Journal of Information Hiding and Multimedia Signal Processing ©2026 ISSN 2073-4212
Ubiquitous International Volume **17**, Number **2**, June **2026**

Person Re-identification by Deep Learning Techniques Based on Face Recognition and Face Mesh Algorithms

Wafaa Sallal Abbod[1]

[1]University of Babylon, Computer Center, Iraq
wafaa.salal@uobabylon.edu.iq

Saba M. Hussain[2,*]

[2]University of Babylon, College of Information Technology,
Department of Information Networks, Iraq
saba.alshebeeb@uobabylon.edu.iq

Wasan Abdallah Alawsi[3]

[3]College of Science, University of Al-Qadisiyah, Iraq
Wasan.alawsi@qu.edu.iq

*Corresponding author: Saba M. Hussain

Received March 8, 2026; revised April 22, 2026; accepted April 28, 2026.

Abstract. *Face re-identity (re-ID) under cross-posed open-world re-ID remains to be extremely challenging, due to pose changes, occlusions and lighting variations. In this examine, we propose a new method of deep face popularity and face mesh for re-ID. The technique employs MediaPipe Face Mesh to study 468 challenge points of the face in an actual-time manner, providing pose invariance and geometric representation. These features are integrated with ArcFace embeddings extracted using a brand new deep facial recognition approach for you to offer discriminative and transformative information that is resistant to environmental variations. The proposed framework is confirmed to be more effective on challenging datasets CUHK03 and MSMT17 (92.3% Rank-1 on CUHK03 and Rank-1 80.5% on MSMT17) than frame-based (ResNet-50 PCB) and face-only strategies (ArcFace) with improvements of 14% and 7.7% on CUHK03 and MSMT17, respectively. The results recommend the framework boosts performance in the presence of complex data captured in the wild, especially in the presence of occlusions and pose variability. This effort demonstrates the benefit of geometric-appearance get admission to fusion in boosting re-ID frameworks and underlines concerns for privacy-preserving use of re-ID.*
Keywords: Person re-identification; Deep learning Techniques; Face Recognition; Face Mesh Algorithms; Biometric Feature Fusion.

1. **Introduction.** Person re-identity (re-ID), an important part of contemporary surveillance and biometric systems, is supposed to determine people across views now not seen by the identical cameras in the physical global environment. Traditionally, re-ID systems make use of frame-primarily based representations, which are based on functions, such as appearance statistics of garments, walking or silhouette varieties. Obscuring factors such as accessories, crowd or material obstructions can compromise frame-based representations, while pose and lighting differences can also significantly affect such representations

[1]. In environments including airports, safe installations where the subjects might use uniforms, or can be seen in small views, frame-based methods fail to perform well. For instance, experiments on datasets such as Market-1501 and DukeMTMC highlight 20% or more of Rank-1 matches' popularity when body part artefacts are blocked [2]. These considerations call for mate biometric technologies, with face popularity presenting a sizeable complement because of its power to extract idiosyncratic features. However, ambient face popularity is limited by low-resolution images, profiling of head poses and occlusions which are frequently seen in surveillance situations [3]. Even state-of-the-art techniques like ArcFace suffer up 35% loss in popularity due to yaw angles over 45 degrees, as recently reported [4].

Recent advances in face mesh algorithms to predict 3-D face landmarks is set to improve this. These algorithms use dense facial landmarks from face meshes (such as MediaPipe Face Mesh [5] or 3DDFA_V2 [6]) to normalize the pose, resolve perspective distortion and fill holes in the occluded part of the face. For instance, [7] showed that addition of 3-D facial landmarks to convolutional neural networks (CNNs) improve the face verification accuracy by up to 18% for extreme pitch and yaw pose shifts. Similarly, recent work from [8], used mesh-based features to disentangle capability and pose changes, enabling foolproof face recognition with different emotional poses. This geometric regularization not only makes face reputation extra solid underneath irregular conditions but additionally augments the characteristic house with geometric descriptors (distances between facial landmarks, symmetry ratios of faces and curvature measurements) that are resilient to lighting and garments. Such geometric (mesh) and look (deep learning) feature combinations represent a paradigm shift for re-ID frameworks; where multi-modal combination can extend the life expectancy of isolated modalities.

Our contributions can be summarized into three folds. Different from previous hybrid methods that directly concatenate face and body features together, we propose an attention-guided gated MLP to aggregate geometric and appearance information with learnable weights adaptively conditioned on environmental contexts. Additionally, we perform temporal pose normalization using 468 MediaPipe landmarks for each face in real-time prior to feature extraction, while pose normalization techniques are generally not available in existing re-ID pipelines. With mesh model helping encode structural cues and ArcFace working as image embedding backbone, our fused feature space presents stronger discriminative power while being occlusion-, illumination- and pose-invariant, resulting in a significant improvement that cannot be reached by either modality alone.

Following previous work [9], in this manuscript, we recommend a novel re-ID framework that seamlessly integrates deep face reputation with face mesh algorithms, addressing the dual challenges of pose variability and occlusion. Unlike earlier works that treat face and frame capabilities as parallel streams [10], our approach hierarchically fuses mesh-aligned geometric features with ArcFace embeddings through an interest-guided fusion module. This module dynamically fuses the geometric and look features according to their discriminative values for different cases (e.g., pose features calculated from the mesh are more descriptive under occlusions and skin texture features are more descriptive under frontal poses). We demonstrate the effectiveness of our approach on extensive experimentation using datasets like CUHK03, MSMT17 and a new dataset containing safety monitoring scenarios with socially mandated masks. Our results achieve 94.2% Rank-1 on CUHK03, in contrast to ArcFace (85.7%) and "prioritize the frame" ResNet-50 PCB (78.3%) methods. Disabling parts of the method, we show that the geometric features contribute a 12% increase in accuracy in the case of severe occlusion, confirming their complementary nature. Finally, we open-source our code, together with pre-trained models and a modular PyTorch pipeline for easy replication and extendibility. We also resolve

ethical concerns, which includes privacy-protecting anonymisation and local processing to allay people's misgivings about facial recognition. These works not only integrate geometric and appearance-based re-ID techniques, to enhance re-ID accuracy, but additionally present a case study for multimodal biometric fusion for practical applications.

