International Journal of Creative and Open Research in Engineering and Management
ISSN: **3108-1754 (Online)**
Volume 02 Issue 04 April-2026 | Impact Factor: 3.5

Deep Learning for Facial Recognition and Detection

VIBHOR

Department of Computer Science Jagan Institute of Management Studies, Rohini

Delhi, India

How to Cite this Article:

VIBHOR, (2026). Deep Learning for Facial Recognition and Detection. International Journal of Creative and Open Research in Engineering and Management, <i>02</i>(04). https://doi.org/10.55041/ijcope.v2i4.443

https://doi.org/10.55041/ijcope.v2i4.443

ABSTRACT

Face recognition is one of the most active and challenging research areas in computer vision and biometric authentication. This paper presents a comprehensive face recognition system built upon machine learning and deep learning techniques, targeting accurate, real-time identification of human faces under unconstrained environments. The proposed system employs a multi-stage pipeline comprising image acquisition, preprocessing, feature extraction using Convolutional Neural Networks (CNNs), and classification using Support Vector Machines (SVMs) and deep neural architectures.

The methodology integrates advanced preprocessing techniques including histogram equalization, facial landmark detection, and geometric normalization to improve robustness across varying illumination, pose, and occlusion conditions. The model is trained and evaluated on benchmark datasets including LFW (Labeled Faces in the Wild), AT&T ORL, and CelebA datasets. Experimental results demonstrate a recognition accuracy of 97.8% on the LFW dataset and 98.9% on the AT&T ORL dataset, with a false acceptance rate (FAR) of 0.7% and a false rejection rate (FRR) of 1.2%, indicating competitive performance against state-of-the-art systems.

The paper discusses model architecture decisions, challenges encountered during training, data augmentation strategies, and comparative analysis against traditional methods such as Eigenfaces and Fisherfaces. The findings demonstrate that CNN-based deep learning models significantly outperform classical feature-based approaches. Future directions including federated learning for privacy-preserving face recognition are discussed.

Keywords: Face Recognition, Convolutional Neural Network, Support Vector Machine, Biometric Authentication, Deep Learning, Computer Vision, Feature Extraction, LFW Dataset.

1. INTRODUCTION

Face recognition technology has emerged as a critical component in modern security infrastructure, human- computer interaction systems, and social media platforms. Unlike traditional authentication methods such as passwords or PINs, biometric face recognition offers a non-intrusive, contactless, and highly user- friendly identification mechanism that is difficult to replicate or steal. The proliferation of surveillance systems, smartphones with facial unlock features, and digital identity verification services has created an enormous demand for accurate, fast, and secure face recognition solutions.

The fundamental goal of a face recognition system is to identify or verify a person from a digital image or video frame by comparing captured facial features against a pre-existing database. Despite decades of research, this problem remains challenging due to several factors including intra-class variability (changes in illumination, pose, expression, aging, occlusion) and inter-class similarity (faces of twins or look-alikes). Traditional handcrafted feature-based methods such as Local Binary Patterns (LBP), Eigenfaces (PCA), and Fisherfaces (LDA) have shown limited generalization in real-world unconstrained environments.

The advent of deep learning, particularly Convolutional Neural Networks (CNNs), has revolutionized the field of face recognition. Deep learning models can automatically learn hierarchical feature representations from raw pixel data, capturing both low-level features (edges, textures) and high-level semantic features (facial structure, identity). Landmark systems such as DeepFace (Facebook), FaceNet (Google), and ArcFace have demonstrated near-human-level recognition accuracy on standard benchmarks.

1.1Objectives

The primary objectives of this research paper are as follows:

- To design and implement a robust end-to-end face recognition pipeline using deep learning.
- To evaluate the performance of CNN-based models against classical algorithms on benchmark datasets.
- To analyze the impact of preprocessing techniques on recognition accuracy under challenging conditions.
- To propose an optimized model architecture that balances accuracy and computational efficiency.
- To identify limitations and propose future research directions for improved real-world deployment.

1.3 Scope of the Paper

This paper is organized as follows: Section 2 reviews existing literature in face recognition. Section 3 formulates the problem statement. Section 4 presents the proposed methodology. Section 5 describes the data analysis and dataset characteristics. Section 6 details the model development process. Section 7 presents and discusses experimental results. Section 8 concludes the paper with future directions.

