IJCERT
Original Research Article

International Journal of Computer Engineering in Research Trends
e-ISSN: 2349-7084 | Volume 13, Issue No: 3, July–September 2026, Pp: 13–25 © 2026,

A Web-Based Smart Campus Identification Framework with YOLOv8n and Quality-Adaptive Temporal Face Fusion

P. Adityalakshmi[1] and K. Karthik[2]

[1]M.Tech Scholar, AI Branch, Department of Computer Science and Engineering, CVR College of Engineering, Hyderabad, India; Email: aithdya10@gmail.com

[2]Associate Professor, Department of Computer Science and Engineering, CVR College of Engineering, Hyderabad, India; Email: karthik.5786@gmail.com

* Corresponding Author: aithdya10@gmail.com

Abstract—Automated student identification in smart-campus environments requires dependable person localization, robust facial matching, rejection of non-enrolled individuals, and practical web-based monitoring. Single-frame recognition can become unreliable when surveillance imagery is affected by blur, low illumination, contrast variation, or over-exposure. This study presents a web-based smart-campus identification framework that combines the nano configuration of You Only Look Once version 8 person detection, FaceNet512 facial representation, and a training-free quality-temporal adaptive fusion mechanism for open-set recognition. The proposed method evaluates consecutive face observations using sharpness, brightness suitability, and contrast, retains the most informative frames, and combines their embeddings using quality-dependent softmax weights. A sequence-quality score also controls an adaptive acceptance threshold for distinguishing enrolled students from unknown individuals. Person detection was evaluated independently on the Common Objects in Context 128 subset, while open-set recognition was assessed using a controlled Labeled Faces in the Wild protocol containing 15 enrolled and 10 unknown identities. Under mixed image-quality conditions, the proposed configuration achieved 100 percent open-set identification accuracy, compared with 90.00 percent for a randomly observed single-frame baseline, 98.33 percent for best-quality single-frame selection, and 93.33 percent for unweighted five-frame averaging. Unknown-person rejection increased from 86.67 percent to 100 percent relative to the random single-frame baseline. The detector obtained a mean average precision at 0.5 intersection-over-union of 0.7606 with approximately 3.73 milliseconds inference time per image in the recorded accelerator experiment. These pilot results indicate that quality-aware temporal fusion can improve recognition robustness without retraining the facial embedding model.

Keywords—smart campus; open-set face recognition; YOLOv8n; FaceNet512; temporal fusion; face image quality

Received: 18 June 2026; Revised: 21 August 2026; Accepted: 18 September 2026; Published: 30 September 2026

DOI: 10.22362/ijcert/2026/v13/i3/v13i3002

1 Introduction

Smart-campus environments increasingly rely on connected cameras, web services, databases, and artificial-intelligence-based analytics to support attendance management, access supervision, and security monitoring. A practical student-identification system must do more than recognize a face under ideal conditions. It must first locate individuals in a video stream, obtain usable facial observations, compare them with an enrolled gallery, reject non-enrolled individuals, record identity events, and expose the results through an administrative interface. These requirements become difficult in unconstrained environments because surveillance imagery is frequently affected by illumination changes, pose variation, low resolution, blur, contrast loss, partial occlusion, and viewing distance.

Deep face recognition has substantially improved biometric identification by learning discriminative facial representations from large image collections. Wang and Deng reviewed the evolution of deep face recognition and showed that advances in architectures, loss functions, training databases, and evaluation protocols have enabled strong verification and identification performance while leaving cross-factor generalization as an important challenge [1]. A further concern is the distinction between closed-set and open-set operation. Geng *et al.* explained that open-set recognition must both classify known samples and reject classes that were not represented at training time [2]. This requirement is directly relevant to campus surveillance, where visitors and other non-enrolled individuals can legitimately appear in the camera stream.

Face-image quality has a strong effect on this decision. Best-Rowden and Jain showed that the utility of a face image for automatic recognition can be estimated and is closely related to matching performance [3]. Su *et al.* later developed a large face-image-quality database and a deep quality-assessment model, further demonstrating the value of estimating recognition utility under diverse acquisition conditions [4]. These studies motivate using quality as an operational signal rather than treating every observation as equally reliable.

Video also provides information that is unavailable in isolated still-image recognition. Lopez-Lopez *et al.* investigated open-set video face recognition in a streaming setting and demonstrated that identity evidence can be updated from sequential observations while maintaining unknown-person rejection [5]. Dong *et al.* likewise showed that temporal face

information can provide identity cues complementary to a single spatial frame [6]. Nevertheless, recurrent or transformer-based video architectures can add training and deployment complexity when a practical application already uses a pre-trained face-embedding model.

The problem becomes more severe when surveillance imagery is degraded. Ullah *et al.* reported substantial performance loss in low-resolution face recognition and proposed a degradation-aware distillation strategy to recover discriminative information [7]. Large-scale benchmarking by Melzi *et al.* also emphasized that pose, occlusion, demographic variation, and enrollment-test mismatch complicate face-recognition generalization [8]. Therefore, results measured under one controlled protocol should not automatically be interpreted as equivalent to full operational performance.

Open-set confidence itself can be unreliable. Jang and Kim showed that conventional classifiers can produce high-confidence predictions for unseen classes and developed a dedicated open-set learning strategy to improve rejection [9]. Such retraining is valuable but is not always convenient for an already deployed recognition backbone. Recognition-oriented quality assessment offers another route: Zhuang *et al.* explicitly considered brightness, contrast, blur, occlusion, and pose when estimating which face image is most useful for recognition [10].

This work therefore introduces a lightweight inference-stage mechanism rather than another recognition backbone. The operational detector is the nano configuration of You Only Look Once (YOLO) version 8, hereafter YOLOv8n. The proposed decision layer is termed Quality-Temporal Adaptive Fusion for Identification (QTAF-ID). It evaluates a short sequence of face observations, selects the most informative frames, fuses their FaceNet512 embeddings with quality-dependent weights, and adapts the open-set acceptance threshold according to fused sequence quality. No gradient update or fine-tuning of the person detector or facial embedding network is required.

The contributions are fourfold. First, a complete web-based smart-campus identification architecture integrates person detection, facial representation, open-set matching, student enrollment, persistent event storage, and administrative monitoring. Second, QTAF-ID combines frame-quality assessment, Top-K selection, quality-weighted temporal fusion, and a quality-adaptive identity threshold. Third, an open-set protocol compares random single-frame recognition, best-quality single-frame selection, unweighted temporal averaging, and the proposed method under identical gallery and test partitions. Fourth, the study reports a reproducible pilot robustness analysis under clean, low-light, blurred, low-contrast, and over-exposed observations while keeping person-detection and identity-recognition evidence explicitly separated.

The remainder of the paper is organized as follows. Section 2 reviews open-set face identification, surveillance recognition, quality assessment, temporal processing, and privacy-aware biometrics. Section 3 presents the proposed framework. Section 4 describes the datasets, experimental protocol, calibration, metrics, and implementation environment. Section 5 reports detection, recognition, statistical, degradation, threshold, and runtime results. Section 6 discusses practical implications and limitations, and Section 7 concludes the study.

2 Related Work

Research relevant to the proposed framework spans open-set face recognition, surveillance person analysis, quality assessment, temporal representation, privacy-preserving biometrics, fairness, and operational acceptance. The common problem is that a deployable identification system must reason about both the quality of its observations and the possibility that the observed identity is not represented in the enrollment gallery.

2.1 Open-Set Recognition and Surveillance Identification

Vareto *et al.* developed an open-set face-recognition approach using maximal-entropy and Objectosphere-based objectives, explicitly targeting unknown individuals that appear at operation time [11]. Their results demonstrate the importance of designing acceptance regions rather than assuming that every probe belongs to an enrolled identity. Irene *et al.* reviewed person-search methods for video surveillance and highlighted the need to coordinate localization, feature representation, retrieval, and computational efficiency in practical systems [12].

Pattnaik *et al.* presented a broad face-recognition taxonomy covering modality, dimensionality, and feature quality and emphasized that deployment conditions can be as important as model architecture [13]. Zennayi *et al.* proposed an unauthorized-access monitoring pipeline combining person detection, tracking, face localization, face recognition, confidence handling, and real-time surveillance processing [14]. Their system-level perspective is closely related to the present architecture, although the proposed QTAF-ID decision formulation is different.