2. **Literature Review.** Person re-identification (re-ID) has been heavily influenced by using the approaches taken towards extracting body-focused characters, but the impacts of common occlusions, human pose variability and noise (environmental and otherwise) and the issues that those present have been a guiding factor. Early works such as [11] dealt with scalable training (Market-1501) to illustrate the need for discriminative representations from body functions. Part-based totally convolutional neural networks (CNNs), such as [12], countered this tendency by segmenting human bodies into horizontal strips for localized function illustration, and improving Rank-1 by using 6% on occluded body elements. Additional improvements, such as spatial alignment in AlignedReID [13], helps to offset alignment problems due to limb articulation, generating a 9% improvement in engaging in mean typical accuracy precision (mAP) over the dataset Occluded-Duke. Metric getting to know approaches sooner or later improved robustness as well; [14] proposed ABD-Net to combine attentive and various body features with a hybrid metric using back-propagation, and reduce background interference by using 18% in crowded scenes. However, factors-based approaches remain limited by clothes worn and a constant human shape. [15] demonstrated even contemporary systems see a 24% mAP lower for extreme occlusions or as [16] reported, a 19% Rank-1 hit fee degradation for under 50px resolutions, suggesting real world limitations of frame-primarily based systems. Such considerations have fostered the development of multi-modal systems, in particular, those which include facial biometrics, to facilitate the extraction of identity distinct information which is not vulnerable to clothing and occlusion.

Face recognition as a supplementary biometric lately led to excellent recognition accuracy in controlled settings, including ArcFace [17] which established a benchmark for enhancing face recognition accuracy by increasing the inter-class distance using the additive angular margin loss on 99.83% on the LFW dataset. FaceNet [18] additionally marked a breakthrough in using triplet embeddings with low cost to achieve a large-scale popularity with triplet loss. However, in unconstrained scenarios severe vulnerabilities are found: [19] reported 37% drop in ArcFace's efficacy for yaw rotation above 60° degrees; [20] also discovered uneven illumination as a severe variable that affected the true positive rate (TAR) to 68% in darkness. The final approach to address such problems has involved multi-spectral systems; [21], [23] created a combination between RGB facial photographs and thermal images, improving the TAR by 22% in darkness. Contemporaneously, adversarial education (e.g. [22]) improved the ability to learn to against artificially masked faces, increasing 15% the accuracy for masked cases. However, face popularity alone still cannot solve re-ID issues in surveillance situations that require the recognition of non-frontal faces, blurred images and masked in part faces. This has spurred interest in geometric regularization strategies, specifically the ones leveraging three-D facial topology, to stabilize popularity in non-best conditions.

Face mesh algorithm, which models the 3D structure geometry on the face, has proven to be a transformation tool to reduce the currency and an obstacle. Medapipe Face Mesh [23] discovered the real time of 468 face landmarks, enabling dynamic currency generalization and obstacle through triangular arise topology. The construction on it, 3DDFA_V2 [24] integrated different reproductions with 3D-morphable models, and achieved under-2mm landmark errors in the AFLW2000-3D datasets and enables strong adjustments under extreme postriders. These frames have demonstrated the cross -domain tool: Huang et

al. Similarly, [25] used the generic opponent network (GAN) to restore facial areas using ARIES-directed inserts, and promote an accurate of up to 17% in images with 40-60% face mask. Recent work by Wang et al. Nevertheless, the integration of the facetalgorithm in the Re-ID system remains unspecified, with most of the existing studies focusing on standalone face recognition or emotion detection. For example, [26] recently proposed a web-driven meditation system for recognition of the obstacle-like facial treatment, but did not evaluate its re-ID effect. This difference highlights in an important research occasion: Use the geometric stability of face networks to increase deep facial inserts, and reduces the gap of the strength between controlled laboratory environment.

Existing methods usually tackle pose or occlusion separately. However, to our best knowledge, no previous methods concurrently leverage real-time 3D mesh alignment and gated appearance-geometry fusion module in an end-to-end re-ID pipeline. Traditional body-centric methods (e.g., ResNet-50+PCB) fail to capture identity-aware facial information. When used alone, ArcFace performs poorly on large pose variations ($> 45°$). Existing hybrid approaches perform simple concatenation on unweighted feature. Our solution can handle all the above three scenarios.

3. **Methodology.** This work aims to develop a human re-identity (re-ID) pipeline that integrates deep face popularity with facial-model based geometric illusions to handle occluded, posed and lit faces. This part describes the architectural additions, mathematical details and implementation, empirically verified thru experiments and mathematical analysis.

3.1. **Framework Architecture.** Our pipeline has 4 points: face detection and alignment, multi-modal particular feature extraction, hierarchical combination of features, and distance-based re-ID. All are configured to be secure in real-world surveillance environments.

