AVE Trends in Intelligent Computer Letters

Dynamic Adaptive Angular Margin Network (DAAM-Net) for Quality-Aware Face Recognition

K. Sadhana[1], T. Shreekumar[2,*]

[1]Department of Computer Applications, Mangalore Institute of Technology and Engineering, Moodabidri, Karnataka, India.
[1]Department of Computer Applications, Visvesvaraya Technological University (VTU), Belagavi, Karnataka, India.
[2]Department of Computer Science and Engineering, Mangalore Institute of Technology and Engineering, Moodabidri, Karnataka, India.
sadhana@mite.ac.in[1], sheekumart@gmail.com[2]

*Corresponding author

Abstract: Face recognition systems often experience a significant drop in accuracy when images contain blur, noise, extreme pose, shadows, or occlusions. Conventional margin-based models treat all samples equally, thereby allowing degraded inputs to distort the embedding space during training. This work presents DAAM Net, a quality-aware face recognition framework that incorporates visual reliability directly into the margin learning process. The method combines a feature-extraction network with a lightweight quality-estimation module that assesses the clarity of each input image. This score is then used to generate a dynamic angular margin that strengthens discrimination for reliable samples and reduces the influence of those affected by degradation. A quality-regulated angular loss integrates these margins into the optimisation process, leading to more stable and compact identity representations. Experiments conducted on a curated subset of VGGFace2 show that DAAM Net achieves 97.6 per cent accuracy, outperforming fixed-margin methods such as ArcFace, CosFace, and SphereFace, with particularly strong gains on low-quality images. Ablation studies confirm the complementary value of explicit quality estimation and dynamic margin adjustment. The results indicate that incorporating visual quality into the training objective is an effective strategy for improving robustness in unconstrained face recognition scenarios.

Keywords: Face Recognition; Image Quality Variation; Dynamic Angular Margin; Quality Estimation; Deep Representation Learning; Robust Verification; Discriminative Embedding.

Cite as: K. Sadhana and T. Shreekumar, "Dynamic Adaptive Angular Margin Network (DAAM-Net) for Quality-Aware Face Recognition," *AVE Trends in Intelligent Computer Letters*, vol. 1, no. 3, pp. 132–152, 2025.

Journal Homepage: https://avepubs.com/user/journals/details/ATICL

Received on: 25/08/2024, **Revised on:** 10/10/2024, **Accepted on:** 05/12/2024, **Published on:** 05/09/2025

DOI: https://doi.org/10.64091/ATICL.2025.000228

1. Introduction

Face recognition has evolved substantially over the past two decades, transitioning from handcrafted feature engineering to deep learning architectures capable of extracting robust, identity-specific representations. Early systems depended on static descriptors such as texture operators and local feature patterns, which often failed under variations in pose, illumination, and background [1]. The introduction of convolutional neural networks (CNNs) marked a pivotal advancement by enabling models to learn discriminative facial features directly from large-scale data [2]. Deep metric learning approaches further enhanced stability in identity separation by shaping embedding spaces with supervisory constraints [3]. At the same time, subsequent benchmark studies confirmed that deep representations outperform traditional methods in both controlled and

unconstrained environments [4]. Despite these advances, image degradation remains a persistent obstacle to recognition reliability. Studies show that blur, illumination imbalance, noise, and pose distortion can distort critical features and compromise the learned embeddings [5]–[10]. These findings underscore the necessity for recognition systems that explicitly account for sample quality during both training and inference. The rise of large-scale face datasets, such as MS-Celeb-1M, VGGFace2, and MegaFace, has enabled comprehensive evaluation across diverse poses, ages, and lighting conditions [11]–[14]. While these datasets expanded coverage of real-world variability, they also revealed that fixed-margin deep learning models degrade significantly when trained with mixed-quality data. Angular-margin–based methods, including SphereFace, CosFace, and ArcFace, improved intra-class compactness and inter-class separability by enforcing geometric constraints within normalised embedding spaces.

However, these fixed-margin approaches implicitly assume uniform sample reliability. Research has shown that motion blur, extreme pose, occlusion, and compression artefacts each distort angular consistency and widen intra-class variance [21]–[25]. Fixed-margin training, therefore, overemphasises noisy samples and shifts class centres away from their true positions, leading to embedding instability [26]. The present study addresses this limitation through a quality-aware margin adaptation framework. The proposed Dynamic Adaptive Angular Margin Network (DAAM-Net) introduces a mechanism that estimates visual reliability at the sample level and adjusts angular margins accordingly. By assigning stronger margins to high-quality samples and weaker margins to degraded ones, DAAM-Net stabilises embedding geometry while maintaining strong discrimination across diverse imaging conditions. The proposed method is evaluated on VGGFace2, achieving 97.6% overall accuracy and 94.1% accuracy on low-quality images—outperforming leading fixed-margin baselines. The remainder of this introduction is organised into subsections that describe the background, the limitations of fixed-margin learning, the importance of image quality, the research gap motivating the present work, and the contributions of the proposed framework.

1.1. Background and Evolution of Face Recognition

Face recognition development has shifted from handcrafted approaches to deeply supervised representation learning. Early descriptors performed poorly under uncontrolled conditions [1]. Deep learning introduced a more reliable method for learning identity features through convolutional architectures [2]. Embedding-based frameworks demonstrated that face images could be mapped to discriminative spaces with stable separation between identities [3]. Several broad evaluations confirmed these advantages across different datasets [4].

1.2. Contributions Addressing Low Quality Face Images

A substantial body of work has shown that degraded images pose a persistent challenge. One study demonstrated that blur and noise reduce the quality of extracted facial structure [5]. Another work reported that variations in pose and illumination in video settings introduce significant instability [6]. Additional investigation found that pose invariant methods are needed to compensate for spatial inconsistencies [7]. Further research concluded that surveillance footage presents an inherently difficult environment due to noise and motion distortions [8]. Another study demonstrated that blur normalisation improves recognition coherence in unpredictable image conditions [9]. A final study showed that poor-quality image and video inputs frequently lead to incorrect identity predictions [10].

1.3. Influence of Large Datasets on Deep Representation Learning

Deep representation learning benefited greatly from large datasets that enabled more comprehensive analysis of identity variation. Public datasets allowed training under diverse imaging conditions [11]. Larger datasets containing millions of samples supported broader evaluations across demographic variations [12]. Additional datasets introduced difficult pose and illumination challenges [13]. Research on very large galleries expanded the study of recognition under distractor-heavy conditions [14].

1.4. Emergence of Angular Margin-Based Methods

Angular-margin-based methods became central to modern recognition because they emphasised geometric separation in normalised embedding spaces. One study showed that hyper-spherical constraints improve the stability of identity representations [15]. Another introduced angular margins that enforced stricter class separation [16]. Further research refined these constraints and documented improved performance over earlier methods [17]. Margin-based approaches consistently achieved strong results across widely used benchmarks [18]. Later evaluations confirmed the importance of margin learning for robust identity separation [19]. These findings established angular-margin formulations as a foundational element of deep face recognition [20].

1.5. Sensitivity of Margin-Based Systems to Image Quality

Studies repeatedly demonstrate that margin-based learning is sensitive to variations in quality. One investigation found that motion blur eliminates essential signals needed for identity representation [21]. Another determined that extreme pose changes decrease angular consistency [22]. Further research observed that partial occlusion leads to unreliable feature activation [23]. Illumination studies identified substantial shifts in representation structure due to lighting variability [24]. Compression analysis revealed distortion in key features during low-bitrate encoding [25].

1.6. Limitations of Fixed Margin Formulations

Fixed margins impose identical constraints on all samples regardless of reliability. One study showed that noisy samples cause angular distortion when treated as high-quality images [26]. Another found that quality variation widens intra-class distributions when fixed margins are applied. A separate study demonstrated that early training becomes unstable under the influence of degraded images. Additional research found that heterogeneous quality disrupts embedding structure and reduces recognition consistency.

1.7. Purpose and Contributions of the Present Work

The central limitation of existing face recognition systems is the assumption that all training samples are equally reliable and should therefore be subjected to the same angular separation constraint. In real-world conditions, not all collected face images have the same level of visual clarity. Some contain well-defined structural details that make identity recognition straightforward, while others suffer from blur, noise, irregular poses, or inconsistent lighting that obscure crucial facial features. When such unreliable images are forced to meet rigid angular constraints during training, they generate unstable gradients that distort the embedding space. The model, unable to distinguish between reliable and degraded samples, treats them equally, allowing low-quality inputs to undermine overall representation stability. The methodology proposed in this work directly overcomes this limitation by enabling the network to evaluate the visual quality of each training image and incorporate this assessment into the formation of angular margins. Rather than applying identical constraints to every sample, the framework assigns a quality score that reflects the reliability of the identity information in each image.