2.2 Privacy, Temporal Information, and Quality

Privacy is a major concern whenever facial imagery is collected continuously. Wahida *et al.* proposed adversarial and distributed learning mechanisms for privacy-preserving smart-city face recognition [15]. In the educational context, Wang *et al.* showed that student acceptance of facial-recognition technology is influenced by trust and perceived privacy, psychological, and performance risks [16]. These findings indicate that technical accuracy alone does not determine whether a campus identification system is appropriate for deployment.

Temporal processing can also support both recognition and privacy. Leong *et al.* extracted temporal facial features while reducing reliance on directly visible facial appearance [17]. Guo *et al.* reviewed federated biometric recognition and identified distributed learning as a promising direction for reducing centralized exposure of biometric data [18]. From the quality-assessment perspective, Babnik *et al.* developed an efficient diffusion-inspired face-image-quality estimator that predicts recognition utility under unconstrained conditions [19].

2.3 Fairness, Security, and Robustness

A trustworthy identification system must also consider demographic and security risks. Kotwal and Marcel reviewed demographic fairness in face recognition and documented persistent performance disparities across demographic groups [20]. Wang *et al.* proposed privacy-preserving faces that retain robust authentication information, illustrating that privacy and identity utility can be treated jointly [21]. Gong *et al.* demonstrated that physical style masks can be optimized to attack face-recognition systems, highlighting the vulnerabil-

ity of recognition pipelines to deliberately manipulated facial appearance [22].

Synthetic data has emerged as another strategy for improving generalization and demographic diversity. DeAndres-Tame *et al.* reported the second FRCSyn-onGoing benchmark and analysed how synthetic data can support face-recognition performance under demographic and challenging-condition constraints [23]. Peng *et al.* proposed privacy-secure federated face-recognition training using random projection, further illustrating the growing importance of decentralized biometric learning [24]. Finally, Elmahmudi and Ugail studied recognition from imperfect or partial faces and showed that missing facial information, rotation, and scale variation can materially affect recognition performance [25].

Table 1 synthesizes the relationship between representative prior studies and the present work before the table is presented.

The literature therefore reveals three recurring strategies: retraining the recognition model, learning a dedicated quality estimator, or developing a more sophisticated temporal architecture. The present work takes a narrower systems-oriented position. It assumes that the detector and face-embedding model are already available and asks whether a training-free inference layer can use image quality to select temporal evidence, fuse embeddings, and adapt the open-set acceptance rule.

3 Proposed QTAF-ID Smart-Campus Identification Framework

The proposed framework retains the original four-layer smart-campus concept of camera acquisition, artificial-intelligence processing, backend data services, and web monitoring while adding QTAF-ID between facial embedding extraction and the final enrolled-or-unknown decision. The high-level schematic is cited in Figure 1 before its appearance.

3.1 Video Acquisition and Person Detection

Let the incoming camera stream be represented by a sequence of frames. Equation 1 defines the stream before it is used in the detection stage.

$$\mathcal{V} = \{F_1, F_2, \ldots, F_T\}, \tag{1}$$

where F_t is the frame captured at time t. The detector operates on a 640×640 resized representation. Equation 2 defines the retained person detections.

$$D_t^{(p)} = \{b_j \mid c_j = \text{person},\ p_j \geq \tau_d\}, \tag{2}$$

where b_j, c_j, and p_j denote the bounding box, class label, and confidence of the jth detection and τ_d is the detector threshold. Only retained person regions are forwarded to face processing.

3.2 Face Representation and Temporal Buffer

Each detected person region undergoes face localization and alignment before FaceNet512 representation. Equation 3 defines the normalized face embedding.

$$\hat{e}_t = \frac{\Phi(A_t)}{\|\Phi(A_t)\|_2 + \varepsilon}, \qquad \hat{e}_t \in \mathbb{R}^{512}, \tag{3}$$

where A_t is the aligned face and $\Phi(\cdot)$ is the pretrained FaceNet512 mapping. QTAF-ID collects a short sequence

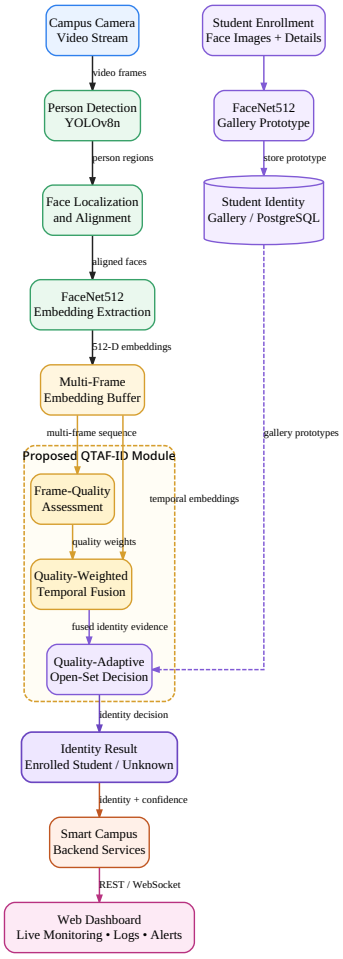

Figure 1. High-level schematic of the proposed web-based smart-campus identification framework integrating person detection, FaceNet512 representation, quality-temporal adaptive fusion, open-set identity decision, enrollment, and web monitoring.

of five observations. Equation 4 defines this buffer.

$$\mathcal{E} = \{\hat{e}_1, \hat{e}_2, \ldots, \hat{e}_5\}. \tag{4}$$

3.3 Frame-Quality Estimation

The quality estimator is deliberately simple and interpretable. For a grayscale face G_t, Equation 5 defines bounded sharpness from Laplacian variance.

$$S_t = \frac{\mathrm{Var}(\nabla^2 G_t)}{\mathrm{Var}(\nabla^2 G_t) + 80}. \tag{5}$$

Equation 6 defines brightness suitability from the mean grayscale intensity μ_t.

$$B_t = \exp\left[-\left(\frac{\mu_t - 127.5}{68}\right)^2\right]. \tag{6}$$

Equation 7 defines normalized contrast using grayscale standard deviation σ_t.

$$C_t = \min\left(\frac{\sigma_t}{58}, 1\right). \tag{7}$$

The combined frame-quality score is cited in Equation 8 before its appearance.

$$q_t = 0.45S_t + 0.35B_t + 0.20C_t. \tag{8}$$

The weighting emphasizes sharpness while retaining complementary brightness and contrast information.

Table 1. Comparative synthesis of representative related work and the gap addressed by QTAF-ID.

Ref.	Primary focus	Main contribution	Limitation relative to the present work
[11]	Open-set face recognition	Explicit unknown-person rejection through adapted open-set learning	Requires adaptation or training of the recognition system rather than a plug-in inference layer
[12]	Surveillance person search	Integrates localization and retrieval perspectives	Does not target quality-weighted temporal face fusion
[14]	Surveillance access control	Multi-stage person and face processing with confidence handling	Uses a different decision formulation and does not implement QTAF-ID
[15]	Privacy-preserving surveillance	Protects facial data in distributed learning	Adds training and distributed-learning complexity
[17]	Temporal private recognition	Exploits temporal information while reducing visible facial exposure	Does not use quality-adaptive embedding fusion
[19]	Face-image quality assessment	Learned recognition-oriented quality estimation	Requires a dedicated learned quality model
[20]	Demographic fairness	Reviews causes, metrics, and mitigation of demographic disparities	Does not provide an operational identification method
[22]	Physical attack robustness	Demonstrates adversarial masked-face vulnerabilities	Attack-focused rather than quality-fusion-focused
[23]	Synthetic face data	Studies generalization, diversity, and bias with synthetic data	Primarily data-generation and training oriented
[24]	Federated private recognition	Privacy-secure model training with random projection	Requires federated learning infrastructure
[25]	Imperfect facial data	Quantifies degradation under partial and manipulated facial inputs	Primarily single-image recognition

3.4 Top-K Selection and Quality-Weighted Fusion

The five observations are ranked by q_t, and the three highest-quality observations are retained. Equation 9 defines the selected index set.

$$\mathscr{K} = \mathrm{TopK}\left(\{q_t\}_{t=1}^{5}, K=3\right). \tag{9}$$

For the retained observations, Equation 10 defines the quality-dependent softmax weights with concentration parameter $\gamma = 6$.

$$w_t = \frac{\exp(\gamma q_t)}{\sum_{j\in\mathscr{K}} \exp(\gamma q_j)}, \qquad t \in \mathscr{K}. \tag{10}$$

Equation 11 then defines the normalized fused embedding.

$$e_f = \frac{\sum_{t\in\mathscr{K}} w_t \hat{e}_t}{\left\|\sum_{t\in\mathscr{K}} w_t \hat{e}_t\right\|_2 + \varepsilon}. \tag{11}$$

The corresponding fused sequence quality is defined in Equation 12.

$$Q = \sum_{t\in\mathscr{K}} w_t q_t. \tag{12}$$

3.5 Enrollment and Gallery Matching

For enrolled student i, multiple enrollment embeddings are averaged and normalized into a gallery prototype. Equation 13 defines the prototype.

$$g_i = \frac{\frac{1}{M_i}\sum_{m=1}^{M_i} \hat{g}_{i,m}}{\left\|\frac{1}{M_i}\sum_{m=1}^{M_i} \hat{g}_{i,m}\right\|_2 + \varepsilon}. \tag{13}$$

The fused probe is compared with every enrolled prototype using cosine similarity. Equation 14 defines this score.

$$s_i = e_f^{\top} g_i. \tag{14}$$

The best candidate is $i^* = \arg\max_i s_i$, with $s_{\max} = \max_i s_i$.

