Scientific Journal of Informatics
Vol. 11, No. 2, May 2024

p-ISSN 2407-7658 https://journal.unnes.ac.id/journals/sji/index e-ISSN 2460-0040

Siamese Neural Networks with Chi-square Distance for Trademark Image Similarity Detection

Suyahman[1*], Sunardi[2], Murinto[3], Arfiani Nur Khusna[4]
[1]Master Program of Informatics, Universitas Ahmad Dahlan, Indonesia
[2]Department of Electrical Engineering, Universitas Ahmad Dahlan, Indonesia
[3,4]Department of Informatics, Universitas Ahmad Dahlan, Indonesia

Abstract.
Purpose: The objective of this study is to address the limitations of existing trademark image similarity analysis methods by integrating a Chi-square distance metric within a Siamese neural network framework. Traditional approaches using Euclidean distance often fail to accurately capture the complex visual features of trademarks, leading to suboptimal performance in distinguishing similar trademarks. This research aims to improve the precision and robustness of trademark comparison by leveraging the Chi-square distance, which is more sensitive to image variations.
Methods: The approach involves modifying a Siamese neural network traditionally employing Euclidean distance with the use the Chi-square distance metric instead. This alteration allows the network to better capture and analyze critical visual features such as color and texture. The modified network is trained and tested on a comprehensive dataset of trademark images, enabling the network to learn and distinguish between similar and dissimilar trademarks based on subtle visual cues.
Result: The findings from this study show a significant increase in accuracy, with the modified network achieving an accuracy rate of 98%. This marks a notable improvement over baseline models that utilize Euclidean distance, demonstrating the effectiveness of the Chi-square distance metric in enhancing the model's ability to discriminate between trademarks.
Novelty: The novelty of this research lies in its application of the Chi-square distance in a deep learning framework specifically for trademark image similarity detection, presenting a novel approach that yields higher precision in image-based comparisons.

Keywords: Trademark, Siamese neural network, Triplet loss, Chi-square
Received May 2024 / **Revised** May 2024 / **Accepted** May 2024

INTRODUCTION

Trademarks (TM) are crucial for distinguishing goods or services produced by different entities, playing a vital role in consumer recognition and business branding. Traditionally, trademarks include unique symbols, words, and images and are registered with governmental intellectual property (IP) offices to protect a company's goodwill and brand equity[1]. According to recent reports, the global volume of TM applications has surged, reaching a record 18 million in 2021, presenting significant challenges in processing and examining these applications accurately and efficiently [2].

The examination of trademarks often requires assessing the similarity of textual and visual elements, which historically has been conducted through manual inspections under systems like the Vienna Classification [3]. However, these manual methods are fraught with subjectivity and inconsistency [4]. Previous studies have explored various computational methods to address these issues, including the use of mathematical models and image feature extraction techniques [5]–[8]. These studies predominantly employed Euclidean distance measures within their models for assessing similarity between images [9], [10]. While effective to a degree, these approaches often fail to capture the nuanced differences in image features that are crucial for trademark comparison, particularly in complex images.

Haddad et al. [11] introduced a trademark recognition and retrieval framework that combines a convolutional neural network (CNN) with a relevance feedback (RF) mechanism. This framework uses

*Corresponding author.
Email addresses: 2307048007@webmail.uad.ac.id (Suyahman)
DOI: [10.15294/sji.v11i2.4654](https://doi.org/10.15294/sji.v11i2.4654)

particle swarm optimization and self-organizing map methods in a preprocessing phase to optimize features and minimize the search space. It was evaluated using the FlickrLogos-32 PLUS dataset. The results highlighted the benefits of integrating CNN with an RF mechanism compared to using sole deep learning models.

CNNs are widely applied to tasks such as image classification, detection, and recognition [12]–[16]. In their study, Perez et al. [17] utilized a pretrained CNN model, VGG19, for the purpose of trademark image retrieval. They trained this network on two distinct databases: VGG19v, which prioritized visual similarities, and VGG19c, which addressed conceptual similarities. VGG19v was trained using a collection of web-sourced logos classified into 151 categories by experts based on visual similarities. On the other hand, VGG19c was trained on USPTO logos that were manually categorized. The integration of these two networks, VGG19v and VGG19c, has been shown to enhance the performance of image detection tasks by leveraging their complementary strengths in classification.

