

Self-supervised Deep Learning Based End-to-End Face Verification Method using Siamese Network

Enoch Solomon
Department of Computer Science
Virginia State University
Richmond, Virginia
esolomon@vsu.edu

Abraham Woubie
Silo AI
Helsinki, Finland
Abraham.zewoudie@siloi.ai

Eyael Solomon Emiru
Department of Information
Engineering and Computer Science
University of Trento, Italy
eyael.emiru@studenti.unitn.it

Abstract—Achieving state-of-the-art results in face verification systems typically hinges on the availability of labeled face training data, a resource that often proves challenging to acquire in substantial quantities. In this research endeavor, we proposed employing Siamese networks for face recognition, eliminating the need for labeled face images. We achieve this by strategically leveraging negative samples alongside nearest neighbor counterparts, thereby establishing positive and negative pairs through an unsupervised methodology. The architectural framework adopts a VGG encoder, trained as a double branch siamese network. Our primary aim is to circumvent the necessity for labeled face image data, thus proposing the generation of training pairs in an entirely unsupervised manner. Positive training data are selected within a dataset based on their highest cosine similarity scores with a designated anchor, while negative training data are culled in a parallel fashion, though drawn from an alternate dataset. During training, the proposed siamese network conducts binary classification via cross-entropy loss. Subsequently, during the testing phase, we directly extract face verification scores from the network's output layer. Experimental results reveal that the proposed unsupervised system delivers a performance on par with a similar but fully supervised baseline.

Index Terms—biometrics, deep learning, face recognition, siamese network, unsupervised method

I. INTRODUCTION

In today's increasingly digital world, safeguarding personal information and ensuring security has become a top priority [1]–[10]. One of the most cutting-edge technologies addressing this concern is face verification, a powerful tool that leverages facial recognition to authenticate individuals [11]. This technology offers immense potential across various industries, from financial services to law enforcement, and even everyday consumer applications.

Deep learning methods are successful in image application specifically in face recognition [12]–[16]. These approaches excel in learning deep features and bottleneck features (BNF) [17], which are used within a conventional GMM-UBM framework [18]. However, these deep learning approaches typically rely on labeled training data.

The authors in [19] reported that PLDA is the efficient back-end for image recognition [19], [20]. Previous works reported PLDA provide a superior performance than cosine scoring. But, this improvement comes with the cost of labeled training data. However, previous works show that

it is not usually easy to have access to large amounts of labeled training data. Thus, the lack of labeled training data results in a significant performance gap between cosine and PLDA scoring techniques [21], [22] in face recognition. Although the authors in [23] proposed automatic labeling techniques, they could not appropriately estimate the true labels. Although these approaches perform reasonably well, the results are still compared to that PLDA that uses actual labels [24]. Whereas, the authors in [25] proposed autoencoder based unsupervised method improve the performance of face recognition systems. Previous studies show that these approaches mainly aim at increasing the discriminative power of face image embeddings. Thus, they can be applied as a back-end in face verification task.

In our proposed work, the main goal is to reduce the reliance on labeled training data for a face verification system. We aim to obtain end-to-end face verification scores without using face image labels. We propose a siamese network [26] consisting of double-branch networks, each with two branches that are CNN encoders. These encoders are inspired by the VGG architecture and adapted for face verification [27]. Traditionally, Siamese networks are trained using pairs of training samples, such as anchor-positive and anchor-negative pairs. However, to avoid using face image labels, we generate these training sample pairs in an unsupervised manner. Positive training data are chosen within one dataset based on their highest cosine scores with the anchor sample. In contrast, negative training data are selected in a similar manner from another dataset, ensuring that the two datasets do not include face images from the same identity.

In our study, we introduced the utilization of a two branch Siamese network. The network processes positive and negative samples, individually paired with anchor samples, and computes the binary cross-entropy loss based on their binary labels. Following the training process, decision scores for face verification trials are derived from the last layer of the network, resulting in an end-to-end face verification system. Our evaluation was performed on the Labeled Faces in the Wild (LFW) dataset [28]. Despite being unsupervised, our results show significant promise, approaching the performance of a fully supervised baseline.