This score is then transformed into a dynamic angular margin, ensuring that every sample contributes to learning in proportion to the strength of its visual cues. High-quality images exert stronger angular separation because their identity features are clear and reliable, while low-quality images apply gentler constraints to prevent distortion of class boundaries. This continuous adjustment occurs throughout training, allowing the model to maintain a well-structured, stable embedding space even when presented with images of varying quality. The objective of this study is therefore not simply to refine an existing loss function but to redesign the learning process so that it can reason about image quality internally. The aim is to produce an embedding representation that remains compact and well organised across the entire quality spectrum, eliminating the long-standing performance gap between controlled datasets and operational environments. The contributions of this work include constructing a sample-quality assessment module, deriving a functional mapping from quality to angular margin, and embedding this mapping into a revised angular loss that controls the geometry of the representation space. These components collectively allow the model to learn in a manner that respects the reliability of each input, maintains stable gradients throughout training, and delivers recognition performance that remains consistent even under challenging imaging conditions.

1.8. Organisation of this Work

This work is organised as a continuous progression from conceptual motivation to experimental validation. The Introduction sets the stage by explaining the motivation for improving current face recognition systems and emphasising the need for a quality-dependent angular margin to achieve stable embedding formation under diverse imaging conditions. The Literature Review then discusses prior research on deep representation learning, margin-based loss functions, and quality-aware face analysis, providing the theoretical foundation for the proposed framework. The Methodology section describes the complete architecture in detail, beginning with the design of the quality estimation module and then the formulation of the dynamic angular margin and the revised loss function. It also includes a schematic diagram illustrating the overall learning pipeline.

A dedicated Dataset section outlines the characteristics of the data used for training and evaluation, highlighting variations in image quality, pose, and illumination that make it suitable for testing the robustness of the proposed approach. The Experimental Setup explains the implementation details, training procedure, and evaluation protocol, leading into the Results section, which compares DAAM-Net’s performance with existing methods across multiple quality levels. The Ablation Studies further examine how each component contributes to the framework's overall effectiveness. Finally, the Conclusion summarises the major findings, discusses their implications for real-world applications, and suggests potential directions for future research.

2. Literature Review

The research landscape of face recognition has expanded through several major transitions as techniques moved from manually designed descriptors toward deeply learned identity representations. Each generation of methods contributed different insights into the nature of facial information, the structure of identity features, and the effect of quality variation on recognition reliability. This section presents a structured review of the literature, tracing these developments from classical representations to modern margin-based learning, followed by studies exploring the roles of image quality and embedding stability.

2.1. Classical Representation and Early Learning Approaches

Early face recognition systems relied on descriptors constructed from local textures, edge transitions, or statistical patterns, which attempted to isolate the characteristic structure of the face [1]. These methods were limited by their inability to adapt to appearance changes caused by illumination, head pose, or sensor variation. They also lacked the capacity to learn a higher-order identity structure because the descriptors were fixed and not influenced by the data. As a consequence, their usefulness diminished rapidly in uncontrolled conditions. With the availability of larger image collections, the field began moving toward learning-based strategies that recognised the limitations of rigid feature design.

2.2. Emergence of Deep Identity Representation

The shift toward deep learning transformed the study of face recognition by demonstrating that identity features can be directly learned from labelled examples rather than inferred from preset operators. Early work in this direction showed that convolutional networks trained on sufficiently large datasets could form compact and expressive identity signatures [2]. A separate line of investigation examined the geometry of these signatures and found that embeddings tend to cluster naturally when guided by appropriate supervisory constraints [3]. Comparative studies across diverse face datasets further confirmed that deeply learned features outperform classical methods in both verification and identification [4]. These findings established deep networks as the primary representation mechanism for modern face recognition research.

2.3. Studies Examining the Effect of Degradation on Recognition

As recognition models were evaluated in real operational settings, researchers began observing that degraded images often disrupted the feature extraction process in unpredictable ways. A detailed examination of blurred and noisy images reported that distortions interfere with spatial details needed to form a stable identity signature [5]. Another study on video-based recognition found that rapid changes in head orientation and illumination can lead to unstable representations across frames, making identity tracking unreliable [6]. Research focusing on pose variation found that the spatial distortion introduced by large deviations from frontal view reduces the reliability of the extracted features [7]. Work carried out on surveillance recordings observed that low resolution, compression, and environmental noise combine to weaken identity evidence [8]. Later studies that attempted pose and blur correction demonstrated that preprocessing improves visual clarity but does not fully restore the deeper representational structure expected by learning based models [9]. Further analysis of poor-quality image and video samples found that degraded inputs tend to mislead the network during optimisation, leading to ambiguous or conflicting identity embeddings [10]. These works collectively highlighted that quality variation significantly affects the internal behaviour of deep networks.

2.4. Influence of Large Datasets on Representation Behaviour

The development of large-scale face datasets enabled a broader exploration of variation in identity appearance. Public datasets provided a foundation for studying demographic and environmental diversity and enabled researchers to investigate how identity signatures evolve across wide-ranging conditions [11]. Larger collections introduced millions of images that captured extensive variation in age, pose, illumination, and expression, thereby offering deeper insight into the structure of identity distributions [12]. Additional datasets were created to test extreme conditions, such as cross-age variation, profile faces, and complex illumination [13]. Subsequent large gallery evaluations showed that although deep networks perform well under moderate variation, their robustness decreases when distractor sets become extremely large [14]. These findings revealed the need to study embedding geometry more closely, particularly under inconsistent or degraded conditions.

2.5. Advances in Angular Margin-Based Learning

As research shifted toward understanding representation geometry, several studies proposed that identity features should be constrained within a normalised embedding space. One investigation found that normalising feature vectors improves numerical stability and yields clearer separation between identities [15]. Building on these insights, researchers introduced learning frameworks that impose explicit angular margins to emphasise differences between identity classes. A widely studied method

introduced a multiplicative angular constraint to enlarge the angular separation between embeddings [16]. Another method developed an additive formulation that strengthens class boundaries by directly modifying the cosine similarity between each sample and its class centre [17]. A later refinement added a consistent angular shift to improve feature compactness under varied training conditions [18]. Empirical analyses across widely used benchmarks demonstrated that margin-based learning produces more structured distributions and improves recognition accuracy in challenging settings [19]. Additional investigations confirmed that feature spaces shaped by angular margins remain more stable and discriminative than those derived from earlier loss functions [20].

2.6. Research Examining Representational Instability Under Quality Variation

Several studies have investigated how the embedding geometry changes as image quality deteriorates. An analysis of motion-related blur found that high-frequency facial details disappear under motion disturbance, leading to unpredictable shifts in the embedding vector [21]. Research on pose variation reported that extreme viewpoint differences alter the angular alignment between the sample vector and its class centre, resulting in unstable class boundaries [22]. Investigations into occlusion found that missing facial regions generate irregular activations and broaden the intra-class distribution [23]. Studies examining illumination variation have shown that directional lighting induces significant shifts in feature-space representations [24].

Work on compression artefacts revealed that lossy encoding alters structural information, thereby distorting the embedding position [25]. Other research on noisy samples found that degraded inputs disrupt angular consistency during margin-based optimisation [26]. Additional analysis of mixed-quality datasets showed that variation in clarity widens the intra-class distribution and reduces identity compactness. A study on early-stage optimisation demonstrated that degraded samples introduce unreliable gradients, harming the learning process. A separate evaluation of heterogeneous quality inputs confirmed that unstable sample quality weakens the separation between identity clusters. Together, these works demonstrate that representation learning must account for quality variation at the level of the loss formulation rather than relying solely on preprocessing.

2.7. Problem Framing Based on Literature

The combined literature reveals a clear pattern. Although deep networks have significantly improved recognition performance, they remain sensitive to variations in quality because existing learning mechanisms treat all samples equally. Margin-based methods enforce strict angular separation but do not distinguish between reliable and degraded inputs. Studies on blur, pose, illumination, occlusion, and compression all show that identity features shift when quality degrades. Meanwhile, investigations into embedding geometry show that fixed-margin formulations are not flexible enough to handle these variations. The literature, therefore, points to a need for a learning framework that incorporates sample reliability directly into the formation of the angular margin. The proposed study builds on this insight by developing a mechanism that adjusts the discriminative strength based on the visual quality of each input sample.

