Scientific Journal of Informatics
Vol. 11, No. 4, Nov 2024
p-ISSN 2407-7658 https://journal.unnes.ac.id/journals/sji/index e-ISSN 2460-0040

Comparative Analysis of CNN Architectures in Siamese Networks with Test-Time Augmentation for Trademark Image Similarity Detection

Suyahman[1], Sunardi[2]*, Murinto[3]
[1]Master Program of Informatics, Universitas Ahmad Dahlan, Indonesia,
[2]Department of Electrical Engineering, Universitas Ahmad Dahlan, Indonesia
[3]Department of Informatics, Universitas Ahmad Dahlan, Indonesia

Abstract.
Purpose: This study aims to enhance the detection of trademark image similarity by conducting a comparative analysis of various Convolutional Neural Network (CNN) architectures within Siamese networks, integrated with test-time augmentation techniques. Existing methods often face challenges in accurately capturing subtle visual similarities between trademarks due to limitations in feature extraction and generalization capabilities. The research seeks to identify the most effective CNN architecture for this task and to assess the impact of test-time augmentation on model performance.
Methods: The study implements Siamese networks utilizing three distinct CNN architectures: VGG16, VGG19, and ResNet50. Each network is trained on a dataset of trademark images to learn deep feature representations that can discriminate between similar and dissimilar trademarks. During the evaluation phase, test-time augmentation (TTA) is applied to enhance model robustness by averaging predictions over multiple augmented versions of the input images. TTA includes transformations such as random rotations (up to 40%), width and height shifts (up to 20%), random shear transformations, zooming (up to 20%), horizontal and vertical flips, and random brightness adjustments.
Result: Experimental findings reveal that the Siamese network based on VGG19 achieves the highest accuracy at 98.82%, outperforming the VGG16-based network with an accuracy of 97.07% and the ResNet50-based network with 50.00% accuracy. The application of TTA has improved performance across all models, with the VGG19 model receiving the highest improvement. The extremely low accuracy of ResNet50 can be attributed to its misinterpretation of original trademark images as close-forged ones, probably due to overfitting or lack of an efficient ability in generalizing very fine visual features.
Novelty: The study conducted a comparative analysis of CNN architectures, namely VGG16, VGG19, and ResNet50 in Siamese networks for trademark image similarity detection.

Keywords: Trademark, CNN, Siamese neural network, Test-time augmentation, Data augmentation
Received September 2024 / **Revised** December 2024 / **Accepted** December 2024

INTRODUCTION

The development of digital media makes trademark infringement easier to detect, but an accurate similarity detection method is needed in calculating the similarity of brand logos to protect intellectual property rights [1]. The manual method of detecting logo similarity by comparing one logo to a logo in the registered mark database is ineffective and error-prone [2]. This problem can lead to missing logos that should be similar but are not detected, so potential infringements that should be known go undetected [3].

Deep learning, particularly Convolutional Neural Networks (CNNs), has transformed image processing by automating feature extraction and hierarchical representation [4]. Because of their capability to learn major feature embeddings, Researchers use Siamese networks, made up of twin CNNs with shared weights, to detect similarity between input pairs [5]. Siamese networks use several CNN designs, including VGGNet and ResNet, to improve image similarity detection [6], [7].

Regarding architectural comparisons, Alshowaish et al. (2022) utilized CNN, a pre-trained deep learning algorithm, to detect identical trademark photographs. They used the VGGNet and ResNet architectures for feature extraction and evaluated the Euclidean distance metric [8]. The results had an average rank of 67,067.788 and an average accuracy of 0.774. This study does not examine the chi-squared distance matrix

*Corresponding author.
Email addresses: sunardi@mti.uad.ac.id (Sunardi)
DOI: 10.15294/sji.v11i4.13811

and only includes VGGNet and ResNet architectures. Later studies improved trademark image detection by delving into the subtleties of CNN structures, building on previous research.

Subsequent studies, building on prior research, examined the complexities of CNN architectures to enhance trademark image detection. Perez et al. investigated CNN models, particularly VGG19, for trademark image retrieval [9]. Two distinct techniques were employed: VGG19v, which concentrates on visual similarities, and VGG19c, which prioritizes logical connections. Combining the two models improved trademark picture detection by using the benefits of both visual and conceptual classification. This shows how greatly specific approaches inside CNN architectures might affect photo similarity detection results.

