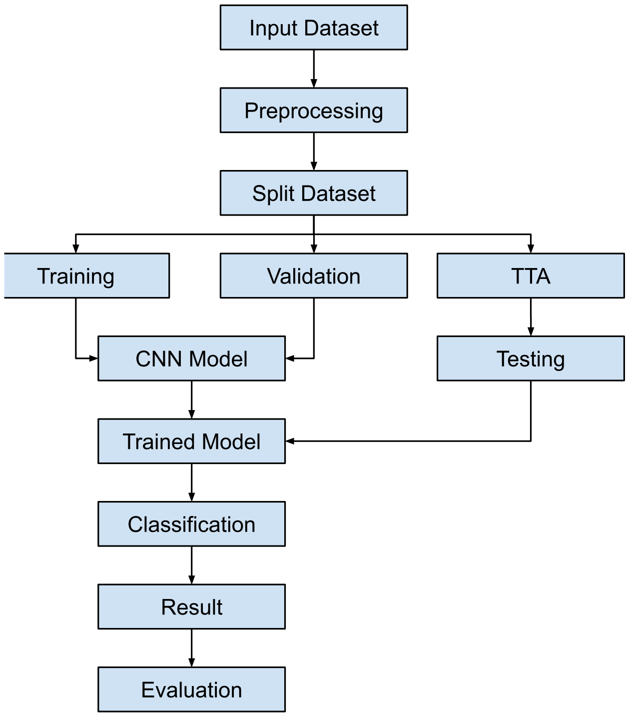

Figure 5. Flowchart for logo classification

2.5. Training

The study utilized deep learning models to identify brand logo trademarks, incorporating VGG16, VGG19, and ResNet50 using a transfer learning-based CNN. The research employed a transfer learning algorithm for the classification of brand logo trademark datasets. The specific configuration of the CNN model during the training process was detailed in Table 3.

Table 3. Parameter configuration.

Configuration	Value
Batch size	32
Epoch	100
Optimizer	Adam
Activation	Softmax
Learning rate	0.0003

2.6. Evaluation

The dataset for brand logo trademarks consists of 10 categories, which necessitated the use of multiclass classification. The values in Table 4 indices are derived from confusion matrix data and represent TP, TN, FN, and FP for each category.

Table 4. Confusion matrix.

		Predicted Label	
		Negative	Positive
Actual Label	Negative	TN	FP
	Positive	FN	TP

Furthermore, the performance of CNN was evaluated using metrics such as Recall, F1-Score, Precision, and Accuracy [30]. Accuracy is a fundamental assessment in classification that denotes the percentage of correctly classified samples. Precision represents the likelihood that a sample is actually positive among all predicted positive samples. Meanwhile, recall pertains to the likelihood of correctly identifying an actual positive sample as positive in comparison to all true positive instances [31]. Additionally, the F1 score is a measure that combines precision and recall into a single value using the harmonic mean.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{1}$$

$$Accuracy = \frac{TP}{TP+FP} \tag{2}$$

$$Accuracy = \frac{TP}{TP+FN} \tag{3}$$

3. RESULTS AND ANALYSIS

Comparing the impact of using TTA on CNN models based on previous literature studies is the main objective of this paper. The performance metrics of the TTA implementation on CNN show the best results on VGG16. Table 5 presents the precision, recall, and F1 Score for each class on the test data. The model training process took approximately three hours, indicating that the transfer learning was sufficiently beneficial in terms of time savings and excellent accuracy.

Table 5. Class and number brand logo trademark images in the dataset.

Classes	Precision	Recall	F1 Score
Akulaku	0.95	1.0	0.97
Ajaib	1.0	1.0	1.0
Alodokter	1.0	0.95	1.0
Alfamart	1.0	1.0	0.97
BCA	1.0	1.0	1.0
Apple	1.0	1.0	1.0
Adidas	1.0	1.0	1.0
Aqua	1.0	0.98	0.99
Miniso	0.83	1.0	0.91
Uniqlo	1.0	1.0	1.0
Macro Average Accuracy	0.98	0.99	0.98

According to Table 4, Miniso logo has the lowest value in precision performance category, with a precision value of 83%. Furthermore, the Akulaku logo has the second lowest value after Aqua, with a precision value of 95%. In the recall performance Alodokter logo has lowest value, with a recall value of 95% recall and in F1 Score, Miniso logo has 91% F1 Score, this is the lowest value in overall performance category.

Figure 6 depicts the confusion matrix of the model training outcomes, with the range 0-50 indicating the quantity of images in the testing phase. The brighter the color exhibited, the higher the image classification result. According to the confusion matrix, some logos are still not perfect in classifying with test data, such as Alfamart logo, which can detect 38 images correctly, but two logos are detected as Miniso. The Aqua logo can detect 42 logos correctly but there is 1 logo that is incorrectly detected as the Akulaku logo.