2. LITERATURE REVIEW

Face recognition has been a subject of extensive academic and industrial research for over five decades. The evolution of techniques can be broadly categorized into three generations: geometric/structural approaches, statistical-based methods, and deep learning-based approaches.

2.1Early Geometric Approaches

The earliest face recognition system was proposed by Bledsoe (1966), which relied on manually measured facial geometry such as eye width, nose length, and chin shape. Kanade (1973) developed the first automated face recognition system using simple image processing techniques. While these approaches were pioneering, they lacked robustness to variations in pose and lighting and required significant manual effort.

2.2Statistical and Appearance-Based Methods

Turk and Pentland (1991) introduced the Eigenfaces approach based on Principal Component Analysis (PCA), which reduced the dimensionality of facial images and represented faces as linear combinations of eigenvectors of the training covariance matrix. This work marked a paradigm shift from geometric to statistical approaches. Belhumeur et al. (1997) proposed Fisherfaces using Linear Discriminant Analysis (LDA), which maximized between-class scatter while minimizing within-class scatter, yielding improved robustness to illumination changes.

Ahonen et al. (2006) introduced Local Binary Patterns Histograms (LBPH) as texture-based descriptors for face recognition. LBPH proved highly robust to monotonic illumination changes and was computationally efficient, making it practical for real-time systems. Wright et al. (2009) proposed Sparse Representation Classification (SRC), which treated face recognition as a sparse coding problem, demonstrating robustness to occlusion and noise.

2.3Deep Learning-Based Approaches

The deep learning era in face recognition began with DeepFace (Taigman et al., 2014), which employed a nine-layer deep CNN and achieved 97.35% accuracy on the LFW benchmark, approaching human-level performance (97.53%). The model used 3D face alignment and demonstrated that deep neural networks could learn powerful identity-preserving representations.

Schroff et al. (2015) proposed FaceNet, which introduced the triplet loss function to learn a compact Euclidean embedding space where distances directly correspond to face similarity. FaceNet achieved 99.63% accuracy on LFW and became the foundation of many production face recognition systems. Sun et al. (2014) developed DeepID series models using verification and identification signals for supervision, achieving competitive performance with smaller models.

Liu et al. (2017) introduced SphereFace, which reformulated the face recognition problem using angular margin in a hyperspherical feature space. Wang et al. (2018) proposed CosFace using large margin cosine loss. Deng et al. (2019) developed ArcFace using additive angular margin loss, achieving state-of-the-art performance on multiple benchmarks.

These margin-based loss functions significantly improved the discriminability of learned face representations.

2.4Summary of Reviewed Works

Author(s)	Year	Method	Dataset	Accuracy
Turk & Pentland	1991	Eigenfaces (PCA)	ORL	96.0%
Belhumeur et al.	1997	Fisherfaces (LDA)	Yale Database	97.5%
Ahonen et al.	2006	LBPH	FERET	97.1%
Taigman et al.	2014	DeepFace (CNN)	LFW	97.35%
Schroff et al.	2015	FaceNet (Triplet)	LFW	99.63%
Deng et al.	2019	ArcFace	LFW, MegaFace	99.83%

Table 1: Summary of Face Recognition Methods and Performance

The literature clearly indicates a progressive improvement in recognition accuracy driven by increasingly sophisticated deep learning architectures and loss functions. The gap between handcrafted and deep-learned features has become insurmountable in most practical scenarios, motivating the deep learning-centric approach adopted in this research.

3. PROBLEM FORMULATION

3.1Problem Statement

Given a database D = {(I1, y1), (I2, y2), ..., (In, yn)} of n labeled facial images, where Ii represents the i- th facial image and yi represents the corresponding identity label from a set of C classes (C distinct individuals), the face recognition problem can be formally decomposed into two sub-problems:

- Face Verification (1:1 Matching): Given two facial images Ii and Ij, determine whether they belong to the same identity, i.e., compute a binary decision f(Ii, Ij) → {same, different} based on a distance or similarity metric in a learned embedding space.
- Face Identification (1:N Matching): Given a probe facial image Ip (possibly unseen during training), identify the most likely identity from the database, i.e., find y* = argmax P(yi | Ip) for i = 1, ..., n.

