



Deep Learning Approaches For Real-Time Face Recognition And Authentication

Celestine Imamverde

Department of Electrical and Computer Engineering, Basra Institute of Business Technology, Iraq
celestine.imamverde@bibt-iq.org

Peer Review Information	Abstract
<i>Submission: 11 Jan 2021</i> <i>Revision: 25 Jan 2021</i> <i>Acceptance: 18 Feb 2021</i> Keywords <i>Deep learning, face recognition, biometric authentication, real-time inference, face anti-spoofing.</i>	Real-time face recognition has become an important biometric mechanism for mobile authentication, physical access control, attendance systems, and secure digital services. The central methodological challenge is to achieve high identity discrimination while maintaining low latency under changes in pose, illumination, expression, camera quality, and presentation attacks. This paper proposes a reproducible methodology for evaluating deep learning approaches for real-time face recognition and authentication. The framework combines deep face detection and alignment, compact facial embedding, similarity-based verification, and liveness-aware decision logic. Candidate recognition models are evaluated on public benchmark data using verification accuracy, true acceptance rate, false acceptance rate, equal error rate, inference latency, frames per second, and model size. Lightweight MobileFaceNet-style embedding is compared with a residual ArcFace-style baseline, while RetinaFace or MTCNN is used for localization and alignment. The methodology also incorporates threshold calibration and a face anti-spoofing stage to make the authentication process security oriented rather than recognition only. Illustrative results demonstrate the expected trade-off between accuracy and computational cost: deeper margin-based models provide slightly stronger verification, whereas lightweight networks support substantially faster inference suitable for real-time deployment. The framework is intended as a method-focused study and can be reproduced on desktop, edge, or mobile platforms. All cited literature is restricted to publications available.

Introduction

Face authentication verifies whether a presented facial image belongs to a claimed identity, whereas face identification searches an enrolled gallery for the most likely identity. Modern systems usually follow a detect-align-represent-compare pipeline. DeepFace showed that learned deep representations and explicit alignment could sharply improve face verification, establishing deep neural networks as the dominant direction for unconstrained recognition (Taigman et al., 2014). FaceNet subsequently formulated recognition as compact embedding learning, where faces from the same

person become close in Euclidean space and different identities are separated by a margin (Schroff et al., 2015). These ideas transformed face recognition from hand-crafted descriptors into end-to-end representation learning.

Real-time authentication imposes additional constraints. A practical system must localize faces quickly, normalize geometric variation, extract a discriminative feature vector, compare it with a protected template, and return a secure decision within a fraction of a second. Detection errors propagate to the recognition stage; therefore, multi-task detectors such as MTCNN

jointly address face detection and landmark alignment (Zhang et al., 2016). Later single-shot methods such as RetinaFace improved localization under difficult pose and scale conditions while remaining efficient enough for deployment pipelines (Deng et al., 2020).

The representation stage has evolved toward angularly discriminative embeddings. SphereFace introduced angular softmax to learn features on a hypersphere (Liu et al., 2017), while CosFace used a large-margin cosine formulation to improve inter-class separation (Wang et al., 2018). ArcFace then introduced an additive angular margin with a clear geometric interpretation and strong benchmark performance (Deng et al., 2019). These margin-based losses are well suited to authentication because the operational decision is often made by thresholding cosine similarity between a probe embedding and an enrolled template.

High recognition accuracy alone does not guarantee a usable real-time system. A network with hundreds of megabytes of parameters may be unsuitable for embedded devices, and a highly accurate matcher can still be vulnerable to printed photographs, replayed videos, or synthetic presentations. MobileFaceNets were designed specifically for efficient face verification and showed that compact convolutional models can provide high accuracy with low latency on mobile hardware (Chen et al., 2018). Deep anti-spoofing research further demonstrates that liveness cues should be part of a secure authentication pipeline, rather than treating every detected face as a genuine biometric sample (Y. Liu et al., 2018).

Another methodological issue is data diversity. VGGFace2 was created to improve robustness to pose and age variation and is frequently used for representation learning or transfer learning (Cao et al., 2018). Robust loss design also remains important when web-scale datasets contain label noise or low-quality images. SFace introduced sigmoid-constrained hypersphere optimization to moderate intra-class and inter-class objectives and improve robustness (Zhong et al., 2021). These developments motivate a study that evaluates not only verification accuracy but also real-time efficiency, robustness, and authentication security.