The remainder of this paper is organized as follows: Section 2 explains the proposed method, Section 3 provides details on the experimental setup and data, Section 4 discusses the results, and we conclude in Section 5.

II. PROPOSED METHOD

The training of a DNN in an end-to-end fashion typically requires access to labeled training data, which can be challenging in real-world scenarios. To address this, especially when labeled data is unavailable, we propose the use of siamese networks [26], characterized by a double-branch architecture. Each branch features a CNN encoder that is motivated by the VGG network [29], which has recently shown promise in face verification [30]. Given our objective of avoiding reliance on face image labels, we introduce a novel approach to generate pairs of training data in an unsupervised manner [31], [32]. Specifically, positive training data are chosen within one dataset based on their similarity score with the anchor. Similarly, negative training data are selected using the same criterion, but from a separate dataset, ensuring that the two datasets do not contain face images from the same identity. This innovative training methodology circumvents the need for labeled data, making the process more accessible and adaptable to real-world scenarios.

The proposed work utilizes a siamese network. It is trained by minimizing binary cross-entropy loss. This network selects the pair of negative-anchor and positive-anchor face images samples, and processes them to generate binary labels of 1 (positive match) or 0 (negative match) at the output. These labels are used to compute the loss during training. Once the network is trained, we can obtain decision scores for face recognition evaluations directly from the network's output. In summary, our proposed siamese network constitutes an end-to-end face recognition system, capable of making binary classification decisions with respect to facial similarity.

A. Training Pairs Selection

The procedure for selecting negative and positive images involves utilizing two datasets, denoted as X and Y . Let C_X and C_Y represent the individuals in datasets X and Y , respectively. It's assumed that individuals in dataset X are distinct from those in dataset Y , i.e., $C_X \cap C_Y = \emptyset$. Algorithm 1 provides an overview of how the selection of negative and positive face image vectors is conducted in an unsupervised manner.

To begin, we extract the face embedding vectors for all face images in datasets X and Y . Next, we compute similarity scores among all face embedding vectors within dataset X by using cosine distance metrics. For each face embedding vector in X , we select a fixed number k of neighbor face embedding as a potential positive face embeddings. Subsequently, we introduce a threshold to select potential k positives face images.

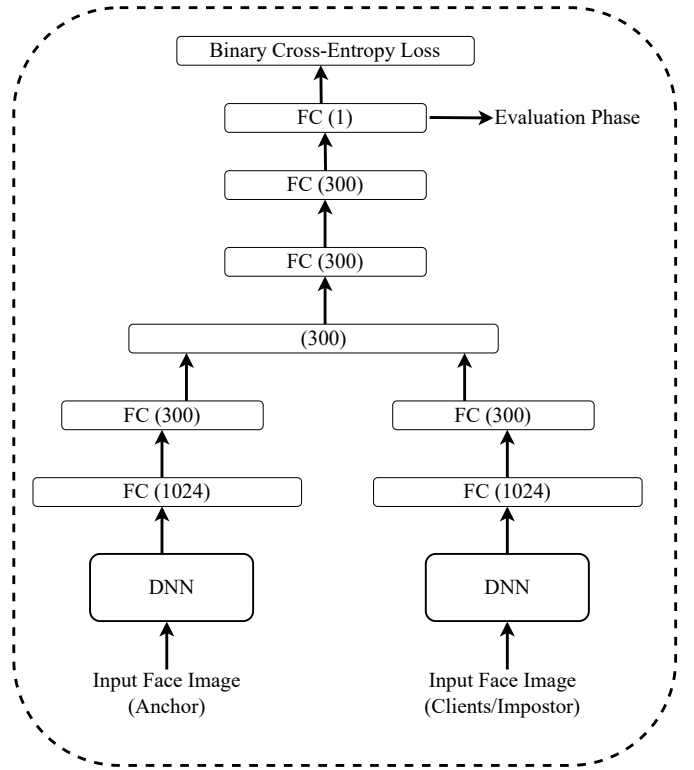


Fig. 1: The architecture of the proposed siamese network. FC denotes fully connected layer. In evaluation phase, decision scores for face recognition are obtained directly from the last layer of the network. This layer computes the final representation of the features extracted from the input face images, which is then used to make binary classification decisions regarding facial similarity.

