applied sciences

MDPI

Article

Hybrid-Margin Softmax for the Detection of Trademark Image Similarity

Chenyang Wang [1], Guangyuan Zheng [2] and Hongtao Shan [1,*]

[1] School of Electrical and Electronic Engineering, Shanghai University of Engineering Science, Shanghai 201620, China; m020220338@sues.edu.cn

[2] College of Information Technology, Shanghai Jianqiao University, Shanghai 201306, China; zhengguangyuan@gench.edu.cn

* Correspondence: 13503581768@163.com

Abstract: The detection of image similarity is critical to trademark (TM) legal registration and court judgment on infringement cases. Meanwhile, there are great challenges regarding the annotation of similar pairs and model generalization on rapidly growing data when deep learning is introduced into the task. The research idea of metric learning is naturally suited for the task where similarity of input is given instead of classification, but current methods are not targeted at the task and should be upgraded. To address these issues, loss-driven model training is introduced, and a hybrid-margin softmax (HMS) is proposed exactly based on the peculiarity of TM images. Two additive penalty margins are attached to the softmax to expand the decision boundary and develop greater tolerance for slight differences between similar TM images. With the HMS, a Siamese neural network (SNN) as the feature extractor is further penalized and the discrimination ability is improved. Experiments demonstrate that the detection model trained on HMS can make full use of small numbers of training data and has great discrimination ability on bigger quantities of test data. Meanwhile, the model can reach high performance with less depth of SNN. Extensive experiments indicate that the HMS-driven model trained completely on TM data generalized well on the face recognition (FR) task, which involves another type of image data.

Keywords: trademark image similarity; metric learning; Siamese neural network; softmax function

check for updates

Citation: Wang, C.; Zheng, G.; Shan, H. Hybrid-Margin Softmax for the Detection of Trademark Image Similarity. *Appl. Sci.* **2024**, *14*, 2865. https://doi.org/10.3390/app14072865

Academic Editor: Atsushi Mase

Received: 21 February 2024
Revised: 24 March 2024
Accepted: 27 March 2024
Published: 28 March 2024

1. Introduction

Trademarks (TMs) are distinctive designations registered to identify products and sources. The exclusivity of TM provides rules for orderly marketing [1]. However, the high incidence of TM misappropriation causes plenty of revenue and reputation loss to legitimate owners. Consumers can be misled to purchase counterfeit products, especially when the right-infringing TM image is similar to a legal one [2]. Meanwhile, the rapidly growing TM image database is massive itself, which brings great pressure on the governing body.

What further complicates this situation is that there are no defined criteria to conduct the test named 'likelihood of confusion' [3]. The test is a critical part of the procedure to determine whether a disputed trademark is similar to another one. Thus, there is a chance that inconsistent judgments are declared by courts of different levels or districts.

Generally, appearance, characters, and sound are taken into consideration during the test [4]. The repetition rate of characters is convenient to assess, and a TM can be pronounced differently among regions. As the most common and important forms of TM, the appearance, by contrast, is more consistent and controversial in judgment.

The feature extraction of TM images is crucial to the above issues. Conventional feature engineering involves manually designed descriptors to detect and match features, e.g., SIFT [5] and ORB [6]. SFIT is a local invariant feature descriptor based on keypoints and local image gradient directions. ORB is a fast binary descriptor based on FAST keypoint detector and binary BRIEF descriptor. These extraction methods focus on some specific image features such as points and edges [7], which makes it expensive to detect TM image similarity comprehensively with several manual descriptors. The great improvement in deep learning, that features which might be omitted by human beings can be extracted efficiently by convolutional kernels, makes introducing computer 'opinions' to the procedure of human judgment on TM image similarity a convincing prospect.

There is a great challenge in building training data when deep learning is introduced in the detection of TM image similarity (TMISD). The performance of the detection model is highly correlated with supervised information, while the annotation of similar TM image pairs takes intensive work with skilled labor involved. Furthermore, the model generalization on millions of new TM designs proposes a higher requirement for training data preparation where extensive TM images are supposed to be covered. Metric learning is naturally suitable for solving the problem of limited training shots [8]. A metric function of similarity can be learned to detect whether inputs are similar, instead of classifying the input samples.

Siamese neural networks (SNNs) are widely used in metric learning to extract pairs of input image features [9,10] and metric functions, e.g., Euclidean distance, Manhattan distance, and cosine similarity are used to compare embedded feature vectors [11,12]. Usually, contrastive loss is used in an SNN to minimize the distance between feature vectors of samples in the same classes and maximize the distance between samples in the different classes. There is a hyper-parameter in contrastive loss to control the threshold of distance. A data-driven triplet network was proposed on the basis of an SNN with an additional CNN branch [13]. The triplet loss function is used to decrease the feature vector distance between the anchor sample and the positive sample, and at the same time increase the distance between the anchor sample and the negative sample. The discrimination ability of the triplet network is improved while the training cost is greatly increased with the combinatorial explosion of the building of triplets, i.e., the input data of the triplet network. In this way, the pressure on training data quantity is transferred to the cost of existing annotated data mining by the more elaborate network architecture.