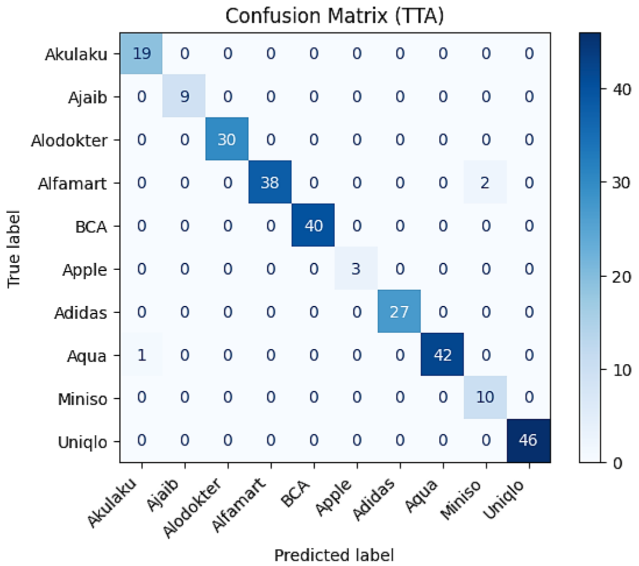

Figure 6. Confusion matrix result

Training and validation learning curves for accuracy and loss are shown in Figure 7. The training and validation losses are minimized to a certain point of equilibrium with a slight difference between the two ultimate loss numbers to figure out a satisfactory fit curve. The model loss on the training dataset is usually often less than the loss on the test dataset. This shows that there is a variance between the training learning curve and the validation losses. Which is referred to as the generalization gap. While model with excessive capacity may learn and fit the training dataset too well. In such instances, the model does not cause a generalization gap [32], [33].

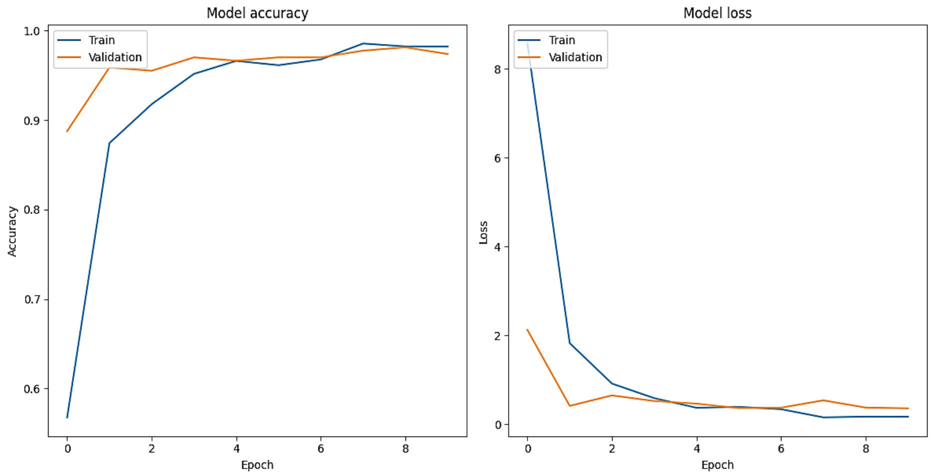

Figure 7. Accuracy and loss curve result

The accuracy of using TTA in CNN shows higher results than the model without TTA in Table 6. Regular CNN experienced an increase in accuracy from 73,78% in the baseline to 75,28% with the use of TTA, VGG16 from 97.37% to 98.87%, VGG19 from 96.25% to 96.62%, and ResNet50 from 80.52% to 82.02%. Although the improvement is not significant, it shows that TTA makes a positive contribution to the model performance.

Table 6. Comparison results of using TTA in CNN architecture

Configuration	Baseline	Using TTA
CNN	0.7378	0.7528
VGG16	0.9737	0.9887
VGG19	0.9625	0.9662
ResNet50	0.8052	0.8202

Regular CNN has the lowest accuracy, while the highest accuracy is achieved by VGG16 as shown in Figure 8.