Research on distance metrics and unique designs continuously advances the field of trademark image detection. Suyahman investigated for trademark picture similarity recognition using Siamese neural networks with the chi-square distance measure [10]. The results showed a considerable improvement in accuracy: the redesigned network attained 98.05% accuracy. This is quite an improvement over baseline models using Euclidean distance. Applying the Chi-square distance metric improves the ability of the model to vary among the same trademarks.

The depth and simplicity of the VGGNet architectures, particularly VGG16 and VGG19, enable efficient feature extraction using consecutive convolutional layers [11]–[14]. ResNet's residual connections fix the problem of vanishing gradients, which makes it easier to train deeper topologies [15], [16]. Notwithstanding their achievements, these architectures possess specific constraints. For example, deeper networks such as ResNet may overfit when trained on limited datasets and may fail to catch subtle visual distinctions essential for trademark similarity identification [17].

By applying diverse alterations to input images during inference and consolidating the predictions, test-time augmentation (TTA) is a promising method that enhances model resilience. Suyahman et al. showed that it can improve the performance of CNN models like VGG16, VGG19, and ResNet50, leading to big gains in classification accuracy across a range of classes in the area of brand logo trademark detection [18]. However, researchers have yet to adequately investigate its utilization in Siamese networks for trademark image similarity identification [19]–[23].

In conclusion, while past research has focused on improving image similarity detection through advanced network architectures or enhanced training methods, there is still a need for systematic comparisons of CNN architectures within Siamese networks for trademark image similarity detection. This study utilizes a dataset sourced from Kaggle, comprising 255 trademark images categorized into five classes: Gojek, Grab, Uniqlo, Miniso, and CircleCI. Each class includes an anchor image, 20 positive images similar to the anchor, and 20 negative images dissimilar to the anchor, presenting challenges due to inter-class visual similarities in terms of shape, color, and background. Furthermore, to improve model robustness and performance, we implement TTA through transformations such as random rotations (up to 40%), width and height shifts (up to 20%), random shear transformations, zooming (up to 20%), horizontal and vertical flips, and random brightness modifications. In order to fill in the gaps in previous research, this study will compare CNN architectures (VGG16, VGG19, and ResNet50) in the context of Siamese networks and look at how adding more test time and using different distance metrics affects the ability to find trademark image similarities.

METHODS

This study shows a new version of the Siamese neural network architecture that uses a variety of Convolutional Neural Network (CNN) frameworks and a lot of different data augmentation techniques, such as random augmentation for training data and test-time augmentation (TTA), to make trademark image similarity detection more accurate. Figure 1 illustrates the research methodology.

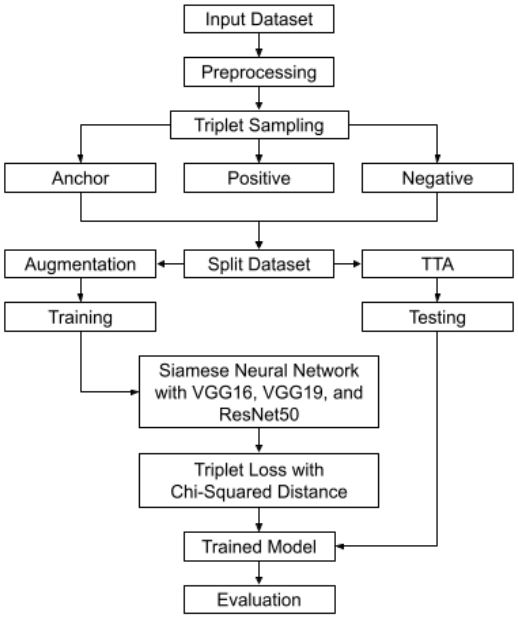

Figure 1. Research methodology

Data collection

The first step involved creating a dataset specifically for this study [24]. The dataset, sourced from Kaggle under the Indonesian Trademark Logo Dataset, comprises a total of 255 trademark images categorized into five classes, namely Gojek, Grab, Uniqlo, Miniso, and CircleCI. A total of 255 trademark images were gathered and categorized into five classes. Each class comprised one anchor image, serving as the reference image for the class, along with 20 positive images that are similar to the anchor and belong to the same class, and 20 negative images that are dissimilar to the anchor and belong to different classes. For testing purposes, 55 images were used to evaluate the performance of the trained models. Figure 2 shows some trademark image samples.