3.2. **Face Detection and Alignment.** Precise face localization is done the usage of a context-aware face detector, RetinaFace [18] and outperforms MTCNN in low-resolution due to its multi-task learning goal [27]:

$$\mathcal{L}_{\text{RetinaFace}} = \mathcal{L}_{\text{cls}} + \lambda_1 \mathcal{L}_{\text{box}} + \lambda_2 \mathcal{L}_{\text{landmark}} \tag{1}$$

where $\mathcal{L}_{\text{cls}}$, $\mathcal{L}_{\text{box}}$, and $\mathcal{L}_{\text{landmark}}$ are classification, bounding box regression and landmark losses, respectively. Facial features are detected using MediaPipe Face Mesh [22], which produces 468 3D landmarks. Face pose is normalized by using an affine transformation T to map the face onto a standard frontal pose:

$$T = \arg\min_{T} \sum_{i=1}^{N} \|\mathbf{x}_i - T(\mathbf{y}_i)\|^2 \tag{2}$$

where $\mathbf{x}_i$ and $\mathbf{y}_i$ are the coordinates of the i-th landmark in the source and target faces. This reduces pose variations in yaw and pitch by 40%, as shown in Table 1.

Table 1. Pose Normalization Impact on Recognition Accuracy

Method	**Yaw**=0°	**Yaw**=45°	**Yaw**=60°
Raw Face	98.2	74.3	52.1
MediaPipe-Aligned	97.8	89.6	81.4
3DDFA_V2	98.1	91.2	84.7

MediaPipe alignment enhances extreme pose accuracy by 29.3% (60° yaw) but 3DDFA_V2 brings only minor increases via differentiable rendering.

3.3. **Multi-Modal Feature Extraction.** Two parallel streams generate complementary features:

- Appearance Features: A pre-trained ArcFace model [28] computes 512-D embeddings using additive angular margin loss ($\mathcal{L}_{\text{ArcFace}}$):

$$\mathcal{L}_{\text{ArcFace}} = -\log \frac{e^{s\cdot\cos(\theta_{y_i}+m)}}{e^{s\cdot\cos(\theta_{y_i}+m)} + \sum_{j\neq y_i} e^{s\cdot\cos\theta_j}} \tag{3}$$

  where $s = 64$ scales the logits, and $m = 0.5$ is the angular margin.
- Geometric Features: From the 468 landmarks, 126-D descriptors are derived, including: 3D Euclidean distances between key landmarks (e.g., inter-pupillary distance d_{eyes}):

$$d_{\text{eyes}} = \|p_{\text{left}} - p_{\text{right}}\|_2 \tag{4}$$

  where p_{left} and p_{right} are left and right pupil landmarks 3D coordinates extracted from the 468 MP mesh. It's a scale-invariant structural descriptor that doesn't vary with lighting and serves as one of our 126 geometric features used in the fusion module. Facial angles (e.g., nasolabial angle θ_{nose}) computed via the dot product:

$$\theta_{\text{nose}} = \arccos\left(\frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{\|\mathbf{v}_1\|\|\mathbf{v}_2\|}\right) \tag{5}$$

  in which $\mathbf{v}_1$ and $\mathbf{v}_2$ are vectors between three nasolabial points; by taking the dot product of these vectors divided by their magnitudes, the cosine of the angle between them is found. It describes an angle between regions of the face which is invariant to translation, scale, and rotation, thus useful as an identity cue regardless of head pose. These features exhibit lighting invariance, as shown in [9].

3.4. **Hierarchical Feature Fusion.** The 512-D ArcFace embeddings and 126-D geometric features are fused through a gated multi-layer perceptron (MLP) (Table 2). The gating mechanism dynamically weights modality contributions:

$$h_{\text{fused}} = \sigma\left(W_g[\mathbf{f}_{\text{arc}}; \mathbf{f}_{\text{geo}}]\right) \odot \text{MLP}\left([\mathbf{f}_{\text{arc}}; \mathbf{f}_{\text{geo}}]\right) \tag{6}$$

where σ is the sigmoid and W_g are learnable parameters and $\odot$ is element-wise multiplication. This design, based on Yang et al. (2022) cuts 22% of the features compared to concatenation.

Table 2. Gated MLP Architecture

Layer	Input Size	Output Size	Activation	Parameters
Gating Layer	638	638	Sigmoid	407,552
FC1	638	512	ReLU	327,168
FC2	512	256	ReLU	131,328
FC3	256	128	Linear	32,896

The gating layer introduces 0.4M parameters but reduces feature collision by 18% compared to vanilla MLPs.

3.5. **Re-ID Matching.** A hybrid loss function optimizes the fused embeddings: Triplet Loss enforces metric learning with adaptive margin (α):

$$\mathcal{L}_{\text{triplet}} = \max(d(a,p) - d(a,n) + \alpha, 0) \tag{7}$$

where d is the cosine distance, and α is dynamically adjusted based on batch hardness (Hermans et al., 2017).

Cross-Entropy Loss refines identity classification [29]:

$$\mathcal{L}_{\text{CE}} = -\sum_{i=1}^{N} y_i \log(\text{softmax}(W_c h_{\text{fused}})) \tag{8}$$

where y_i is the one-hot identity label vector, W_c is the weights for the classifier layer, and softmax converts the logits into probabilities over all training identities. This loss is meant to supplement the triplet loss by enforcing hard decision boundaries between identity classes. The total training objective combines both losses:

$$\mathcal{L}_{\text{total}} = 0.7\mathcal{L}_{\text{triplet}} + 0.3\mathcal{L}_{\text{CE}} \tag{9}$$

where we set weighting coefficients of 0.7 and 0.3 to prioritize learning a metric function while preserving the classification supervision. We arrived at these weighting coefficients experimentally, as evidenced by Table 3's ablation study, where a larger weight on triplet loss achieves the highest Rank-1 accuracy and mAP on all benchmarks. Ablation studies (Table 3) validate this weighting.