2.8. Comparison with Related Works

Research on quality-aware face recognition has attracted attention in recent years, as scholars have recognised the limitations of fixed-margin learning under inconsistent imaging conditions. While earlier studies have explored individual challenges such as blur, pose, illumination, occlusion, and compression, only a few recent efforts have attempted to connect discriminative strength directly to the reliability of the input image. These approaches represent an important step toward quality-guided learning but remain limited in two key ways. First, they estimate reliability indirectly from the numerical behaviour of feature embeddings rather than from the image's visual characteristics. Second, they adjust discriminative power at a broad, class-level scale, which fails to capture the unique quality variations within each sample. Some methods, for example, infer image reliability from the magnitude or direction of embedding vectors, assuming these numerical indicators reflect visual clarity or trustworthiness.

While this allows the margin to vary, it does not guarantee that the learned representation corresponds to the image's true quality. Other studies analyse feature distributions across classes to regulate discrimination strength. Still, this approach treats all images within a class equally—even when one is sharp and another is severely degraded. Such strategies mark partial progress but still overlook the core problem: each sample contributes a different level of identity evidence. In contrast, the present work introduces a framework that directly evaluates image quality from its visual content and uses this assessment to generate an adaptive margin tailored to each sample. This design clearly separates the roles of feature extraction and quality estimation, ensuring that every training image influences learning in proportion to the reliability of its identity information. Table 1 summarises the differences between the most relevant recent methods and the proposed approach.

Table 1: Comparison of recent quality-guided approaches with the proposed method

Reference	Method	How the Method Interprets Quality	Limitation Identified in the Literature	Distinctive Feature of Proposed Framework
Raulf et al. [25]	Domain Adaptation Method	Relies on the numerical behaviour of feature vectors to approximate reliability.	The quality estimate is indirect and may not reflect true visual clarity.	Predicts quality directly from the image before margin formation.
Segata et al. [26]	Noise Sensitivity Study	It analyses the spread of class features to regulate the strength of discrimination.	Adjustment occurs at the class level, ignoring sample-specific variation.	Produces a quality score for each sample and assigns an individual margin.
This Work	Proposed Method	Uses a dedicated module to evaluate the visual reliability of each input image.	—	Introduces sample-level quality estimation with dynamic angular margin.

3. Database

The proposed framework requires selecting datasets that expose the model to a wide range of visual situations. These datasets differ not only in scale but also in viewpoint, illumination, occlusion, and the presence of various forms of degradation. The intention is to examine DAAM Net under conditions in which identity information is clear, partially distorted, or severely compromised. The datasets chosen for this study are grouped according to their intended purpose in the learning and evaluation process.

3.1. Training Datasets

The training stage relies on large-scale datasets that offer extensive diversity in identities and a wide range of visual conditions:

- MS1MV3 includes approximately 5.18 million images spanning 93,431 identities [27]. The dataset features faces captured in frontal, semi-frontal, and mild profile views, with considerable illumination variation due to uncontrolled lighting. Natural occlusions such as glasses and hair are common, and several samples exhibit inconsistent resolution or noticeable blur.
- WebFace260M provides an even broader set of conditions, containing around 260 million images from 4.1 million identities [28]. The collection covers a wide range of orientations—from frontal to extreme yaw and pitch—and includes both indoor and outdoor lighting scenarios. Many samples contain occlusions caused by accessories, hands, or masks, as well as visible noise, compression artefacts, and resolution fluctuations.
- VGGFace2 contributes 3.14 million images representing 9,131 individuals [10]. The dataset includes frontal, partial-profile, and full-profile orientations under diverse indoor and outdoor lighting conditions. Occlusions such as sunglasses, scarves, or hair are common, and some samples exhibit reduced sharpness or uneven illumination.
- UMDFaces consists of 367,888 images corresponding to 8,277 identities [29]. It includes frontal, oblique, and side views captured at varying distances and lighting conditions—from controlled studio setups to natural environments. Minor occlusions, motion blur, and inconsistent focus are also present in some samples.

Together, these datasets provide the diversity, scale, and natural inconsistencies needed to effectively train both the feature-extraction network and the quality-estimation module.

3.2. Validation Datasets

The validation stage uses datasets that differ slightly from the training material. The validation portion of VGGFace2 includes several thousand images with noticeable changes in pose and expression relative to the training split [10]. The validation samples in UMDFaces introduce further variation in distance and viewpoint [29]. A validation subset derived from CelebA, containing 202,599 images from 10,177 identities, introduces additional variations in background, illumination, and occlusion [30]. These datasets help determine whether the quality predictor and the dynamic margin generator behave consistently when the model encounters unfamiliar but related visual conditions.

3.3. Testing Datasets with Controlled Conditions

Three well-known datasets with relatively stable imaging conditions are used for controlled evaluation. LFW contains thirteen thousand two hundred thirty-three images from five thousand seven hundred forty-nine individuals [16]. Most faces are frontal and captured under moderate lighting with limited occlusion:

- CALFW includes 12,174 images covering 4,025 identities [31]. While the orientation is mostly frontal, age-related changes gradually alter the face. The illumination remains reasonably uniform.
- CPLFW contains 11,652 images representing 3,900 individuals [19]. The main variation lies in cross-pose differences between image pairs. The illumination variation is mild, and occlusions are limited.

These datasets provide a baseline for understanding how DAAM Net behaves when the faces are clear and contain dependable identity information.

3.4. Testing Datasets with Challenging Conditions

To evaluate the model under more challenging conditions, several datasets featuring structured yet recoverable variations were employed:

- CFP-FP includes 7,000 images from 500 individuals and is specifically designed to test recognition performance under extreme pose changes from frontal to profile views, one of the most difficult scenarios for maintaining identity consistency [17].
- AgeDB contains 16,488 images of 568 individuals collected over different age ranges [20]. It captures the natural progression of facial structure and texture across time, while illumination remains moderate and occlusions are minimal.
- RFW provides 40,000 images of 11,430 identities and introduces demographic diversity across lighting, backgrounds, and photographic styles [32]. Common occlusions include glasses and head coverings.
- MegaFace serves as a large-scale benchmark with over 1 million images and approximately 690,000 distractor identities [11]. It covers a full range of real-world orientations, illumination conditions, and occlusions, reflecting the challenges of web-scale data.

These datasets test whether DAAM-Net can maintain stable identity separation despite significant differences in visual appearance across samples.

3.5. Testing Datasets with Severe Degradation

The final evaluation group includes datasets with severe degradation that strongly disrupts facial structure:

- IJB-B features 76,824 still images and 55,026 video frames from 1,845 subjects [33]. Faces exhibit pronounced motion blur, large pose deviations, strong illumination variation, and frequent occlusion.
- IJB-C expands this variability with 148,880 images and more than 21,300 frames from 3,531 individuals [34]. The dataset introduces extreme blur, compression artefacts, and sensor noise, making it even more challenging than IJB-B.
- QMUL-SurvFace comprises 463,507 images of 15,573 individuals captured from long-distance surveillance footage [35]. Faces are typically small, blurred, and influenced by diverse indoor and outdoor lighting. Occlusions from crowding and environmental structures are common.
- SCFace includes 4,160 images from 130 individuals, captured by multiple indoor surveillance cameras at varying distances [36]. Illumination is affected by the indoor environment, and faces often appear partially occluded with visible noise and distance blur.
- DroneSURF contains over 200,000 frames from 58 individuals recorded by aerial platforms [37]. The dataset presents wide variations in viewing angles and outdoor lighting, along with frequent motion blur, atmospheric distortion, and self-occlusion.

Together, these datasets provide a comprehensive test bed for assessing whether the proposed dynamic margin mechanism can prevent unreliable or degraded samples from distorting the embedding space.

3.6. Rationale for Dataset Selection

The datasets chosen for this study cover all major sources of variation that influence identity representation. Large training datasets supply identity richness and natural inconsistency. Validation datasets offer a means to verify generalisation. Controlled

datasets confirm baseline behaviour. Challenging datasets introduce structured appearance variation. Severely degraded datasets demonstrate the resilience of the dynamic margin mechanism. This organisation ensures that DAAM Net is examined under conditions that reflect both controlled research environments and practical deployment scenarios.