Figure 2. Trademark image from kaggle

Data preprocessing

All images were resized to 128 × 128 pixels to standardize the input dimensions for the CNN architectures [25]. Image normalization was performed by scaling pixel intensity values to the range [0, 1], which aids in accelerating the convergence of the neural networks during training [26], [27].

Data augmentation

To enhance the diversity of the training and testing data, extensive data augmentation techniques were applied [28]. Random augmentation was performed using the ImageDataGenerator class in Keras [29]. The augmentation techniques used in the research can be seen in Table 1.

Table 1. Augmentation techniques

Techniques	Range
Random Rotations	40%
Width and Height Shifts	20%
Shear Transformations	Random
Zoom	20%
Horizontal and Vertical Flips	Horizontally or Vertically
Brightness Adjustments	Random

These augmentations were applied in real-time during training, effectively increasing the dataset size and helping the models generalize better to unseen data.

Triplet generation

To train the Siamese networks, triplets consisting of an anchor image (A), a positive image (P), and a negative image (N) were created [30]. The anchor (A) serves as the foundational image from the training dataset. The positive (P) is an enhanced image analogous to the anchor and belongs to the identical class. The negative (N) is an enhanced image that differs from the anchor and originates from a distinct class. By creating all feasible combinations of positive and negative pairs inside each class and implementing random augmentations, a total of 400 triplets per class were generated, culminating in 2,000 triplets across all five classes. Because of the large number of triplets, the model was able to distinguish between similar and dissimilar trademarks [31].

Siamese network architecture

The Siamese network architecture consists of two identical subnetworks, each utilizing one of the following CNN architectures: the VGG16, VGG19, or ResNet50. The VGG16 is a 16-layer deep convolutional neural network known for its ability to efficiently extract features from either "VGG19" or 'ResNet50'. On the other hand, VGG19 is a convolutional neural network with 19 layers that has increased support for feature representation. ResNet50 is a fifty-layer convolutional neural network (CNN) that uses residual learning to train deeper networks. Each individual subnetwork is responsible for transforming an input image into a feature embedding, identified by the symbol $f(x)$, where x is the input image. A distance metric is used to compare anchor embeddings with embeddings from other photos in order to assess the degree of similarity [32].

Loss function

The networks were trained using the triplet loss function [33], which encourages the model to learn embeddings where the distance between the anchor and positive images is minimized and the distance between the anchor and negative images is maximized. The triplet loss function, as shown in Equation (1), is defined as:

$$L(A,P,N) = max\ (0, D(A,P) - D(A,N) + \alpha \quad (1)$$

where D is the Chi-square distance, A is the anchor, P is the positive example, N is the negative example, and α is the margin set at 1.0. The selection of the Chi-square distance is supported by evidence from prior research, which reported a significant improvement in accuracy, with the modified network achieving a 98% accuracy rate. This improvement, compared to baseline models employing Euclidean distance, underscores the Chi-square distance metric's superior capability in distinguishing between similar and dissimilar trademarks [10].

Model training

The training process was conducted separately for each CNN architecture within the Siamese network framework. The training parameter used can be seen in Table 2.

Table 2. Training parameters

Parameters	Value
Epoch	15
Batch Size	128
Learning Rate	1e-3
Optimizer	Adam

After the training process is over, the model will encode new inputs into a feature space. The distances that it has learned will reflect whether the inputs are similar or dissimilar, depending on the triplet arrangement.

Evaluation

Finally, the model's performance is assessed by calculating the accuracy, which results from the confusion matrix [34]. The confusion matrix shows model performance by summarizing the number of true negatives, false negatives, and correct and incorrect predictions [35]. Numerous metrics are calculated from this information, including recall, accuracy, precision, and F1 score. Precision measures the accuracy percentage; recall evaluates the model's ability to identify all true positives. One fair assessment of memory

and accuracy is the F1-score. These criteria allow one to evaluate the model's capacity to identify exact and unique trademark images methodically and comprehensively.

RESULTS AND DISCUSSIONS

This section compares various Convolutional Neural Network (CNN) architectures within Siamese networks. Test-Time Augmentation (TTA) is employed to assess the level of resemblance among trademark images. The TTA technique improves model performance by employing data augmentation during the testing phase. The outcomes of this enhancing method are depicted in Figure 3, which can be seen here.