Alshowaish et al. [18] developed a system for detecting trademark image similarity using deep learning techniques, specifically employing two pretrained convolutional neural networks, VGGNet and ResNet. The system automatically extracted image features and used Euclidean distance to assess similarity between trademarks, achieving promising results on the Middle East Technical University dataset, with an average rank of 67,067.788, a normalized average rank of 0.0725, and a mean average precision of 0.774. This demonstrates the system's potential to enhance the accuracy and efficiency of trademark examination processes.

Recent advancements in image retrieval have benefited from deep learning technologies, specifically convolutional neural networks (CNNs), which have substantially improved the accuracy of feature extraction and image similarity assessments. Techniques such as the triplet loss method have been utilized to refine the learning process by comparing an anchor image with positive and negative examples, thereby enhancing the model's discriminatory capabilities [19]. However, there remains a significant gap in the adaptation of these methods to the specific needs of TM image comparison, particularly regarding the suitability of distance metrics used in these models.

The aim of this study is to overcome the limitations of current trademark image similarity analysis methods by incorporating a Chi-square distance metric within a Siamese neural network framework. This approach deviates from the traditional Euclidean distance, aiming to better accommodate the specific characteristics of TM logos which vary by shape, color, and textural differences [20]. By modifying the distance algorithm to incorporate Chi-square distance, this research seeks to capture more effectively the subtle dissimilarities between logos that other models might overlook.

METHODS

This study introduces a modified Siamese neural network architecture employing the Chi-square distance metric, aimed at enhancing the accuracy of image-based trademark similarity analysis. This research applies the research methodology shown in Figure 1.

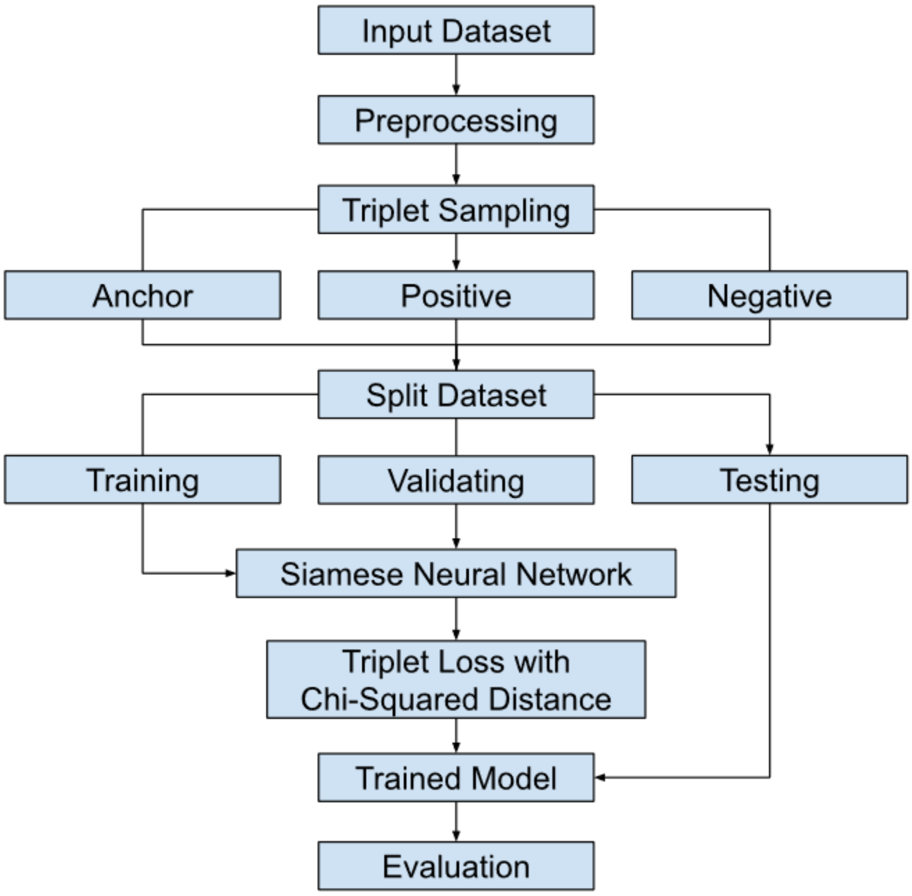

Figure 1. Research method

In the initial stages of the research, a dataset is assembled for the purpose of subsequent analyses. At this stage, a total of 255 images featuring five distinct trademarks have been collated. The complete image dataset comprises training data with a single anchor image, 20 positive images, and 20 negative images. For purposes of model testing, the dataset utilized is comprised of 55 test data. The collected dataset is then imported into Google Drive for convenient data processing and storage. Each brand in the dataset will be used as a label in the model that will be created. Figure 2 shows some trademark image samples.