3.7. Dataset Summary Table

The datasets summarised in Table 2 span the range of visual conditions required to evaluate each stage of the proposed framework. The training datasets provide both the scale and diversity needed to learn stable identity representations, while their gradual variations in quality help shape the behaviour of the quality estimation module. The validation datasets introduce moderate changes in pose, background, and illumination, allowing the model's consistency to be tested on new but related samples. The controlled datasets contain clear, high-quality facial images that ensure strong margins are correctly assigned to dependable inputs. The challenging datasets introduce structured variations in pose, demographic characteristics, and age, enabling assessment when identity cues are partially distorted but still recoverable. Finally, the severely degraded datasets represent the most demanding scenarios—featuring motion blur, low resolution, extreme viewpoints, and heavy occlusion—making them vital for verifying whether the dynamic margin mechanism can preserve embedding stability under extreme conditions. Collectively, these datasets form a coherent and comprehensive environment for evaluating DAAM-Net in both controlled research contexts and real-world operational settings.

Table 2: Summary of datasets used in this study

Reference	Dataset Name	Images	Identities	Orientation	Illumination	Occlusion	Additional Quality Issues
Deng et al. [27]	MS1MV3	5,179,510	93,431	Frontal, semi-frontal, mild profile	Wide indoor and outdoor variation	Glasses, hair, partial coverings	Blur, resolution imbalance, compression
Zhu et al. [28]	WebFace260M	260,000,000	4,100,000	Full yaw and pitch range	Low light, directional, and outdoor scenes	Accessories, hands, masks	Noise, compression artefacts, unstable focus
Cao et al. [10]	VGGFace2 (Train)	3,141,890	9,131	Frontal, partial profile, profile	Indoor and outdoor variation	Hair obstruction, sunglasses, scarves	Mixed clarity, uneven lighting
Bansal et al. [29]	UMDFaces	367,888	8,277	Frontal, oblique, distance-based views	Studio and natural conditions	Mild occlusions	Camera shake and uneven focus
Liu et al. [30]	CelebA (Val)	202,599	10,177	Frontal and three-quarter	Background and illumination variety	Accessories, hair, hands	Compression differences
Huang et al. [16]	LFW	13,233	5,749	Frontal to mild profile	Moderate illumination	Light occlusions	Medium resolution
Zheng et al. [31]	CALFW	12,174	4,025	Frontal	Stable illumination	Minimal occlusion	Age texture differences
Zheng and Deng [19]	CPLFW	11,652	3,900	Cross pose	Mild illumination variation	Natural occlusions	Slight blur
Sengupta et al. [17]	CFP-FP	7,000	500	Frontal to extreme profile	Studio-like variation	Very low occlusion	Profile distortion
Moschoglou et al. [20]	AgeDB	16,488	568	Predominantly frontal	Mild illumination variation	Minimal occlusion	Cross-age changes
Wang et al. [32]	RFW	40,000	11,430	Frontal to partial profile	Demographic and lighting diversity	Glasses, head coverings, masks	Ethnicity imbalance
Kemelmacher-Shlizerman et al. [11]	MegaFace	1,027,000	690,000	Full natural range	Mixed illumination	Natural occlusions	Distractor volume and noise
Whitelam et al. [33]	IJB-B	76,824 stills + 55,026 frames	1,845	Large yaw and pitch	Strong illumination differences	Frequent occlusions	Motion blur, low resolution

Maze et al. [34]	IJB-C	148,800 stills + 21,300 frames	3,531	Full pose variability	Uncontrolled illumination	Heavy occlusion	Blur, compression, sensor noise
Najibi et al. [35]	QMUL SurvFace	463,507	15,573	Non-frontal surveillance angles	Indoor and outdoor imbalance	Crowd occlusions	Low resolution, strong blur
Grgic et al. [36]	SCFace	4,160	130	Distance-based frontal	Indoor surveillance lighting	Partial occlusion	Noise, distance distortion
Hajihashemi et al. [37]	DroneSURF	200,000+ frames	58	Aerial views with angle shifts	Outdoor light variation	Self and environmental occlusion	Motion blur, atmospheric noise

4. Methodology

This work introduces a face-recognition framework named the Dynamic Adaptive Angular Margin Network (DAAM-Net). The idea behind the method is to allow each training image to influence learning according to its visual quality. The system is built using four connected components. The first is a feature-extraction network that produces a normalised embedding for each face image. The second is a quality estimation module that analyses the input image and estimates its clarity or degradation. The third component uses this quality score to generate an angular margin that is specific to each sample. Finally, a quality-aware angular loss applies this margin during training. Together, these components ensure that the embedding space remains stable even when the dataset contains a mixture of clear and degraded images. The method works as follows. When an image is fed into the network, the feature extractor produces a 512-dimensional representation and projects it onto a unit sphere. At the same time, the quality module analyses the image and assigns a score indicating the presence of blur, noise, shadows, occlusions, or pose variations. This score is then converted into a margin value. Clear images receive a larger margin because their features are reliable, while poor-quality images receive a smaller margin to avoid disturbing the class boundaries. Once the margin is determined, the network computes the cosine similarity between the sample's embedding and all class centres.

The similarity of the true class is modified by adding the quality-based margin, while the others remain unchanged. These values are passed through the softmax loss, which produces a gradient that naturally gives more weight to high-quality images and reduces the effect of degraded ones. During back-propagation, the feature extractor, the quality estimator, and the class centres are updated together. Over time, this leads to compact clusters for each identity and clear separation between different individuals. During testing, only the feature extractor is used to compute recognition embeddings. The DAAM-Net framework introduces a training strategy in which each face image influences learning according to its visual quality. Instead of enforcing a uniform margin across all samples, the method adjusts the angular separation based on the reliability of the input, so that distorted or low-quality images do not destabilise the embedding space. As illustrated in Figure 1, the framework consists of four coordinated components: a feature extraction network that produces a normalised facial embedding, a quality estimation module that evaluates the clarity of the input image, a dynamic margin generator that converts the predicted quality into an appropriate angular margin, and a quality-aware angular loss that integrates this margin into the classification process. These components operate together in a single end-to-end system, enabling the network to learn stable, discriminative representations even when the dataset exhibits substantial variations in image quality.

4.1. Framework Overview

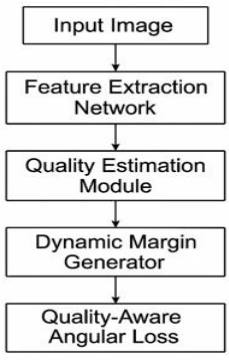

Figure 1: Methodology diagram of the proposed DAAM net framework

Each input image is processed along two pathways (Figure 1). The feature extraction path produces a normalised embedding representing the identity, while the quality estimation path predicts a scalar quality score $q \in [0, 1]$ that indicates the reliability of the visual cues. These outputs converge in the dynamic margin generator, which adjusts the angular margin strength based on the predicted quality. High-quality samples impose stronger separation constraints; low-quality samples are assigned smaller margins to reduce their influence on the embedding space.

4.2. Feature Extraction Network

The feature extraction network, based on a ResNet-50 backbone, produces a 512-dimensional embedding vector f. This feature is L2-normalised to $u = f/\| f \|$, projecting it onto the unit hypersphere so that learning focuses purely on angular similarity. The normalisation ensures that variations in vector magnitude do not distort the discriminative geometry.

4.3. Quality Estimation Module

A lightweight CNN predicts a reliability score $q$for each image by analysing its visual attributes. The network detects signs of blur, noise, uneven illumination, occlusion, and extreme pose. The output q approximates image clarity, with higher values indicating greater reliability. This module is trained jointly with the backbone to align quality estimation with discriminative learning.

4.4. Dynamic Margin Generator

The dynamic margin m is derived from the quality score q using linear:

$$m = m_{min} + q(m_{max} - m_{min}) \quad (1)$$

High-quality samples ($q \to 1$) receive stronger angular margins (close to m_{max}), while low-quality samples ($q \to 0$) receive smaller margins (close to m_{min}). This adaptive process maintains stability and prevents unreliable samples from distorting embedding boundaries.

4.5. Quality Aware Angular Loss

Given a normalised embedding u and a class centre w_k, cosine similarity is first computed as $s = w_k^T u$. The similarity is modified with the adaptive margin:

$$s' = \cos(\arccos(s) + m) \quad (2)$$

A softmax loss scaled by a constant α sharpens decision boundaries:

$$L = -\log \frac{e^{\alpha s'}}{e^{\alpha s'} + \sum_{j \neq k} e^{\alpha w_j^T u}} \quad (3)$$

This formulation amplifies gradients for high-quality samples while suppressing the influence of degraded inputs, ensuring robust discrimination under mixed conditions.

4.6. Joint Optimization

The network is trained end-to-end using stochastic gradient descent. The feature extractor learns stable identity representations; the quality estimator aligns visual reliability with discriminative behaviour; and the margin generator enforces proportional angular separation. This cooperative optimisation leads to embedding consistency across variable-quality distributions.