Figure 3. Examples of augmented images for the training and testing datasets

During the testing process, three distinct CNN architectures were evaluated: VGG16, VGG19, and one additional bespoke model. The training period for each model was one period, and then the models were evaluated based on their training loss and testing accuracy. Throughout fifteen epochs, the loss function for the VGG16, VGG19, and ResNet50 models demonstrates unique trends, indicative of their ability to reduce the disparity between individual image pairs. VGG16 exhibits a fast decline in loss, attaining zero by the sixth epoch. This indicates that the model quickly learns to differentiate between images that are similar and images that are different. This persistent zero-loss from the fifth epoch onward implies that VGG16 was successful in optimizing the feature space, which enabled accurate distance measures for picture pairs that were comparable as well as those that were distinct.

In a similar manner, VGG19 displays a declining loss trajectory, albeit one that takes slightly more time to reach near-zero, which occurs about by the ninth epoch. This slower convergence in comparison to VGG16 may be due to the greater complexity of VGG19, which necessitates additional epochs in order to improve the picture representations and perfect its distance calculations for similarity identification.

On the other hand, ResNet50 displays a loss of 1.00000 over all epochs, which indicates that the model had difficulty learning an efficient measure for differentiating between images that are similar and those that are not similar. The absence of loss reduction suggests that the architecture of ResNet50 was not well-suited for this task or that it had difficulties learning from the dataset. The lack of loss reduction implies either ResNet50's design was not appropriate for this work or that it struggled learning from the dataset. These problems can have resulted from overfitting or underfitting dynamics or improper hyperparameter values.

As Figure 4 shows, the loss behavior of VGG16 and VGG19 is much different from ResNet50.

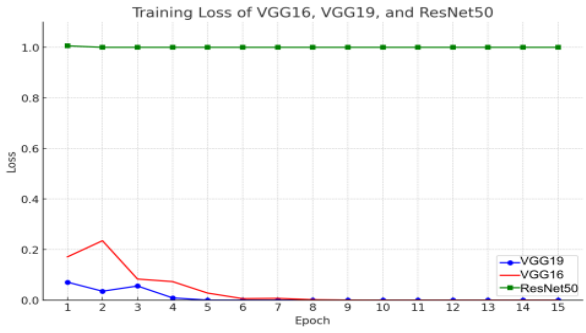

Figure 4. Training loss comparison

Over the epochs, VGG16, VGG19, and ResNet50's accuracy in spotting image similarity emphasizes even more their different performance. Reaching 100% accuracy by the fifth epoch and preserving flawless similarity detection for the whole training, VGG16 has the most effective learning curve. VGG16's great capacity in precisely detecting both similar and dissimilar pairs reflects in this fast improvement and continuous accuracy.

Though somewhat longer to develop, VGG19 reaches similar accuracy—more than 97% following the third epoch and peaks almost at 98.52% by the sixth epoch. VGG19 still shows great dependability in separating picture pairs according to their similarity measures, even if it does not reach the ideal accuracy seen in VGG16.

ResNet50 begins with a lower accuracy of 66.91% and varies just little over the epochs, peaking at 72.35% in the seventh epoch. Given ResNet50's large and continuous loss values, this modest accuracy improvement points to challenges in learning effective similarity measures. The poor performance of the model could result from a difficulty in capturing complex image features required for exact similarity identification, thereby requiring more tuning or changes.

Figure 5 shows the notable performance difference in terms of picture similarity between VGG16, VGG19, and ResNet50 among their accuracy comparison.