3.6 Quality-Adaptive Open-Set Decision

A fixed threshold treats all observations as equally reliable. QTAF-ID instead makes the threshold stricter when fused quality decreases. Equation 15 defines the adaptive threshold.

$$\tau(Q) = \tau_0 + \lambda(1-Q), \tag{15}$$

Table 2. Principal configuration of the proposed QTAF-ID framework.

Component	Configuration
Person detector	YOLOv8n, 640×640 input
Facial representation	FaceNet512, 512 dimensions
Sequence length	5 observations
Quality indicators	Sharpness, brightness, contrast
Quality coefficients	0.45, 0.35, 0.20
Selected observations	Top-K, $K=3$
Softmax concentration	$\gamma = 6$
Similarity	Cosine similarity
Base threshold	$\tau_0 = 0.51$
Quality penalty	$\lambda = 0.225$
Output	Enrolled identity or Unknown
Model retraining	Not required

Algorithm 1 Quality-Temporal Adaptive Fusion for Open-Set Identification

Require: Five facial observations $\{X_t\}_{t=1}^{5}$ and enrolled gallery $\mathscr{G}$
Ensure: Enrolled identity or Unknown
for $t = 1$ to 5 **do**
  compute S_t, B_t, and C_t
  compute $q_t = 0.45S_t + 0.35B_t + 0.20C_t$
  extract and normalize FaceNet512 embedding $\hat{e}_t$
end for
select the three highest-quality observations
compute softmax weights with $\gamma = 6$
compute fused embedding e_f and fused quality Q
compare e_f with all gallery prototypes using cosine similarity
compute $\tau(Q) = 0.51 + 0.225(1-Q)$
if $s_{\max} \geq \tau(Q)$ **then**
  return candidate identity i^*
else
  return Unknown
end if

where validation selected $\tau_0 = 0.51$ and $\lambda = 0.225$. Equation 16 gives the final open-set rule.

$$\hat{y} = \begin{cases} i^*, & s_{\max} \geq \tau(Q), \\ \text{Unknown}, & s_{\max} < \tau(Q). \end{cases} \tag{16}$$

A top-1/top-2 margin was explored during validation, but the selected QTAF-ID margin was zero; it therefore has no functional role and is excluded from the final method.

Table 2 consolidates the principal settings before the table is presented.

Algorithm 1 summarizes the complete inference procedure.

3.7 Web and Data-Service Integration

After recognition, the identity result, confidence, camera identifier, and timestamp are passed to a Python FastAPI artificial intelligence service, then to a Node.js/TypeScript backend, PostgreSQL storage, and a Next.js dashboard. The dashboard exposes live monitoring, event logs, student enrollment, and unknown-person alerts. Docker-based containerization separates deployment concerns from recognition logic. The proposed contribution is therefore compatible with the existing web architecture while remaining independent of interface technology.

3.8 Operational Data Flow and Event Handling

The recognition pipeline is intentionally separated from persistent data management. During enrollment, the student record contains administrative attributes such as identifier, name, branch, year, and the normalized gallery prototype. During live monitoring, the artificial-intelligence service does not need to reload raw enrollment photographs for every decision; it retrieves the stored prototype representation required by the matcher. This reduces repeated image decoding and keeps the operational decision path focused on compact numerical features.

Each recognition event is handled as a transactional record that contains the camera identifier, event time, predicted identity state, top similarity, and fused quality. For an enrolled match, the backend can associate the event with the student record; for an Unknown outcome, the event can be stored without forcing an enrolled identity assignment. This distinction is important because open-set recognition should preserve uncertainty rather than convert every observation into a known label.

The dashboard is designed as an observer of the recognition service rather than part of the recognition algorithm itself. Consequently, temporary dashboard unavailability does not alter the numerical identity decision, and the recognition service can continue to write events to persistent storage. Likewise, the database and web interface can be modified without retraining the vision models. This separation improves maintainability and allows the artificial-intelligence component to be evaluated independently from user-interface behavior.

A production implementation should also apply temporal de-duplication so that repeated frames from the same person do not create a large number of nearly identical log entries. A short event-cooldown window or a person-tracking identifier can be used to merge repeated observations into one monitored visit. Such logic is outside the present pilot evaluation but is compatible with the architecture in Figure 1.

3.9 Computational Characteristics of the Decision Layer

The computational work added by QTAF-ID is small relative to facial embedding extraction. For a sequence of L observations with P image pixels per observation, the handcrafted quality calculation is linear in the number of processed pixels, $O(LP)$. Top-K ranking is negligible for $L = 5$, and temporal fusion requires only weighted operations on 512-dimensional vectors. Gallery matching scales linearly with the number of enrolled identities because one cosine similarity is evaluated for each stored prototype.

This decomposition clarifies where optimization effort should be directed. Increasing the gallery from tens to thousands of students increases similarity comparisons, but the vector dimension remains fixed. Approximate-nearest-neighbor indexing can therefore be introduced if gallery search later becomes a bottleneck. In contrast, the present pilot shows that the expensive operation is FaceNet512 feature extraction. The proposed method is consequently best understood as a low-overhead decision layer around a comparatively heavier embedding stage rather than as a replacement for the embedding network.

Table 3. Experimental data partition used in the reproducible pilot.

Item	Value
Detection dataset	COCO128
Detection images	128
Person instances	254
Recognition dataset	LFW
Enrolled identities	15
Unknown identities	10
Gallery images per enrolled identity	3
Total gallery images	45
Known validation probes	30
Unknown validation probes	30
Known test probes	30
Unknown test probes	30
Random seed	42

4 Experimental Setup and Evaluation Protocol

The experimental design evaluates the person detector and open-set face-identification mechanism separately because the two components use different public datasets. This prevents a misleading camera-to-identity performance claim assembled from unrelated samples.

4.1 Person-Detection Protocol

The YOLOv8n detector was validated on the person class of Common Objects in Context (COCO128) 128-image subset. The run used 128 images containing 254 person instances, a 640×640 input, batch size 16, and a pretrained YOLOv8n model. Precision, recall, mean average precision at 0.5 intersection-over-union, mean average precision averaged from 0.5 to 0.95, and per-image timing were recorded.

4.2 Open-Set Recognition Protocol

The facial experiment used Labeled Faces in the Wild (LFW) colour images through a public dataset interface. Fifteen identities were designated as enrolled and ten different identities were reserved exclusively as unknown. Each enrolled identity contributed three gallery images, two validation probes, and two test probes. Each unknown identity contributed three validation and three test probes. Thus, the gallery contained 45 images and both validation and final test partitions contained 60 probes with a balanced 30 known and 30 unknown composition.

Table 3 summarizes the data partition before its appearance.

4.3 Controlled Mixed-Quality Sequences

Each probe image was transformed into five deterministic pseudo-temporal observations: clean, low light, Gaussian blur, low contrast, and over-exposure. The order of conditions was deterministically shuffled for every probe so the random single-frame baseline did not systematically receive the clean image. The low-light transformation is defined in Equation 17 before

Table 4. Compared open-set recognition configurations.

Method	Definition
Random single frame	First observation after deterministic sequence shuffling
Best-quality single frame	Observation with the largest quality score
Unweighted five-frame mean	Equal average of all normalized embeddings
QTAF-ID	Top-3 quality selection, weighted fusion, adaptive threshold

Table 5. Validation-selected decision parameters.

