JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

NETWORK-AWARE, PRIVACY-PRESERVING, AND RESOURCE-EFFICIENT FACE RECOGNITION FOR IOT, EDGE, AND CLOUD-BASED AUTHENTICATION: A SYSTEMATIC SURVEY

Mazen Abdullah Bahashwan
Computer Engineering, Hadhramout University , Mukalla , Yemen
m.bahashwan@hu.edu.ye

ABSTRACT

Modern face recognition algorithms have progressed from being computer vision tasks to a broad networked biometrics solution, used flexibly on mobile devices, Internet-of-Things (IoT) cameras, edge servers, obstructed infrastructure, security cameras, access control systems, etc. Recently, with the development of various social, and indoor and/or out-door infrastructure security electronic devices, the restricted use of algorithms vision tasks, on computers, has become in adequate. Thus, to broaden their applications, on portable and security infrastructure devices, it necessitates to optimize latency, bandwidth, energy consumption, scalability, security, privacy protection, and bias at the same time. This survey paper aims at reviewing, the previous advances in network-aware, privacy-preserving, and computationally efficient face recognition with the emphasis on the use of face recognition in IoT, edge, mobile, and obstructed environments. To this respect, here firstly we briefly review the modern face recognition, that is based on sensing, face detection, alignment, representation learning, template comparison, decision-making, and data storage/transmission steps. Then, we describe how the algorithms can be applied, based on some technologies such as convolutional neural networks, deep metric learning, margin-based loss functions, Vision Transformer architecture, and alignment-invariant position encoding. We also here review lightweight models and efficient inference schemes tailored for edge-constrained devices, to proceed on discussing privacy protection solutions, to which this involves synthetic data generation, template masking, homomorphic encryption, secure multiparty computation, differential privacy, and federated learning approaches. Besides that, here we discuss the robustness of face recognition solutions to match the low-quality images; mask, scarfs and glasses obstructed faces, misalignment issues, as well as artificial faces generated by GANs and StyleGANs. Lastly, we tackle open problems in face recognition as a crucial security network service at a personal level and infrastructure building security, proposing a classification framework for mapping the recognition algorithms and solutions to various networked architectures and evaluation metrics.

.

***KEYWORDS*:** Face Recognition; Biometric Authentication; Network Security; IoT; Edge Computing; Privacy-Preserving Machine Learning; Synthetic Data; Lightweight Deep Learning

1. INTRODUCTION

Face recognition (FR) has become one of the widely used biometric technology, that provides verification and identification of images and videos without using passwords or physical tokens (e.g., smart cards, keys etc). The introducing of novel models such as FaceNet, VGGFace, ArcFace, CosFace, AdaFace, and transformer-based

methods [1]-[8], in deep learning contributed in many ways in the improvement of FR. Those, models can also be involved in various applications such as the opening of smart H/P screens, e-ID Cards, e-passports, border checks, access to services (i.e., health care, finances etc.), surveillance, smart premises to control exit/entry. [17]-[22].

Nevertheless, although in the past the applications of FR was only a desk top vision . Today, this

technology is continuously varying, growing and developing fast, invading most of our immobile and mobile security and communication devices, such as surveillance cameras/mobile phones etc. They can capture and process images. Then, the detection and embedding may be done at the edge node, and the template exchange can happen over wireless and wired networks, as matched by edge-cloud services. Thus, initiating the security/applications' actions; but as, the computing is a limit in those mobile /immobile network device memory resources, and the private biometric information. Those FR-devices entails the improvement of quality service (QoS), by tackling latency transmissions, bandwidth, security of data channel, template privacy, energy-efficient inference, reliable edge-cloud services availability, and accountability. The improvement of those aspects is crucial in the Internet of Things (IoT) and mobile/immobile networks devices.

Conventional surveys have widely covered various FR-technologies topics. Of them CNN architectures, databases, performance metrics, FR with masks, and privacy preserving face recognition [11], [12], [18]-[22] etc. Those reports in the past mostly focused on algorithms, considering FR as a computer-based vision. In the literature, there is very limited studies that paid attention to FR, as a networked service with accuracy, latency, throughput, communication overhead, partitioning of tasks between the edge and cloud nodes, energy efficiency, security, and privacy issues. Therefore, this present survey aims to fill this gap with an integrative approach that considers networking aspects of FR together with algorithmic design, security, and deployments.