The core challenge lies in learning a feature mapping function F: R^(H×W×C) → R^d that maps raw facial images of arbitrary dimensions to a compact d-dimensional embedding vector, such that:

- Intra-class distance is minimized: $||F(Ii) - F(Ij)||_2$ is small when yi = yj
- Inter-class distance is maximized: $||F(Ii) - F(Ij)||_2$ is large when yi ≠ yj

3.2Intra-class Variability

A single individual's face can appear drastically different due to illumination changes (direct/indirect lighting, shadows), pose variations (frontal, profile, 3/4 view), facial expressions (neutral, happy, angry), aging, makeup, accessories (glasses, beard), and partial occlusion (masks, scarves).

3.3Inter-class Similarity

Faces of identical twins or closely related individuals may share highly similar biometric patterns, making discrimination extremely difficult. The system must learn fine-grained discriminative features that capture identity-specific characteristics beyond gross facial geometry.

3.4Data Scarcity and Privacy

Collecting large-scale labeled facial datasets is resource-intensive and raises significant privacy and ethical concerns, particularly with increasingly stringent data protection regulations such as GDPR. The system must be capable of learning from limited labeled examples and potentially generalizing to new identities without retraining.

4. METHODOLOGY

The proposed face recognition system adopts a modular pipeline architecture consisting of five sequential stages: image acquisition, face detection and alignment, preprocessing, deep feature extraction, and classification/matching. The overall system architecture is illustrated and described in detail below.

4.1System Pipeline Overview

Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6
Image Acquisition	Face Detection (MTCNN)	Preprocessing & Alignment	Feature Extraction (CNN)	Embedding & Matching	**Identity Decision**

Table 2: System Pipeline Stages

4.2Face Detection

Face detection is the first critical step, responsible for localizing and extracting facial regions from input images. We employ the Multi-task Cascaded Convolutional Network (MTCNN) proposed by Zhang et al. (2016), which jointly performs face detection and facial landmark localization in a three-stage cascade: P- Net (Proposal Network), R-Net (Refinement Network), and O-Net (Output Network).

MTCNN achieves high recall rates across diverse image conditions and outputs five facial landmarks (two eyes, nose tip, two mouth corners) used for subsequent alignment. The detection operates at multiple image scales to handle faces of varying sizes within the same frame.

4.3Preprocessing and Alignment

Robust preprocessing is essential for reducing the impact of intra-class variability. The following steps are applied sequentially:

1. Face Alignment: Using the detected facial landmarks, an affine transformation matrix is computed to normalize face orientation. All faces are warped to a canonical frontal pose of size 112×112 pixels.

2. Histogram Equalization: Adaptive Histogram Equalization (CLAHE) is applied to normalize illumination variations while preserving local contrast details.
3. Pixel Normalization: Pixel values are normalized to the range [-1, 1] using mean subtraction and standard deviation scaling computed on the training set.
4. Data Augmentation: During training, random horizontal flipping, random rotation (±15°), random brightness/contrast jitter, and Gaussian noise injection are applied to improve generalization.

4.4Feature Extraction — CNN Architecture

The feature extraction backbone is a modified ResNet-50 architecture adapted for face recognition tasks. The network accepts preprocessed 112×112×3 facial images and produces 512-dimensional embedding vectors. Key architectural modifications include:

- Batch Normalization after every convolutional layer to stabilize training.
- Dropout (p=0.4) before the final embedding layer to prevent overfitting.
- L2-normalized embeddings to ensure consistent scale in the embedding space.
- Removal of the original classification head and addition of a BN-Dropout-FC-BN block.

The network is trained using the ArcFace loss function, which adds an additive angular margin $m = 0.5$ in the angular space between the sample embedding and the weight vector of the corresponding class:

$$L = -\log \left[e^{(s\cdot\cos(\theta_{yi} + m))} / \left(e^{(s\cdot\cos(\theta_{yi} + m))} + \Sigma_{j\neq yi}\, e^{(s\cdot\cos(\theta_j))}\right) \right]$$

where $s = 64$ is the feature scale, θ_{yi} is the angle between the embedding and the weight of the correct class, and $m = 0.5$ is the angular margin. This formulation enforces greater intra-class compactness and inter-class discrepancy compared to softmax cross-entropy.