Figure 2. Trademark image

Pre-processing is performed to clean and prepare the data, including tasks such as resizing images to a 128 x 128 scale and normalizing values to fit the neural network's input requirements [21]. Image resizing ensures uniformity, while normalization adjusts pixel values to a range suitable for neural network processing, enhancing model performance and convergence [22].

The triplet sampling stage is next. This stage involves organizing the data into triplets. Each triplet consists of an anchor point, a positive example that is similar to the anchor, and a negative example that is dissimilar to the other two points [23]. Triplet creation is done randomly by creating all possible pairs of positive and negative images, resulting in 400 triplets per class. This method ensures a diverse range of training samples, improving the model's ability to distinguish between similar and dissimilar trademarks [24].

The next stage is to divide the dataset into two distinct subsets: training data and test data. The ratio of these two subsets is 80:20. Training data is data that will later be used in the model training process. Test data is data that will later be used to reduce the possibility of errors occurring. Splitting the dataset in this way allows for a comprehensive evaluation of the generalization ability of the model [25].

Training involves a Siamese neural network, which leverages a triplet loss function [26]. This function, possibly incorporating a Chi-square distance metric [27], ensures that model learn to minimise the distance between the anchor and the positive image while maximising the distance between the anchor and the negative image. The Chi-square distance metric is particularly effective for comparing histogram data, such as color histograms in images [28]. This metric is sensitive to the variance in image features such as color and texture, making it suitable for trademark image comparison. The Chi-square distance as a distance function in triplet loss calculations uses the following formula:

$$D_{\chi^2}(\mathrm{X}, \mathrm{Y}) = \sum \frac{(X_i - Y_i)^2}{X_i + Y_i} \tag{1}$$

where X_i and Y_i are values of element of the two compared feature vectors. Meanwhile, the triplet loss function is formulated to minimize the distance between the anchor and positive image and maximize the distance between the anchor and negative image with a certain margin. The triplet loss formula is as follows:

$$L(A, P, N) = \max\,(0, D(A, P) - D(A, N) + \alpha \tag{2}$$

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 Siamese neural network used in this research is a CNN Xception architecture consisting of three identical sub-networks that share the same weights. Each sub-network consists of several convolution layers that are followed by pooling and fully connected layers [29]. The Xception architecture, known for its depthwise separable convolutions, offers efficient and robust feature extraction [30].

Models are compared using Chi-square and Euclidean distance metrics, employing the same hyperparameters to ensure comparable evaluations. This approach highlights the effectiveness of different distance metrics in enhancing model performance. The specific values of the hyperparameters used can be seen in Table 1.

Table 1. Hyperparameters

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

Once trained, the trained model is used to encode new inputs into a feature space where the learned distances indicate similarity or dissimilarity as defined by the triplet configuration.

At the evaluation stage, the effectiveness of the model was evaluated using a confusion matrix and other metrics, thereby providing a detailed picture of the model's performance by illustrating true positives, false positives, true negatives, and false negatives [31]. After obtaining the confusion matrix, accuracy, precision, recall, F1-score, and MAP (mean average precision) are calculated [32].

The term accuracy refers to the proportion of instances that have been correctly classified, expressed as a percentage. In contrast, precision indicates the ratio of true positive instances to the total number of positive predictions, including both false positives and false negatives. Recall, on the other hand, is defined as the ratio of true positives to the sum of true and false positives, representing the model's capability to identify relevant events. F1-score combines precision and recall into one metric by taking their harmonic average.

MAP is the average precision score at each threshold, useful for evaluating ranking tasks. This comprehensive approach, from data preparation to detailed evaluation, ensures a robust assessment of the model's ability to differentiate between different classes of data.

RESULTS AND DISCUSSIONS

This section outlines and discusses the results of the motif classification process involving trademark images using the Siamese neural network with Euclidean and Chi-square distances. The results of the similarity analysis and evaluation are elucidated in this section. The main focus of this section is the accuracy and model evaluation results for each trademark image.

Following the image reading and preprocessing steps, the images are randomly generated into triplet samples, resulting in 400 image triplets derived from 20 positive images and 20 negative images. An example of the triplet sampling results can be seen in Figure 3.