Another research idea in metric learning is loss-driven training methods such as recent works on face recognition (FR) [14,15]. Instead of building a large-scale dataset for training, these metric learning methods transformed the softmax function to conduct a margin penalty on the decision boundary, aiming to develop the discrimination ability of SNNs. The typical SphereFace [16], CosFace [17,18], and ArcFace [19] are all designed to expand intraclass space and reduce interclass space. Some of the reasons are that the performance of a data-driven model relies on the quality of information contained in training data excessively, and manual annotation is a major expenditure of human efforts. Furthermore, the SNN trained on close-set data shows bad performance in generalizing on open-set data [19].

More specifically, for the TMISD task, Setchi proposed a TM similarity analysis system to conduct the 'likelihood of confusion' test with three models [20]. Global and local shape feature descriptors, i.e., Zernike moment and an edge gradient co-occurrence matrix are used to extract TM image features. Euclidean distance is used to compute similarity. On this basis, Trappey introduced SNNs into the feature extraction of TM images [21]. VGG16 is used to build an SNN. Alshowaish used pre-trained CNNs to build an SNN including VGG16 and ResNet50 [22]. Most of these works focused on data-driven metric learning methods. However, the training database encounters a great challenge of annotation and covering the rapid growth in new TM designs.

We choose to research the TMISD task from the perspective of a loss-driven metric learning method. Here is a brief introduction to the frequently used loss function, i.e., the softmax function in the classification. The expression of softmax is as follows:

$$L_1 = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum\limits_{j=1}^{n} e^{W_j^T x_i + b_j}} \tag{1}$$

where N is the class number, W and b are weight and bias terms, x_i is the embedded feature vector belonging to the y_i-th class, and W_j is the j-th column of the weight W. Then, by fixing the weight $\|W\| = 1$ and feature $\|x_i\| = 1$ by ℓ_2 normalization, and by fixing the bias $b_j = 0$, the decision boundary is transferred to the angular space.

The transformed softmax function is as follows:

$$L_2 = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{\cos\theta_{y_i}}}{e^{\cos\theta_{y_i}} + \sum\limits_{j=1, j\neq y_i}^{n} e^{\cos\theta_j}} \tag{2}$$

where θ is the included angle between the normalized weight and feature vector. The prediction will depend only on the angle, and the decision boundary can be optimized by margin penalties.

The main contributions of this study are as follows:

(1) We researched the TMISD task with prevalent methods in metric learning including data-driven and loss-driven. The performance of these methods was investigated from several evaluation aspects regarding the TMISD task, including accuracy, F1 score, training cost, and generalization ability.

(2) According to the peculiarity of TM images, a hybrid-margin softmax (HMS) is proposed. Two additive margins are attached to the cosine term and the angular term of softmax, respectively, to expand the decision boundary in the angular space. The magnitudes of the weight and feature vector are preserved to retain the input information as much as possible. The metric function used to calculate the similarity is replaced by a classifier, i.e., a fully connected layer.

(3) Experiments indicate that the detection model penalized by HMS can be trained on small numbers of annotated data and reaches high detection accuracy with fewer layers of SNN. Furthermore, the HMS detection model trained completely on TM data generalizes well on the face recognition (FR) task, which indicates that the model trained on HMS has great input image discrimination ability.

2. Materials and Methods

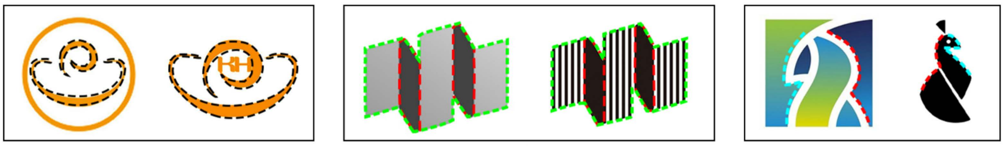
2.1. Hybrid-Margin Softmax

The peculiarity of TM images is crucial to the TMISD task. We compared the FR and the TMISD task to have a better view of the latter:

(1) The compositions of images in an FR task are constant. The principal parts of the input pairs of samples are human faces that always come from one exact person or different ones. The features extracted from the input are fixed generally, such as the shapes of faces, eyes, and noses. Plus, there are external interfering terms that should be considered including gestures, illuminations, ages, image noises, etc.

(2) The TMISD task is aimed at detecting the similarity of TM images. Generally, a TM design consists of a single element or several ones. The elements of the disputed TM image will not be identical to the legal one but partly similar in contours, colors, and textures, as shown in Figure 1. It is common for there to be both similar and different elements between two TM images in disputed cases. It should be noted that new outlines can be formed by the varying placements of elements. Furthermore,

interfering terms mentioned in the FR task are no longer to be considered, since TM images are artificially designed in most cases.



(a) (b) (c)

Figure 1. Some examples of TM images: (**a**) Similar pairs. Similar in element shapes and general color. (**b**) Similar pairs. Similar in contour, some elements. (**c**) Dissimilar pairs. Similar in partial contour, dissimilar in other factors.

