2024 IEEE Conference on Artificial Intelligence (CAI)

Mining Contrastive Loss for Kinship Verification

Boxuan Hu, Yunhao Xu, Junlin Hu
School of Software, Beihang University, Beijing, China
hujunlin@buaa.edu.cn

***Abstract*—Facial kinship verification is a task that aims to recognize biological relationships between individuals based on facial images. Kinship verification is challenging because it requires identifying subtle similarities between relatives. The supervised contrastive loss applied to kinship verification generates a large number of redundant pair-wise samples. As an improvement, we propose a mining contrastive (MC) loss to enhance the discriminative ability of contrastive loss through a hard sample mining strategy which emphasizes the balance between positive and negative samples, improving overall verification accuracy. Compared with supervised contrastive loss, our proposed MC loss achieves better performance on FIW dataset.**

***Index Terms*—Kinship verification, contrastive loss, deep learning, sample mining**

I. Introduction

Face kinship verification [1], [2] is a challenging domain of computer vision. The focus of this task is to identify whether there is a family relationship between individuals based on pair-wise facial images, as shown in Figure 1. This task attracts a lot of attention due to its applications in areas such as locating missing individuals, mining social relationships, genetic studies, and behavioral analysis [3].

With the development of deep learning, convolutional neural networks become the main feature extraction techniques. Li et al. [4] introduced two network structures, employing star-shaped and hierarchical graphs respectively for relation inference. Huang et al. [5] proposed the adaptive weighted k-order triplet metric network, a method that combines local and global features by engaging in high-order feature interactions and synergizing multi-layer convolutional features. This approach effectively captures discriminative features and emphasizes challenging negative samples, thereby enhancing performance metrics.

In this paper, we propose a mining contrastive (MC) loss to optimize the feature extraction network. Our MC loss function combines hard sample mining strategy and weight balance between positive and negative samples, effectively improving the problem of supervised contrastive loss [6], [7] being unable to handle sample pair redundancy. The main contributions of this paper include: (1) we propose the mining contrastive loss and adopt a hard sample mining strategy and weight balance to overcome the limitations of supervised contrastive loss in handling sample pair redundancy; and (2) better results are achieved on FIW dataset in most relationships compared to the supervised contrastive loss.

This work was supported by the National Natural Science Foundation of China under Grant 62006013. *(Corresponding author: Junlin Hu.)*

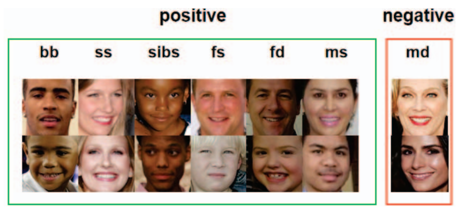

Fig. 1. The relations include bb (brothers), ss (sisters), sibs (siblings), fs (father-son), fd (father-daughter), ms (mother-son) and md (mother-daughter). If two samples have a kinship, the result is positive, otherwise it is negative.

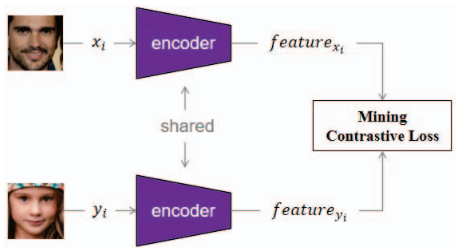

Fig. 2. The framework of our MC method.

II. Method

The overall framework of our MC method is illustrated in Figure 2. The framework includes two main components: neural network encoder and mining contrastive loss function.

A. Neural Network Encoder

The backbone model of the encoder f is ResNet18, with a batch size N. The inputs include N pairs kinship images $(\boldsymbol{x}_i, \boldsymbol{y}_i)$, and the outputs consist of N pairs of feature vectors $(f(\boldsymbol{x}_i), f(\boldsymbol{y}_i))$. The input images $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ have dimensions of $[112, 112, 3]$, and dimensions of the output feature vectors $f(\boldsymbol{x}_i)$ and $f(\boldsymbol{y}_i)$ are 512. All N pairs of input images are positive samples, indicating the presence of a kinship relationship between the two individuals.

B. Mining Contrastive Loss Function

When computing the loss, the N pairs feature vectors are combined into N positive pairs and $N^2 - N$ negative pairs. We use the supervised contrastive (SC) loss function [6] as baseline, which has shown good performance in the facial kinship verification task. Its formulation is expressed as follows:

$$L = \frac{1}{2N} \sum_{i=1}^{N} (S(f(\boldsymbol{x}_i), f(\boldsymbol{y}_i)) + S(f(\boldsymbol{y}_i), f(\boldsymbol{x}_i))), \tag{1}$$

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00167

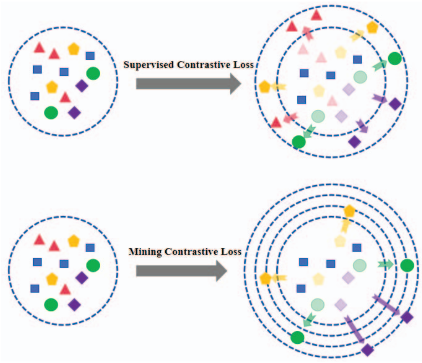

Fig. 3. Compared to supervised contrastive loss, mining contrastive loss can first filter out invalid samples (red triangles) in the new feature space.

where $S(.,.)$ is computed by:

$$S(\boldsymbol{z}_i, \boldsymbol{z}_j) = -\log \frac{e^{sim(\boldsymbol{z}_i, \boldsymbol{z}_j)/\tau}}{\sum_{k=1,k\neq i}^{2N} e^{sim(\boldsymbol{z}_i, \boldsymbol{z}_k)/\tau}}, \tag{2}$$

in which $sim(.,.)$ represents cosine similarity between two feature vectors, and τ is a free parameter.