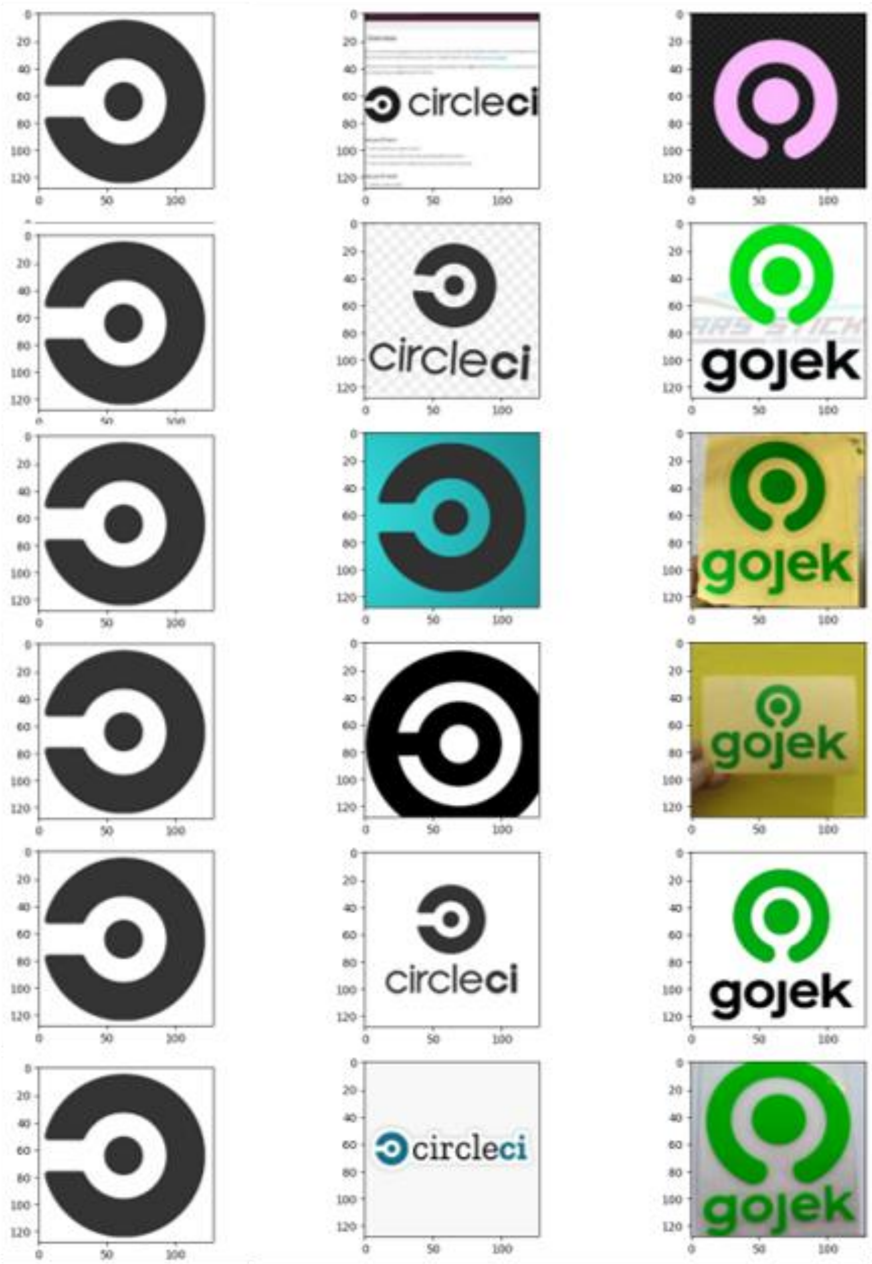

Figure 3. Result of preprocessing and triplet sampling

The sampled image triplets are trained using a Siamese neural network with the Euclidean and Chi-square distance metrics. The performance of the Siamese neural network with the Euclidean distance metric, as illustrated in Figure 4, is characterized by fluctuating training loss and inconsistent testing accuracy. The training loss shows a downward trend overall but experiences a significant spike at epoch 6, indicating potential overfitting or instability during the training process. The testing accuracy also varies considerably across epochs, with notable peaks and troughs, suggesting that the model struggles to generalize well to the testing data, resulting in lower overall accuracy.