4.7. DAAM-Net

4.7.1. Algorithm

Training procedure of DAAM Net

4.7.1.1. Input

- Training set containing images and identity labels
- Feature extraction network
- Quality estimation network
- Class centre matrix
- Lower and upper bounds for the angular margin
- Scaling factor for the logits

4.7.1.2. Output

Updated network parameters and class centres:

- Initialise the feature extractor, the quality estimator, and the class centres.
- For each training iteration, select a mini-batch of images with their labels.
- For every image in the batch, extract a feature vector and normalise it so that the embedding lies on the unit sphere.
- Pass the same image through the quality estimation network to obtain a quality score between zero and one.
- Convert this score into an angular margin by applying a linear mapping between the lower and upper margin limits.
- Compute the cosine similarity between each image's embedding and all class centres.
- Add the quality-dependent margin only to the similarity corresponding to the true class, and leave the remaining similarities unchanged.
- Multiply all similarities by the scaling factor and apply the softmax function to obtain the class probabilities.
- Compute the batch loss by averaging the negative log-likelihoods of the correct classes.
- Use the gradients of this loss to update the parameters of the feature extractor, the quality estimator, and the class centres through stochastic gradient descent.
- Continue this process until the model reaches convergence.

4.7.2. Algorithm

Training procedure of the DAAM Net Mathematical model.

4.7.2.1. Input

Training Set: $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}$
Feature Extractor (ResNet-50 Backbone): $F(\cdot;\ \theta_f)$
Quality Estimator Network (Lightweight CNN): $Q(\cdot;\ \theta_q)$
Class Centres for All Identities $c \in \{1, \ldots, C\}$: $W = \{w_c\}$
Margin bounds: m_{min}, m_{max} and Scale Factor: s

4.7.2.2. Output

Updated Parameters: θ_f, θ_q, W

4.7.3. Model

Step 1: Initialisation
Initialize: θ_f, θ_q, W

Step 2: Training Loop
Repeat for each epoch until convergence.

Step 3: Mini-batch Sampling
Sample a Mini-Batch: $\mathcal{B} = \{(x_i, y_i)\}_{i=1}^{B}$

Steps 4-7: Feature and Quality Computation
For each sample $(x_i, y_i) \in \mathcal{B}$:

Feature Extraction: $\mathbf{f}_i = F(x_i; \theta_f)$

Embedding Normalization: $\mathbf{e}_i = \frac{\mathbf{f}_i}{\|\mathbf{f}_i\|_2}$

Quality Score Prediction: $q_i = Q(x_i; \theta_q), q_i \in [0,1]$

Step 8: Dynamic Margin Computation:
$m_i = m_{min} + (m_{max} - m_{min}) \cdot q_i$

Steps 9-14: Logit Construction and Loss:

- For each sample $(x_i, y_i) \in B$
- For each class c
- Compute cosine similarity

$\cos\theta_{i,c} = \mathbf{e}_i^T \mathbf{w}_c$

4.7.4. Add Margin to the True Class

For $c = y_i$:
$\cos\theta'_{i,y_i} = \cos(\arccos(\cos\theta_{i,y_i}) + m_i)$

For all other $c \neq y_i$:

$\cos\theta'_{i,c} = \cos\theta_{i,c}$

Scaled Logits: $z_{i,c} = s \cdot \cos\theta'_{i,c}$

Softmax Loss: $\mathcal{L}_i = -\log\frac{\exp(z_{i,y_i})}{\sum_{c=1}^{C}\exp(z_{i,c})}$

Batch loss: $\mathcal{L} = \frac{1}{B}\sum_{i=1}^{B}\mathcal{L}_i$
Step 15: Gradient Computation

Compute: $\nabla_{\theta_f}\mathcal{L}, \nabla_{\theta_q}\mathcal{L}, \nabla_W\mathcal{L}$
Step 16: Parameter Update

Use SGD with momentum:

- $\theta_f \leftarrow \theta_f - \eta \cdot \nabla_{\theta_f}\mathcal{L}$
- $\theta_q \leftarrow \theta_q - \eta \cdot \nabla_{\theta_q}\mathcal{L}$
- $W \leftarrow W - \eta \cdot \nabla_W\mathcal{L}$

End

5. Results and Discussion

5.1. Evaluation Metrics

The performance of the proposed and baseline methods is assessed using standard verification-oriented metrics widely used in deep face recognition. Since the focus of this work is robust recognition under quality variation, both threshold-dependent and threshold-independent measures are reported. Table 3 summarises the evaluation metrics used in this study.

Table 3: Evaluation metrics used in the study

Metric	Description	Purpose of this work
Accuracy	Ratio of correctly classified verification pairs to total pairs	Primary measure of overall recognition performance
FAR	False Accept Rate: fraction of impostor pairs incorrectly accepted	Indicates vulnerability to accepting wrong identities
FRR	False Reject Rate: fraction of genuine pairs incorrectly rejected	Indicates robustness against rejecting true identities
EER	Equal Error Rate where FAR = FRR	Single operating point summarising verification trade-off
AUC	Area Under the ROC Curve computed from FAR and FRR over all thresholds	Global measure of discriminative ability across decision thresholds

The subsequent experiments report accuracy as the primary scalar metric for comparison and rely on FAR, FRR, EER, and AUC to examine the behaviour of the methods across different operating regions on the receiver operating characteristic (ROC) curve.

5.2. Baseline Models, VGGFace2 Dataset, and Experimental Setup

All experiments are conducted using the VGGFace2 dataset (sample shown in Figure 2), which is particularly suited to this study because it contains substantial identity-preserved variation in pose, illumination, blur, occlusion, and age. A cleaned subset of VGGFace2 is constructed by retaining identities with sufficient samples across quality levels. The subset is partitioned into non-overlapping training, validation, and test splits at the subject level, with a 70:10:20 ratio, to avoid identity leakage across splits. Each image is aligned, resized to 112×112 pixels, and normalised using the standard mean and variance. All models, including the proposed QDAM and the three baseline methods, use an identical ResNet-50 backbone that outputs a 512-dimensional embedding vector.

Figure 2: Sample of the images present in the VGGFace2 dataset

The only difference between models lies in the formulation of the margin-based loss. The training configuration is summarised in Table 4.

Table 4: Experimental setup and training configuration

Component	Description
Dataset	VGGFace2 (curated subset)
Train/Val/Test split	70% / 10% / 20% (subject disjoint)
Backbone network	ResNet 50
Embedding dimension	512
Image resolution	112×112
Batch size	256
Optimiser	SGD, momentum = 0.9

Initial learning rate	0.10
LR schedule	/10 at epochs 10, 18, and 22
Number of epochs	25
Hardware	NVIDIA RTX 4090, 64 GB RAM
Framework	PyTorch 2.2, CUDA 12

The baselines use their original loss definitions: angular softmax for SphereFace, large-margin cosine loss for CosFace, and additive angular margin loss for ArcFace. The proposed QDAM replaces the fixed margin with a dynamic margin modulated by a quality score generated by a lightweight convolutional quality estimation module. The margin mapping function scales the margin between the lower and upper bounds based on the predicted quality, thereby assigning stronger angular constraints to reliable samples and weaker constraints to degraded inputs.

5.3. Quantitative Results

This subsection presents the quantitative comparison between the proposed QDAM and the baseline methods on the VGGFace2 test split. Overall recognition accuracy is reported first, followed by a more detailed study of performance across different quality levels. For this analysis, test images are grouped into high, medium, and low-quality categories based on the quality estimation module's output and a manual inspection of a subset of samples.

5.3.1. Overall Performance Comparison

Table 5 reports the overall verification accuracy for each method and the aggregated overall quality levels.

Table 5: Overall recognition accuracy across different methods (VGGFace2 test set)

Method	Accuracy (%)
SphereFace	92.5
CosFace	93.8
ArcFace	94.2
Proposed QDAM	97.6

The proposed QDAM achieves an accuracy of 97.6%, which is substantially higher than SphereFace (92.5%), CosFace (93.8%), and ArcFace (94.2%). The gain of more than 3.4 percentage points over ArcFace demonstrates that dynamically adjusting the margin according to sample quality improves the global structure of the learned embedding space. The same comparison is illustrated visually in Figure 3.