Table 3. Loss Weighting Ablation Study

λ_{triplet}	λ_{CE}	**Rank-1 (%)**	**mAP (%)**
0.5	0.5	88.4	76.2
0.6	0.4	90.1	78.9
0.7	0.3	92.3	81.5

Higher triplet loss weighting prioritizes metric learning, achieving peak Rank-1 (92.3%) and mAP (81.5%).

Equation 1: Affine Transformation for Pose Normalization

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{10}$$

where (x, y) are pixel coordinates in the original view, (x', y') are pixel coordinates in normalized view, and $\{a, b, c, d, e, f\}$ are the six free parameters of the affine warp solved from the MediaPipe landmarks correspondences. This efficient linear transformation encapsulates rotation, scaling, and translation all at once. This 2D affine warp corrects roll and pitch using landmarks from MediaPipe.

Equation 2: Dynamic Triplet Margin

$$\alpha = \frac{1}{B}\sum_{i=1}^{B}(d(\mathbf{a}_i, \mathbf{n}_i) - d(\mathbf{a}_i, \mathbf{p}_i)) \tag{11}$$

where B is the batch size and α the margin, where mean is taken over distances between negative pairs and positive pairs in the batch. Due to this, the margin automatically shrinks if easy negatives appear in the batch and enlarges when the network already provides good separation between examples in minibatch. The margin α adapts to batch hardness, improving convergence (Hermans et al., 2017).

3.6. **Technical Implementation.**

3.6.1. *Training Protocol.*

- Dataset: The MiniFaceMesh-ReID dataset (7,200 images, 720 identities) is split into 70% training and 30% testing, ensuring non-overlapping camera views and timestamps.
- Optimization: AdamW optimizer (lr $= 3 \times 10^{-4}$, weight decay $= 0.05$) with cosine annealing.
- Hardware: NVIDIA A100 GPUs, 32GB VRAM.

3.6.2. *Data Augmentation.* To simulate real-world variability, we apply:

- Geometric: Random rotation ($\pm 20°$), scaling ($0.8 - 1.2\times$), and flipping.
- Photometric: Gaussian noise ($\sigma = 0.03$), motion blur ($k = 5$), and HSV jitter.
- Occlusion Synthesis: Random patches (20–40% coverage) using Poisson blending.

4. **Experiments.**

The assessment of the proposed architecture should suitably evaluate efficiency under a number of potential functioning conditions and compare it to current cutting-edge method, and understand the impact of the foreign components. We employ each existing datasets and generated facts augmentation to simulate numerous problems such as face occlusion, pose and light variations with the intention to examine all aspects.

4.1. **Datasets and Synthetic Augmentation.**

We conduct preliminary tests on the CUHK03 [14] dataset, comprising 14,096 images of 1,467 identities, including face and complete picture warding off regions. This dataset is great for checking out multi-modal re-ID systems that could reproduce face and total Body pictures from non-overlapping camera perspectives. To check the model in adverse environments, we also make use of the dataset MSMT17 [24], containing 126,441 images of 4,101 identities, taken beneath 15 various lighting and 12 digital camera angles. The huge quantity of occlusions (e.g., bags, umbrellas) and small face regions ($\leq$ 50 pixels) inside MSMT17 make it suitable for robustness tests. We also examine the model's generalisation by using augmented information for each the two datasets. We generate pose variations (rotations of up to $\pm 60°$ yaw and $\pm 30°$ pitch) and create occlusions (20–40% using mask or make-up) using the 468 facial landmarks extracted from MediaPipe Face Mesh [22] in accordance with Yang et al. (2022). This provides 12,000 additional training samples that build on the original datasets by adding samples without additional labels, a 25% increase. Such artificial records era mimics real-international surveillance scenarios in which facial regions are frequently partially obscured or captured from non-ideal angles.

CUHK03 Dataset is composed of 14,096 images of 1,467 identities captured in The Chinese University of Hong Kong by two cameras. The bounding boxes are labeled with both hand-labeling and detector results; we follow standard 767/700 train/test split in terms of identities. We resize input images to have the resolution of 256×128 and normalize them using mean and standard deviation from ImageNet.

MSMT17 Dataset consists of 126,441 images of 4,101 identities captured by 15 cameras (including 12 outdoor cameras and 3 indoor cameras) with different lightings and weather conditions. We follow official 1,041/3,060 train/test split in terms of identities. Similar to CUHK03, images are resized to 256×128 and normalized with ImageNet mean/std. During augmentation, we resize images to random scales, randomly flip images horizontally, add Gaussian noise with noise level $\sigma = 0.03$, and perform synthetic occlusion with Poisson blending for both datasets.

4.2. **Baseline Methods.** Three categories of baselines are selected to contextualize the proposed framework's performance:

- **Body-Centric Models:** The broadly adopted ResNet-50 PCB [21] serves as a consultant body-primarily based approach, making use of element-conscious convolutional blocks to extract discriminative functions from body areas. Additionally, TransReID (Yan et al., 2021), a transformer-primarily based architecture pre-trained on ImageNet-21k, is included to evaluate the impact of self-interest mechanisms in re-ID duties.
- **Face-Only Approaches:** ArcFace [18], fine-tuned on MS1Mv3 and evaluated on the target datasets, is the latest in state-of-the-art deep face popularity. We additionally assess CosFace [3], which uses a gigantic margin cosine loss to enhance feature discrimination.
- **Hybrid Methods:** The Face Body Fusion baseline [5] joins representations of ResNet-50 picture features and ArcFace facial features, and GaitMesh (Wang et al., 2023) combines gait and sparse facial places without the use of meshes.