Method	Threshold terms	Margin
Random single frame	$\tau = 0.53$	0.03
Best-quality single frame	$\tau = 0.56$	0
Unweighted mean	$\tau = 0.25$	0.20
QTAF-ID	$\tau_0 = 0.51$, $\lambda = 0.225$	0

its appearance.

$$X_{\mathrm{ll}} = \mathrm{clip}(0.52X - 10). \tag{17}$$

The Gaussian-blur transformation is defined in Equation 18.

$$X_{\mathrm{bl}} = G_{7\times7,\sigma=2.0} * X. \tag{18}$$

The low-contrast transformation is defined in Equation 19.

$$X_{\mathrm{lc}} = \mathrm{clip}(0.58X + 48). \tag{19}$$

The over-exposure transformation is defined in Equation 20.

$$X_{\mathrm{oe}} = \mathrm{clip}(1.42X + 28). \tag{20}$$

These sequences are controlled stress tests derived from still images and are not presented as genuine consecutive campus-video frames.

4.4 Compared Recognition Configurations

Four methods were compared under the same gallery and partitions: a random observed single frame, the best-quality single frame, an unweighted mean of all five embeddings, and QTAF-ID. Table 4 defines the baselines before the table is presented.

4.5 Validation-Based Calibration and Metrics

All decision parameters were selected on validation data only. The fixed-threshold baselines were searched over similarity thresholds from 0.25 to 0.85. QTAF-ID searched the base threshold, quality penalty, and an optional similarity-margin term. The final QTAF-ID validation selection was $\tau_0 = 0.51$, $\lambda = 0.225$, and zero margin. Table 5 gives the selected settings.

For N probes, Equation 21 defines open-set accuracy.

$$A_{\mathrm{os}} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}(\hat{y}_i = y_i). \tag{21}$$

Equation 22 defines known-identity Top-1 accuracy for N_K enrolled probes.

$$A_K = \frac{1}{N_K}\sum_{i\in\mathcal{K}} \mathbb{I}(\hat{y}_i = y_i). \tag{22}$$

Equation 23 defines unknown-person rejection for N_U non-enrolled probes.

$$R_U = \frac{1}{N_U}\sum_{i\in\mathcal{U}} \mathbb{I}(\hat{y}_i = \mathrm{Unknown}). \tag{23}$$

Table 6. Recorded implementation environment for the pilot evaluation.

Component	Recorded setting
Execution platform	Google Colab
Accelerator	Nvidia Tesla T4
Accelerator memory	approximately 14.9 gigabytes
Python	3.13.15
PyTorch	2.11.0 with CUDA 12.8
Ultralytics	8.4.155
Face representation	FaceNet512 via DeepFace
Embedding dimension	512
Sequence length	5
Top-K	3
Random seed	42
Test probes	60

The false-acceptance rate is $1 - R_U$. Macro harmonic mean of precision and recall (F1) is also reported across enrolled identities and the unknown class. Paired correctness differences were assessed with an exact McNemar/binomial comparison, and a significance threshold of $p < 0.05$ was used.

4.6 Implementation Environment and Reproducibility

The executed notebook reported Python 3.13.15, PyTorch 2.11.0 with Compute Unified Device Architecture (CUDA) 12.8 support, Ultralytics 8.4.155, and a Nvidia Tesla accelerator (T4) graphics processing unit (GPU) with approximately 14.9 gigabytes of memory. Random, NumPy, and PyTorch seeds were fixed to 42. Face embeddings were generated with FaceNet512 through DeepFace; because LFW images are already face-centered benchmark samples, the controlled experiment bypassed an additional face detector during embedding extraction. This implementation detail is intentionally distinguished from the operational architecture in Figure 1.

Table 6 summarizes the implementation environment.

4.7 Reproducibility Controls

Reproducibility was treated as part of the experimental design rather than as a post-processing step. Identity selection, gallery/validation/test assignment, condition ordering, and numerical randomization were generated from the fixed seed 42. Gallery images were not reused as validation or test probes, and threshold calibration was performed on the validation partition only. These controls reduce the risk of identity-instance leakage and prevent the final test set from influencing the selected open-set threshold.

The controlled degradations were generated with fixed image-processing parameters instead of manually chosen per-image adjustments. Therefore, every recognition method received the same underlying probe and the same five transformed observations. The only difference among the compared methods was how those observations were selected or fused and how the final acceptance rule was applied. This paired design is the reason an exact paired correctness test is more informative than comparing independent proportions.

The experiment archive also records the selected known and unknown identities, gallery indices, validation indices, test indices, detector metrics, recognition metrics, condition-level outputs, and timing measurements. Keeping these intermediate artifacts is important because a single aggregate accuracy value cannot reveal whether an apparent gain results from known-identity correction, unknown rejection, or both.

Table 7. YOLOv8n person-detection results on COCO128.

Metric	Result
Precision	0.8120
Recall	0.6693
Mean average precision at 0.5	0.7606
Mean average precision at 0.5:0.95	0.5323
Pre-processing time	0.605 ms/image
Inference time	3.729 ms/image
Post-processing time	0.984 ms/image

4.8 Runtime Measurement Scope

Timing values are interpreted at the component level. Detector preprocessing, inference, and post-processing were reported directly by the detection framework. FaceNet512 wall-clock embedding time was measured separately, and the QTAF-ID decision overhead was timed only after the required embeddings were already available. This prevents the very small vector-fusion cost from being mistaken for complete multi-frame recognition latency.

The reported detector and recognition times should also not be added together to create a synthetic end-to-end latency because they were obtained from different dataset loops and processing paths. A true deployment benchmark should begin at camera frame acquisition and end after the database event and dashboard update, while measuring queueing, decoding, face localization, embedding extraction, network communication, and storage. That end-to-end experiment is reserved for future campus-video validation.

5 Results and Analysis

The measured results are reported in the same component-wise manner as the experimental design. Person detection is interpreted on COCO128, whereas open-set recognition is interpreted on the fixed protocol derived from LFW. No artificial end-to-end accuracy is formed by combining these different datasets.

5.1 Person-Detection Performance

Table 7 presents the recorded YOLOv8n person-detection results before the table appears.

The detector provides a computationally economical person-localization front end, although recall is lower than precision and therefore remains an upstream limitation. A missed person cannot subsequently be recovered by facial matching. Recent work has likewise explored attention-enhanced YOLOv8 models in recognition-oriented applications [26]. Nimma *et al.* combined attention and transformer processing with YOLOv8 for real-time surveillance object detection, further illustrating the suitability of lightweight single-stage detection for monitoring pipelines [27]. Their numerical results are not treated as direct baselines because datasets and architectures differ.

5.2 Open-Set Identification Performance

Table 8 reports the four recognition strategies before the table is presented.

The random observed single-frame baseline correctly classified 54 of 60 probes, while QTAF-ID correctly classified all 60 in this controlled pilot. Best-quality single-frame selection was already strong at 98.33 percent, indicating that frame selection itself explains a substantial part of the gain. Unweighted temporal averaging reached 93.33 percent, demonstrating that adding more observations without quality control is not sufficient.

Figure 2 is cited before it visualizes the same results.

The profile emphasizes that QTAF-ID is most clearly distinguished from arbitrary single-frame recognition, while the best-quality single-frame baseline remains close. Open-set classifier research similarly focuses on constraining the acceptance region for known classes; Zhang *et al.* proposed adversarial compact wrapping to improve rejection of unseen samples [28]. Quality-aware training also supports the principle that low-quality faces should not contribute equally; Zheng *et al.* developed quality-adaptive class centers for large-scale face recognition [29].

5.3 Paired Statistical Comparison

Table 9 presents the exact paired comparisons against QTAF-ID.

Against the random single-frame baseline, six errors were corrected and no new error was introduced, producing $p = 0.03125$. This supports improvement under the specific controlled protocol. The comparisons with the two stronger baselines did not reach the predefined 0.05 threshold; therefore, the experiment does not establish universal superiority over every quality-selected or temporal alternative.

5.4 Condition-Wise Robustness

Table 10 reports the single-frame robustness results under each controlled visual condition.

Low light produced the smallest mean quality score, and low light and blur both reduced accuracy to 88.33 percent. Figure 3 is cited before it presents the degradation profile.

Fan *et al.* similarly treated low illumination as a recognition problem and developed a face-recognition-driven low-light enhancement framework [30]. He *et al.* addressed recognition under pose and occlusion using detachable self-supervised bypass networks [31]. Xin *et al.* further investigated robustness under noisy and hard samples through adaptive mining and margining [32]. The present experiment is simpler, but the condition-level behaviour is consistent with the broader observation that visual degradation affects identity evidence.