4.5Classification and Matching

For the closed-set identification task, a Support Vector Machine (SVM) classifier with RBF kernel is trained on the extracted 512-dimensional embeddings. For the open-set verification task, cosine similarity is computed between the probe embedding and gallery embeddings, and a threshold $\tau = 0.65$ (tuned on a validation set) is applied to make the accept/reject decision.

The system also supports a k-Nearest Neighbor (k-NN) matching mode where the probe embedding is compared against all gallery embeddings and the top-k nearest neighbors vote for the final identity prediction.

5. DATA ANALYSIS

5.1Datasets Used

Three publicly available benchmark datasets were used for training, validation, and evaluation:

5.1.1 Labeled Faces in the Wild (LFW)

LFW is the most widely used benchmark for unconstrained face recognition. It contains 13,233 facial images of 5,749 individuals collected from web news articles, with 1,680 individuals having two or more images. The dataset exhibits significant variation in illumination, pose, expression, and background clutter. The standard evaluation protocol uses 6,000 matched/mismatched pairs for verification performance measurement.

5.1.2 AT&T ORL Dataset

The AT&T ORL dataset contains 400 images of 40 distinct individuals, with 10 images per subject captured under varying lighting, facial expressions, and facial details (with/without glasses). Images are grayscale, 92×112 pixels. The controlled nature of this dataset makes it suitable for evaluating algorithm performance under limited variation.

5.1.3 CelebA Dataset

CelebFaces Attributes (CelebA) contains over 200,000 celebrity images with 40 binary attribute annotations per image. We used a subset of CelebA for training the deep model on a large-scale face distribution, providing diverse appearances across age, ethnicity, and gender.

5.2Dataset Statistics

Dataset	Total Images	Identities	Images/Identity	Split (Train/Val/Test)
LFW	13,233	5,749	~2.3 avg	70% / 15% / 15%
AT&T ORL	400	40	10	60% / 20% / 20%
CelebA (subset)	150,000	8,500	~17.6 avg	80% / 10% / 10%

Table 3: Dataset Statistics and Splits

5.3Data Preprocessing Analysis

Exploratory data analysis revealed several key observations:

- Class Imbalance: The LFW dataset exhibits severe class imbalance, with 80% of identities having fewer than 5 images. This was addressed through oversampling using augmentation and weighted sampling during training.
- Illumination Distribution: Analysis of pixel intensity histograms revealed bimodal distributions in unprocessed images, confirming the importance of CLAHE normalization.
- Pose Distribution: Approximately 35% of LFW images contain non-frontal faces (yaw > 30°), necessitating robust alignment strategies.
- Image Quality: 8.2% of dataset images were identified as low-quality (blurred, heavily occluded, or extremely low-resolution) and were excluded from training to prevent label noise.

5.4Feature Visualization

Principal Component Analysis (PCA) was applied to visualize the distribution of extracted 512-dimensional embeddings in 2D space using t-SNE (t-Distributed Stochastic Neighbor Embedding). The resulting visualization showed well-separated clusters corresponding to distinct identities in the embedding space, with minimal overlap between inter-class clusters, validating the discriminative power of the learned representations. Intra-class clusters exhibited high compactness, confirming the effectiveness of ArcFace loss in reducing within-class variation.

6. MODEL DEVELOPMENT

6.1Baseline Models

Three baseline models were implemented for comparative analysis:

- Eigenfaces + SVM: PCA applied to flattened 112×112 images retaining 150 principal components, followed by SVM classification.
- LBPH + SVM: LBP histogram features extracted from 8-neighborhood with radius 2, concatenated across a 4×4 grid of image cells, followed by SVM classification.
- MobileNetV2 + Softmax: Lightweight CNN pre-trained on ImageNet, fine-tuned with standard cross-entropy loss for face identification.

6.2Proposed Model: ResNet-50 + ArcFace

The proposed model consists of the following components:

6.2.1 Backbone Architecture

ResNet-50 was chosen as the backbone due to its favorable accuracy-efficiency trade-off and proven performance in face recognition literature. The network employs residual connections that mitigate the vanishing gradient problem and enable training of very deep networks. The backbone produces 2048- dimensional feature maps which are globally average-pooled and projected to a 512-dimensional embedding via a fully connected layer.