Algorithm 1 The proposed algorithm to select positive and negative face images for each face image in the training dataset.

Require: Training face images t_i , $1 < i < n$, $x_i \in X$ & $y_i \in Y$, $1 \leq i \leq n$, and threshold

Ensure: Positive and Negative face images P_{ij} and N_{ij} , $1 \leq i \leq m$ and $1 \leq j \leq k$

for each training face image x_i **do**

for each training face image x_p , $1 < p < m$ **do**

 Compute Positive Scores $i,p = \text{cosine}(x_i, x_p)$

end for

 From Positive Scores i,p , select k highest face images.

if Positive Scores $i,p < \text{threshold}$

$P_{i,j} = x_p$

for each training face image y_m , $1 < n < m$ **do**

 Compute Negative Scores $i,n = \text{cosine}(y_i, y_n)$

end for

 From Negative Scores i,n , select k highest face images.

if Negative Scores $i,n < \text{threshold}$

$N_{i,j} = y_n$

end for

To select the negative face image embeddings, we score all the face embeddings in X with those in Y , by using cosine similarity scores. For each face embedding in X , we select k number of face embeddings from Y that by using their cosine distance scores. Given that the person in dataset X does not appear in dataset Y , the k selected face embeddings are considered potential negative face images. Subsequently, we apply a threshold to their corresponding cosine scores to identify the most challenging negative face images. This process ensures that each face embedding in X is paired with k most positive and k negative face image embeddings.

Assuming there are n face images in each dataset X and Y , then we will have a total of n plus n times k face image training data. For example, if we have 100 training data in each datasets with $k=10$, then we will have $(100+100)*10 = 1000$ training data. This way, we were able to increase intrinsically the number training dataset. In our experiments, training pairs consist of two samples, namely anchor-positive and anchor-negative.

To select the negative face embedding vectors, we compute cosine similarity scores between all face embedding vectors in set X and those in set Y . For each face embedding vector in X , we choose K face embedding vectors from Y that have the closest scores. Given that individuals in X do not appear in Y , these K selected face embeddings represent potential negative face images. Subsequently, we apply a threshold to the corresponding cosine scores to identify the most challenging negatives face images among them. As a result, each face embedding vector in X is assigned K positive and K negative face embedding vectors.

B. Proposed Siamese Network

The training of a DNN in an end-to-end manner usually requires labeled face images for the training data, which can be challenging to obtain in practical scenarios. When such labels are unavailable, we propose the use of siamese networks. These networks operate in a fully unsupervised manner, distinguishing them from a typical DNN classifiers. Traditionally, a siamese network is trained using pairs of samples, consisting of an anchor, a positive, and negative. To bypass the need for face image labels, we propose generating these training pairs in an unsupervised manner, which are then fed into the network. Specifically, we advocate for a siamese network utilizing binary cross-entropy loss. Following the training process, we extract decision scores from the output of the proposed network for the evaluation. This approach allows us to perform face verification in an unsupervised manner.

C. End-to-End Face Verification System

Figure 1 depicts the architecture of the proposed siamese network. This network comprises two identical branches, both acting as CNN encoders. During training, a pair consisting of an anchor face image along with either a positive or negative face image is inputted into each branch

of the network respectively. These two branches share the same parameters. Following the CNN encoder, the encoded vectors output from the two branches are concatenated. Then, the concatenated vector is fed into the FC layers. The last layer is a binary class layer. It will be one for the case of anchor and positive face images, whereas it will be zero for the case of anchor and negative face images. Throughout the training process, the network minimizes binary cross-entropy loss.

After training with selected positive and negative samples in an unsupervised manner, evaluations are conducted in an end-to-end way. During evaluation, pairs of test face images are inputted into the network, and decision scores are directly obtained from the last layer. This approach establishes an unsupervised end-to-end face verification system using the proposed Siamese network.