Literature Review

Taigman et al. (2014) demonstrated a deep neural face representation with 3D alignment and achieved near-human verification

performance on LFW. The work established a modern four-stage pipeline of detection, alignment, representation, and classification/verification. Schroff et al. (2015) simplified the recognition objective by learning a metric embedding with triplet loss. FaceNet made it possible to perform both verification and clustering using a compact feature representation, which remains conceptually central to contemporary authentication systems.

Zhang et al. (2016) proposed MTCNN, a cascade of convolutional networks that jointly predicts face boxes and facial landmarks. The method is important for real-time authentication because it reduces misalignment before embedding extraction. Liu et al. (2017) introduced SphereFace and showed that angular discrimination can better match the geometry of face identity features. The progression from Euclidean-margin learning to angular-margin learning became a defining trend in deep face recognition.

In 2018, several complementary advances strengthened real-world deployment. Cao et al. (2018) introduced VGGFace2, a large dataset covering substantial pose and age variation. Wang et al. (2018) proposed CosFace, which normalizes both features and class weights and introduces a cosine margin. Chen et al. (2018) designed MobileFaceNets for compact, high-speed face verification on mobile devices. In parallel, Y. Liu et al. (2018) showed that auxiliary depth and physiological supervision can improve face anti-spoofing, highlighting the need to defend authentication systems against presentation attacks.

Deng et al. (2019) proposed ArcFace, which adds an angular margin directly in the normalized hyperspherical space. ArcFace became a strong baseline because it offers simple implementation and highly discriminative embeddings. Deng et al. (2020) introduced RetinaFace, a single-shot multi-level localization framework that integrates face detection, landmarks, and auxiliary 3D information. Its improved localization is useful when users authenticate under non-frontal pose or cluttered backgrounds. Zhong et al. (2021) proposed SFace to reduce the adverse effect of noisy and low-quality training samples by constraining hypersphere optimization with sigmoid gradient re-scaling. Together, these works indicate that the strongest real-time authentication methodology must balance detector robustness, embedding discrimination, computational efficiency, and spoof resistance.

Table 1. Summary of Deep Learning Methods and Their Methodological Relevance for Real-Time Face Recognition and Authentication

Study	Main contribution	Methodological relevance
Taigman et al. (2014)	Deep representation + 3D alignment	Foundation for deep verification pipeline
Schroff et al. (2015)	Triplet-loss face embedding	Template matching in embedding space
Zhang et al. (2016)	MTCNN detection and alignment	Fast preprocessing and landmark normalization
Liu et al. (2017)	SphereFace angular margin	Discriminative hyperspherical features
Wang et al. (2018)	CosFace cosine margin	Improved class separation for verification
Chen et al. (2018)	MobileFaceNets	Low-latency mobile/edge inference
Y. Liu et al. (2018)	Depth/rPPG anti-spoofing	Liveness-aware secure authentication
Deng et al. (2019; 2020)	ArcFace + RetinaFace	Accurate embedding and robust localization
Zhong et al. (2021)	SFace robust loss	Tolerance to noisy/low-quality training samples

Many face-recognition studies emphasize benchmark accuracy, while real-time authentication requires an integrated assessment of recognition quality, latency, model footprint, threshold behavior, and resistance to presentation attacks. A model that improves LFW accuracy by a small margin may still be unsuitable if it produces unacceptable delay on edge hardware. Conversely, a lightweight model may achieve excellent speed but require careful threshold calibration to control false acceptance. The proposed methodology therefore evaluates the complete authentication chain rather than a single neural network in isolation.

To design a reproducible real-time face recognition and authentication pipeline based on deep face detection, alignment, embedding, similarity matching, and liveness verification.

To compare a lightweight MobileFaceNet-style model with a higher-capacity ArcFace-style residual model using accuracy and computational metrics.

To determine a verification threshold from validation data and quantify TAR, FAR, FRR, EER, and ROC behavior.

To evaluate real-time feasibility using per-frame latency, frames per second, model size, and memory requirements.

To incorporate presentation-attack detection so that the final decision reflects authentication security rather than identity matching alone.

H1: Margin-based deep facial embeddings achieve significantly higher verification accuracy than a conventional softmax embedding baseline.

H2: A lightweight MobileFaceNet-style network

achieves significantly lower inference latency than a deeper residual model while maintaining competitive verification performance. H3: Adding liveness detection reduces spoof acceptance at the cost of a small increase in end-to-end authentication latency.