5.5 Adaptive-Threshold Behaviour

Across QTAF-ID test sequences, fused quality ranged approximately from 0.539 to 0.838. Equation 24 summarizes the corresponding threshold range before it appears.

$$0.5466 \leq \tau(Q) \leq 0.6136. \tag{24}$$

Known probes produced a mean top similarity of approximately 0.8296, while unknown probes averaged approximately 0.3677. Figure 4 is cited before showing the relationship between sequence quality and top similarity.

The adaptive rule does not simply increase every similarity requirement. It increases the threshold for weaker sequences and approaches the base threshold for higher-quality sequences. The observed separation between known and unknown samples occurred under the present small test protocol and should not be assumed to persist unchanged when the gallery is expanded to thousands of identities.

5.6 Runtime and System Implications

Table 11 reports the recorded runtime components.

The QTAF-ID decision stage adds negligible overhead after embeddings exist, but the five-frame configuration necessar-

Table 8. Open-set identification results on the fixed 60-probe test partition.

Method	Open-set accuracy (%)	Known Top-1 (%)	Unknown rejection (%)	False acceptance (%)	Macro-F1
Random observed single frame	90.00	93.33	86.67	13.33	0.9102
Best-quality single frame	98.33	100.00	96.67	3.33	0.9864
Unweighted five-frame mean	93.33	93.33	93.33	6.67	0.9333
QTAF-ID	**100.00**	**100.00**	**100.00**	**0.00**	**1.0000**

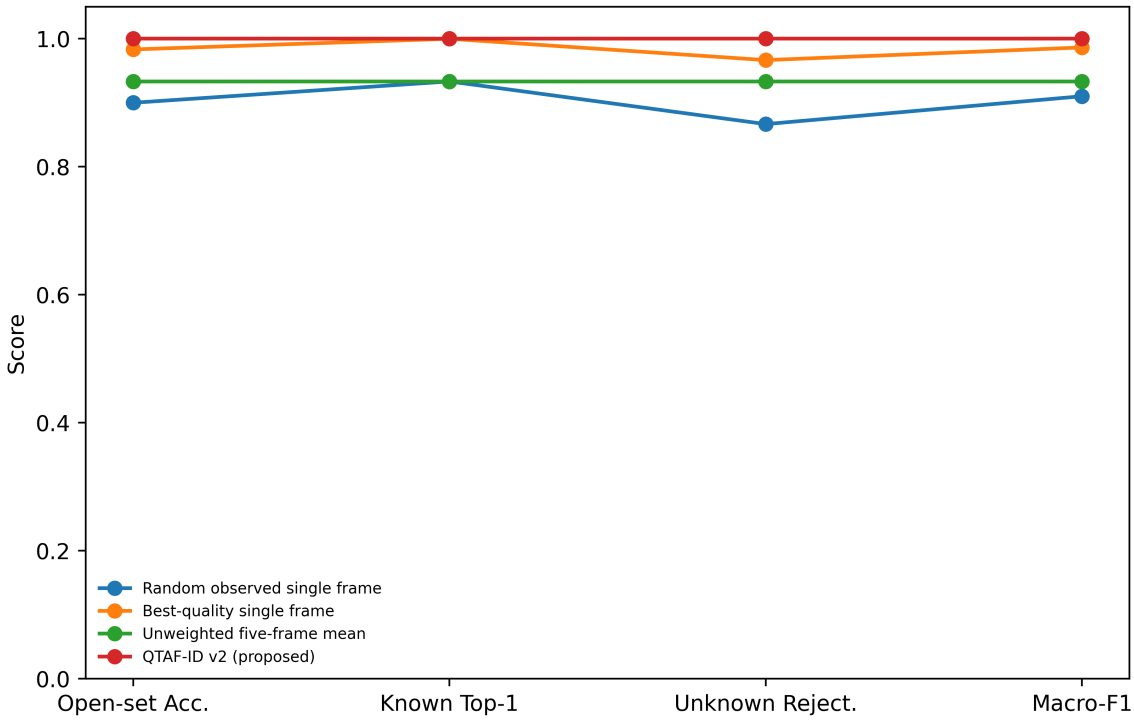

Figure 2. Comparison of random single-frame recognition, best-quality single-frame selection, unweighted five-frame averaging, and QTAF-ID across open-set accuracy, known Top-1 identification, unknown rejection, and macro-F1.

Table 9. Exact paired correctness comparison with QTAF-ID.

Baseline	Corrected	Worsened	Exact p
Random single frame	6	0	0.03125
Best-quality single frame	1	0	1.000
Unweighted five-frame mean	4	0	0.125

Table 10. Condition-wise single-frame robustness.

Condition	Accuracy (%)	Mean quality	Mean top similarity
Clean	93.33	0.748	0.595
Low light	88.33	0.355	0.544
Gaussian blur	88.33	0.516	0.554
Low contrast	95.00	0.599	0.589
Over-exposure	95.00	0.618	0.579

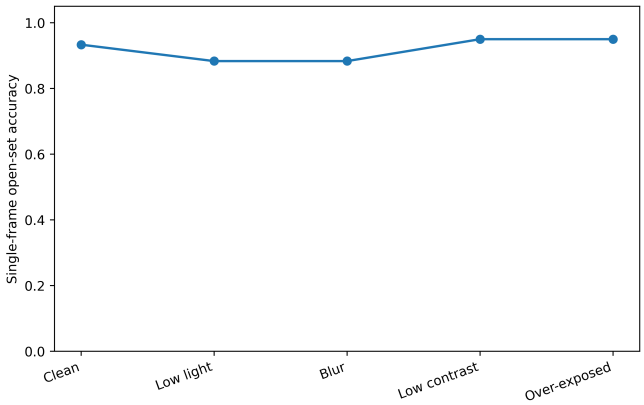

Figure 3. Condition-wise single-frame open-set identification accuracy under clean, low-light, blurred, low-contrast, and over-exposed observations.

ily incurs more embedding computation than a single-frame method. Accordingly, the contribution is improved decision robustness rather than lower end-to-end latency. Distributed processing can become important as the number of cameras increases; Martínez *et al.* demonstrated a fog-computing architecture for surveillance, weapon detection, and face recognition [33].

Privacy must also be considered at deployment time. Tao *et al.* proposed privacy-preserving outsourcing for face recognition using locally linear embedding and secure external computation [34]. Luo *et al.* proposed differential-privacy obfuscation for facial sensitive regions while balancing recognition utility [35]. These methods are not experimental baselines in the present study, but they motivate encryption, access control, retention policies, and privacy-aware processing in future campus deployment.

6 Discussion

6.1 Why Quality-Aware Fusion Helps

The results indicate that the main benefit of QTAF-ID comes from how observations are selected and combined rather than from replacing the underlying detector or facial representa-

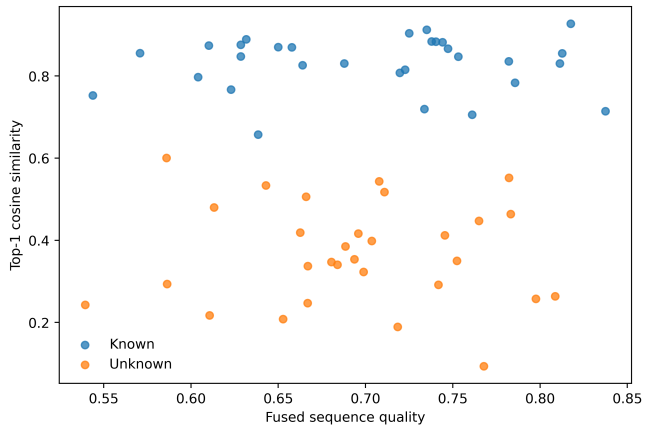

Figure 4. Relationship between fused sequence quality and top cosine similarity for enrolled and unknown probes under QTAF-ID.

Table 11. Recorded runtime characteristics.

Operation	Mean runtime
YOLOv8n preprocessing	0.605 ms/image
YOLOv8n inference	3.729 ms/image
YOLOv8n post-processing	0.984 ms/image
FaceNet512 embedding	269.764 ms/frame
QTAF-ID decision overhead	0.0142 ms/sequence

tion. The random single-frame baseline reached 90.00 percent open-set accuracy, whereas selecting the best-quality single frame increased accuracy to 98.33 percent. This difference alone shows that the quality of the observation is a major source of recognition variability. The proposed fusion step then combined several high-quality observations and removed the remaining error in this pilot.