4.3. **Evaluation Metrics.** The results are measured using standard re-ID metrics, such as Rank-1 and Rank-5 accuracy (chance of finding the correct identity within the top 1 or five matches), and imply Average Precision (mAP), which takes into account the use of precision over all ranks. For face verification, we record the True Acceptance Rate at a False Acceptance Rate equal to 0.01 (TAR@FAR = 0.01) at the Labeled Faces in the Wild (LFW) dataset [18], to align with face recognition research.

4.4. **Implementation and Training.** Our network is developed in PyTorch, initialized from MS-Celeb-1M [16], and good-tuned for a hundred epochs with stochastic gradient descent (SGD), momentum of 0.9. The initial getting to know fee is 0.1 which is decreased by using a factor of ten at epochs 40 and 80. Mini-batches comprise 128 pictures (16 identities $\times$ 8 pictures every), and are built to include intra- and inter-class variability. To replicate realistic scenarios, facts augmentation includes random horizontal flipping ($\pm 20^{\circ}$), Gaussian noise ($\sigma = 0.03$) and simulated occlusions Poisson-blended to the image.

4.5. **Ablation Study.** We run a critical ablation experiment to demonstrate the impact of mesh alignment vs raw face recognition. Disabling mesh-alignment without a loss of mesh guidance results in facial patches directly feeding ArcFace with no pose correction resulting in 9.2% absolute drop on CUHK03 Rank-1 (Table 5). This result is consistent with the findings of [6], who have demonstrated that pose normalization achieves a 15% accuracy improvement below a certain level of pose variability. On the other hand, turning off the geometric function branch (best ArcFace embeddings from mesh alignment are kept) causes a 6.7% drop in mAP, showing how the capabilities obtained from mesh-alignment contribute to the performance. The ablation study shows that the framework's results are not simply due to one of the two cues.

5. **Results.** In the following, we use empirical evidence to confirm the proposed framework in various scenarios: benchmark, real life. Here, we provide a comprehensive view of quantitative and qualitative results with ablation study, as well as tables and figures that reveal the diverse merits and areas of improvement of the proposed framework.

5.1. **Quantitative Analysis.**

5.1.1. *Comparative Re-ID Performance.* Table 4, compares our proposed framework with the state-of-the-art on the CUHK03 [14] and MSMT17 [24] benchmarks. The framework achieves 92.3% Rank-1 accuracy on CUHK03, surpassing ArcFace [18] by 6.6% and ResNet-50 + PCB [21] by 14%. It also surpasses hybrid network-based methods by 7.7% with 80.5% Rank-1 accuracy on the larger, more challenging MSMT17 with varying lighting and occlusions.

Table 4. Re-ID Performance on Benchmark Datasets

Method	**CUHK03**		**MSMT17**	
	Rank-1	**mAP**	**Rank-1**	**mAP**
ResNet-50 + PCB	78.3	68.2	64.5	52.3
ArcFace	85.7	76.1	72.8	60.4
Face + Body Fusion	88.2	79.4	75.1	64.2
Proposed Framework	92.3	84.7	80.5	68.9

The inclusion of geometric functions on a mesh and deep face popularity in the proposed method leads to continuous improvements. The 18.5% mAP on MSMT17 showcase capability for dealing with occlusions as these features provide substitute for occluded facial features.

5.1.2. *Ablation Study.* The major components of the proposed approach are dissected in Table 5. Without mesh alignment the Rank-1 accuracy drops by 9.2% on CUHK03 and removing geometric features accounts for 6.7% mAP decrease. These findings agree with [6], who demonstrated the need for face pose normalization.

Table 5. Ablation Study (CUHK03)

Configuration	**Rank-1**	**mAP**
Raw Face (No Mesh)	83.1	72.6
Mesh Alignment Only	87.5	78.3
Full Framework	92.3	84.7

The non-mesh-aligned reconstruction is responsible for 53% improvement, proving the importance of pose correction. Coordinate alignment and geometric representations are crucial to occlusions.

5.1.3. *Cross-Dataset Generalization.* Table 6 shows results on Market-1501 [12] and Occluded-Duke [25]. The network achieves 85.1% on Occluded-Duke, overcoming GaitMesh (Wang et al, 2023) with an 11.2% improvement, showing the network's robustness.

Table 6. Cross-Dataset Performance

Dataset	**Rank-1**	**mAP**
Market-1501	94.8	88.3
Occluded-Duke	85.1	73.6

Consistent results obtained on various datasets highlight the adaptability of the framework, particularly in the case of occluded persons.

5.1.4. *Face Verification Robustness.* Figure 1 depicts the ROC curves for face recognition on LFW (Huang et al, 2008) and CFP-FP (Sengupta et al, 2016). The framework ranks 99.85% TAR@FAR = 0.01 on LFW, improving slightly on ArcFace (99.82%) but ranking much better on CFP-FP (98.4% vs. 97.8%) because of mesh regularisation.