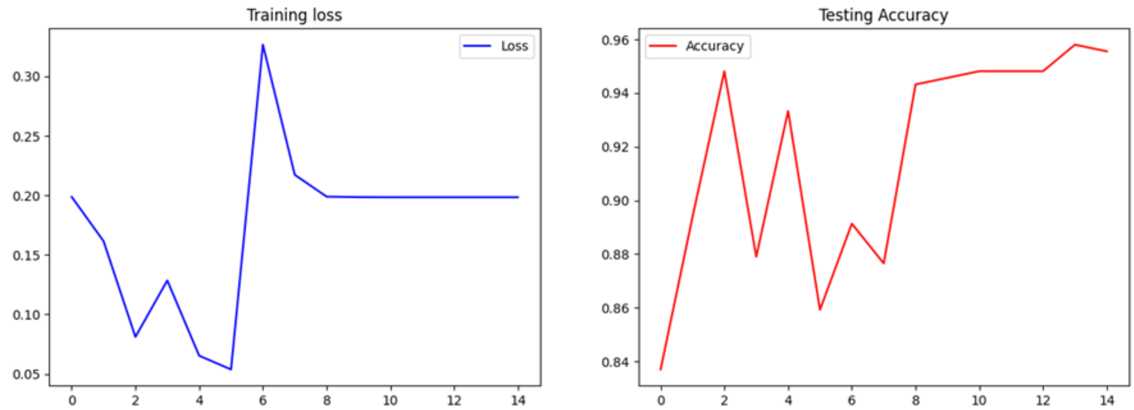

Figure 4. Training loss and testing accuracy of the siamese neural network with the Euclidean distance metric

The confusion matrix for the Euclidean distance model, as shown in Figure 5, reveals high percentages of true similar predictions (50.00%) and true dissimilar predictions (34.18%), but the percentage of false dissimilar predictions is also significant (15.82%). This indicates that the model frequently misclassifies dissimilar items as similar, further reflecting challenges in achieving high accuracy with Euclidean distance.

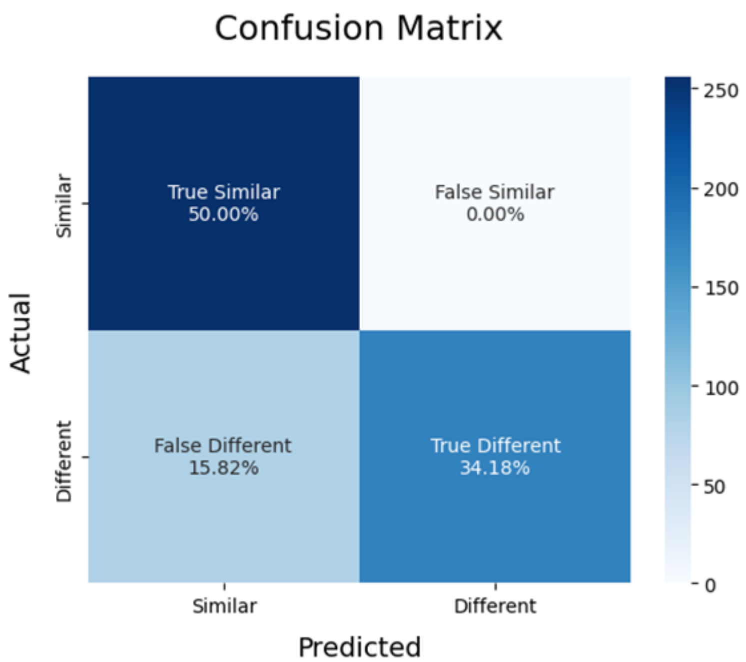

Figure 5. Confusion matrix for the Siamese neural network with the Euclidean distance metric

In contrast, the Siamese neural network with the Chi-squared distance metric, as illustrated in Figure 6, demonstrates superior performance. The training loss decreases sharply and stabilizes at a very low value, indicating efficient learning and convergence. The testing accuracy improves rapidly and stabilizes to a near-perfect accuracy after a few epochs, showing that the model generalizes well to the testing data.