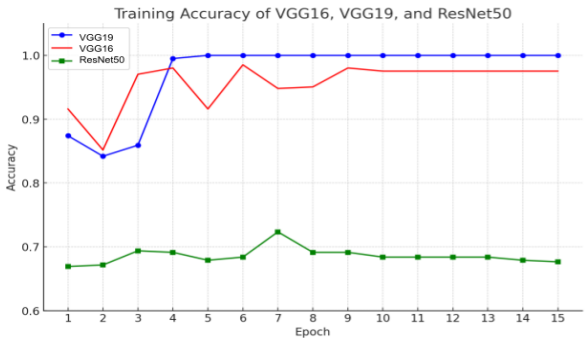

Figure 5. Training accuracy comparison

The confusion matrices for ResNet50, VGG16, and VGG19 models offer understanding of their performance in varying between "similar" and "different" images. With VGG16's first confusion matrix, the model achieves 97.07% accuracy. The model's great efficacy in precisely detecting both similar and dissimilar image pairs is shown by the true positive rate (True Similar) of 48.44% and the true negative rate (True Different) of 48.63%. The misclassification rates are minimal, with only 1.56% of similar images misclassified as dissimilar (False Similar) and 1.37% of dissimilar images misclassified as similar (False Different). These low error rates reflect VGG16's strong ability to recognize similarities and differences with very little confusion, as shown in Figure 6.

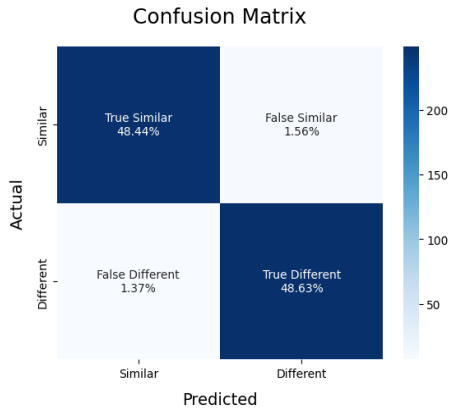

Figure 6. Confusion matrix of VGG16

The second confusion matrix for VGG19, with an accuracy of 98.82%, shows even better performance compared to VGG16. The True Similar rate is 50%, and the True Different rate is 48.83%, with no misclassification of similar pairs (False Similar = 0%) and a small misclassification of dissimilar pairs (False Different = 1.17%). This reflects the model's near-perfect ability to distinguish between image pairs, making almost no errors in its predictions, as seen in Figure 7.

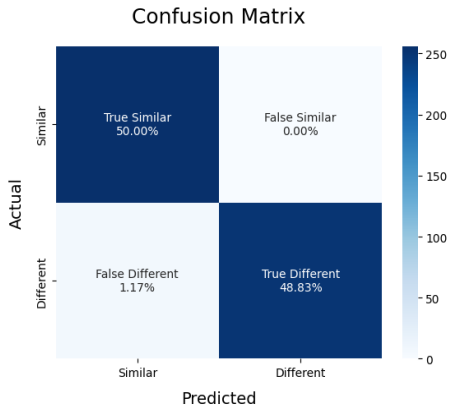

Figure 7. Confusion matrix of VGG19

In contrast, the third confusion matrix for ResNet50, which shows an accuracy of 50%, reveals significant challenges in the model's performance. While the True Similar rate is 50%, the True Different rate is 0%, indicating that the model failed to correctly identify any dissimilar pairs. Instead, 50% of the dissimilar pairs were misclassified as similar (false different), and 50% of the predictions were correct for similar pairs. This suggests that ResNet50 struggled to capture the appropriate features to differentiate dissimilar images, leading to high misclassification rates and poor overall performance, as depicted in Figure 8.

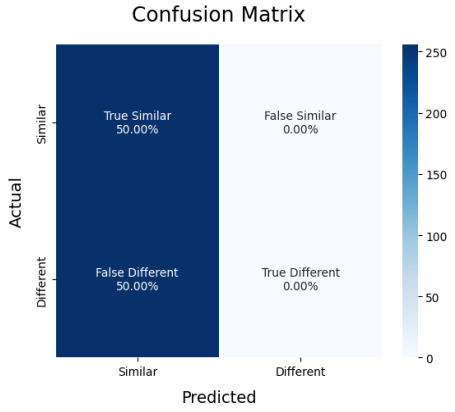

Figure 8. Confusion matrix of ResNet50

The testing accuracy results for VGG16, VGG19, and ResNet50, compared to Xception from prior research by Suyahman et al. [10], reveal clear distinctions in their performance on the image similarity task. As shown in Table 3, VGG19 achieved the highest accuracy at 98.82%, followed by VGG16 at 97.07%. Xception, as reported in earlier studies, achieved an accuracy of 98.05%, demonstrating its strong capability in handling trademark image similarity detection.