The unweighted five-frame mean is especially informative. It achieved 93.33 percent accuracy, which is higher than arbitrary single-frame recognition but substantially lower than best-quality selection and QTAF-ID. Therefore, the availability of more frames is not sufficient by itself. When low-light or blurred frames are averaged with reliable frames at equal weight, they can shift the aggregate representation away from the most informative identity evidence. Top-K selection and softmax weighting are intended to limit that effect.

6.2 Statistical Interpretation

The paired analysis also provides a more cautious interpretation than raw percentages alone. Six errors were corrected relative to the random single-frame baseline and the exact paired test reached $p = 0.03125$. In contrast, only one error separated QTAF-ID from the best-quality single-frame method. The present evidence therefore supports the claim that quality-aware temporal fusion is beneficial relative to arbitrary single-frame recognition under mixed-quality observations, but it does not justify claiming universal superiority over every carefully selected single-frame or trainable temporal system.

6.3 Adaptive Threshold and Gallery Scale

The adaptive threshold offers a second layer of reliability control. A fixed threshold assumes that similarities derived from strong and weak sequences have equivalent evidential value. QTAF-ID raises the acceptance requirement as fused quality falls. This is useful conceptually for open-set operation, where a low-quality probe that produces a moderately large gallery similarity should be treated more cautiously than a clean sequence with the same score. Nevertheless, threshold calibration remains population dependent. A campus gallery containing thousands of visually similar students can narrow the separation between enrolled and unknown similarities and therefore requires re-calibration.

6.4 Detection-to-Recognition Error Propagation

The detector remains an upstream constraint. Precision of 0.8120 and mAP at 0.5 of 0.7606 are useful for a lightweight front end, but recall of 0.6693 indicates that some persons are missed in the COCO128 component evaluation. QTAF-ID cannot recover identities that never reach the face-processing stage. Future end-to-end testing must therefore consider crowded scenes, partial visibility, long camera-to-subject distances, and detector-to-recognizer error propagation.

6.5 Scalability and Computational Trade-Offs

The runtime analysis reveals a clear robustness-computation trade-off. The QTAF-ID decision rule itself requires only around 0.014 ms once embeddings have been computed, whereas FaceNet512 extraction averaged roughly 269.76 ms per frame in the recorded environment. A production system could reduce redundant extraction by accumulating embeddings asynchronously, batching multiple faces, or incorporating tracking so a subject remains associated across several frames without repeatedly starting a new recognition transaction.

6.6 Operational Governance and Human Oversight

The web architecture also requires governance beyond model accuracy. An Unknown result indicates that the system cannot match a person confidently to the enrolled gallery; it should not be interpreted automatically as suspicious or unauthorized behaviour. Likewise, facial embeddings remain sensitive biometric information even when raw enrollment images are not consulted during every recognition event. Encryption, role-based access, retention limits, audit logging, informed institutional policy, and template-revocation procedures are therefore necessary for any operational deployment.

6.7 Deployment Scenarios and Decision Policy

The experimental comparison also suggests that one deployment policy need not be used for every recognition event. The best-quality single-frame baseline already reached 98.33 percent accuracy in the pilot, which means that a system with strict latency constraints could first attempt recognition from the strongest available observation. Temporal fusion can then be activated when the first observation is below a quality threshold, when the top similarity is close to the acceptance boundary, or when the same tracked subject remains visible for several frames. Such a staged policy would preserve most of the robustness benefit while avoiding unnecessary repeated embedding extraction for easy cases.

A second practical choice concerns the interpretation of Unknown. The current decision rule treats Unknown as a valid open-set outcome, not as a failed software state. In an attendance application, an Unknown result may simply mean that the current frame is unsuitable or that the person is not enrolled in the course. In an access-control setting, the same result may trigger a request for a secondary credential rather than an automatic denial. In a security-monitoring setting, the event may be stored for authorized review without assuming

malicious intent. Separating recognition output from administrative action reduces the risk of converting model uncertainty into an unjustified operational conclusion.

The threshold should therefore be governed by the cost of the two principal errors. A false acceptance assigns an unknown individual to an enrolled identity, while a false rejection labels an enrolled student as Unknown. The acceptable balance between these errors differs by application. Attendance logging may tolerate a moderate number of false rejections if the user can confirm attendance manually, whereas a high-security entry point may require a stricter false-acceptance limit. The adaptive threshold in QTAF-ID changes with image quality, but its base value and penalty coefficient still need to be calibrated for the intended operational risk.

Camera placement also affects the usefulness of temporal fusion. At a doorway, a person may remain visible for only a short interval, making rapid acquisition of several usable observations important. In a corridor, a longer track can provide more frames but also larger pose changes. In a classroom, the same person may appear repeatedly with partial occlusion from other students. A future implementation can therefore vary the buffer duration and Top-K policy by camera role while preserving the same underlying quality and fusion formulation.

The gallery update policy deserves similar attention. Student appearance changes over semesters, and enrollment photographs may become less representative because of hairstyle, facial hair, spectacles, or other benign changes. A production system should not update a gallery prototype automatically from every accepted observation because one erroneous acceptance could contaminate the stored identity representation. Safer strategies include administrator-approved re-enrollment, periodic controlled capture, or high-confidence template updating with rollback support. These operational safeguards are outside the present pilot but are necessary for a reliable longitudinal campus system.

6.8 Recommended Full-Scale Validation Design

A full-scale study should preserve the leakage controls of the pilot while replacing the pseudo-temporal sequences with real video. At minimum, enrolled students and unknown participants should be recorded across several cameras, indoor and outdoor locations, lighting periods, and subject-camera distances. Enrollment sessions should be separated temporally from evaluation sessions so that the test set reflects realistic appearance and environmental change rather than near-duplicate captures.

The evaluation should report performance at three levels. The first level is person detection, using precision, recall, and localization measures. The second is face processing, including successful face localization and the proportion of detections that yield usable embeddings. The third is open-set identification, including known Top-1 accuracy, unknown rejection, false acceptance, false rejection, and calibration across quality ranges. Reporting these stages separately will make it possible to identify whether an end-to-end failure originated in person detection, face processing, or gallery matching.

A larger gallery is also important because open-set difficulty increases when more enrolled prototypes compete with an unknown probe. Experiments should therefore include progressively larger galleries rather than only one fixed population size. Threshold calibration can then be studied as a function of gallery scale, camera domain, and quality. This will reveal whether the current linear quality penalty remains adequate or whether a non-linear or camera-specific calibration model is required.

Runtime evaluation should similarly move from isolated component timing to concurrent multi-camera load testing. The useful quantities are not only average inference times but also throughput per camera, queue delay, memory consumption, sustained processing rate, and tail latency under simultaneous streams. Because the present framework separates the artificial-intelligence service from the web interface and database, each service can be profiled independently before measuring the complete camera-to-dashboard path.

Finally, full-scale validation should include structured error analysis rather than only aggregate averages. Misidentifications should be categorized by illumination, blur, pose, occlusion, subject distance, demographic group, camera source, and gallery similarity. Such analysis would determine whether the handcrafted quality score captures the dominant failure modes and would provide direct evidence for extending the quality model with pose, resolution, occlusion, or uncertainty information.

6.9 Limitations and Threats to Validity

Several limitations constrain external validity. The recognition protocol is small, with 15 enrolled and 10 unknown identities. The five-frame sequences are deterministic transformations of one LFW probe rather than true consecutive surveillance frames. COCO128 and LFW evaluate different components, so the paper does not claim a single camera-to-identity accuracy. The gallery uses only three enrollment images per known subject. The quality score does not explicitly model pose, occlusion, facial resolution, landmark confidence, or motion. Finally, demographic fairness was not evaluated and should be tested explicitly before operational use.

The controlled design also creates several specific threats to validity. Internal validity is influenced by the chosen degradation parameters because a darker or more severely blurred transformation would change the difficulty of the test. The deterministic transformations nevertheless preserve fairness among methods because every strategy is evaluated on the same observations. Construct validity is limited by the handcrafted quality score: sharpness, brightness, and contrast are useful cues, but they do not fully represent recognition utility. Pose, occlusion, face size, landmark confidence, and motion are omitted and can be incorporated in future versions.

External validity is the strongest limitation. LFW was created for unconstrained face-recognition research but is not a direct substitute for corridor or classroom surveillance. Camera height, focal length, compression, crowd density, motion, and subject distance can all change the distribution of facial observations. The current five-frame sequence also reuses one identity image under deterministic transformations rather than modelling natural viewpoint change over time. A future field study should therefore use true tracked video sequences and evaluate whether quality scores remain stable under motion and identity association errors.