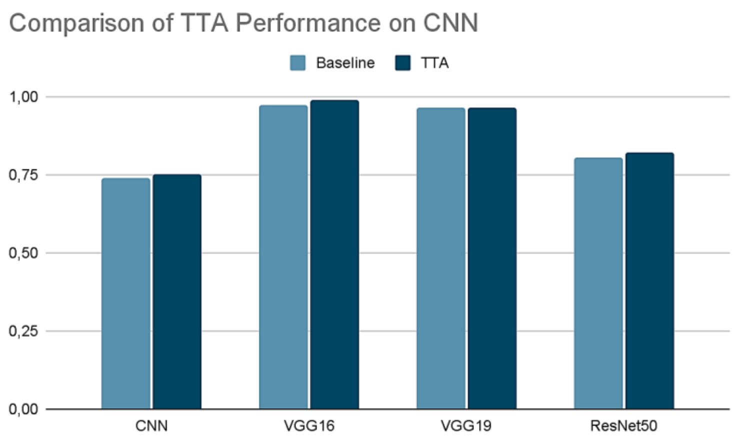

Figure 8. Comparison results of using TTA in CNN architecture

4. CONCLUSION

This study introduces a CNN design for identifying brand logos belonging to 10 different classes. The use of TTA is proven to be able to answer problems by improving the performance of CNN models such as VGG16, VGG19, and ResNet50. Not only does it increase prediction accuracy, TTA also shows a positive impact on overfitting and model robustness to variations in input data. The VGG16 and TTA models demonstrated superior performance in accuracy, finishing the training process within two and a half hours, thus validating the effectiveness of the proposed model compared to other CNN models utilized for logo detection. The application of TTA has the highest performance on the VGG16 model with 98% precision, 99% recall, and 98% F1-score, and 98.87% accuracy. The research results provide a better understanding of the relevance and impact of TTA in the use of CNN architectures. It is hoped that further research can be carried out by building future logo brand datasets by increasing the diversity of logo brands as well as the number of classes. This contributes to the development of a CNN architecture capable of making more accurate predictions in difficult situations.

REFERENCES

[1] Trappey, C. V., Trappey, A. J. C., & Lin, S. C.-C. (2020). Intelligent trademark similarity analysis of image, spelling, and phonetic features using machine learning methodologies. Advanced Engineering Informatics, 45, 101120.
[2] Lou, G., & Shi, H. (2020). Face image recognition based on convolutional neural network. China Communications, 17(2), 117–124.
[3] Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J. N., Wu, Z., & Ding, X. (2020). Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. Medical Image Analysis, 63, 101693.
[4] Moreno-Barea, F. J., Jerez, J. M., & Franco, L. (2020). Improving classification accuracy using data augmentation on small data sets. Expert Systems with Applications, 161, 113696.
[5] Rahman, A., Winarko, E., & Mustofa, K. (2022). Product image retrieval using category-aware siamese convolutional neural network feature. Journal of King Saud University - Computer and Information Sciences, 34(6), 2680–2687.
[6] Nurkhasanah, N., & Murinto, M. (2022). Klasifikasi penyakit kulit wajah menggunakan metode convolutional neural network. Sainteks, 18(2), 183.
[7] Nugroho, A. S. (2020). Sistem pengenalan botol plastik berdasarkan label Merek Menggunakan faster-rcnn. Techno (Jurnal Fakultas Teknik, Universitas Muhammadiyah Purwokerto), 21(2), 111.
[8] Furqon, D. T., & Murinto, M. (2022). masked face identification using the convolutional neural network method. Jurnal Informatika, 16(2), 82.
[9] Sunardi, Fadlil, A., & Prayogi, D. (2023). Room security system using machine learning with face recognition verification. Revue d'Intelligence Artificielle, 37(5), 1187–1196.
[10] Sunardi, Yudhana, A., & Fahmi, M. (2023). SVM-CNN hybrid classification for waste image using morphology and HSV color model image processing. Traitement Du Signal, 40(4), 1763–1769.
[11] Murinto, M., Rosyda, M., & Melany, M. (2023). Klasifikasi Jenis Biji Kopi Menggunkan Convolutional Neural Network dan Transfer Learning pada Model VGG16 dan MobileNetV2. JRST (Jurnal Riset Sains Dan Teknologi), 7(2), 183.
[12] Wang, G., Li, W., Aertsen, M., Deprest, J., Ourselin, S., & Vercauteren, T. (2019). Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks. Neurocomputing, 338, 34–45.
[13] Moshkov, N., Mathe, B., Kertesz-Farkas, A., Hollandi, R., & Horvath, P. (2019). Test-time augmentation for deep learning-based cell segmentation on microscopy images. Cold Spring Harbor Laboratory.
[14] Son, J., & Kang, S. (2023). Efficient improvement of classification accuracy via selective test-time augmentation. Information Sciences, 642, 119148.
[15] Sharma, S., Guleria, K., Tiwari, S., & Kumar, S. (2022). A deep learning based convolutional neural network model with VGG16 feature extractor for the detection of Alzheimer Disease using MRI scans. Measurement: Sensors, 24, 100506.
[16] Dey, N., Zhang, Y.-D., Rajinikanth, V., Pugalenthi, R., & Raja, N. S. M. (2021). Customized VGG19 architecture for pneumonia detection in chest x-rays. Pattern Recognition Letters, 143, 67–74.
[17] Rajpal, S., Lakhyani, N., Singh, A. K., Kohli, R., & Kumar, N. (2021). Using handpicked features in conjunction with ResNet-50 for improved detection of COVID-19 from chest X-ray images. Chaos, Solitons & Fractals, 145, 110749.
[18] Takahashi, R., Matsubara, T., & Uehara, K. (2020). Data augmentation using random image cropping and patching for Deep Cnns. IEEE Transactions on Circuits and Systems for Video Technology, 30(9), 2917–2931.
[19] Tursun, O., Denman, S., Sridharan, S., & Fookes, C. (2022). Learning test-time augmentation for content-based image retrieval. Computer Vision and Image Understanding, 222, 103494.
[20] Direktorat Jenderal Kekayaan Intelektual. (2024). Pangkalan Data Kekayaan Intelektual. https://pdki-indonesia.dgip.go.id/
[21] Nawrocka, A., Nawrocki, M., & Kot, A. (2023, June 12). Research study of image classification algorithms based on Convolutional Neural Networks. 2023 24th International Carpathian Control Conference (ICCC).
[22] Gabbasov, R., & Paringer, R. (2020, May 26). Influence of the receptive field size on accuracy and performance of a convolutional neural network. 2020 International Conference on Information Technology and Nanotechnology (ITNT).
[23] Khan, A., Sohail, A., Zahoora, U., & Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. Artificial Intelligence Review, 53(8), 5455–5516.
[24] Sukegawa, S., Yoshii, K., Hara, T., Yamashita, K., Nakano, K., Yamamoto, N., Nagatsuka, H., & Furuki, Y. (2020). Deep neural networks for dental implant system classification. Biomolecules, 10(7), 984.