D. DNN Encoder

The DNN encoder block draws inspiration from the VGG architecture [33], which has been recently tailored for face verification in [27]. This encoder is structured with three primary blocks, with each block comprising two convolutional layers and one max-pooling layer. Following these blocks, there are two Fully Connected (FC) layers, having 1024 and 300 neurons, respectively. The function of the DNN encoder is to encode the input face image, reducing it to a 300-dimensional vector.

III. EXPERIMENTAL SETUP AND DATASET

A. Experimental Setup

The training phase utilized the CelebA dataset [34], while the evaluation was conducted on the Labeled Faces in the Wild (LFW) dataset [28]. The performance metrics assessed were EER and accuracy.

During the selection of positives and negatives, we divided the validation partition of the CelebA dataset into two equal parts, creating datasets denoted as A and B, as explained previously. The CNN encoder was identical for both networks. The threshold values for the selection of positive and negative samples were 0.3 and 0.1, respectively. The activation function used for both the DNN and fully connected layers was ReLU, while the final layer of the network employed sigmoid activation function. The network is trained for 300 epochs or until the error is no longer decreasing using a batch size of 64 images. It uses stochastic gradient descent, momentum of 0.91, weight decay of 0.00001 and a logarithmically decaying learning rate from 10^{-2} to 10^{-8} .

The baseline system underwent training utilizing the entire validation partition of the CelebA dataset. Its architecture entails a DNN encoder followed by a classification layer. To ensure an equitable comparison, the architecture of the CNN encoder was precisely replicated from that of the siamese networks. The training process involved minimizing cross-entropy loss, employing the Adam optimizer with identical parameters as those in the siamese networks.

TABLE I: It presents a comparison of the proposed method with different values of k and the baseline methods. The evaluation metrics used is Equal Error Rate (EER). The results demonstrate the performance of the proposed method across various k values, showcasing its effectiveness compared to the baseline methods.

Model	k	EER
Baseline	-	7.93
Proposed	2	7.91
Proposed	5	7.83
Proposed	10	6.90
Proposed	12	7.01
Proposed	15	8.02

During evaluation, face image embeddings were extracted from the CNN encoder of this network.

B. Datasets

The CelebA dataset [34], initially introduced by Liu et al., comprises a vast collection of over 200,000 face images showcasing 10,177 celebrities. This dataset incorporates various elements like pose variations and background clutter, offering a comprehensive representation of real-world scenarios. To facilitate model training and evaluation, the dataset is divided into distinct sets for training, validation, and testing purposes. In this work, we utilized the training, validation, and test segments of the CelebA dataset as our unlabeled dataset for training the autoencoder.

The Labeled Faces in the Wild dataset (LFW) [28] is a compilation of 13,233 images, featuring 5,749 distinct individuals. For testing, the dataset is randomly and uniformly divided into ten subsets. Each subset consists of 300 matched pairs (depicting the same person) and 300 mismatched pairs (depicting different individuals), which results in a total of 3,000 matched pairs and 3,000 mismatched pairs used for testing [28].

IV. RESULTS

A. Results on LFW

In Table I, we present a comprehensive comparison between our proposed systems and the baseline, evaluating them based on the Equal Error Rate (EER) expressed as a percentage. The table encompasses results for different values of k . Throughout the experimental trials, face recognition scores are extracted from the output of the proposed network. The findings reveal that increasing the value of k leads to enhanced system performance. The optimal EER of was achieved when k was set to 10. While higher values of k were explored, no significant improvement has been observed, suggesting a balance between performance and computational cost. This underscores the promising potential of the proposed approach, particularly in unsupervised scenarios, demonstrating its effectiveness in face verification tasks without the reliance on labeled training data.

TABLE II: Comparison of the proposed method result with some of state-of-the-art methods on LFW testing dataset.