To sum up, compared to the FR task, there are supposed to be more margin penalties on the decision boundary to tolerate a wide variety of element design changes in pairs of similar TM images. Meanwhile, the detection model should extract more information from input images to have a full understanding of the similarity degree and avoid false alarms. Therefore, the detection model should be further penalized, and the learnable parameters should be preserved as much as possible.

Given the characteristics of the TMISD task, a hybrid-margin softmax (HMS) is proposed as follows:

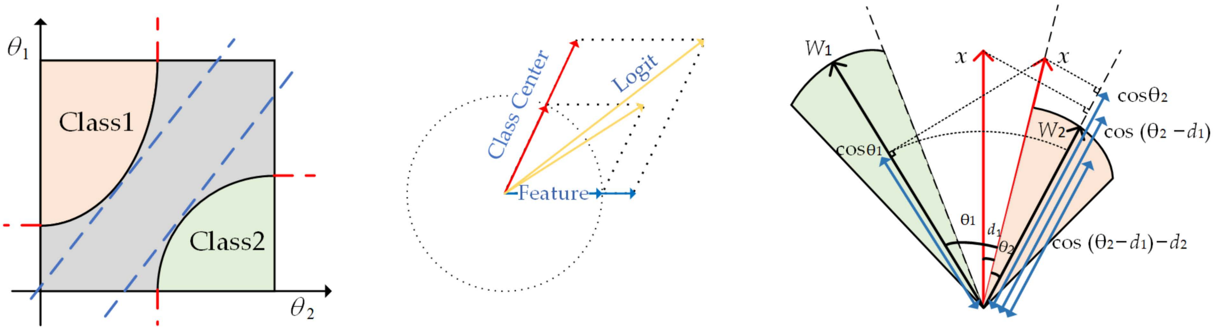
$$L_3 = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s\ \|W_{y_i}^T\|\ \|x_i\|\ (\cos(\theta_{y_i}-d_1)-d_2)}}{e^{s\ \|W_{y_i}^T\|\ \|x_i\|\ (\cos(\theta_{y_i}-d_1)-d_2)}+\sum_{j=1,j\neq y_i}^{n}e^{s\ \|W_j^T\|\ \|x_i\|\ \cos\theta_j}} \tag{3}$$

where s is a global scale factor. W_{y_i} are the weights of the fully connected layer, and x_i is the feature vector of i-th sample extracted by SNN. The weights and feature vectors are not normalized, and the biases are set to zero. Additive margin d_1 and d_2 are attached to the angle term and the cosine term, respectively.

The decision boundary of HMS loss is as follows:

$$\|W_1\|\cos\theta_1 = \|W_2\|(\cos(\theta_2-d_1)-d_2) \tag{4}$$

The decision boundary still can be considered as laying in the angular space with a varying amplitude of the cosine curve, as shown in Figure 2a.



(**a**) (**b**) (**c**)

Figure 2. Interpretations of HMS. (**a**) Decision boundary; (**b**) logit normalization; (**c**) penalties of HMS.

2.2. Interpretation of HMS

Several margins are attached to the inner product form of the logit to tolerate unpredictable little changes between similar elements of TM images. The magnitudes of extracted feature vectors, and the weight vectors, which are learnable parameters, are preserved to retain information from the input as much as possible. The model can benefit from this information when similar elements are contained in the pairs of dissimilar TM images. Plus, the normalization changes the magnitude and direction of the logit, as shown in Figure 2b. The SNN has to constantly adapt to these changes during the training.

Considering a group of TM images with an anchor sample, a similar one and a dissimilar one are given, suppose the class center of the anchor in the feature space is W_2, the class center of the dissimilar one is W_1, and the feature vector of a similar one is x, as shown in Figure 2c. The model can make the right prediction by the following calculation:

$$\|W_1\| \cos\theta_1 \leq \|W_2\|(\cos(\theta_2 - d_1) - d_2) \tag{5}$$

where $\theta_i (i = 1, 2)$ denote the angle between feature x and class center W_i.

The value of the cosine term is decreased by two margins d_1 and d_2. The model is penalized further to improve discrimination ability. The magnitude of the weight vector can be scaled for better prediction during model training.

There is a toy experiment, as shown in Figure 3, to describe the distribution of features extracted by the SNN trained on different transformations of the softmax function. These features are sent to the classifier to give a prediction. The SNN discrimination ability of TM images is described visually in this way. Red and blue spots are visualized features of two input TM images. Spots in the first row are from dissimilar TM images and spots in the second row are from similar TM images. In the first row, the first four feature spots are loose and chaotic. It is not solid enough for the classifier to judge they are not similar. The last feature spots in the first row extracted by the SNN trained on HMS are oriented intensively and separable in the meantime. The spot distributions in the second row also indicate that features learned from the SNN with HMS are compact and adequate for making a judgment.