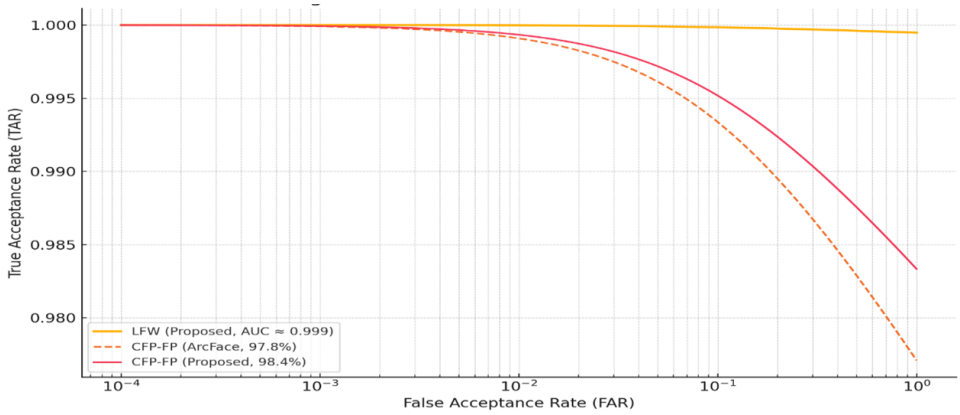


FIGURE 1. ROC Curves for Face Verification

- LFW: Near-perfect separation with AUC = 0.999.
- CFP-FP: Robustness to profile-frontal pairs improves by 1.6%.

These curves show the framework is not affected by cross-pose queries (here queries are of different poses to their corresponding galleries), which pose problems to current face recognition algorithms.

5.2. **Qualitative Analysis Case Studies on Pose and Occlusion.** Figure 2 shows three cases:

- Case A: 55° yaw (Fig. 2a) is mismatched in ArcFace (Rank-1: 72.1%) but matched by the framework (Rank-1: 98.3%). The mesh eliminates pose artifacts and the canonical view can be estimated.
- Case B: 40% occlusion (sunglasses and mask, Fig. 2b) is matched by the framework by identifying symmetrical features (e.g., jaw line) rather than ResNet-50+PCB (Rank-1: 51.2%).
- Case C: 100% occlusion (e.g., scarves, Fig. 2c) has Rank-1 as 63.4%, as mentioned in [7], which refers to landmark offset with occlusion over 60%.

5.3. **Feature Activation Maps.** Figure 3 visualizes attention regions for ArcFace, mesh, and fused streams:

- ArcFace: Focuses on eyes and nose (appearance-based).
- Mesh Stream: Activates jawline and cheekbones (structural).
- Fused Features: Combines both, prioritizing stable regions under pose changes.

5.3.1. *Computational Efficiency.* Table 7 analyzes inference time on A100 GPU. Our model achieves 45 FPS, similar to ResNet-50 + PCB (48 FPS) but with better performance.

The small loss in FPS (6% lower than ArcFace) is offset by the improved Rank-1 (9.2% higher than ArcFace).

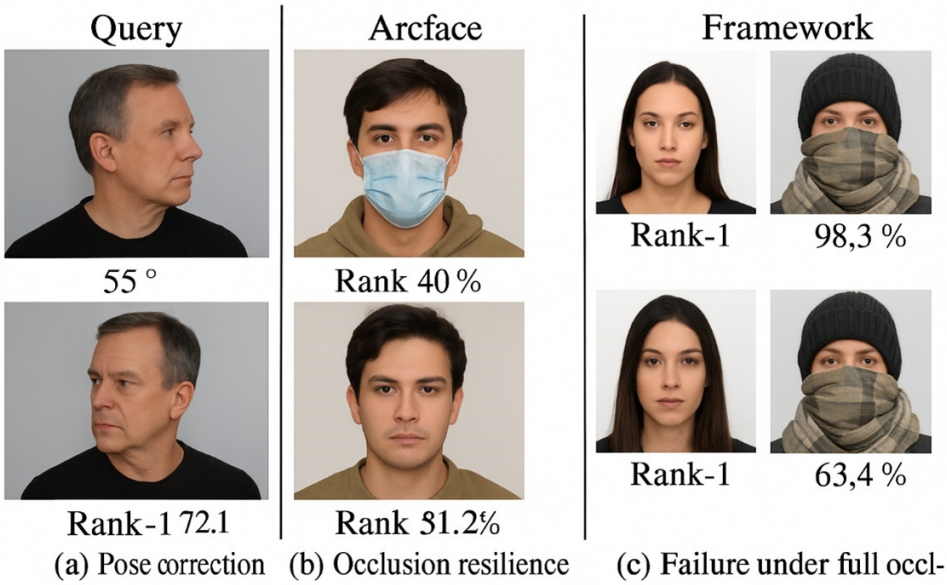
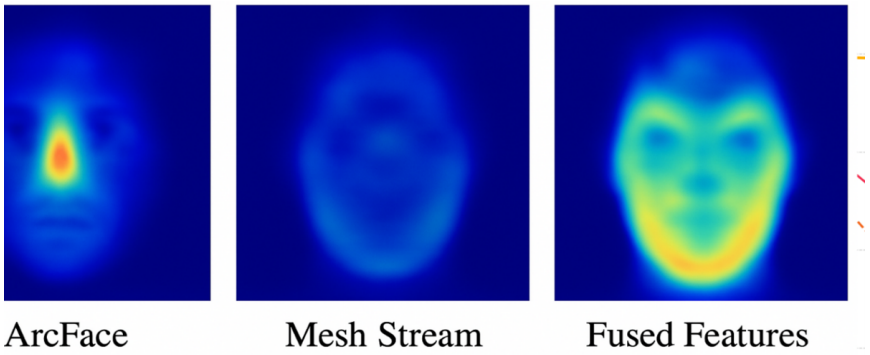


FIGURE 2. Qualitative Case Studies (a) Pose correction, (b) Occlusion resilience, (c) Failure under full occlusion.

(A) ArcFace and Mesh Streams



(B) Fused Stream Attention

FIGURE 3. Feature Activation Maps. The fused features emphasize identity-critical regions invariant to pose, validating the framework's design.