Model	Training size	Labeled/Unlabeled	Testing size	Accuracy(%)
UniformFace	6.1M	Labeled	6K	99.80 [35]
ArcFace	5.8M	Labeled	6K	99.82 [36]
GroupFace	5.8M	labeled	6K	99.85 [37]
CosFace	5M	Labeled	6K	99.73 [22]
Marginal Loss	4M	Labeled	6K	99.48 [38]
CurricularFace	3.8M	Labeled	6K	99.80 [39]
RegularFace	3.1M	Labeled	6K	99.61 [40]
MDCNN	1M	Labeled	6K	99.38 [41]
PSO AlexNet TL	14M	Labeled	6K	99.57 [42]
Ben Face	0.5M	Labeled	6K	99.20 [43]
F ² C	5.8M	Labeled	6K	99.83 [44]
PCCycleGAN	0.5M	Unlabeled	6K	99.52 [45]
CAPG GAN	1M	Unlabeled	6K	99.37 [46]
UFace	200K	Unlabeled	6K	99.40 [31]
Proposed	200K	Unlabeled	6K	99.67

Table II provides a comprehensive comparison between our proposed method and state-of-the-art approaches. It's important to note that our method is compared with both supervised and unsupervised training techniques. Notably, our proposed training method demonstrates competitive performance without the need for an extensive labeled dataset. While some methods like ArcFace, GroupFace, Marginal Loss, and CosFace exhibit slightly higher accuracy, it's worth highlighting that our approach is trained on a significantly smaller dataset (approximately 200K images), whereas most state-of-the-art methods rely on millions of training images. This showcases the efficiency and effectiveness of our proposed method, especially in scenarios where labeled data is limited.

V. CONCLUSION

In this work, we introduced a siamese network for face verification that operates without the reliance on labeled data. This network features two branches, each functioning as a CNN encoder, and was trained as a binary classifier. Given our objective of avoiding the need for face image labels, the training sample pairs were generated using an unsupervised approach. Positive training images were selected within a single dataset based on the highest cosine scores with a designated anchor, while negative training images followed the same selection criteria but were drawn from a different dataset.

Following the training process, decision scores were obtained using the proposed network, and the evaluation was conducted using the LFW dataset. Notably, our experimental results demonstrated that our proposed system, despite being unsupervised, achieved results that closely paralleled those of fully supervised baselines. This highlights the effectiveness of our approach in face verification, even in the absence of labeled data.