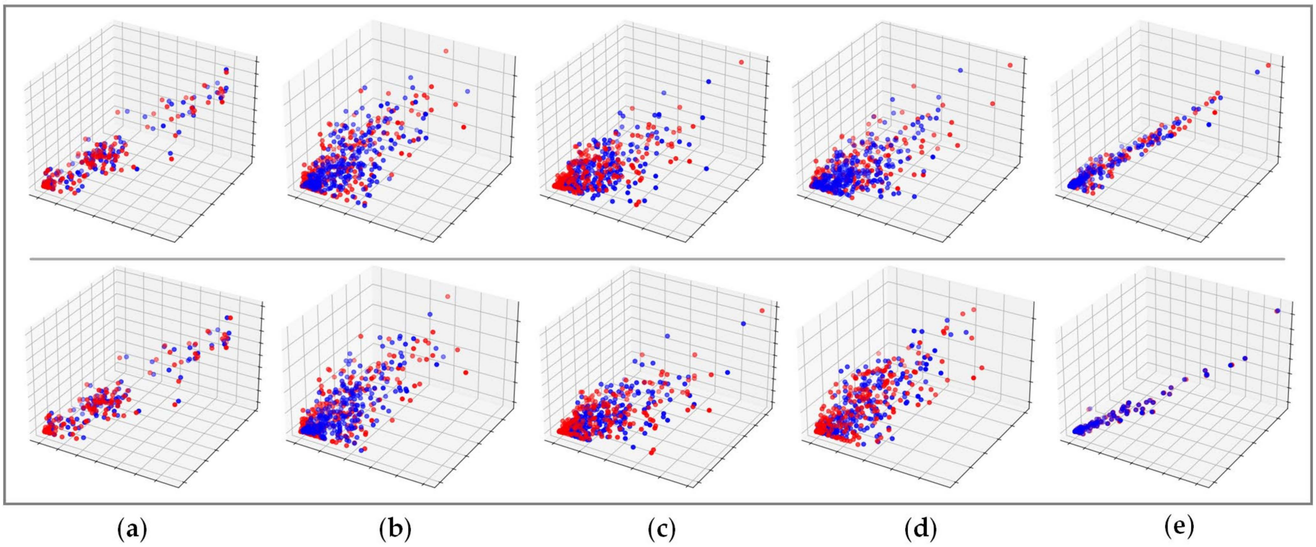


Figure 3. Visualized feature spots extracted by SNN trained on different transformations of softmax function. (**a**) SphereFace; (**b**) CosFace; (**c**) ArcFace; (**d**) HMS (normalized); (**e**) HMS.

3. Results

We conducted two branches of experiments. The comparison of loss-driven methods includes detection models based on an SNN trained on different transformations of softmax. The comparison of data-driven methods includes detection models based on the typical SNN (trained by contrastive loss), triplet network, and fine-tuning method.

3.1. Datasets

There were two types of image data involved in the experiments: TM images and human faces. The face data were used to train the feature extractor in the fine-tuning method and used for testing in the loss-driven methods. The TM image training data were compiled from real-world trademarks and annotated, consisting of 300 pairs of similar samples. The TM test data were collected from real court-disputed cases and cleaned manually, consisting of 1000 pairs of similar samples. The dissimilar TM image pairs were randomly selected and paired, consisting of equivalent numbers of similar pairs. The public LFW dataset was used as human face data, consisting of 3000 pairs of positive samples and 3000 pairs of negative samples [23,24].

The data preprocessing included input images cropped to a fixed-size shape and preserved colors. The TM training data were normalized with corresponding statistics. The TM test data and LFW data were normalized with mean value $\overline{X} = 0.5$ and variance $\sigma = 0.5$.

3.2. Experimental Setup

The details of the detection process are given in this section. The detection based on the loss-driven methods includes feature extracting and classifying.

The process of feature extracting is as follows: the backbone of the feature extractor is an SNN consisting of two identical CNNs that have the same structure and weight, as shown in Figure 4. Images input through the SNN can be encoded to vectors in the same feature space. Several CNN structures are implemented, including a simple self-defined six-layer CNN and resnet18, 34, 50, 101, and 152 [25]. The fully connected layer in resnet is removed, and the six-layer CNN has a similar structure to resnet, including a batchnorm (BN) layer and pooling layer, as shown in Figure 5. There is no residual module in the six-layer CNN.