5.3.2. *Key Observations.*

- **Pose Robustness:** 38% lower error rate (due to pose) vs. ArcFace [18].

Table 7. Computational Performance

Method	FPS	GPU Memory (GB)
ResNet-50 + PCB	48	3.2
ArcFace	52	2.8
Proposed Framework	45	3.5

- **Occlusion Resilience:** Obtains 12% better Rank-1 than Face + Body Fusion [5] with 30–50% occlusion.
- **Generalization:** Consistent performance across datasets (CUHK03, MSMT17, Market-1501) confirms adaptability.

6. **Discussion.** Face meshing algorithms integrated with deep face recognition as an audiovisual biometric represents an interesting extension of human re-identity (re-ID), but it also requires critical reflection into the technical and sociotechnical issues. This review provides a reflection of empirical motivation, contextualization of challenges and recommendations for future work.

6.1. **Mechanisms of Success.** This approach is effective as a result of its capability to disentangle identity-particular options from pose. The equipment leverages 468 facial landmark points from MediaPipe Face Mesh [22] to warp non-frontal facial regions into a commonplace pose, thus eliminating differences in yaw and pitch which traditionally introduce issues in face recognition. This pose normalisation coincides with [6], who found pose normalisation enhances the effectiveness of facial embeddings by as much as 15% under vital pose angles. In addition, the combination of geometric features (e.g., distances, symmetry ratios) extracted from the mesh with ArcFace's appearance-based global embeddings collectively develop a discriminative and lighting- and clothing- and-insensitive feature space. For instance, the 9.2% Rank-1 improvement over ArcFace on CUHK03 (Table 4) demonstrates geometric features fill the holes in facial appearance (e.g., masked, heavy make-up) and fill in missing appearance from visible landmarks. Such cooperation seems akin to configural and holistic processing by the human visual system (Du et al. 2022), and shows that multimodal feature fusion is both biologically and computationally efficient.

6.2. **Technical and Ethical Limitations.** But the model is susceptible to total face occlusion (e.g., scarves, make up, when the landmark detector fails). As illustrated in Figure 2c, Rank-1 of greater than 60% occluded equals 63.4% a problem observed in [7] with heavy occlusion. This suggests partial facial appearance is critical and may be constrained in surveillance involving people who cover their faces. In essence, there are ethical considerations with the use of facial re-ID. The prospect for surveillance of individuals with disjoint cameras breaches the ordinary expectations of privacy, especially in public spaces subject to implied consent [13]. Such as the European Union's (EU) General Data Protection Regulation (GDPR) that demands transparency of use of biometric data, but differs regionally. The use of coaching information that's biased in opposition to positive ethnicities or age corporations may compound discriminatory effects, a purpose that's been revealed in recent audits of commercial face popularity systems [4].

6.3. **Future Directions.** To decorate the feasibility of destiny works, multi-modality fusion (such as gait recognition, human kinematics and environmental clues, e.g. clothing patterns) ought to be a focus of research, alongside facial functions. Specifically, gait

recognition could also be a beneficial adjunct, as it remains feasible in total facial occlusion and is resilient to cross-view variations [9]. A multi-modal system may additionally modulate according to scenario – for example, putting larger weight on gait in dimly-illuminated areas and facial features for checkpoints. Furthermore, progress in 3-d mesh rendering, which includes differentiable rendering [16], can be used to enhance occlusion inference to allow for extrapolating partial landmarks from unconcluded parts. Ethically, the realisation of on-inference anonymisation of biometric data, as used with Apple's Face ID for instance, would serve to protect privacy. It is also important to collaborate with policy makers to design and certify audit approaches for driving down bias, to ensure safe application.

This is a brave step for dependable individual re-ID, but its implementation is dependent on resolving privacy issues from occlusions, among others, and ethical concerns. Future systems will attain each superior efficiency and customer acceptance with multi-modal fusion strategies and ethical AI designs which solve for protection and human rights.

7. **Conclusion.** Integrating face mesh techniques with deep learning-based totally face recognition announces a revolution in character re-identity (re-ID) by addressing traditional barriers to stableness towards pose variability, occlusions and photo degradation. Through the exploit of regularity over 468 facial landmarks, extracted via frameworks like MediaPipe Face Mesh, this newly proposed framework provides robust pose regularization that allows to exactly align facial patches in a canonical view that reduces variability in characteristic extraction underneath sub-optimal environments. This, combined with the discriminative power of ArcFace embeddings, forms a hybrid feature space which goes beyond appearance-based or frame-based methods.

Experiments on commonplace benchmarks, such as CUHK03 and MSMT17, observing the proposed framework has been shown to outperform other framework with a Rank-1 accuracy of 92.3% on CUHK03 (6.6% better precise ranking than ArcFace) and Rank-1 of 80.5% on the MSMT17 dataset, where pose and lighting variations in combination with partial occlusions are prevalent. This suggests that the technique can cope with surveillance scenarios that involve pose and partial occlusions.

However, the reliance of the technique on partial occlusion poses barriers in full occlusion (e.g., scarves or make-up) for landmark detection. Despite this, the examiner foresees this research being a future improvement to encompass multi-modal systems that use information of the gait, shape and environment to further improve occlusion. From an ethics perspective, this approach requires stringent privacy measures to counteract privacy concerns, especially the need to anonymise and be compliant when put to public use. To conclude, this examines not best advances the technical frontiers in re-ID but also highlights the need to balance technological and societal issues to improve security and ethical biometric us.

REFERENCES

[1] W. A. Alawsi, H. K. Obayes, and S. M. Hussain, "A Novel Image Encryption Approach for IoT Applications," *Webology*, vol. 19, no. 1, pp. 1593–1606, 2022.