Statistical conclusion validity is constrained by the small 60-probe final test set. Perfect empirical accuracy corresponds to zero observed errors, not zero true error probability. Confidence intervals and larger test populations are required before

deployment-level claims are made. The one-error difference between QTAF-ID and best-quality single-frame selection is especially important: it is scientifically more appropriate to describe the result as encouraging than to overstate a broad superiority claim.

Implementation validity is also limited because the controlled LFW embedding path bypassed an additional face detector and alignment stage. In the operational architecture, face localization and alignment can fail independently of identity matching. End-to-end campus testing must therefore include these stages, and the resulting errors should be reported separately so the source of performance loss is identifiable.

Finally, institutional deployment introduces policy validity questions that cannot be resolved by recognition metrics. Attendance, access control, and security monitoring have different risk tolerances. A false match may have very different consequences from a missed attendance event. The decision threshold, review workflow, data-retention period, and escalation policy should therefore be selected for the intended administrative purpose rather than copied directly from the pilot configuration.

These limitations define the study as a controlled robustness pilot. The findings are most useful as evidence that a training-free quality-aware decision layer can improve the use of an existing face-embedding system, not as proof of production-scale campus readiness.

7 Conclusion

This study presented a web-based smart-campus identification framework that combines lightweight person detection, FaceNet512 facial representation, and a training-free quality-temporal adaptive fusion mechanism for open-set identity decisions. The person detector achieved a precision of 0.8120, recall of 0.6693, mean average precision at 0.5 of 0.7606, and approximately 3.73 milliseconds inference time per image in the recorded accelerator environment. The controlled open-set recognition protocol contained 15 enrolled identities, 10 unknown identities, 45 gallery images, and 60 final test probes. The proposed temporal fusion configuration achieved 100 percent open-set accuracy and 100 percent unknown-person rejection in this pilot, compared with 90.00 percent and 86.67 percent, respectively, for random single-frame recognition. The strongest single-frame baseline reached 98.33 percent accuracy, showing that frame-quality selection itself accounts for a substantial portion of the benefit. The exact paired comparison against random single-frame recognition corrected six errors without introducing new ones. The decision-layer overhead was approximately 0.014 milliseconds after embeddings were available, while feature extraction remained the dominant computational cost. Future work should evaluate genuine multi-camera campus video, larger galleries, multi-object tracking, pose and occlusion-aware quality measures, demographic fairness, biometric-template protection, and complete end-to-end system latency. The reported results should be interpreted as controlled pilot evidence of robustness rather than as a claim of full-campus deployment performance.

Declarations

Acknowledgements. The authors acknowledge the Department of Computer Science and Engineering, CVR College of Engineering, Hyderabad, India, for providing the academic environment and support for this research work.

Funding. The authors declare that no specific grant from any funding agency in the public, commercial, or not-for-profit sectors was received for this study.

Author Contributions. P. Adityalakshmi: conceptualization, methodology, implementation, experimentation, data analysis, visualization, and manuscript preparation. K. Karthik: supervision, methodology review, technical validation, interpretation of results, manuscript review, and overall research guidance. The contribution statement should be confirmed by both authors before final journal submission.

Conflict of Interest / Competing Interests. The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this manuscript.

Data Availability. The reported evaluation uses publicly available benchmark datasets. Person-detection validation uses COCO128, and open-set facial identification uses Labeled Faces in the Wild. The experimental partitions, seed, controlled degradation procedure, and calibration settings are documented in the reproducible notebook used for this study.

Code Availability. The reproducible experimental notebook used to create the pilot partitions, perform calibration, compute metrics, evaluate controlled degradation conditions, and generate experimental outputs is available from the authors for research and reproducibility purposes.

Ethics Approval. No new human participants were recruited and no new biometric data were collected specifically for the reported benchmark-based experiments. Any future live campus deployment should be conducted under applicable institutional ethics, privacy, consent, and data-protection requirements.

Informed Consent. Not applicable to the reported benchmark-based experiments because no new participants were recruited.

Consent for Publication. The manuscript does not report newly collected identifiable participant information; separate participant publication consent was therefore not required for the reported benchmark experiments.

Generative AI and AI-Assisted Technologies. Artificial-intelligence-assisted tools were used for language refinement, formatting assistance, code organization, and preparation of reproducible experimental documentation. The authors remain responsible for the scientific methodology, verification of experimental results, interpretation, references, and final manuscript content.

Additional Information / Supplementary Material. The executed experimental notebook and result archive may be supplied as supplementary reproducibility material.

References

[1] Mei Wang, Weihong Deng, “Deep face recognition: A survey," *Neurocomputing*, vol. 429, pp. 215–244, 2021. doi: https://doi.org/10.1016/j.neucom.2020.10.081.

[2] Chuanxing Geng, Sheng-Jun Huang, Songcan Chen, "Recent advances in open set recognition: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 10, pp. 3614–3631, 2021. doi: https://doi.org/10.1109/TPAMI.2020.2981604.

[3] Lacey Best-Rowden, Anil K. Jain, "Learning face image quality from human assessments," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 12, pp. 3064–3077, 2018. doi: https://doi.org/10.1109/TIFS.2018.2799585.

[4] Shaolin Su, Hanhe Lin, Vlad Hosu, Oliver Wiedemann, Jinqiu Sun, Yu Zhu, Hantao Liu, Yanning Zhang, Dietmar Saupe, "Going the extra mile in face image quality assessment: A novel database and model," *IEEE Transactions on Multimedia*, vol. 26, pp. 2671–2685, 2024. doi: https://doi.org/10.1109/TMM.2023.3301276.

[5] Eric Lopez-Lopez, Xose M. Pardo, Carlos V. Regueiro, "Incremental learning from low-labelled stream data in open-set video face recognition," *Pattern Recognition*, vol. 131, Art. no. 108885, 2022. doi: https://doi.org/10.1016/j.patcog.2022.108885.

[6] Xingbo Dong, Jiewen Yang, Andrew Beng Jin Teoh, Dahai Yu, Xiaomeng Li, Zhe Jin, "Video-based face outline recognition," *Pattern Recognition*, vol. 152, Art. no. 110482, 2024. doi: https://doi.org/10.1016/j.patcog.2024.110482.

[7] Mohsin Ullah, Imtiaz Ahmad Taj, Rana Hammad Raza, "Degradation model and attention guided distillation approach for low resolution face recognition," *Expert Systems with Applications*, vol. 243, Art. no. 122882, 2024. doi: https://doi.org/10.1016/j.eswa.2023.122882.

[8] Pietro Melzi, Ruben Tolosana, Ruben Vera-Rodriguez, Minchul Kim, Christian Rathgeb, Xiaoming Liu, Ivan DeAndres-Tame, Aythami Morales, Julian Fierrez, Javier Ortega-Garcia, Weisong Zhao, Xiangyu Zhu, Zheyu Yan, Xiao-Yu Zhang, Jinlin Wu, Zhen Lei, Suvidha Tripathi, Mahak Kothari, Md Haider Zama, Debayan Deb, Bernardo Biesseck, Pedro Vidal, Roger Granada, Guilherme Fickel, Gustavo Führ, David Menotti, Alexander Unnervik, Anjith George, Christophe Ecabert, Hatef Otroshi Shahreza, Parsa Rahimi, Sébastien Marcel, Ioannis Sarridis, Christos Koutlis, Georgia Baltsou, Symeon Papadopoulos, Christos Diou, Nicolò Di Domenico, Guido Borghi, et al., "FRCSyn-onGoing: Benchmarking and comprehensive evaluation of real and synthetic data to improve face recognition systems," *Information Fusion*, vol. 107, Art. no. 102322, 2024. doi: https://doi.org/10.1016/j.inffus.2024.102322.

[9] Jiwon Jang, Chong O. Kim, "Teacher–Explorer–Student learning: A novel learning method for open set recognition," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36, no. 1, pp. 767–780, 2025. doi: https://doi.org/10.1109/TNNLS.2023.3336799.

[10] Ning Zhuang, Qiang Zhang, Chunhong Pan, Bo Ni, Yi Xu, Xiaokang Yang, Wenjun Zhang, "Recognition oriented facial image quality assessment via deep convolutional neural network," *Neurocomputing*, vol. 358, pp. 109–118, 2019. doi: https://doi.org/10.1016/j.neucom.2019.04.057.

[11] Rafael H. Vareto, Yao Linghu, Terrance E. Boult, William Robson Schwartz, Marcel Günther, "Open-set face recognition with maximal entropy and Objectosphere loss," *Image and Vision Computing*, vol. 141, Art. no. 104862, 2024. doi: https://doi.org/10.1016/j.imavis.2023.104862.