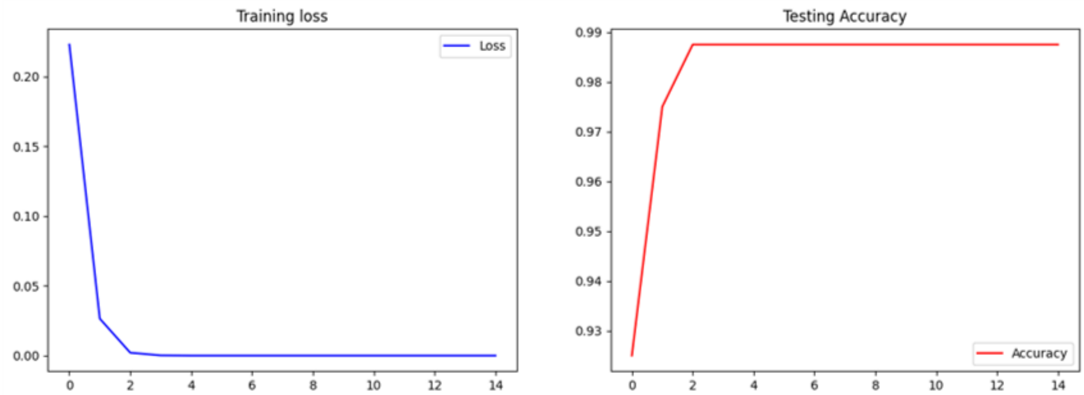

Figure 6. Training loss and testing accuracy of the siamese neural network with the Chi-square distance metric

The confusion matrix for the Chi-square distance model, based on Figure 7, indicates an extremely low percentage of false dissimilar predictions (1.95%) and a high percentage of true dissimilar predictions (48.05%), which underscores the model's improved ability to distinguish between similar and dissimilar items.

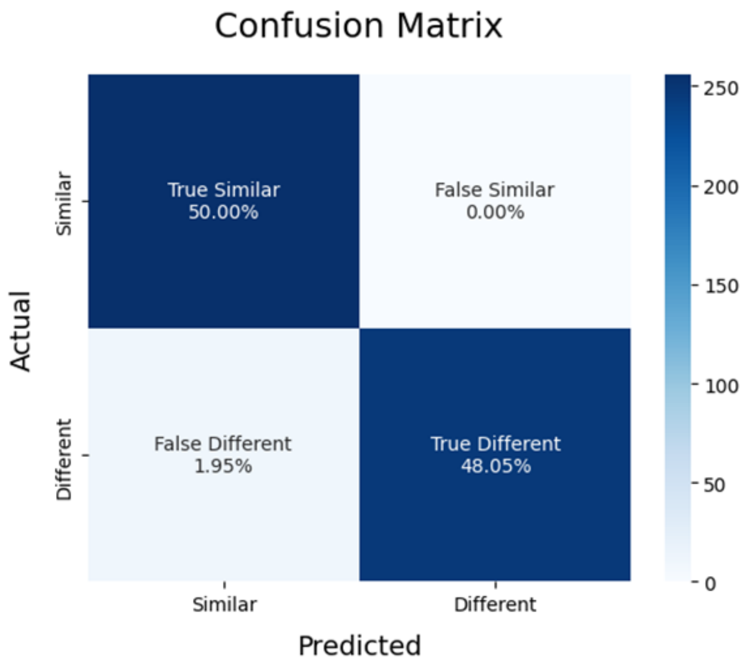

Figure 7. Confusion matrix for the siamese neural network with the chi-square distance metric

In comparison to a similar model utilizing the Euclidean distance metric, with a recorded accuracy of 84%, the Chi-square distance metric provides a significant improvement, registering an accuracy of 98%. This substantial enhancement can be attributed to the Chi-square distance's sensitivity to the variance in image features such as color and texture, which are critical in the visual differentiation of trademark images. Comparison of the performance of the Siamese neural network with the Euclidean and Chi-square distance metrics can be seen in Table 1.

Table 1. Comparison of the siamese neural network with the euclidean and chi-square distance metrics

Metric Distance	Accuracy	Precision	Recall	F1-Score	MAP
Euclidean	0.84	1	0.75	0.86	1
Chi-square (Purpose)	0.98	1	0.96	0.98	0.98

The use of the Chi-square distance metric in Siamese neural networks for trademark image similarity detection brings several advantages over traditional distance metrics. The Chi-square distance metric is particularly effective in handling discrepancies in image histograms, which is pivotal when comparing images that may vary slightly in hue or saturation due to photographic conditions or alterations. This capability makes this model exceptionally suitable for legal frameworks of trademark similarity, where minute differences can determine the outcomes of infringement cases.