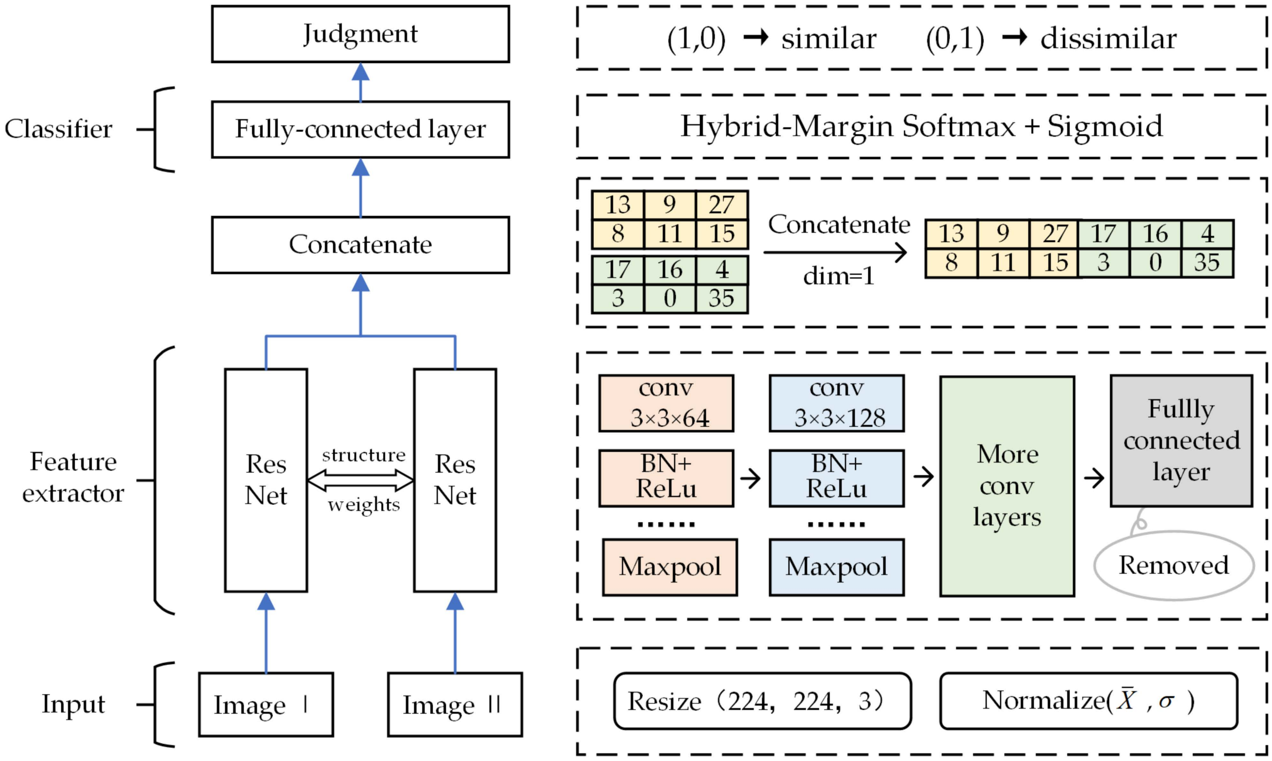


Figure 4. The procedure of TM image similarity detection.

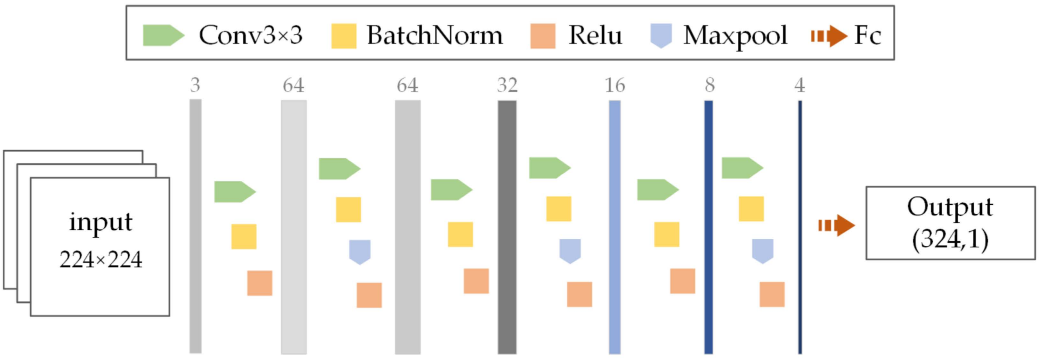


Figure 5. The structure of the 6-layer CNN.

The process of classifying is as follows: output vectors of the feature extractor are concatenated, activated by the transformation of softmax, and then sent to the fully connected layer, i.e., the classifier with the sigmoid activation function. Output judgment of similarity, the same as the input label, is one-hot encoded.

The detection processes of the data-driven methods are as follows:

(1) Fine-tuning method: An SNN to be transferred is trained on the LFW training dataset. The backbone of the SNN is composed of an original series of resnet. When the SNN reaches 95% or more accuracy on the LFW test dataset, the fully connected layer is removed and the rest of the weights are frozen. The trained and frozen SNN and a new fully connected layer compose the TM detection model. Then, the model is further trained on the TM training dataset and tested on the TM test dataset.

(2) Triplet model: Each input consists of two similar TM images, a dissimilar one, and corresponding labels. The triplet is built from the TM training dataset by attaching a random TM image to the pairs of similar samples.

The six-layer CNN is excluded from the data-driven methods since a shallow-depth CNN is not able to meet the demand of fitting in the triplet network model and fine-tuning method.

Other experiment setups are as follows:

The scale factor s in softmax was set to 90, the angular margin in SphereFace was 4, the additive margin of cosine term in CosFace was 0.006, the additive margin of angular term in ArcFace was 0.003, the additive margins of cosine term and angular term in HMS were 0.006 and 0.003, respectively. All experiments were conducted in the pytorch framework. Cuda was used to accelerate training. The learning rate was set to 0.001, the batch size was 16, the optimizer in the triplet network was Adam, and the optimizer in other methods was SGD (momentum was 0.9).