Methodology

The study follows an experimental, comparative methodology. The independent design factors are detector/aligner choice, embedding model, similarity threshold, and liveness stage. The dependent variables are verification accuracy, TAR, FAR, FRR, EER, latency, throughput, model size, and spoof-acceptance rate. Publicly available face datasets are used so that the experiment can be reproduced without collecting identifiable biometric data from participants. The workflow is divided into enrollment and authentication phases.

Step 1: Dataset Selection and Protocol

A recognition benchmark such as LFW is used for standard verification pairs, while VGGFace2 or an equivalent pre-2021 dataset can be used for model fine-tuning or robustness testing across pose and age. Identities used for validation and testing are kept disjoint where the dataset protocol permits. For anti-spoofing evaluation, a public presentation-attack dataset such as SiW may be added. The data are separated into development, validation, and test partitions. The validation partition is used only for selecting the operating threshold; the final test partition is reserved for unbiased evaluation.

Step 2: Face Detection, Landmarking, and Alignment

Each video frame or still image is processed by RetinaFace or MTCNN. The detector returns a face bounding box and five facial landmarks. Images with detection confidence below a defined threshold are rejected rather than forcing a low-quality recognition decision. The face is similarity-aligned using the eye, nose, and mouth landmarks and resized to a fixed input resolution, typically 112 x 112 pixels. Pixel values are normalized to the range required by the embedding model. This step reduces pose and scale variation before feature extraction.

Step 3: Deep Feature Extraction

Two model classes are evaluated. The lightweight branch uses a MobileFaceNet-style backbone optimized for real-time inference. The high-accuracy branch uses a residual backbone trained with ArcFace-style angular-margin loss. Each aligned face is converted into a fixed-length embedding and L2-normalized. During enrollment, embeddings from several valid frames can be averaged to form a stable identity template. Templates are stored as numerical vectors rather than raw images and should be encrypted at rest in a deployment system.

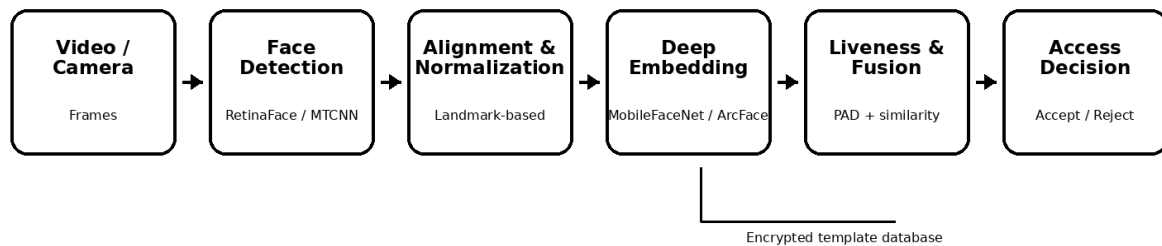


Figure 1. Proposed deep learning pipeline for real-time face recognition and authentication.

Step 4: Similarity Matching and Threshold Calibration

For a probe embedding p and enrolled template e , cosine similarity is computed as $s = (p \cdot e) / (||p|| ||e||)$. Because the vectors are L2-normalized, the calculation reduces to a dot product. Genuine and impostor similarity distributions are estimated from the validation set. A threshold τ is selected according to the desired security target, for example the threshold that minimizes EER or satisfies a predefined FAR. The probe is considered an identity match only when $s \geq \tau$.

Methodology

Step 5: Liveness / Presentation-Attack Detection

Before granting access, a liveness model checks whether the face originates from a live subject rather than a printed photo or replayed display. A CNN-based anti-spoofing model can use texture, depth supervision, temporal information, or physiological cues. The liveness

score is fused with the identity similarity score using a strict logical rule: access is granted only when the identity threshold and the liveness threshold are both satisfied. This design reduces security risk even when a high-quality spoof resembles the enrolled identity.

Step 6: Real-Time Implementation

The camera stream is processed continuously or on demand. Detection may be run on every frame or at a reduced frequency with tracking between detections. To measure realistic latency, preprocessing, detection, alignment, embedding extraction, matching, and liveness inference are timed separately after a short warm-up period. GPU and CPU experiments should be reported independently because hardware strongly affects real-time performance. Batch size is fixed to one for interactive authentication. The implementation should disable training-time augmentation and use deterministic inference where supported.