CONCLUSION

This study demonstrates the effectiveness of the Chi-square distance metric in improving the performance of Siamese neural networks for distinguishing between similar and dissimilar items. Numerical results highlight the superior accuracy and stability of the Chi-square distance model, which has achieved a testing accuracy that stabilizes to a near-perfect level at approximately 98%, compared to the fluctuating and lower accuracy of around 90% achieved with the Euclidean distance model. The Chi-square distance model also exhibits a significantly low false dissimilar prediction rate of 1.95%, compared to the 15.82% rate exhibitedas by the Euclidean distance model. Additionally, the true dissimilar prediction rate is higher at 48.05% for the Chi-square distance model, indicating better classification performance. These findings suggest that the Chi-square distance metric is more suitable for tasks involving histogram-like data representations, offering a more robust measure for comparing features. The research's impact lies in providing a reliable and efficient method for improving classification accuracy in Siamese neural networks, which can be beneficial for various applications in machine learning and pattern recognition.

REFERENCES

[1] H. Heniyatun, P. Sulistyaningsih, B. T. Iswanto, Y. Asiyah, and C. B. E. Praja, "Kajian Yuridis Perlindungan Merek Terhadap Gugatan Merek Nama Orang Terkenal," *Borobudur Law Rev.*, vol. 2, no. 2, pp. 137–149, Aug. 2020, doi: 10.31603/burrev.4648.

[2] W. I. P. O. (WIPO), "World Intellectual Property Indicators 2023," *Genewa: WIPO*, 2023. https://www.wipo.int/publications/en/details.jsp?id=4678 (accessed May 10, 2024).

[3] C. V. Trappey, A. J. C. Trappey, and S. C.-C. Lin, "Intelligent trademark similarity analysis of image, spelling, and phonetic features using machine learning methodologies," *Adv. Eng. Informatics*, vol. 45, p. 101120, Aug. 2020, doi: 10.1016/j.aei.2020.101120.

[4] S. Katyal and A. Kesari, "Trademark Search, Artificial Intelligence and the Role of the Private Sector," *Berkeley Technol. Law J.*, 2024, [Online]. Available: https://ssrn.com/abstract=3760112

[5] T. Lan, X. Feng, Z. Xia, S. Pan, and J. Peng, "Similar Trademark Image Retrieval Integrating LBP and Convolutional Neural Network," 2017, pp. 231–242. doi: 10.1007/978-3-319-71598-8_21.

[6] A. J. C. Trappey, C. V. Trappey, and S. Shih, "An intelligent content-based image retrieval methodology using transfer learning for digital IP protection," *Adv. Eng. Informatics*, vol. 48, p. 101291, Apr. 2021, doi: 10.1016/j.aei.2021.101291.

[7] X. Peng, X. Zhang, Y. Li, and B. Liu, "Research on image feature extraction and retrieval algorithms based on convolutional neural network," *J. Vis. Commun. Image Represent.*, vol. 69, p. 102705, May 2020, doi: 10.1016/j.jvcir.2019.102705.

[8] 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," *ndonesian J. Artif. Intell. Data Min.*, vol. 7, no. 2, 2024.

[9] A. Agarwal, D. Sharma, and D. Agrawal, *Trademark Image Retrieval using Color and Shape Features and Similarity Measurement*. 2021.

[10] R. K. Tripathi and S. C. Agrawal, "A shape and texture features fusion to retrieve similar Trademark