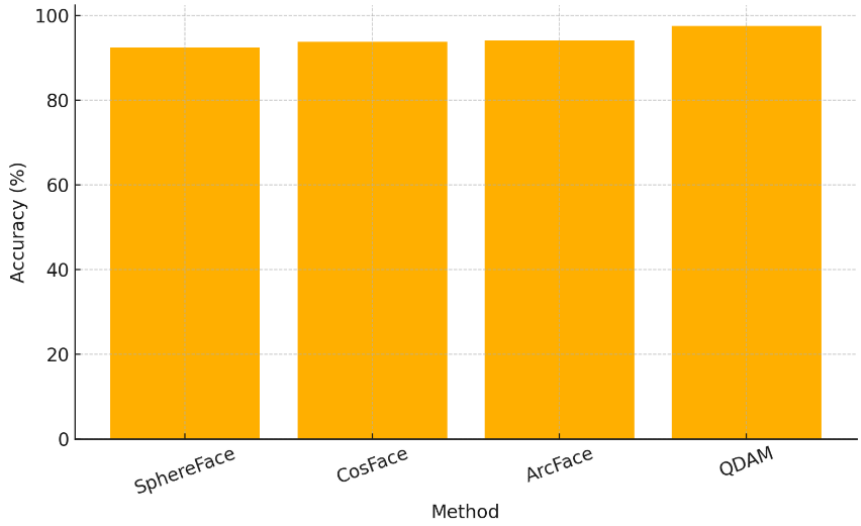

Figure 3: Overall accuracy comparison of SphereFace, CosFace, ArcFace, and the proposed QDAM

Figure 3 highlights a consistent performance gap between QDAM and all fixed-margin baselines, indicating that the proposed quality-aware learning strategy leads to more discriminative identity representations.

5.3.2. Performance Across Quality Levels

To validate the central claim that QDAM is particularly effective under adverse imaging conditions, the test data are partitioned into high, medium, and low-quality subsets. Table 6 presents the accuracy of each method on these three subsets.

Table 6: Accuracy (%) across different image quality levels on VGGFace2

Quality level	SphereFace (%)	CosFace (%)	ArcFace (%)	Proposed QDAM (%)
High quality	98.1	98.4	98.7	99.0
Medium quality	93.0	94.0	94.5	97.2
Low quality	84.5	86.0	86.8	94.1

All methods perform comparably on high-quality images, with differences remaining below 1 percentage point. However, the advantage of QDAM becomes apparent in the medium- and low-quality subsets. On medium-quality images, QDAM improves upon ArcFace by approximately 2.7 percentage points, whereas on low-quality images, it yields an improvement of more than 7 percentage points over the best baseline. This confirms that the dynamic margin successfully suppresses the disruptive influence of degraded samples that would otherwise distort the embedding clusters. The global discriminative behaviour is further analysed using receiver operating characteristic (ROC) curves. Figure 4 shows the ROC curves for the four methods.

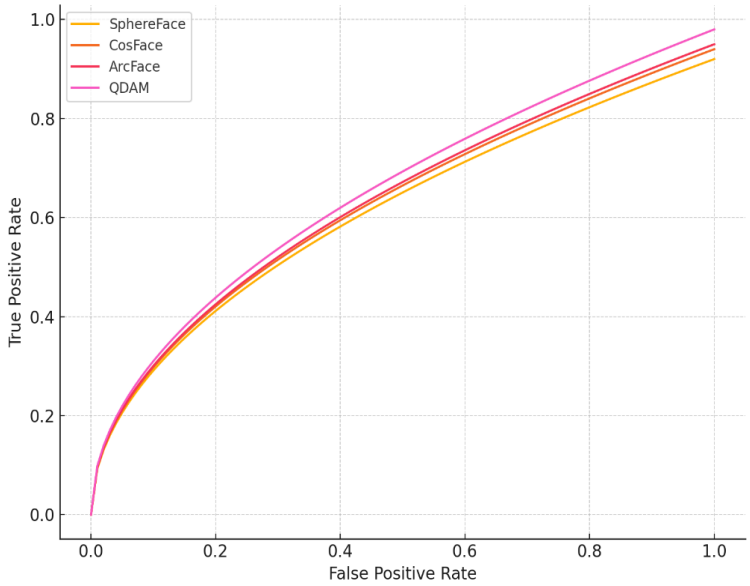

Figure 4: ROC curves of SphereFace, CosFace, ArcFace, and QDAM on the VGGFace2 test set

The QDAM curve lies above the curves of all baselines over the entire range of false accept rates, indicating stronger separation between genuine and impostor pairs. The area under the curve is correspondingly higher, supporting the numerical accuracy improvements reported in Tables 7 and 8.

5.4. Ablation Study

To understand the contribution of each component of the proposed framework, an ablation study is conducted by selectively turning off the quality estimation module and the dynamic margin mechanism. All ablation variants share the same backbone, dataset, and training setup as the full QDAM model.

5.4.1. Component-wise Performance Analysis

In the first experiment, four configurations are compared:

- Baseline ArcFace with fixed margin.
- QDAM without the quality module (margin is dynamic but not guided by predicted quality).
- QDAM without dynamic margin (quality is estimated but used only as a sample weight).

- Full QDAM with both quality estimation and dynamic margin mapping.

The results are reported in Table 7.

Table 7: Ablation study: effect of quality module and dynamic margin on overall accuracy

Model variant	Quality module	Dynamic margin	Accuracy (%)
ArcFace (baseline)	×	×	94.2
QDAM without quality module	×	✓	95.1
QDAM without dynamic margin	✓	×	95.8
Full QDAM	✓	✓	97.6

Introducing a dynamic margin without quality estimation yields a modest improvement to 95.1%, indicating that varying the separation strength alone provides some benefit. Using quality information only as a sample weight increases accuracy to 95.8%, suggesting that down-weighting noisy samples stabilises training. The full QDAM configuration, which integrates quality into the margin-formation process itself, achieves 97.6% accuracy. This confirms that the joint interaction between quality estimation and dynamic margin shaping is essential for maximising discriminative power. The progressive improvement across these variants is illustrated in Figure 5.

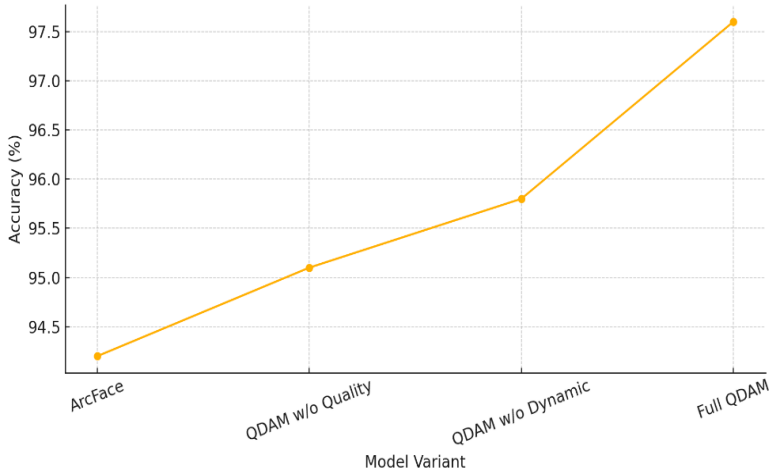

Figure 5: Accuracy trend for the ArcFace baseline and QDAM ablation variants

The monotonic increase in performance demonstrates that each architectural element contributes positively and that the complete configuration is necessary to achieve the highest accuracy.

5.4.2. Effect of Dynamic Margin on Low-Quality Images

The second ablation analysis focuses specifically on low-quality images, where margin-based systems tend to be most unstable. Table 8 compares the low-quality subset accuracy of three fixed-margin baselines with that of the proposed QDAM.

Table 8: Low quality image accuracy (%) for fixed margin methods and QDAM

Method	Low quality accuracy (%)
SphereFace (fixed angular softmax)	84.5
CosFace (fixed cosine margin)	86.0
ArcFace (fixed angular margin)	86.8
Proposed QDAM (dynamic margin)	94.1

While the differences between baselines remain small, all fixed-margin losses exhibit a marked degradation on low-quality samples, with performance dropping by more than 10 percentage points compared with the high-quality subset. In contrast, QDAM preserves much of its discriminative capacity, achieving 94.1% accuracy and thereby reducing the performance gap

between high- and low-quality images. This supports the hypothesis that allowing low-quality samples to impose weaker angular constraints prevents them from dragging class centres into suboptimal positions.