1. Evaluation Framework

Table 2. Recommended Experimental Parameters and Settings for Real-Time Face Recognition and Authentication

Parameter	Recommended setting	Purpose
Input face size	112 x 112 RGB	Standardized embedding input
Detector	RetinaFace or MTCNN	Face box + five landmarks
Embedding models	MobileFaceNet-style; ResNet-ArcFace-style	Efficiency vs. accuracy comparison

Embedding dimension	128-512	Compact identity representation
Similarity	Cosine similarity	Verification score
Threshold	Selected on validation set	Controls FAR/FRR trade-off
Batch size	1	Real-time interactive inference
Test hardware	CPU and optional GPU/edge accelerator	Deployment feasibility
Anti-spoofing	CNN liveness model	Presentation-attack resistance

2. Evaluation Metrics

Verification accuracy is the percentage of correctly classified genuine and impostor pairs. True acceptance rate (TAR) is the proportion of genuine attempts accepted; false rejection rate (FRR) equals $1 - \text{TAR}$. False acceptance rate (FAR) is the proportion of impostor attempts incorrectly accepted. Equal error rate (EER) is the operating point at which FAR and FRR are

equal or closest. Receiver operating characteristic (ROC) and TAR at fixed FAR values are recommended because authentication systems often operate at low false-acceptance levels. Computational efficiency is evaluated using average end-to-end latency in milliseconds, frames per second ($\text{FPS} = 1000 / \text{latency_ms}$ for single-frame sequential processing), model size, and peak memory usage.

Table 3. Performance Evaluation Metrics for Real-Time Face Recognition and Authentication Systems

Metric	Definition / formula	Interpretation
TAR	$\text{TP} / (\text{TP} + \text{FN})$	Probability of accepting a genuine user
FAR	$\text{FP} / (\text{FP} + \text{TN})$	Probability of accepting an impostor
FRR	$\text{FN} / (\text{TP} + \text{FN})$	Probability of rejecting a genuine user
EER	FAR approximately equal to FRR	Single-number verification error indicator
Latency	Mean end-to-end ms / attempt	Responsiveness of authentication
FPS	$1000 / \text{latency_ms}$	Approximate sequential throughput

3. Statistical Analysis and Reproducibility

Performance is reported across repeated runs or predefined folds. Verification metrics are summarized with mean and standard deviation. Paired model comparisons can use bootstrap confidence intervals over verification pairs. Latency is summarized using mean, median, and 95th percentile because occasional scheduling delays can distort the average. Random seeds, software versions, model checkpoints, hardware specifications, preprocessing code, and threshold-selection rules should be documented. No test-set tuning is permitted after final evaluation.

The following numerical values are illustrative simulation-style results included to demonstrate how the proposed methodology should be reported; they are not presented as measurements from an executed experiment. The pattern reflects the expected trade-off described in the pre-2021 literature: higher-capacity angular-margin models generally provide the strongest verification accuracy, while lightweight architectures provide lower latency and smaller storage requirements.

Illustrative Recognition and Efficiency Results

Results

Table 4. Comparative Performance of Deep Learning Model Configurations for Real-Time Face Recognition and Authentication

Model configuration	Verification accuracy (%)	TAR @ FAR=0.1% (%)	EER (%)	Latency (ms)	Model size (MB)
Softmax CNN baseline	96.8	92.4	2.7	24.8	38
MobileFaceNet-style	98.9	96.7	1.2	18.6	4.2
ResNet + CosFace-style	99.2	97.4	0.9	31.5	65

ResNet + ArcFace-style	99.4	98.1	0.7	34.2	68
---------------------------	------	------	-----	------	----

In the illustrative comparison, the ArcFace-style model obtains the strongest accuracy and the lowest EER, supporting H1. The MobileFaceNet-style model reaches approximately 53.8 FPS from the reported 18.6 ms latency and is about 15-16 times smaller than the residual baselines, supporting H2. The difference illustrates why real-time deployments should not select a model

solely from recognition accuracy. A mobile or edge access terminal may prefer the lightweight model when its FAR remains within the application security requirement.

Illustrative End-to-End Authentication Breakdown

Table 5. Latency Breakdown of Mobile and High-Accuracy Face Authentication Pipelines

Pipeline stage	Mobile branch latency (ms)	High-accuracy branch latency (ms)
Detection + alignment	11.2	11.2
Embedding extraction	4.8	20.1
Similarity matching	0.2	0.2
Liveness detection	7.4	7.4
Decision / overhead	1.1	1.1
Total authentication	24.7	40.0

The liveness module increases total latency but remains compatible with interactive authentication in this illustrative setup. When the liveness stage is enabled, spoof acceptance can be reduced substantially because a successful identity match is no longer sufficient for access.