Image Material," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1116, no. 1, p. 012026, Apr. 2021, doi: 10.1088/1757-899X/1116/1/012026.
[11] M. Haddad, A. Saleh, and Y. Jaber, "Trademark image recognition utilizing deep convolutional neural network," *Ann. Electr. Electron. Eng.*, vol. 2, no. 3, pp. 13–20, 2019.
[12] M. F. Alhakim, B. Prasetiyo, J. Unjung, and M. Muslim, *CNN Model for the Small Data in Vehicle Miniature Classification*. 2023. doi: 10.1109/ICONNIC59854.2023.10467559.
[13] M. Murinto, M. Rosyda, and M. Melany, "Klasifikasi Jenis Biji Kopi Menggunkan Convolutional Neural Network dan Transfer Learning pada Model VGG16 dan MobileNetV2," *JRST (Jurnal Ris. Sains dan Teknol.*, vol. 7, no. 2, p. 183, Sep. 2023, doi: 10.30595/jrst.v7i2.16788.
[14] M. Fahmi, A. Yudhana, and S. Sunardi, "Pemilahan Sampah Menggunakan Model Klasifikasi Support Vector Machine Gabungan dengan Convolutional Neural Network," *JURIKOM (Jurnal Ris. Komputer)*, vol. 10, no. 1, pp. 76–81, 2023, doi: 10.30865/jurikom.v10i1.5468.
[15] M. Murinto, S. Winiarti, and I. Faisal, "Particle Swarm Optimization Algorithm for Hyperparameter Convolutional Neural Network and Transfer Learning VGG16 Model," *J. Comput. Sci.*, vol. 5, no. 1, pp. 474–480, 2024.
[16] A. Maftukhah, A. Fadlil, and S. Sunardi, "Butterfly Image Classification using Convolution Neural Network with AlexNet Architecture," *J. INFOTEL*, vol. 16, no. 1, Feb. 2024, doi: 10.20895/infotel.v16i1.1004.
[17] 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)*, IEEE, Jul. 2018, pp. 1–7. doi: 10.1109/IJCNN.2018.8489045.
[18] 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.
[19] W. Yang, L. Xu, X. Chen, F. Zheng, and Y. Liu, "Chi-Squared Distance Metric Learning for Histogram Data," *Math. Probl. Eng.*, vol. 2015, pp. 1–12, 2015, doi: 10.1155/2015/352849.
[20] 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.
[21] H. A. Salman and A. Kalakech, "Image Enhancement using Convolution Neural Networks," *Babylonian J. Mach. Learn.*, vol. 2024, pp. 30–47, Jan. 2024, doi: 10.58496/BJML/2024/003.
[22] H. Rajoli, F. Lotfi, A. Atyabi, and F. Afghah, "Triplet Loss-less Center Loss Sampling Strategies in Facial Expression Recognition Scenarios," Feb. 2023, doi: https://doi.org/10.48550/arXiv.2302.04108.
[23] Y. Wang, X. Xu, Z. Wang, R. Li, Z. Hua, and H. Song, "ShuffleNet-Triplet: A lightweight RE-identification network for dairy cows in natural scenes," *Comput. Electron. Agric.*, vol. 205, p. 107632, Feb. 2023, doi: 10.1016/j.compag.2023.107632.
[24] P. Xu *et al.*, "LVLM-eHub: A Comprehensive Evaluation Benchmark for Large Vision-Language Models," Jun. 2023, [Online]. Available: http://arxiv.org/abs/2306.09265
[25] A. F. Abdul Fadlil, I. Riadi, and F. Andrianto, "Improving Sentiment Analysis in Digital Marketplaces through SVM Kernel Fine-Tuning," *Int. J. Comput. Digit. Syst.*, vol. 15, no. 1, pp. 159–171, Jul. 2024, doi: 10.12785/ijcds/160113.
[26] B.-R. Cha and B. Vaidya, "Enhancing Human Activity Recognition with Siamese Networks: A Comparative Study of Contrastive and Triplet Learning Approaches," *Electronics*, vol. 13, no. 9, p. 1739, May 2024, doi: 10.3390/electronics13091739.
[27] H. Sadeghi and A.-A. Raie, "HistNet: Histogram-based convolutional neural network with Chi-squared deep metric learning for facial expression recognition," *Inf. Sci. (Ny).*, vol. 608, pp. 472–488, Aug. 2022, doi: 10.1016/j.ins.2022.06.092.
[28] N. Banic and N. Elezovic, "TVOR: Finding Discrete Total Variation Outliers Among Histograms," *IEEE Access*, vol. 9, pp. 1807–1832, 2021, doi: 10.1109/ACCESS.2020.3047342.
[29] K. Shaheed *et al.*, "DS-CNN: A pre-trained Xception model based on depth-wise separable convolutional neural network for finger vein recognition," *Expert Syst. Appl.*, vol. 191, p. 116288, Apr. 2022, doi: 10.1016/j.eswa.2021.116288.
[30] R. Zhang, F. Zhu, J. Liu, and G. Liu, "Depth-Wise Separable Convolutions and Multi-Level Pooling for an Efficient Spatial CNN-Based Steganalysis," *IEEE Trans. Inf. Forensics Secur.*, vol. 15, pp. 1138–1150, 2020, doi: 10.1109/TIFS.2019.2936913.
[31] W. Ayadi, W. Elhamzi, I. Charfi, and M. Atri, "Deep CNN for Brain Tumor Classification," *Neural Process. Lett.*, vol. 53, no. 1, pp. 671–700, Feb. 2021, doi: 10.1007/s11063-020-10398-2.

[32] A. Ahmad, D. Saraswat, V. Aggarwal, A. Etienne, and B. Hancock, "Performance of deep learning models for classifying and detecting common weeds in corn and soybean production systems," *Comput. Electron. Agric.*, vol. 184, p. 106081, May 2021, doi: 10.1016/j.compag.2021.106081.