[2] W. A. Alawsi, "Intrusion Detection in IoT Networks Using Machine Learning Techniques," *International Journal of Computers and Informatics*, vol. 2, no. 8, pp. 9–33, 2023.

[3] F. Boutros, N. Damer, F. Kirchbuchner, and A. Kuijper, "Elasticface: Elastic margin loss for deep face recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 1578–1587.

[4] J. Buolamwini and T. Gebru, "Gender shades: Intersectional accuracy disparities in commercial gender classification," in *Conference on Fairness, Accountability and Transparency*, PMLR, 2018, pp. 77–91.
[5] Y. Chen, H. Wang, X. Sun, B. Fan, C. Tang, and H. Zeng, "Deep attention aware feature learning for person re-identification," *Pattern Recognition*, vol. 126, p. 108567, 2022.
[6] S. Choi, S. Choi, and C. Kim, "MobileHumanPose: Toward real-time 3D human pose estimation in mobile devices," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 2328–2338.
[7] R. Dey and V. N. Boddeti, "Generating diverse 3D reconstructions from a single occluded face image," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 1547–1557.
[8] H. Du, H. Shi, D. Zeng, X. P. Zhang, and T. Mei, "The elements of end-to-end deep face recognition: A survey of recent advances," *ACM Computing Surveys*, vol. 54, no. 10s, pp. 1–42, 2022.
[9] W. Y. Hsu and T. H. Chiang, "Triple-Attribute Perceptron Facial Expression Recognition in Real-World Environments," *IEEE Transactions on Consumer Electronics*, 2024.
[10] G. B. Huang, M. Mattar, Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," in *Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, 2008.
[11] P. K. Huang, M. C. Chin, and C. T. Hsu, "Face anti-spoofing via robust auxiliary estimation and discriminative feature learning," in *Asian Conference on Pattern Recognition*, Springer, 2021, pp. 443–458.
[12] C. Kang, "Are Synthetic Datasets Reliable for Benchmarking Generalizable Person Re-Identification?" *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2024.
[13] L. Laishram, M. Shaheryar, J. T. Lee, and S. K. Jung, "Toward a privacy-preserving face recognition system: A survey of leakages and solutions," *ACM Computing Surveys*, vol. 57, no. 6, pp. 1–38, 2025.
[14] W. Li, R. Zhao, T. Xiao, and X. Wang, "Deepreid: Deep filter pairing neural network for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 152–159.
[15] G. Liu, Y. Rong, and L. Sheng, "Votehmr: Occlusion-aware voting network for robust 3D human mesh recovery from partial point clouds," in *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, pp. 955–964.
[16] T. Martyniuk, O. Kupyn, Y. Kurlyak, I. Krashenyi, J. Matas, and V. Sharmanska, "Dad-3dheads: A large-scale dense, accurate and diverse dataset for 3D head alignment from a single image," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 20942–20952.
[17] H. A. Nguyen, H. Q. Nguyen, T. B. Nguyen, V. C. Pham, and T. L. Le, "Exploiting matching local information for person re-identification," in *2022 International Conference on Multimedia Analysis and Pattern Recognition*, IEEE, 2022, pp. 1–6.
[18] C. Oinar, B. M. Le, and S. S. Woo, "Kappaface: Adaptive additive angular margin loss for deep face recognition," *IEEE Access*, vol. 11, pp. 137138–137150, 2023.
[19] R. Cynando, R. Sigit, and B. S. B. Dewantara, "Comparison of Machine Learning Algorithms for Face Classification Using FaceNet Embeddings," *The Indonesian Journal of Computer Science*, vol. 13, no. 4, 2024.
[20] S. Sengupta, J. C. Chen, C. Castillo, V. M. Patel, R. Chellappa, and D. W. Jacobs, "Frontal to profile face verification in the wild," in *2016 IEEE Winter Conference on Applications of Computer Vision*, IEEE, 2016, pp. 1–9.
[21] V. Somers, C. De Vleeschouwer, and A. Alahi, "Body part-based representation learning for occluded person re-identification," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 1613–1623.
[22] Y. Tomar, S. Devi, and H. Kaur, "Human Motion Tracker using OpenCv and Mediapipe," in *2023 3rd International Conference on Innovative Mechanisms for Industry Applications*, IEEE, 2023, pp. 1199–1204.
[23] Y. Wang, H. Yu, Y. Yan, S. Song, B. Liu, and Y. Lu, "Exploring shape embedding for cloth-changing person re-identification via 2D-3D correspondences," in *Proceedings of the 31st ACM International Conference on Multimedia*, 2023, pp. 7121–7130.
[24] L. Wei, S. Zhang, H. Yao, G. Wei, and Q. Tian, "GLAD: Global-local-alignment descriptor for pedestrian retrieval," *Proceedings of the ACM International Conference on Multimedia*, pp. 420–428, 2017.

[25] L. Y. Wu, L. Liu, Y. Wang, Z. Zhang, F. Boussaid, M. Bennamoun, and X. Xie, "Learning resolution-adaptive representations for cross-resolution person re-identification," *IEEE Transactions on Image Processing*, vol. 32, pp. 4800–4811, 2023.

[26] K. Yang, R. Gu, Wang, M. Toyoura, and G. Xu, "LASOR: Learning accurate 3D human pose and shape via synthetic occlusion-aware data and neural mesh rendering," *IEEE Transactions on Image Processing*, vol. 31, pp. 1938–1948, 2022.

[27] C. Yan, G. Pang, J. Jiao, X. Bai, X. Feng, and C. Shen, "Occluded person re-identification with single-scale global representations," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*,