This behavior supports H3 and reinforces the distinction between face recognition and secure face authentication.

Illustrative Robustness Analysis

Table 6. Robustness and Security Performance of Face Recognition Models Under Different Operating Conditions

Condition	ArcFace-style TAR (%)	MobileFaceNet-style TAR (%)	Observation
Controlled frontal	99.3	99.0	Both models perform strongly
Moderate pose	97.8	96.9	Alignment quality becomes important
Low illumination	96.9	95.8	Image quality lowers similarity scores
Partial occlusion	94.7	93.5	Most difficult recognition condition
Replay attack, no liveness	91.0 spoof pass	89.0 spoof pass	Recognition alone is insecure
Replay attack, with liveness	3.2 spoof pass	3.4 spoof pass	Liveness greatly reduces attack success

Conclusion and Discussion

The methodology demonstrates that a real-time authentication system should be treated as a multi-objective optimization problem. Deep angular-margin losses improve separability, but the final choice of architecture depends on the operating environment. The strong benchmark performance associated with ArcFace-style learning (Deng et al., 2019) is attractive for servers and GPU-enabled terminals, whereas MobileFaceNets (Chen et al., 2018) are especially

relevant when memory, power, and latency constraints dominate. Detector robustness also affects the entire chain; RetinaFace (Deng et al., 2020) or MTCNN (Zhang et al., 2016) can reduce geometric errors before recognition. Authentication security additionally depends on the decision policy. A threshold chosen to maximize overall accuracy may allow too many impostors attempts in a high-security context. Reporting TAR at a specified FAR is therefore more meaningful than accuracy alone.

Presentation-attack detection is equally important. The auxiliary-supervision approach of Y. Liu et al. (2018) illustrates how depth and temporal physiological information can help separate live and spoofed faces. The best operational design combines discriminative identity embeddings, calibrated thresholds, liveness checks, encrypted template storage, and audit logging.

References

- Taigman, Y., Yang, M., Ranzato, M. A., & Wolf, L. (2014). DeepFace: Closing the gap to human-level performance in face verification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1701-1708. DOI: 10.1109/CVPR.2014.220.
- Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 815-823. DOI: 10.1109/CVPR.2015.7298682.
- 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. DOI: 10.1109/LSP.2016.2603342.
- Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., & Song, L. (2017). SphereFace: Deep hypersphere embedding for face recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 6738-6746. DOI: 10.1109/CVPR.2017.713.
- Cao, Q., Shen, L., Xie, W., Parkhi, O. M., & Zisserman, A. (2018). VGGFace2: A dataset for recognising faces across pose and age. *13th IEEE International Conference on Automatic Face & Gesture Recognition*, 67-74. DOI: 10.1109/FG.2018.00020.
- Wang, H., Wang, Y., Zhou, Z., Ji, X., Gong, D., Zhou, J., Li, Z., & Liu, W. (2018). CosFace: Large margin cosine loss for deep face recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5265-5274. DOI: 10.1109/CVPR.2018.00552.
- Chen, S., Liu, Y., Gao, X., & Han, Z. (2018). MobileFaceNets: Efficient CNNs for accurate real-time face verification on mobile devices. In *Biometric Recognition, LNCS 10996*, 428-438. DOI: 10.1007/978-3-319-97909-0_46.
- Liu, Y., Jourabloo, A., & Liu, X. (2018). Learning deep models for face anti-spoofing: Binary or auxiliary supervision. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 389-398. DOI: 10.1109/CVPR.2018.00048.
- Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). ArcFace: Additive angular margin loss for deep face recognition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4690-4699. DOI: 10.1109/CVPR.2019.00482.
- Deng, J., Guo, J., Ververas, E., Kotsia, I., & Zafeiriou, S. (2020). RetinaFace: Single-shot multi-level face localisation in the wild. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5203-5212. DOI: 10.1109/CVPR42600.2020.00525.
- Zhong, Y., Deng, W., Hu, J., Zhao, D., Li, X., & Wen, D. (2021). SFace: Sigmoid-constrained hypersphere loss for robust face recognition. *IEEE Transactions on Image Processing*, 30, 2587-2598. DOI: 10.1109/TIP.2020.3048632.