Supervised contrastive loss (1) exists two issues [8]. Firstly, it adopts a uniform weight allocation strategy to allocate equal weights to all sample pairs, which ignores the difference in the weight of positive and negative sample pairs and fails to consider the actual contribution and importance of each pair. Secondly, it lacks an effective sample mining strategy and cannot eliminate the negative impact of invalid samples.

To address them, we propose a mining contrastive (MC) loss. Let $\mathbf{X} = (\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_N})^T$, $Y = (\boldsymbol{y_1}, \boldsymbol{y_2}, ..., \boldsymbol{y_N})^T$, and $\mathbf{M} = \mathbf{X}\mathbf{Y}^T$. For positive samples, we first conduct sample mining to filter samples that satisfy the condition: $\mathbf{M}_{pos} - \varepsilon < max(\mathbf{M}_{neg})$. Then the positive loss can be formulated as:

$$l_{pos} = \frac{1}{N}\frac{1}{\alpha} \log \left(1 + \sum_{i=1}^{N} e^{-\alpha(\mathbf{M}_{ii} - \lambda)} \right). \tag{3}$$

Similarly, for negative samples, samples that satisfy the criteria $\mathbf{M}_{neg} + \varepsilon > min(\mathbf{M}_{pos})$ are selected, and the negative loss is:

$$l_{neg} = \frac{1}{N}\frac{1}{\beta} \log \left(1 + \sum_{i=1}^{N} \sum_{j=1,j\neq i}^{N} e^{\beta(\mathbf{M}_{ij} - \lambda)} \right), \tag{4}$$

where ϵ, α, β, and λ are hyperparameters.

Finally, we formulate the loss function of our MC method over each training batch as:

$$L_{MC} = l_{pos} + l_{neg}. \tag{5}$$

Figure 3 depicts the basic ideas of supervised contrastive (SC) loss and our mining contrastive (MC) loss.

III. Experiments

Dataset. FIW dataset [9] comprises images from 1000 families, ensuring that each family member has at least one photograph. This dataset provides a comprehensive reflection of real-world facial image information. The dataset includes seven types of relationships: bb, ss, sibs, fs, fd, ms, and md. We select the first 2000 positive samples and the last 2000 negative samples for each relationship as the test set. Except for the test set, the first 200 positive samples and the last 200 negative samples in the remaining data are selected as the validation set. Since our model training only requires positive samples, all remaining positive samples are used as the training set. Before training, image preprocessing is performed, resizing the images to 112×112 pixels and applying random cropping and normalization. We set hyperparameters $\alpha = 0.2$, $\beta = 3.1$, $\lambda = 1.5$ and $\epsilon = 1$. The default batch size is 32. We measure the accuracy of our method using the ROC curve.

TABLE I
Accuracy (%) of supervised contrastive (SC) loss and mining contrastive (MC) loss on different kinship relations.

Method	bb	ss	sibs	fs	fd	ms	md	Mean
SC [6]	85.0	76.3	81.0	84.0	**80.3**	74.5	71.6	78.9
MC	**86.8**	**77.3**	**83.7**	**86.3**	79.5	**75.8**	**74.8**	**80.6**

Results. The accuracies for kinship verification in different relationships are shown in Table I. Our method achieves an average accuracy of 80.6%. Importantly, the MC loss outperforms supervised contrastive (SC) in the majority of relationships, demonstrating its effectiveness.

IV. Conclusion

We have proposed a mining contrastive loss to consider the mining of effective samples and the balance of positive and negative samples. Our method achieves better results than supervised contrastive loss on FIW dataset.

References

[1] R. Fang, K. D. Tang, N. Snavely, and T. Chen, "Towards computational models of kinship verification," in *IEEE International Conference on Image Processing*, 2010, pp. 1577–1580.
[2] J. Lu, J. Hu, X. Zhou, Y. Shang, Y. Tan, and G. Wang, "Neighborhood repulsed metric learning for kinship verification," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2594–2601.
[3] G. Kaminski, S. Dridi, C. Graff, and E. Gentaz, "Human ability to detect kinship in strangers' faces: effects of the degree of relatedness," *Proceedings of the Royal Society B: Biological Sciences*, vol. 276, no. 1670, pp. 3193–3200, 2009.
[4] W. Li, J. Lu, A. Wuerkaixi, J. Feng, and J. Zhou, "Reasoning graph networks for kinship verification: From star-shaped to hierarchical," *IEEE Transactions on Image Processing*, vol. 30, pp. 4947–4961, 2021.
[5] S. Huang, J. Lin, L. Huangfu, Y. Xing, J. Hu, and D. D. Zeng, "Adaptively weighted k-tuple metric network for kinship verification," *IEEE Transactions on Cybernetics*, vol. 53, no. 10, pp. 6173–6186, 2023.
[6] T. Chen, S. Kornblith, M. Norouzi, and G. E. Hinton, "A simple framework for contrastive learning of visual representations," in *International Conference on Machine Learning*, 2020, pp. 1597–1607.
[7] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan, "Supervised contrastive learning," 2021.
[8] X. Wang, X. Han, W. Huang, D. Dong, and M. R. Scott, "Multi-similarity loss with general pair weighting for deep metric learning," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5022–5030.
[9] J. P. Robinson, M. Shao, Y. Wu, and Y. Fu, "Families in the wild (FIW): large-scale kinship image database and benchmarks," in *ACM Conference on Multimedia*, 2016, pp. 242–246.