[12] S. Irene, A. John Prakash, V. Rhymend Uthariaraj, "Person search over security video surveillance systems using deep learning methods: A review," *Image and Vision Computing*, vol. 143, Art. no. 104930, 2024. doi: https://doi.org/10.1016/j.imavis.2024.104930.

[13] I. Pattnaik, A. Dev, A. K. Mohapatra, "A face recognition taxonomy and review framework towards dimensionality, modality and feature quality," *Engineering Applications of Artificial Intelligence*, vol. 126, Art. no. 107056, 2023. doi: https://doi.org/10.1016/j.engappai.2023.107056.

[14] Youssef Zennayi, Souad Benaissa, Hassane Derrouz, Zakaria Guennoun, "Unauthorized access detection system to the equipments in a room based on the persons identification by face recognition," *Engineering Applications of Artificial Intelligence*, vol. 124, Art. no. 106637, 2023. doi: https://doi.org/10.1016/j.engappai.2023.106637.

[15] Farah Wahida, M. A. P. Chamikara, Ibrahim Khalil, Mohammed Atiquzzaman, "An adversarial machine learning based approach for privacy preserving face recognition in distributed smart city surveillance," *Computer Networks*, vol. 254, Art. no. 110798, 2024. doi: https://doi.org/10.1016/j.comnet.2024.110798.

[16] Min Wang, Yifan Kang, Bailu Deng, Xi Lan, "Exploring college students' risk perception and acceptance intention of facial recognition technology in China," *Telematics and Informatics*, vol. 95, Art. no. 102193, 2024. doi: https://doi.org/10.1016/j.tele.2024.102193.

[17] Shu Min Leong, Raphaël C. W. Phan, Vishnu Monn Baskaran, Chee Pun Ooi, "Privacy-preserving facial recognition based on temporal features," *Applied Soft Computing*, vol. 96, Art. no. 106662, 2020. doi: https://doi.org/10.1016/j.asoc.2020.106662.

[18] Jian Guo, Hengyu Mu, Xingli Liu, Hengyi Ren, Chong Han, "Federated learning for biometric recognition: A survey," *Artificial Intelligence Review*, vol. 57, Art. no. 208, 2024. doi: https://doi.org/10.1007/s10462-024-10847-7.

[19] Žiga Babnik, Peter Peer, Vitomir Štruc, "eDifFIQA: Towards efficient face image quality assessment based on denoising diffusion probabilistic models," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 6, no. 4, pp. 458–474, 2024. doi: https://doi.org/10.1109/TBIOM.2024.3376236.

[20] Kshitij Kotwal, Sébastien Marcel, "Review of demographic fairness in face recognition," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 8, no. 1, pp. 20–45, 2026. doi: https://doi.org/10.1109/TBIOM.2025.3601217.

[21] Tao Wang, Wenying Wen, Xiangli Xiao, Zhongyun Hua, Yu-Shu Zhang, Yuming Fang, "Beyond privacy: Generating privacy-preserving faces supporting robust image authentication," *IEEE Transactions on Information Forensics and Security*, vol. 20, pp. 2564–2576, 2025. doi: https://doi.org/10.1109/TIFS.2025.3541859.

[22] Hao Gong, Mengqi Dong, Shiqing Ma, Seyit Camtepe, Surya Nepal, Chang Xu, "Stealthy physical masked face recognition attack via adversarial style optimization," *IEEE Transactions on Multimedia*, vol. 26, pp. 5014–5025, 2024. doi: https://doi.org/10.1109/TMM.2023.3330089.

[23] Ivan DeAndres-Tame, Ruben Tolosana, Pietro Melzi, Ruben Vera-Rodriguez, Minchul Kim, Christian Rathgeb, Xiaoming Liu, Luis F. Gomez, Aythami Morales, Julian Fierrez, Javier Ortega-Garcia, Zhizhou Zhong, Yuge Huang, Yuxi Mi, Shouhong Ding, Shuigeng Zhou, Shuai He, Lingzhi Fu, Heng Cong, Rongyu Zhang, et al., "Second FRCSyn-onGoing: Winning solutions and post-challenge analysis to improve face recognition with synthetic data," *Information Fusion*, vol. 120, Art. no. 103099, 2025. doi: https://doi.org/10.1016/j.inffus.2025.103099.

[24] Jialiang Peng, Huiting Sun, Dong Yang, Ahmed A. Abd El-Latif, Joel J. P. C. Rodrigues, "PriSecFedFR: Privacy-secure face recognition model training via federated learning and random projection," *Expert Systems with Applications*, vol. 297, Art. no. 129383, 2026. doi: https://doi.org/10.1016/j.eswa.2025.129383.

[25] Ali Ahmed Elmahmudi, Hassan Ugail, "Deep face recognition using imperfect facial data," *Future Generation Computer Systems*, vol. 99, pp. 213–225, 2019. doi: https://doi.org/10.1016/j.future.2019.04.025.

[26] Reham Hosney, Fatma M. Talaat, Eman M. El-Gendy, Mahmoud M. Saafan, "AutYOLO-ATT: An attention-based YOLOv8 algorithm for early autism diagnosis through facial expression recognition," *Neural Computing and Applications*, vol. 36, pp. 17199–17219, 2024. doi: https://doi.org/10.1007/s00521-024-09966-7.

[27] Divya Nimma, Omaia Al-Omari, Rahul Pradhan, Zoirov Ulmas, R. V. V. Krishna, Ts. Yousef A. Baker El-Ebiary, Vuda Sreenivasa Rao, "Object detection in real-time video surveillance using attention based transformer-YOLOv8 model," *Alexandria Engineering Journal*, vol. 118, pp. 482–495, 2025. doi: https://doi.org/10.1016/j.aej.2025.01.032.

[28] L. Zhang, M. Wan, P. Huang, G. Yang, "Adversarial compact wrapping classifier learning for open set recognition," *Information Sciences*, vol. 680, Art. no. 121114, 2024. doi: https://doi.org/10.1016/j.ins.2024.121114.

[29] Z. Zheng, Z. Liu, Y.-W. Si, X. Yuan, J. Duan, X. Li, X. Zhang, X. Gong, "Quality adaptive class center for lightweight large-scale face recognition," *Information Sciences*, vol. 732, Art. no. 122944, 2026. doi: https://doi.org/10.1016/j.ins.2025.122944.

[30] Yihua Fan, Yongzhen Wang, Dong Liang, Yiping Chen, Hui Xie, Fu Lee Wang, Jianxin Li, Mingqiang Wei, "Low-FaceNet: Face recognition-driven low-light image enhancement," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–13, 2024. doi: https://doi.org/10.1109/TIM.2024.3372230.

[31] Mingjie He, Jie Zhang, Shiguang Shan, Xilin Chen, "Enhancing face recognition with detachable self-supervised bypass networks," *IEEE Transactions on Image Processing*, vol. 33, pp. 1588–1599, 2024. doi: https://doi.org/10.1109/TIP.2024.3364067.

[32] Yang Xin, Xiang Zhong, Yu Zhou, Jianmin Jiang, "Robust face recognition via adaptive mining and margining of noise and hard samples," *IEEE Transactions on Image Processing*, vol. 34, pp. 8114–8129, 2025. doi: https://doi.org/10.1109/TIP.2025.3634979.

[33] H. Martínez, F. J. Rodríguez-Lozano, F. León-García, J. M. Palomares, J. Olivares, "Distributed Fog computing system for weapon detection and face recognition," *Journal of Network and Computer Applications*, vol. 232, Art. no. 104026, 2024. doi: https://doi.org/10.1016/j.jnca.2024.104026.

[34] Y. Tao, Y. Li, F. Kong, Y. Shi, M. Yang, J. Yu, H. Zhang, "Privacy-preserving outsourcing scheme of face recognition based on locally linear embedding," *Computers & Security*, vol. 144, Art. no. 103931, 2024. doi: https://doi.org/10.1016/j.cose.2024.103931.

[35] Yuling Luo, Tinghua Hu, Tao Hu, Xue Ouyang, Junxiu Liu, Qiang Fu, Qin Sheng, Shengfeng Qin, Zhen Min, XiaoGuang Lin, "DPO-Face: Differential privacy obfuscation for facial sensitive regions," *Computers & Security*, vol. 154, Art. no. 104434, 2025. doi: https://doi.org/10.1016/j.cose.2025.104434.