REFERENCES

- [1] Kumar, J. & Patel, D. A survey on internet of things: Security and privacy issues. *International Journal Of Computer Applications*. **90** (2014)
- [2] Solomon, E. & Cios, K. FASS: Face Anti-Spoofing System Using Image Quality Features and Deep Learning. *Electronics*. **12**, 2199 (2023)
- [3] Solomon, E. & Cios, K. HDLHC: Hybrid Face Anti-Spoofing Method Concatenating Deep Learning and Hand-Crafted Features. *2023 IEEE 6th International Conference On Electronic Information And Communication Technology (ICEICT)*. pp. 470-474 (2023)
- [4] Chekole, E., Chattopadhyay, S., Ochoa, M., Guo, H. & Cheramangalath, U. CIMA: Compiler-Enforced Resilience Against Memory Safety Attacks in Cyber-Physical Systems. *Computers Security*. **94** pp. 101832 (2020), <https://www.sciencedirect.com/science/article/pii/S0167404820301061>
- [5] Chekole, E., Ochoa, M. & Chattopadhyay, S. SCOPE: Secure Compiling of PLCs in Cyber-Physical Systems. *International Journal Of Critical Infrastructure Protection*. **33** pp. 100431 (2021), <https://www.sciencedirect.com/science/article/pii/S1874548221000238>
- [6] Chekole, E. & Huaqun, G. ICS-SEA: Formally Modeling the Conflicting Design Constraints in ICS. *Proceedings Of The Fifth Annual Industrial Control System Security (ICSS) Workshop*. pp. 60-69 (2019), <https://doi.org/10.1145/3372318.3372325>
- [7] Chekole, E., Castellanos, J., Ochoa, M. & Yau, D. Enforcing Memory Safety in Cyber-Physical Systems. *Computer Security*. pp. 127-144 (2018)
- [8] Chekole, E., Chattopadhyay, S., Ochoa, M. & Huaqun, G. Enforcing Full-Stack Memory-Safety in Cyber-Physical Systems. *Engineering Secure Software And Systems*. pp. 9-26 (2018)
- [9] Chekole, E., Thulasiraman, R. & Zhou, J. EARIC: Exploiting ADC Registers in IoT and Control Systems. *Applied Cryptography And Network Security Workshops*. pp. 245-265 (2023)
- [10] Wong, A., Chekole, E., Ochoa, M. & Zhou, J. On the Security of Containers: Threat Modeling, Attack Analysis, and Mitigation Strategies. *Computers Security*. **128** pp. 103140 (2023), <https://www.sciencedirect.com/science/article/pii/S0167404823000500>
- [11] Kumar, N., Berg, A., Belhumeur, P. & Nayar, S. Attribute and simile classifiers for face verification. *2009 IEEE 12th International Conference On Computer Vision*. pp. 365-372 (2009)
- [12] Serengil, S. & Ozpinar, A. Lightface: A hybrid deep face recognition framework. *2020 Innovations In Intelligent Systems And Applications Conference (ASYU)*. pp. 1-5 (2020)
- [13] Abraham Woubie and Tom Bäckström: Federated learning for privacy-preserving speaker recognition, In *IEEE Access*, 2021; pp. 149477–149485.
- [14] Abraham Woubie and Tom Bäckström: Voice-quality Features for Deep Neural Network Based Speaker Verification Systems, In *29th European Signal Processing Conference (EUSIPCO)*, 2021; pp. 176–180.
- [15] Abraham Woubie and Tom Bäckström: Federated Learning for Privacy Preserving On-Device Speaker Recognition, In *ISCA Symposium on Security and Privacy in Speech Communication*, 2021; pp. 1–5.
- [16] Abraham Woubie, Pablo Zarazaga and Tom Bäckström: The Use of Audio Fingerprints for Authentication of Speakers on Speech Operated Interfaces, In *ISCA Symposium on Security and Privacy in Speech Communication*, 2021; pp. 6–9.
- [17] Prasad, P., Pathak, R., Gunjan, V. & Ramana Rao, H. Deep learning based representation for face recognition. *ICCCE 2019: Proceedings Of The 2nd International Conference On Communications And Cyber Physical Engineering*. pp. 419-424 (2020)
- [18] Mallouh, A., Qawaqneh, Z. & Barkana, B. New transformed features generated by deep bottleneck extractor and a GMM-UBM classifier for speaker age and gender classification. *Neural Computing And Applications*. **30** pp. 2581-2593 (2018)
- [19] Ioffe, S. Probabilistic linear discriminant analysis. *Computer Vision-ECCV 2006: 9th European Conference On Computer Vision, Graz, Austria, May 7-13, 2006, Proceedings, Part IV 9*. pp. 531-542 (2006)
- [20] El Shafey, L., McCool, C., Wallace, R. & Marcel, S. A scalable formulation of probabilistic linear discriminant analysis: Applied to face recognition. *IEEE Transactions On Pattern Analysis And Machine Intelligence*. **35**, 1788-1794 (2013)
- [21] Hafed, Z. & Levine, M. Face recognition using the discrete cosine transform. *International Journal Of Computer Vision*. **43** pp. 167-188 (2001)