3.3. Loss-Driven Method Experiments

To prove that HMS enables the SNN to learn separable enough features of a TM image, we compared the detection models based on different transformations of softmax under the same experimental conditions. The accuracy and F1 score of detecting similar and dissimilar TM image pairs are shown in Tables 1 and 2. We also tested the detection model on the LFW dataset while the SNN was still trained with TM image data. The results are shown in Tables 3 and 4.

For the TMISD task, SphereFace achieves up to 96.39% accuracy, which outperforms CosFace and ArcFace greatly. However, when the depth of SNN increases, the model is overfitted severely. The accuracy of HMS regarding normalization of feature and weight vectors is slightly better than that of CosFace and ArcFace. HMS achieves the best performance, with up to 98.97% accuracy, which is a 2.58% improvement over SphereFace. Another notable thing is that a simple six-layer SNN trained on HMS works well on the TMISD task, with 97.45% accuracy and an F1 score of 0.9516.

Table 1. The accuracy of the TMISD task.

Transformations of Softmax	**Accuracy (%)**					
	6-Layer	**ResNet18**	**ResNet34**	**ResNet50**	**ResNet101**	**ResNet152**
SphereFace [16]	96.25	95.36	96.39	— *	—	—
CosFace [17,18]	43.81	51.55	52.06	52.58	55.15	58.76
ArcFace [19]	46.39	47.94	53.09	52.06	56.70	52.58
HMS (normalized)	54.12	53.61	54.12	57.73	52.06	53.09
HMS	97.45	97.42	97.94	**98.39**	98.97	98.52

Note: *—indicates that SNN is overfitted, same below.

Table 2. The F1 score of the TMISD task.

Transformations of Softmax	**F1 Score**					
	6-Layers	**ResNet18**	**ResNet34**	**ResNet50**	**ResNet101**	**ResNet152**
SphereFace [16]	0.9539	0.9516	0.9574	—	—	—
CosFace [17,18]	0.4171	0.4778	0.5131	0.5306	0.5915	0.5789
ArcFace [19]	0.4800	0.4294	0.5381	0.5373	0.6216	0.5534
HMS (normalized)	0.5189	0.5714	0.4671	0.6339	0.5373	0.5646
HMS	0.9516	0.9735	0.9798	**0.9749**	0.9746	0.9897

Table 3. The accuracy of the FR task.

Transformations of Softmax	**Accuracy (%)**					
	6-Layers	**ResNet18**	**ResNet34**	**ResNet50**	**ResNet101**	**ResNet152**
SphereFace [16]	—	—	—	—	—	—
CosFace [17,18]	45.63	47.25	46.97	47.38	47.82	46.38
ArcFace [19]	46.05	48.23	49.07	48.63	46.87	46.90
HMS (normalized)	50.13	50.50	49.06	50.55	48.30	51.18
HMS	80.45	**90.57**	82.45	82.30	84.17	82.90

Table 4. The F1 score of the FR task.

Transformations of Softmax	**F1 Score**					
	6-Layers	**ResNet18**	**ResNet34**	**ResNet50**	**ResNet101**	**ResNet152**
SphereFace [16]	—	—	—	—	—	—
CosFace [17,18]	0.3932	0.4468	0.4348	0.4411	0.5229	0.4389
ArcFace [19]	0.4564	0.4461	0.5789	0.5466	0.5173	0.4927
HMS (normalized)	0.3828	0.5380	0.3794	0.5524	0.3872	0.5978
HMS	0.8173	**0.9002**	0.8464	0.8449	0.8517	0.8483

For the FR task, the detection model trained on SphereFace with TM data is overfitted. The performance of HMS (normalized) is also better than that of CosFace and ArcFace. The model trained on HMS generalizes well on the LFW dataset with up to 90.57% accuracy and an F1 score of 0.9002. A simple six-layer SNN trained on HMS can reach 80.45% accuracy and an F1 score of 0.8173.

The performances of HMS with different depths of SNN were tested, as shown in Tables 1–4. ResNet18 was adequate for meeting the demand of the TMISD task and generalizing on the FR task. This also indicates that the SNN penalized by HMS is adequate to learn sufficient and critical information with fewer network layers. The training expenses are reduced as a result.

The detection accuracy of the model (resnet18) trained on HMS with different hyperparameters is shown in Figure 6. The accuracy fluctuation caused by scale factor s is higher than the margin d_1 and d_2.