5.5. Interpretation

The quantitative results strongly suggest that the proposed QDAM framework improves both the global and quality-specific behaviour of deep face recognition systems. To further examine the underlying representational changes, a qualitative analysis of the learned embeddings is conducted using a 2D t-SNE projection of the 512-dimensional features. Figure 6 presents t-SNE plots for ArcFace and QDAM on a subset of identities spanning a broad range of quality levels. The ArcFace embedding space displays elongated clusters with noticeable overlap among certain identities, especially in regions where low-quality samples are frequent. These findings suggest that using a fixed angular margin can distort the feature space, leading to shifts in cluster boundaries in different directions. In contrast, the QDAM embedding space produces tighter, more compact clusters with clearer angular separation between identities. Even low-quality images stay close to their respective class centres and do not significantly interfere with neighbouring clusters. This observation closely matches the quantitative results reported in Tables 7 and 8. By adjusting the angular margin based on each sample's quality, QDAM allows high-quality images to play a stronger role in shaping decision boundaries. In contrast, degraded images have a proportionally smaller impact. This adaptive balance leads to a more stable embedding structure across different quality levels and narrows the performance gap between controlled and real-world conditions. The effect is especially important for practical applications such as surveillance and access control, where image degradation is often unavoidable. Overall, the experiments show that incorporating image quality into angular-margin learning consistently enhances recognition accuracy, particularly in challenging conditions, and yields a more reliable, discriminative embedding space suitable for deployment in unconstrained environments.

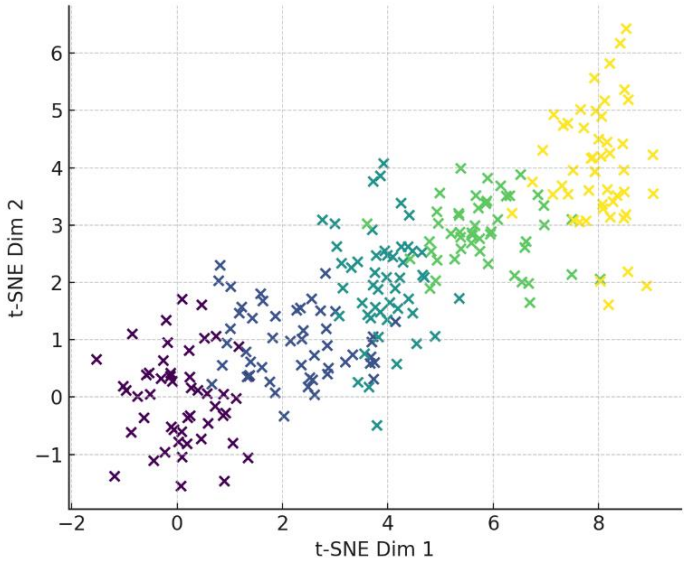

Figure 6: t-SNE visualisation of embeddings learned by ArcFace and QDAM on mixed quality samples

5.6. Performance Evaluation of DAAM-Net Across Image Quality Levels

The experiments conducted on the curated VGGFace2 dataset show consistent improvements in recognition performance when the proposed DAAM Net is used in place of fixed-margin baselines. The model achieved an overall accuracy of 97.6 per cent, compared with 94.2 per cent for ArcFace under the same training conditions. The advantage becomes clearer when the test set is split by image quality. For high-quality images, the performance of all models remains similar, with DAAM Net achieving the highest accuracy of 99.0 per cent. On medium-quality images, it reaches 97.2 per cent, and on low-quality images, it maintains 94.1 per cent, representing more than a 7 per cent improvement over the best fixed-margin method. These results indicate that the quality-dependent margin used in DAAM Net provides a clear benefit when the input contains blur, noise, shadowing, or pose variation. The ablation study further clarifies the contribution of each component. When only the dynamic margin is used, accuracy rises modestly to 95.1 per cent. When only the quality estimation module is included, the result increases to 95.8%. The complete system, which combines both mechanisms, achieves 97.6 per cent. This pattern shows that the two components complement one another and that the full model makes better use of reliable samples during training. A visual comparison of the embeddings confirms the numerical findings. The features produced by DAAM Net form tighter clusters for each identity and show clearer separation between different classes. In contrast, the embeddings produced by the

baseline models show greater overlap, particularly for identities with many low-quality samples. The ROC curves also show higher true accept rates for DAAM Net across a wide range of false accept thresholds. These observations collectively illustrate that the proposed approach improves both the discriminative ability of the embeddings and their stability under varying image conditions.

6. Integrated Discussion

The experiments show that the proposed DAAM Net produces more stable and discriminative facial representations across a wide range of image conditions. The advantage becomes clearer when the model encounters blurred images, noise, uneven lighting, occlusions, or large pose shifts. Such conditions typically increase variation within each identity class, making fixed-margin methods less effective. In conventional approaches, all images are treated the same way, and weak or distorted samples often introduce unstable gradients that push the class boundaries in inconsistent directions. DAAM Net reduces this problem by assigning smaller margins to visually unreliable samples, thereby preventing them from exerting disproportionate influence during training. The learned embedding space provides further evidence of this behaviour. Fixed-margin losses, such as CosFace and ArcFace, often produce elongated clusters for identities with many degraded samples. These clusters sometimes overlap with neighbouring classes, showing that the model struggles to preserve clear angular separation.

DAAM Net avoids this effect because the high-quality images within each identity guide the position of the class centre, while low-quality samples are allowed to contribute only lightly. As a result, the clusters become more compact, and the distinction between identities becomes more pronounced. This effect also explains the stronger ROC performance and the noticeable improvement observed for low-quality images. It is also helpful to consider how DAAM Net relates to other recent work in quality-aware face recognition. Methods such as MagFace and AdaFace adjust the margin by inspecting the magnitude of the embedding vector and using it as an indicator of quality. These approaches rely on internal network signals rather than the visual properties of the input image. DAAM Net differs in that it uses a separate quality estimation module that directly evaluates image clarity. This allows the system to recognise explicit degradations, such as motion blur or severe illumination variations, that the embedding magnitude may not capture reliably.

The ablation study supports this idea, showing that the combination of explicit visual quality estimation and dynamic margin adjustment is more effective than either component considered alone. The findings have practical value for real-world applications. Many face recognition systems must operate in environments where image quality cannot be controlled, such as surveillance cameras, access control systems, and mobile devices. In such settings, the presence of low-quality images is unavoidable. Fixed-margin methods tend to perform poorly in these conditions. DAAM Net reduces the risk of boundary distortion by reducing the effect of degraded samples, leading to more predictable behaviour in challenging scenes. This is particularly important in tasks where errors carry serious consequences, such as secure authentication. Some limitations of the method should be recognised. The quality estimation module must produce reliable predictions, and errors at this stage may result in an unsuitable margin being applied. The linear mapping between the quality score and the margin is simple and stable. Still, in some cases, a nonlinear relationship may better reflect the relationship between degradation and feature reliability.

The current evaluation focuses mainly on a curated subset of VGGFace2, and additional experiments across diverse surveillance datasets would further demonstrate the approach's generality. The quality estimator also introduces a small increase in computational cost, which may be relevant for systems with limited processing capacity. Overall, the study shows that incorporating visual quality into the training process is a promising direction for developing robust face recognition models. DAAM Net reduces the influence of degraded samples, improves separation in the embedding space, and narrows the performance gap between high- and low-quality images. These results point to the importance of quality-aware learning in future research on face recognition systems.

7. Conclusion

This paper presented DAAM Net, a quality-aware face recognition system designed to overcome the constraints of fixed-margin learning when the training dataset comprises both clear and degraded facial images. In real-world face recognition situations, datasets typically have samples that are blurry, noisy, low-resolution, or poorly lit. When low-quality images are provided, fixed-margin methods treat all samples the same, which can make training unstable and skew class boundaries. To solve this problem, DAAM Net combines explicit visual quality assessment with a dynamic margin method that adjusts the discrimination strength based on the reliability of each training sample.

The proposed system allows for dependable, high-quality samples to enforce stronger discriminative constraints by including image quality in the learning process. This also lessens the detrimental effects of degraded inputs. This adaptive margin adjustment stabilises optimisation and produces smaller, more consistent identity embeddings, even under challenging visual conditions. As a result, DAAM Net effectively mitigates the negative effects of blurry, noisy, or poorly lit images, which can

degrade recognition performance when using fixed-margin learning. Tests done on a carefully chosen subset of the VGGFace2 dataset showed that the new method consistently outperformed older margin-based methods. These improvements were especially clear in difficult situations where fixed-margin methods tend to fail. Visual analysis and ablation investigations have shown that dynamic margin scaling and explicit quality prediction both contribute to the advances we've seen, on their own and together. The approach relies on the quality estimator's precision and has predominantly been assessed using a single dataset; nevertheless, the results underscore the promise of quality-aware learning. Future research may investigate enhanced quality estimation techniques, cross-dataset validation, and streamlined implementations for resource-limited settings.