6.2.2 Training Configuration

Hyperparameter	**Value**
Optimizer	SGD with Momentum (m=0.9)
Initial Learning Rate	0.1
LR Schedule	Step decay: ×0.1 at epoch 20, 28, 32
Weight Decay	5×10^{-4}
Batch Size	256 (distributed across 4 GPUs)
Total Epochs	35
ArcFace Margin (m)	0.5
Feature Scale (s)	64
Embedding Dimension	512

Table 4: Training Hyperparameters

6.2.3 Transfer Learning Strategy

The ResNet-50 backbone was initialized with ImageNet pre-trained weights. A two-phase training strategy was adopted: (1) feature extractor frozen, only embedding layer trained for 5 epochs; (2) end-to-end fine- tuning with a reduced learning rate of 0.01. This strategy improved convergence speed and final accuracy compared to training from scratch.

6.3Model Compression and Optimization

For deployment on resource-constrained devices, model compression techniques were applied: (1) Post- training quantization to INT8 precision using TensorFlow Lite, reducing model size from 94MB to 24MB with < 0.5% accuracy degradation; (2) Structured pruning removing 30% of filters with lowest L1-norm magnitudes, achieving 1.8× inference speedup; (3) Knowledge distillation training a lightweight MobileNetV3 student model using soft labels from the ResNet-50 teacher model.

7. RESULTS AND DISCUSSION

7.1 Evaluation Metrics

System performance was measured using the following metrics:

- Accuracy (%): Proportion of correctly identified faces in the test set.
- False Acceptance Rate (FAR): Rate of accepting impostors as genuine subjects.
- False Rejection Rate (FRR): Rate of rejecting genuine subjects.
- Equal Error Rate (EER): Point where FAR = FRR, representing the operating point trade-off.
- ROC AUC: Area under the Receiver Operating Characteristic curve.

7.2 Performance Comparison

Model	Dataset	Accuracy	FAR (%)	FRR (%)	ROC AUC
Eigenfaces + SVM	LFW	85.2%	8.3	9.1	0.912
LBPH + SVM	LFW	88.7%	6.2	6.8	0.941
MobileNetV2 + Softmax	LFW	93.4%	3.9	4.2	0.969
ResNet-50 + ArcFace (Proposed)	LFW	**97.8%**	**0.7**	**1.2**	**0.997**
Eigenfaces + SVM	AT&T ORL	91.0%	5.2	7.1	0.948
LBPH + SVM	AT&T ORL	94.5%	3.8	4.0	0.972
ResNet-50 + ArcFace (Proposed)	AT&T ORL	**98.9%**	**0.5**	**0.8**	**0.999**

Table 5: Comparative Performance Results

7.3 Discussion

The experimental results conclusively demonstrate the superiority of the proposed ResNet-50 + ArcFace model over all baseline methods. Key observations from the results are as follows:

- The proposed model achieved 97.8% accuracy on LFW, representing a 4.4% absolute improvement over the best classical method (LBPH + SVM at 88.7%) and a 4.4% improvement over MobileNetV2, highlighting the benefit of deeper architectures with margin-based loss functions.
- The FAR of 0.7% and FRR of 1.2% on LFW indicate strong security-usability balance, meeting practical deployment requirements for access control systems (typically requiring FAR < 1%).
- On the AT&T ORL dataset, the proposed model reached 98.9% accuracy, confirming its effectiveness even on small controlled datasets where the advantage of deep learning over classical methods is less pronounced.
- The near-perfect ROC AUC of 0.997 on LFW confirms that the model maintains excellent discrimination across all operating thresholds, crucial for adaptive security systems.

7.4 Ablation Study

An ablation study was conducted to quantify the contribution of each component:

Configuration	CLAHE	Augmentation	ArcFace	Accuracy
ResNet-50, No extras	No	No	No	93.1%
+ CLAHE Normalization	Yes	No	No	94.8%
+ Data Augmentation	Yes	Yes	No	96.2%
+ ArcFace Loss (Full Model)	**Yes**	**Yes**	**Yes**	**97.8%**

Table 6: Ablation Study Results on LFW Dataset

The ablation study clearly establishes that each component contributes incrementally to the final performance. CLAHE normalization alone yields a 1.7% improvement, data augmentation adds 1.4%, and the ArcFace loss function provides the most significant boost of 1.6%, confirming the importance of discriminative loss design in face recognition.