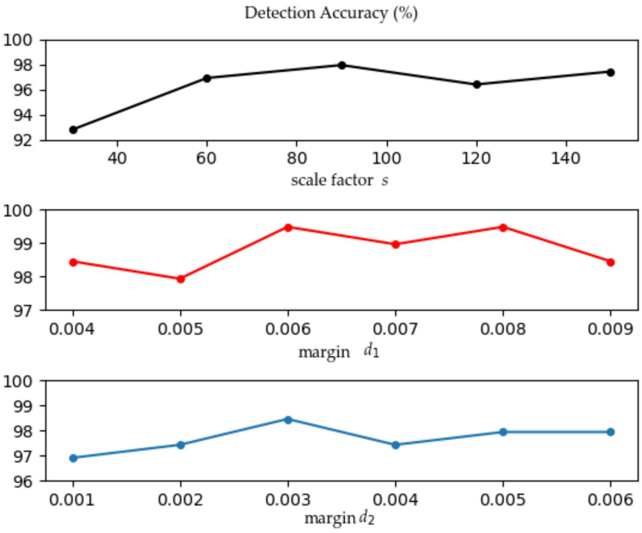


Figure 6. The accuracy of the detection model (resnet18) trained with different hyper-parameters in HMS.

3.4. Data-Driven Method Experiments

The performance of detection models based on the typical SNN, triplet network, and fine-tuning methods is shown in Table 5.

Table 5. The performance of data-driven methods on the TMISD task.

	SNN (Contrastive Loss) Method		Triplet Network Method		Fine-Tuning Method	
	Accuracy (%)	F1	Accuracy (%)	F1	Accuracy (%)	F1
ResNet18	41.53	0.4237	85.05	0.8449	53.61	0.5588
ResNet34	46.73	0.4535	**92.27**	**0.9282**	**70.65**	**0.7149**
ResNet50	47.18	0.4654	58.25	0.6897	56.19	0.5685
ResNet101	46.53	0.4549	54.13	0.6642	47.42	0.5049
ResNet152	46.84	0.4431	59.28	0.6802	46.39	0.4851

The triplet network achieved 92.27% accuracy with an F1 score of 0.9282 on the TMISD task, which is a significant improvement over a typical SNN. However, the performance gap came with greatly increased training costs in terms of memory and time. The performance of the fine-tuning method on the TMISD task was not satisfactory considering the training cost of the transferred knowledge. But when a model for a similar task is readily available, the fine-tuning method makes for a good choice with minimal training cost and fine performance.

For the detection models based on the triplet method and fine-tuning method, the performance changes rapidly with the depths of the network. These data-driven methods obtain a large gain in performance only under the condition that an adaptive depth CNN is employed for the task.

3.5. Discussion

Metric learning is an appropriate research idea for the TMISD task since the requirement for annotation data during the training is reduced. With the same numbers of similar TM pairs, the triplet network and fine-tuning data-driven methods can improve performance greatly compared to a simple SNN model. The triplet model enhances the discrimination ability with an additional input during training. The fine-tuning method transfers the learned information from other tasks and alleviates the pressure of data annotation.

The advantage of the detection models based on the data-driven methods is not prominent compared to the typical loss-driven models since the training is complicated and expensive. The performance gaps between the SphereFace model and the other two models, CosFace and ArcFace, are huge, but the performance cannot be sustained when SNN depth is increased or a new type of image is input for detection. The SphereFace model can be damaged by the diversity of TM images.

The HMS model outperformed other methods in the following aspects: (1) the compactness of similar TM pairs is tightened obviously; (2) the discrimination ability for another type of image, i.e., face data, is improved, which indicates the model trained on HMS is robust; (3) the training cost is reduced as a result of the requirements of annotated data and deep SNN depth being loosened.

In general, the introduction of the loss-driven model training idea is meaningful to the TMISD task. The challenges of training data-building and generalization on new data are dealt with in a low-cost way.

4. Conclusions

The detection of TM image similarity (TMISD) is an essential procedure for court judgments on TM infringement cases and TM legal registration, while the training data-building of similar TM pairs and model generalization on fast-growing numbers of new TM designs are huge challenges for the task. To address these issues, similarity detection models based on loss-driven metric learning methods were researched. Compared to data-driven methods, including the triplet network model and fine-tuning method, the optimization of the softmax loss function had a larger gain in performance, with less data preparation and training cost.

A hybrid-margin softmax (HMS) is proposed based on the peculiarity of TM images. Additive margins are attached to the cosine and angular term of softmax in the angular space to tolerate the slight differences between the similar parts of similar TM image pairs. The weights of the classifier and extracted feature vectors in the softmax are not normalized, aiming to best preserve the information of input images.

The detection model trained on HMS is further penalized to improve the discrimination ability of TM images. The model can be trained on small numbers of TM training data. Experiments indicate that the model trained on HMS achieves the best performance on the TMISD task with up to 98.97% accuracy and an F1 score of 0.9746, compared to other transformations of softmax. The model can also achieve high performance with fewer SNN layers. Furthermore, the HMS-driven model trained completely on TM image data generalized well on the FR task, with up to 90.57% accuracy and an F1 score of 0.9002.