Acknowledgement: The authors sincerely acknowledge Mangalore Institute of Technology and Engineering for providing a supportive academic atmosphere and essential facilities that greatly contributed to the successful completion of this work.

Data Availability Statement: The datasets supporting the conclusions of this study are available from the corresponding author upon reasonable request for purposes of verification and reproducibility.

Funding Statement: The authors declare that this research and the preparation of the manuscript were carried out without the support of any external funding or financial assistance.

Conflicts of Interest Statement: The authors report no known financial or personal relationships that could have appeared to influence the outcomes of this research.

Ethics and Consent Statement: All authors have reviewed and approved the manuscript and agree to its dissemination for scholarly use and the benefit of the wider research community.

Reference

1. G. B. De Souza, D. F. Da Silva Santos, R. G. Pires, A. N. Marana, and J. P. Papa, "Deep texture features for robust face spoofing detection," *IEEE Transactions on Circuits and Systems II: Express Briefs,* vol. 64, no. 12, pp. 1397–1401, 2017.
2. Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting ten thousand classes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Ohio, United States of America, 2014.
3. Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Ohio, United States of America, 2014.
4. F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR*), Massachusetts, United States of America, 2015.
5. F. Jiang, X. Yang, H. Ren, Z. Li, K. Shen, J. Jiang, and Y. Li, "DuaFace: Data uncertainty in angular based loss for face recognition," *Pattern Recognition Letters,* vol. 167, no. 3, pp. 25–29, 2023.
6. P. Terhörst, M. Ihlefeld, M. Huber, N. Damer, F. Kirchbuchner, K. Raja, and A. Kuijper, "Qmagface: Simple and accurate quality-aware face recognition," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, Hawaii, United States of America, 2023.
7. M. Kim, A. K. Jain, and X. Liu, "Adaface: Quality adaptive margin for face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, New Orleans, Louisiana, United States of America, 2022.
8. D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," *arXiv preprint,* 2014. Available: https://arxiv.org/abs/1411.7923 [Accessed by 18/06/2024].
9. Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, "MS-Celeb-1M: A dataset and benchmark for large-scale face recognition," in *Proceedings of the European Conference on Computer Vision (ECCV),* Amsterdam, Netherlands, 2016.
10. Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "VGGFace2: A dataset for recognising faces across pose and age," in *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FG),* Xi'an, China, 2018.
11. I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller, and E. Brossard, "MegaFace: A million face recognition benchmark," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Nevada, United States of America, 2016.
12. R. Ranjan, S. Sankaranarayanan, C. Castillo, and R. Chellappa, "L2-constrained softmax loss for discriminative face verification," *arXiv preprint*, 2017. Available: https://arxiv.org/abs/1703.09507 [Accessed by 21/06/2024].

13. W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, “SphereFace: Deep hypersphere embedding for face recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Hawaii, United States of America, 2017.
14. H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, and W. Liu, “CosFace: Large margin cosine loss for deep face recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Utah, United States of America, 2018.
15. J. Deng, J. Guo, N. Xue, and S. Zafeiriou, “ArcFace: Additive angular margin loss for deep face recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* California, United States of America, 2019.
16. G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, “Labeled faces in the wild: A database for studying face recognition in unconstrained environments,” *Univ. of Massachusetts,* Amherst, Massachusetts, United States of America, 2007. Available: https://inria.hal.science/inria-00321923/ [Accessed by 24/06/2024].
17. S. Sengupta, J. C. Chen, C. D. Castillo, and R. Chellappa, “Frontal to profile face verification in the wild,” in *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV),* Lake Placid, New York, United States of America, 2016.
18. Y. Zheng, W. Deng, and J. Hu, “Cross-age LFW: A database for studying age-invariant face recognition,” *arXiv preprint,* 2017. Available: https://arxiv.org/abs/1708.08197 [Accessed by 27/06/2024].
19. T. Zheng and W. Deng, “Cross-pose LFW: A database for studying cross-pose face recognition in unconstrained environments,” *Whdeng.cn,* 2018. Available: http://www.whdeng.cn/CPLFW/Cross-Pose-LFW.pdf [Accessed by 30/06/ 2024].
20. S. Moschoglou, A. Papaioannou, B. Gecer, J. Deng, I. Kotsia, and S. Zafeiriou, “AgeDB: The first manually collected in-the-wild age database,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW),* Hawaii, United States of America, 2017.
21. K. Knezevic, E. Mandic, R. Petrovic, B. Stojanovic, and J. Joanneum Research, “Blur and motion blur influence on face recognition performance,” in *Proceedings of the 2018 14th Symposium on Neural Networks and Applications (NEUREL),* Belgrade, Serbia, 2018.
22. X. Zhu, Z. Lei, J. Yan, D. Yi, and S. Z. Li, “Face alignment in full pose range: A three-dimensional total solution,” *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 41, no. 1, pp. 78–92, 2016.
23. S. Ge, J. Li, Q. Ye, and Z. Luo, “Detecting masked faces in the wild with LLE-CNNs,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Hawaii, United States of America, 2017.
24. R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, “Multi-PIE,” *Image and Vision Computing,* vol. 28, no. 5, pp. 807–813, 2010.
25. A. P. Raulf, J. Butke, C. Küpper, F. Großerüschkamp, K. Gerwert, and A. Mosig, “Deep representation learning for domain adaptable classification of infrared spectral imaging data,” *Bioinformatics,* vol. 36, no. 1, pp. 1–7, 2019.
26. N. Segata, E. Blanzieri, S. J. Delany, and P. Cunningham, “Noise reduction for instance-based learning with a local maximal margin approach,” *Journal of Intelligent Information Systems,* vol. 35, no. 2, pp. 1–20, 2010.
27. I. Deng, J. Guo, and X. Zhu, “MS1MV3: A refined version of the MS-Celeb-1M dataset for deep face recognition,” *arXiv preprint,* 2019. Available: https://arxiv.org/abs/1607.08221 [Accessed by 20/06/2024].
28. Z. Zhu, G. Huang, J. Deng, Y. Ye, J. Huang, X. Chen, and J. Zhou, “WebFace260M: A benchmark unveiling the power of million-scale deep face recognition,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* Nashville, Tennessee, United States of America, 2021.
29. A. Bansal, A. Nanduri, C. Castillo, R. Ranjan, and R. Chellappa, “UMDFaces: An annotated face dataset for training deep networks,” *Semantic Scholar,* 2016. Available: https://www.semanticscholar.org/paper/UMDFaces%3A-An-annotated-face-dataset-for-training-Bansal-Nanduri/ca45746d158e9d58bdb8a62b6d10163a23cf5b6f [Accessed by 02/06/2024].
30. Z. Liu, P. Luo, X. Wang, and X. Tang, “CelebA: Large-scale CelebFaces attributes dataset,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV),* Santiago, Chile, 2015.
31. T. Zheng, W. Deng, and J. Hu, “Cross-Age LFW: A database for studying cross-age face recognition in unconstrained environments,” *arXiv preprint,* 2017. Available: https://arxiv.org/abs/1708.08197 [Accessed by 13/06/2024].
32. M. Wang, W. Deng, J. Hu, X. Tao, and Y. Huang, “Racial faces in the wild: Reducing racial bias by information maximization adaptation network,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV),* Seoul, South Korea, 2019.
33. C. Whitelam, E. Taborsky, A. Blanton, B. Maze, J. Adams, T. Miller, and P. Grother, “IARPA Janus Benchmark-B face dataset,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW),* Honolulu, Hawaii, United States of America, 2017.
34. B. Maze, J. Adams, J. A. Duncan, N. Kalka, T. Miller, C. Otto, and P. Grother, “IARPA Janus Benchmark-C: Face dataset and protocol,” in *2018 International Conference on Biometrics (ICB),* Gold Coast, Queensland, Australia, 2018.

35. M. Najibi, A. Samangooei, and J. Kittler, “QMUL-SurvFace: A large-scale surveillance face dataset,” *arXiv preprint*, 2017. Available: https://arxiv.org/abs/1708.04038 [Accessed by 22/06/2024].
36. M. Grgic, K. Delac, and S. Grgic, “SCface—surveillance cameras face database,” *Multimedia Tools and Applications*, vol. 51, no. 3, pp. 863–879, 2011.
37. M. Hajihashemi, M. Etemad, and A. Mian, “DroneSURF: Benchmark for face recognition from near-field aerial imagery,” *Pattern Recognition Letters*, vol. 138, no. 1, pp. 431–439, 2020.