- [22] 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. *Proceedings Of The IEEE Conference On Computer Vision And Pattern Recognition*. pp. 5265-5274 (2018)
- [23] Gao, Y., Ma, J. & Yuille, A. Semi-supervised sparse representation based classification for face recognition with insufficient labeled samples. *IEEE Transactions On Image Processing*. **26**, 2545-2560 (2017)
- [24] Muslihah, I. & Muqorobin, M. Texture characteristic of local binary pattern on face recognition with probabilistic linear discriminant analysis. *International Journal Of Computer And Information System (IJCIS)*. **1**, 22-26 (2020)
- [25] Hammouche, R., Attia, A., Akhrouf, S. & Akhtar, Z. Gabor filter bank with deep autoencoder based face recognition system. *Expert Systems With Applications*. **197** pp. 116743 (2022)
- [26] Chen, X. & He, K. Exploring simple siamese representation learning. *Proceedings Of The IEEE/CVF Conference On Computer Vision And Pattern Recognition*. pp. 15750-15758 (2021)
- [27] Cheng, S. & Zhou, G. Facial expression recognition method based on improved VGG convolutional neural network. *International Journal Of Pattern Recognition And Artificial Intelligence*. **34**, 2056003 (2020)
- [28] Huang, G., Mattar, M., Berg, T. & Learned-Miller, E. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. *Workshop On Faces InReal-LifeImages: Detection, Alignment, And Recognition*. (2008)
- [29] Sengupta, A., Ye, Y., Wang, R., Liu, C. & Roy, K. Going deeper in spiking neural networks: VGG and residual architectures. *Frontiers In Neuroscience*. **13** pp. 95 (2019)
- [30] Gwyn, T., Roy, K. & Atay, M. Face recognition using popular deep net architectures: A brief comparative study. *Future Internet*. **13**, 164 (2021)
- [31] Solomon, E., Woubie, A. & Cios, K. UFace: An Unsupervised Deep Learning Face Verification System. *Electronics*. **11**, 3909 (2022)
- [32] Solomon, E. Face Anti-Spoofing and Deep Learning Based Unsupervised Image Recognition Systems. (2023)
- [33] Chang, X., Wu, J., Yang, T. & Feng, G. Deepfake face image detection based on improved VGG convolutional neural network. *2020 39th Chinese Control Conference (CCC)*. pp. 7252-7256 (2020)
- [34] Liu, Z., Luo, P., Wang, X. & Tang, X. Large-scale celebfaces attributes (celeba) dataset. *Retrieved August*. **15**, 11 (2018)
- [35] Duan, Y.; Lu, J.; Zhou, J. Uniformface: Learning deep equidistributed representation for face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, 15-20 June 2019; pp. 3415–3424.
- [36] Deng, J.; Guo, J.; Xue, N.; Zafeiriou, S. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, 15-20 June 2019; pp. 4690–4699.
- [37] Kim, Y.; Park, W.; Roh, M.; Shin, J. Groupface: Learning latent groups and constructing group-based representations for face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 13-19 June 2020; pp. 5621–5630.
- [38] Deng, J.; Zhou, Y.; Zafeiriou, S. Marginal loss for deep face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Honolulu, HI, USA, 21-26 July 2017; pp. 60–68.
- [39] Huang, Y.; Wang, Y.; Tai, Y.; Liu, X.; Shen, P.; Li, S.; Li, J.; Huang, F. Curricularface: Adaptive curriculum learning loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 13-19 June 2020; pp. 5901–5910.
- [40] Zhao, K.; Xu, J.; Cheng, M. Regularface: Deep face recognition via exclusive regularization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, 15-20 June 2019; pp. 1136–1144.
- [41] Huang, X.; Zeng, X.; Wu, Q.; Lu, Y.; Huang, X.; Zheng, H. Face Verification Based on Deep Learning for Person Tracking in Hazardous Goods Factories. *Processes* **2022**, *10*, 380.
- [42] Elagougne, H.; Belahcene, M.; Bourennane, S. Hybrid descriptor and optimized CNN with transfer learning for face recognition. *Multimed. Tools Appl.* **2022**, *81*, 9403–9427.

- [43] Ben Fredj, H.; Bouguezzi, S.; Souani, C. Face recognition in unconstrained environment with CNN. *Vis. Comput.* **2021**, *37*, 217–226.
- [44] Wang, K.; Wang, S.; Zhang, P.; Zhou, Z.; Zhu, Z.; Wang, X.; Peng, X.; Sun, B.; Li, H.; You, Y. An efficient training approach for very large scale face recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, Louisiana, USA, 21-24 June 2022; pp. 4083–4092.
- [45] Liu, Y.; Chen, J. Unsupervised face frontalization for pose-invariant face recognition. *Image Vis. Comput.* **2021**, *106*, 104093.
- [46] Hu, Y.; Wu, X.; Yu, B.; He, R.; Sun, Z. Pose-guided photorealistic face rotation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18-23 June 2018; pp. 8398–8406.