[25] Yang, Y., Zhang, L., Du, M., Bo, J., Liu, H., Ren, L., Li, X., & Deen, M. J. (2021). A comparative analysis of eleven neural networks architectures for small datasets of lung images of COVID-19 patients toward improved clinical decisions. Computers in Biology and Medicine, 139, 104887.

[26] Singhal, A., & Bedi, P. (2021). Universal quantitative steganalysis using deep residual networks. In Advances in Intelligent Systems and Computing (pp. 465–475). Springer Singapore.

[27] Cohen, S., Goldshlager, N., Rokach, L., & Shapira, B. (2023). Boosting anomaly detection using unsupervised diverse test-time augmentation. Information Sciences, 626, 821–836.

[28] Oza, P., Sharma, P., Patel, S., Adedoyin, F., & Bruno, A. (2022). Image augmentation techniques for mammogram analysis. Journal of Imaging, 8(5), 141.

[29] Oh, Y., Park, S., & Ye, J. C. (2020). Deep Learning COVID-19 features on CXR using limited training data sets. IEEE Transactions on Medical Imaging, 39(8), 2688–2700.

[30] Alakus, T. B., & Turkoglu, I. (2020). Comparison of deep learning approaches to predict COVID-19 infection. Chaos, Solitons & Fractals, 140, 110120.

[31] Wang, J., Yang, L., Huo, Z., He, W., & Luo, J. (2020). Multi-Label classification of fundus images with efficientnet. IEEE Access, 8, 212499–212508.

[32] J. Brownlee, “How to use learning curves to diagnose machine learning model performance,” MachineLearningMastery.com. Aug-2019.

[33] J. Brownlee, “How to avoid overfitting in Deep Learning Neural Networks,” MachineLearningMastery.com. Aug-2019.

BIBLIOGRAPHY OF AUTHORS

Suyahman currently a student in Master Program of Informatics, Universitas Ahmad Dahlan, Yogyakarta, Indonesia. He has interested in Deep Learning, Pattern Recognition, Image Processing, Computer Vision, Intelligent Systems. He can be contacted at the email address: 2307048007@webmail.uad.ac.id

Sunardi received his Ph.D. from Universiti Teknologi Malaysia, Johor Bahru, Malaysia, in 2011 in electrical engineering. Currently, he is a lecturer at the Department of Electrical Engineering, Ahmad Dahlan University, Yogyakarta. His research interests include data, information break, and communication. He can be contacted at the email address: sunardi@mti.uad.ac.id

Murinto received his Doctoral Degree in computer science from Universitas Gadjah Mada, Indonesia. He currently is a lecturer in Department of Informatics, Universitas Ahmad Dahlan, Yogyakarta, Indonesia. He has a research interest in Computer Graphics, Image Processing, Computer Vision, Intelligent Systems. He can be contacted at the email address: murintokusno@tif.uad.ac.id

Arfiani Nur Khusna received his Master in informatic from STMIK AMIKOM Yogyakarta Indonesia. He currently is a lecturer in Department of Informatics, Universitas Ahmad Dahlan, Yogyakarta, Indonesia. He has a research interest in information system and information retrieval. He can be contacted at the email address: arfiani.khusna@tif.uad.ac.id