Author Contributions: Methodology, C.W.; software, C.W.; validation, C.W. and H.S.; data curation, H.S.; writing—original draft preparation, C.W.; writing—review and editing, G.Z. and H.S.; visualization, C.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by National Natural Science Foundation of China (Grant: 62173222).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The TM dataset used in this research can be obtained from the following link: https://pan.baidu.com/s/11gIv9yj327xKCyq4v5TyeQ?pwd=tmid, accessed on 26 March 2024. If the link fails, you can contact the corresponding author for a new link.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Duch-Brown, N.; Martens, B.; Mueller-Langer, F. The Economics of Ownership, Access and Trade in Digital Data. *SSRN J.* **2017**. [CrossRef]
2. Johnson, S. Trademark Territoriality in Cyberspace: An Internet Framework for Common-Law Trademarks. *Berkeley Technol. Law J.* **2014**, *29*, 1253–1300.
3. Simon, D.A. The Confusion Trap: Rethinking Parody in Trademark Law. *Wash. Law Rev.* **2013**, *88*, 1021.

4. Besen, S.M.; Raskind, L.J. An Introduction to the Law and Economics of Intellectual Property. *J. Econ. Perspect.* **1991**, *5*, 3–27. [CrossRef]
5. Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
6. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: An Efficient Alternative to SIFT or SURF. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 2564–2571.
7. Sabry, E.S.; Elagooz, S.S.; El-Samie, F.E.A.; El-Bahnasawy, N.A.; El-Banby, G.M.; Ramadan, R.A. Evaluation of Feature Extraction Methods for Different Types of Images. *J. Opt.* **2023**, *52*, 716–741. [CrossRef]
8. Li, S.; Jin, J.; Li, D.; Wang, P. Research on Transductive Few-Shot Image Classification Methods Based on Metric Learning. In Proceedings of the 2023 7th International Conference on Communication and Information Systems (ICCIS), Virtual, 15–17 October 2023; pp. 146–150.
9. Bromley, J.; Bentz, J.W.; Bottou, L.; Guyon, I.; Lecun, Y.; Moore, C.; Säckinger, E.; Shah, R. Signature verification using a "siamese" time delay neural network. *Int. J. Patt. Recogn. Artif. Intell.* **1993**, *7*, 669–688. [CrossRef]
10. Koch, G.; Zemel, R.; Salakhutdinov, R. Siamese Neural Networks for One-Shot Image Recognition. In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 7–9 July 2015; Volume 37.
11. Melekhov, I.; Kannala, J.; Rahtu, E. Siamese Network Features for Image Matching. In Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 4–8 December 2016; pp. 378–383.
12. Nandy, A.; Haldar, S.; Banerjee, S.; Mitra, S. A Survey on Applications of Siamese Neural Networks in Computer Vision. In Proceedings of the 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 5–7 June 2020; pp. 1–5.
13. Hoffer, E.; Ailon, N. Deep Metric Learning Using Triplet Network. In *Similarity-Based Pattern Recognition, Proceedings of the Third International Workshop, SIMBAD 2015, Copenhagen, Denmark, 12–14 October 2015*; Springer: Berlin/Heidelberg, Germany, 2018.
14. Boutros, F.; Damer, N.; Kirchbuchner, F.; Kuijper, A. ElasticFace: Elastic Margin Loss for Deep Face Recognition. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), New Orleans, LA, USA, 19–20 June 2022; pp. 1577–1586.
15. Choi, J.; Kim, Y.; Lee, Y. Robust Face Recognition Based on an Angle-Aware Loss and Masked Autoencoder Pre-Training. In Proceedings of the ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Seoul, Republic of Korea, 14–19 April 2024; pp. 3210–3214.
16. Liu, W.; Wen, Y.; Yu, Z.; Li, M.; Raj, B.; Song, L. SphereFace: Deep Hypersphere Embedding for Face Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
17. Wang, H.; Wang, Y.; Zhou, Z.; Ji, X.; Gong, D.; Zhou, J.; Li, Z.; Liu, W. CosFace: Large Margin Cosine Loss for Deep Face Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
18. Wang, F.; Cheng, J.; Liu, W.; Liu, H. Additive Margin Softmax for Face Verification. *IEEE Signal Process. Lett.* **2018**, *25*, 926–930. [CrossRef]
19. Deng, J.; Guo, J.; Yang, J.; Xue, N.; Kotsia, I.; Zafeiriou, S. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 5962–5979. [CrossRef] [PubMed]
20. Setchi, R.; Anuar, F.M. Multi-Faceted Assessment of Trademark Similarity. *Expert Syst. Appl.* **2016**, *65*, 16–27. [CrossRef]
21. Trappey, C.V.; Trappey, A.J.C.; Lin, S.C.-C. Intelligent Trademark Similarity Analysis of Image, Spelling, and Phonetic Features Using Machine Learning Methodologies. *Adv. Eng. Inform.* **2020**, *45*, 101120. [CrossRef]
22. Alshowaish, H.; Al-Ohali, Y.; Al-Nafjan, A. Trademark Image Similarity Detection Using Convolutional Neural Network. *Appl. Sci.* **2022**, *12*, 1752. [CrossRef]
23. Huang, G.B.; Ramesh, M.; Berg, T.; Learned-Miller, E. *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments*; University of Massachusetts: Amherst, MA, USA, 2007.
24. Huang, G.B.; Learned-Miller, E. *Labeled Faces in the Wild: Updates and New Reporting Procedures*; University of Massachusetts: Amherst, MA, USA, 2014.
25. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.