7.5Error Analysis

Analysis of misclassified samples revealed that the majority of errors occurred in cases involving extreme pose variation (yaw > 60°), severe occlusion (> 50% of face occluded), and very low image resolution (< 32×32 effective face region). Identical-twin pairs also posed a significant challenge, with a twin confusion rate of 34% — an inherent limitation of appearance-based methods. Future work addressing 3D structural features may mitigate twin confusion rates.

8. CONCLUSION AND FUTURE WORK

8.1Conclusion

This paper presented a comprehensive face recognition system built upon a ResNet-50 backbone trained with ArcFace loss and equipped with a robust preprocessing pipeline incorporating MTCNN-based face detection, CLAHE illumination normalization, and geometric alignment. The system demonstrated competitive recognition performance, achieving 97.8% accuracy on the challenging LFW benchmark and 98.9% on the AT&T ORL dataset, with favorable FAR/FRR characteristics suitable for practical access control applications.

The proposed system represents a pragmatic, deployment-ready solution that balances accuracy, computational efficiency, and robustness. Model compression experiments demonstrated that accuracy can be maintained within 0.5% while reducing computational requirements by up to 4× through quantization and pruning, enabling deployment on edge devices.

8.2Future Work

Several directions for future research are identified:

- Federated Learning: Implementing privacy-preserving federated learning protocols to train face recognition models on distributed datasets without centralizing raw biometric data, addressing growing privacy regulation concerns.
- 3D Face Recognition: Extending the system to incorporate 3D depth information from structured light or ToF sensors to improve robustness to pose variation and identical twin confusion.
- Anti-Spoofing Integration: Integrating liveness detection modules using texture analysis, depth estimation, or rPPG-based physiological signal detection to prevent presentation attacks using photographs or 3D masks.Few-Shot Learning: Developing few-shot and zero-shot recognition capabilities using meta-learning approaches (e.g., MAML, Prototypical Networks) to enable recognition of new identities from a single enrollment image.
- Continual Learning: Addressing the catastrophic forgetting problem to enable incremental enrollment of new identities without full model retraining.

- Bias and Fairness Analysis: Conducting systematic analysis of demographic bias (gender, age, ethnicity) in recognition performance and developing debiasing training strategies to ensure equitable system behavior.

REFERENCES

[1] Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. Journal of Cognitive Neuroscience, 3(1), 71-86.

[2] Belhumeur, P. N., Hespanha, J. P., & Kriegman, D. J. (1997). Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7), 711-720.

[3] Ahonen, T., Hadid, A., & Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(12), 2037-2041.

[4] Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., & Ma, Y. (2009). Robust face recognition via sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(2), 210-227.

[5] Taigman, Y., Yang, M., Ranzato, M. A., & Wolf, L. (2014). DeepFace: Closing the gap to human-level performance in face verification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1701-1708.

[6] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 815-823.

[7] Sun, Y., Wang, X., & Tang, X. (2014). Deep learning face representation by joint identification- verification. Advances in Neural Information Processing Systems (NeurIPS), 27, 1988-1996.

[8] Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., & Song, L. (2017). SphereFace: Deep hypersphere embedding for face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 212-220.

[9] Wang, H., Wang, Y., Zhou, Z., Ji, X., Gong, D., Zhou, J., & Liu, W. (2018). CosFace: Large margin cosine loss for deep face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 5265-5274.

[10] Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). ArcFace: Additive angular margin loss for deep face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 4690-4699.

[11] Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters, 23(10), 1499-1503.

[12] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770- 778.

[13] Huang, G. B., Ramesh, M., Berg, T., & Learned-Miller, E. (2007). Labeled Faces in the Wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst.

[14] Samaria, F., & Harter, A. (1994). Parameterisation of a stochastic model for human face identification. In Proceedings of the IEEE Workshop on Applications of Computer Vision, 138-142.

[15] Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). Deep learning face attributes in the wild. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), 3730-3738.