These findings highlight the superior performance of VGG19 among the current models being tested, showcasing its effectiveness in extracting relevant features due to its deeper architecture. Xception, while not part of this study's experiments, serves as a benchmark from previous research, providing a useful point of comparison for evaluating the improvements achieved by the VGG models.

In contrast, ResNet50 performed significantly worse, with an accuracy of only 50%, equivalent to random guessing. This suggests that ResNet50, in its current configuration, struggled to learn meaningful

distinctions between similar and dissimilar images in the dataset. Possible reasons for this include overfitting or an inability to generalize effectively from the training data.

As indicated in Table 3, the results demonstrate that VGG16 and VGG19 outperform ResNet50 in the current experiments, with VGG19 emerging as the most reliable architecture for image similarity tasks. Compared to Xception, the accuracy of VGG19 also represents a slight improvement, reinforcing its potential for further refinement and application in this domain.

Table 3. Comparison of result from VGG16, VGG19, and ResNet50

Model	Accuracy	Precision	Recall	F-1 Score
Xception	0.9805	1	0.9600	0.8600
VGG16	0.9707	0.9689	0.9726	0.9707
VGG19	0.9882	1	0.9771	0.9884
ResNet50	0.5000	1	0.5000	0.6666

The application of TTA techniques improved the accuracy of each model. The inclusion of test data helped to make the models more resistant to fluctuations, improving their ability to detect visual similarities. Greater reliability of results is achieved through the use of TTA, which averages predictions over many advanced input versions. The integration of a powerful CNN architecture, specifically VGG19, in conjunction with TTA has the potential to significantly improve the accuracy of trademark picture similarity recognition. It is possible that future studies will concentrate on further optimizing these designs, investigating the influence of increasing the number of training epochs, and looking into the possibility of including more diverse datasets in order to improve the robustness of algorithms.

CONCLUSION

In this study, VGG16, VGG19, and ResNet50 were tested for brand image similarity. The VGG19 had the highest accuracy at 98.82%, followed by the VGG16 at 97.07%. Both models were well able to distinguish brand images. Both models were able to distinguish between similar and different brand images. However, ResNet50 performed poorly, with an accuracy of just 50%, indicating that it is difficult to obtain suitable similarity measurements for this application. All models performed better when equipped with increased test time (TTA). These devices became more resistant to changes in brand data. The VGG19 architecture improved the most in the process. The results show that VGG-based architectures are better at detecting similarities between brands, particularly with TTA. It is possible that ResNet50 still needs to be optimized to detect inter-brand similarities well. Models based on VGG19 can be used in real-life applications, such as automatic trademark verification, image retrieval, and proving trademark infringement in court.

ACKNOWLEDGEMENT

This work is supported by the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia on the basis of the following contracts: 107/E5/PG.02.00.PL/2024, 0609.12/LL5-INT/AL.04/2024, and 121/PTM/LPPM UAD/VI/2024. The authors would also like to thank the reviewers for their helpful comments and suggestions, which helped to improve the presentation.

REFERENCES

[1] A. Pokrovskaya, “Liability for trademark infringement on e-commerce marketplaces,” *Int. J. Law Chang. World*, vol. 2, no. 1, pp. 87–101, Jun. 2023, doi: 10.54934/ijlcw.v2i1.40.

[2] D. Vesnin, D. Levshun, and A. Chechulin, “Trademark Similarity Evaluation Using a Combination of ViT and Local Features,” *Information*, vol. 14, no. 7, p. 398, Jul. 2023, doi: 10.3390/info14070398.

[3] G. S. L. Jr., *Research Handbook on the Law and Economics of Trademark Law*. Edward Elgar Publishing, 2023.

[4] M. Krichen, “Convolutional Neural Networks: A Survey,” *Computers*, vol. 12, no. 8, p. 151, Jul. 2023, doi: 10.3390/computers12080151.

[5] Suyahman, Sunardi, Murinto, and A. N. Khusna, “Siamese Neural Network Optimization Using Distance Metrics for Trademark Image Similarity Detection,” *Preprint*, vol. July 2024, 2024, [Online]. Available: https://www.researchgate.net/publication/381886654_Siamese_Neural_Network_Optimization_Using_Distance_Metrics_for_Trademark_Image_Similarity_Detection#fullTextFileContent

[6] P. Ranjan and A. Girdhar, “Deep Siamese Network with Handcrafted Feature Extraction for

Hyperspectral Image Classification," *Multimed. Tools Appl.*, vol. 83, no. 1, pp. 2501–2526, Jan. 2024, doi: 10.1007/s11042-023-15444-4.
[7] J. Zhang, H. Huang, X. Jin, L.-D. Kuang, and J. Zhang, "Siamese visual tracking based on criss-cross attention and improved head network," *Multimed. Tools Appl.*, vol. 83, no. 1, pp. 1589–1615, Jan. 2024, doi: 10.1007/s11042-023-15429-3.
[8] H. Alshowaish, Y. Al-Ohali, and A. Al-Nafjan, "Trademark Image Similarity Detection Using Convolutional Neural Network," *Appl. Sci.*, vol. 12, no. 3, p. 1752, Feb. 2022, doi: 10.3390/app12031752.
[9] C. A. Perez *et al.*, "Trademark Image Retrieval Using a Combination of Deep Convolutional Neural Networks," in *2018 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2018, pp. 1–7. doi: 10.1109/IJCNN.2018.8489045.
[10] A. Nur Khusna, "Siamese Neural Networks with Chi-square Distance for Trademark Image Similarity Detection," *Sci. J. Informatics*, vol. 11, no. 2, pp. 429–438, 2024, doi: 10.15294/sji.v11i2.4654.
[11] M. Feng and J. Su, "Learning Multi-Layer Attention Aggregation Siamese Network for Robust RGBT Tracking," *IEEE Trans. Multimed.*, vol. 26, pp. 3378–3391, 2024, doi: 10.1109/TMM.2023.3310295.
[12] A. B. Perdana and A. Prahara, "Face Recognition Using Light-Convolutional Neural Networks Based On Modified Vgg16 Model," in *2019 International Conference of Computer Science and Information Technology (ICoSNIKOM)*, Nov. 2019, pp. 1–4. doi: 10.1109/ICoSNIKOM48755.2019.9111481.
[13] S. R and S. Patilkulkarni, "Visual speech recognition for small scale dataset using VGG16 convolution neural network," *Multimed. Tools Appl.*, vol. 80, no. 19, pp. 28941–28952, Aug. 2021, doi: 10.1007/s11042-021-11119-0.
[14] A. Faghihi, M. Fathollahi, and R. Rajabi, "Diagnosis of skin cancer using VGG16 and VGG19 based transfer learning models," *Multimed. Tools Appl.*, vol. 83, no. 19, pp. 57495–57510, Dec. 2023, doi: 10.1007/s11042-023-17735-2.
[15] B. Koonce, *Convolutional Neural Networks with Swift for Tensorflow*. Berkeley, CA: Apress, 2021. doi: 10.1007/978-1-4842-6168-2.
[16] S. A. Hasanah, A. A. Pravitasari, A. S. Abdullah, I. N. Yulita, and M. H. Asnawi, "A Deep Learning Review of ResNet Architecture for Lung Disease Identification in CXR Image," *Appl. Sci.*, vol. 13, no. 24, p. 13111, Dec. 2023, doi: 10.3390/app132413111.
[17] M. Razavi, S. Mavaddati, and H. Koohi, "ResNet deep models and transfer learning technique for classification and quality detection of rice cultivars," *Expert Syst. Appl.*, vol. 247, p. 123276, Aug. 2024, doi: 10.1016/j.eswa.2024.123276.
[18] S. Suyahman, S. Sunardi, M. Murinto, and A. N. Khusna, "Data Augmentation Using Test-Time Augmentation on Convolutional Neural Network-Based Brand Logo Trademark Detection," *Indones. J. Artif. Intell. Data Min.*, vol. 7, no. 2, p. 266, 2024, doi: 10.24014/ijaidm.v7i2.28804.
[19] M. Kimura, "Understanding Test-Time Augmentation," 2021, pp. 558–569. doi: 10.1007/978-3-030-92185-9_46.
[20] D. Shanmugam, D. Blalock, G. Balakrishnan, and J. Guttag, "Better Aggregation in Test-Time Augmentation," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2021, pp. 1194–1203. doi: 10.1109/ICCV48922.2021.00125.
[21] J. Son and S. Kang, "Efficient improvement of classification accuracy via selective test-time augmentation," *Inf. Sci. (Ny).*, vol. 642, p. 119148, Sep. 2023, doi: 10.1016/j.ins.2023.119148.
[22] S. Cohen, N. Goldshlager, L. Rokach, and B. Shapira, "Boosting anomaly detection using unsupervised diverse test-time augmentation," *Inf. Sci. (Ny).*, vol. 626, pp. 821–836, May 2023, doi: 10.1016/j.ins.2023.01.081.
[23] W. Wang, Y. Chen, X. He, and Z. Li, "Soft Augmentation-Based Siamese CNN for Hyperspectral Image Classification With Limited Training Samples," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2021.3103180.
[24] Suyahman, "Indonesian Trademark Logo Dataset," *Kaggle*, 2024. https://www.kaggle.com/dsv/8401698
[25] N. K. C. Pratiwi, Y. N. Fu'adah, and E. Edwar, "Early Detection of Deforestation through Satellite Land Geospatial Images based on CNN Architecture," *J. INFOTEL*, vol. 13, no. 2, pp. 54–62, May 2021, doi: 10.20895/infotel.v13i2.642.
[26] X. Pei *et al.*, "Robustness of machine learning to color, size change, normalization, and image enhancement on micrograph datasets with large sample differences," *Mater. Des.*, vol. 232, p.

112086, Aug. 2023, doi: 10.1016/j.matdes.2023.112086.
[27] J. Alman, J. Liang, Z. Song, R. Zhang, and D. Zhuo, "Bypass Exponential Time Preprocessing: Fast Neural Network Training via Weight-Data Correlation Preprocessing," *Conf. Neural Inf. Process. Syst.*, no. NeurIPS, 2023, [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2023/file/9690d4746230cfea3d067fca695ba648-Paper-Conference.pdf
[28] L. Huang, J. Qin, Y. Zhou, F. Zhu, L. Liu, and L. Shao, "Normalization Techniques in Training DNNs: Methodology, Analysis and Application," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 8, pp. 10173–10196, Aug. 2023, doi: 10.1109/TPAMI.2023.3250241.
[29] A. K. Godishala, H. Yassin, R. Veena, and D. T. Ching Lai, "Breast Cancer Tumor Image Classification Using Deep Learning Image Data Generator," in *2022 7th International Conference on Image, Vision and Computing (ICIVC)*, Jul. 2022, pp. 418–423. doi: 10.1109/ICIVC55077.2022.9886202.
[30] N. Shaffi, F. Hajamohideen, M. Mahmud, A. Abdesselam, K. Subramanian, and A. Al Sariri, "Triplet-Loss Based Siamese Convolutional Neural Network for 4-Way Classification of Alzheimer's Disease," 2022, pp. 277–287. doi: 10.1007/978-3-031-15037-1_23.
[31] Y. Liu *et al.*, "Causal Triplet: An Open Challenge for Intervention-centric Causal Representation Learning," *Proc. Mach. Learn. Res.*, vol. 213, pp. 553–573, 2023, [Online]. Available: https://proceedings.mlr.press/v213/liu23a/liu23a.pdf
[32] P. Li, H. Yan, and X. Lu, "A Siamese neural network for learning the similarity metrics of linear features," *Int. J. Geogr. Inf. Sci.*, vol. 37, no. 3, pp. 684–711, Mar. 2023, doi: 10.1080/13658816.2022.2143505.
[33] A. Dehghan, P. Razzaghi, K. Abbasi, and S. Gharaghani, "TripletMultiDTI: Multimodal representation learning in drug-target interaction prediction with triplet loss function," *Expert Syst. Appl.*, vol. 232, p. 120754, Dec. 2023, doi: 10.1016/j.eswa.2023.120754.
[34] N. K. E. Sapitri, U. Sa'adah, and N. Shofianah, "Knowledge Discovery from Confusion Matrix of Pruned CART in Imbalanced Microarray Data Ovarian Cancer Classification," *Sci. J. Informatics*, vol. 11, no. 1, pp. 227–236, 2024, doi: 10.15294/sji.v11i1.50077.
[35] J. Unjung and M. R. Ningsih, "Optimized Handwriting-based Parkinson's Disease Classification Using Ensemble Modeling and VGG19 Feature Extraction," *Sci. J. Informatics*, vol. 10, no. 4, p. 489, 2023, doi: 10.15294/sji.v10i4.47108.