Our core research hypothesis is that robust FR deployment requires an approach that integrates accuracy/robustness of FR, network-aware deployment, security/privacy protection, and efficiency of edge and mobile devices. The former two first aspects are connected to each other. Thus, the sending of face images to the cloud allows to control the processing capabilities with increased usage of bandwidth, and more potential privacy leaks to which it can be resolved by performing of FR on a local mobile with efficient architectures e.g., Ghost Face Nets [9]. Using the synthetic data can decrease the dependency on web-crawled biometric data with domain gaps and fairness problems [13]-[16], that necessitates an integrated taxonomy helping to design a system from several points of view

1.1 Contributions of this survey

- Taxonomy of FR systems that are aware of the network and cover the sensing, edge, network, and cloud layers.
- Comparative study of the current FR pipeline that involves detection, alignment, representation, matching, decision-making, and storage/transportation stages.
- State-of-the-art FR techniques to process low quality, occluded, masked, stylized, and misaligned face images such as AdaFace and relative keypoint position encoding.
- Resource-efficient FR approaches and methods of inference applicable to mobile devices and the Internet of Things.
- Consideration of privacy and security, synthetic data generation, templates protection, fairness issues, and other aspects.
- Future-oriented research direction for secure, fair, privacy-preserving, and efficient FR on future networks..

2. REVIEW METHODOLOGY AND SCOPE

In this study, an unstructured literature review approach has been selected in lieu of a chronological narrative. For the purpose of reviewing the literature, priority is being given to relevant survey papers, benchmarks, challenge papers, and representative system papers published mainly during 2021 to 2026. However, seminal face recognition (FR) techniques are also considered for their relevance to understand current research trends.

The inclusion criteria included (i) research papers that cover topics related to face verification, face identification, detection-alignment-recognition pipeline, privacy-preserving FR, generation of synthetic face images, lightweight FR, and FR in access control and surveillance; (ii) papers dealing with methodological, benchmarking, and deployment-related information relevant to networked systems; and (iii) papers highlighting any challenges in face recognition, such as poor quality imagery, face occlusions, biases, face template security, and FR systems deployed on edge

devices. Papers not related to human FR, focusing purely on image classification, and irrelevant to methodology were excluded from the study.

The literature has been classified not only by algorithm families but based on various system-level concerns. Therefore, the survey categorizes the literature based on the following seven dimensions: stage in the pipeline, family of models, deployment environment architecture, resources limitation, privacy or security considerations, choice of dataset, and evaluation methodology. This organization of literature is aimed at helping researchers design FR systems for IoT, edge, mobile, and cloud computing environments.

Table 1. Review dimensions used to organize the surveyed literature.

Dimension	Representative questions	Examples in the literature
Pipeline stage	Which part of the FR workflow is addressed?	Detection, alignment, embedding extraction, template matching, decision, storage/transmission [10], [31], [32].
Model learning	What model or learning strategy is used?	CNNs, margin-based losses, ViTs, synthetic-data training, multimodal VIS/IR fusion [4]-[8], [17].
Deployment architecture	Where are computation and matching performed?	On-device, edge, cloud, hybrid edge-cloud, federated, surveillance network.
Resource constraint	Which operational cost is optimized?	FLOPs, memory, latency, energy, bandwidth, throughput [9], [33], [34].
Security/privacy	Which risk is mitigated?	Template inversion, raw-image leakage, spoofing, adversarial attacks, secure inference [11], [12].
Dataset strategy	What data source and benchmark are used?	Real web-crawled datasets, synthetic datasets, stylized data, masked datasets [13]-[16], [18], [30].
Evaluation	Which metrics are reported?	Accuracy, TAR@FAR, EER, ROC/AUC, latency, FLOPs, fairness gap, privacy leakage.

3. FACE RECOGNITION PIPELINE IN NETWORKED ENVIRONMENTS

A modern face recognition system consists of four main technical components: face detection, alignment, representation, and verification/identification. A recent benchmark study of face recognition pipelines analyzed different combinations of recognition models, detectors, distance metrics, and alignment types, showing that performance depends on interactions within the entire pipeline rather than on the embedding model [10]. This observation is important in the context of networked face recognition as errors in early stages can be propagated into later matching decision made in the cloud.

Capturing stage involves capturing face images or videos using cameras, mobile phones, CCTV networks, or dedicated sensors. In practice, capture stage is affected by factors such as lighting conditions, pose, motion blur, resolution, occlusions, masks, and sensor quality. Face Detection stage involves face localization, and possibly also face landmark estimation. Well-known face detectors such as MTCNN or RetinaFace output face bounding boxes and face landmarks, which can be used for face alignment. Face alignment stage aims to normalize the face scale, rotations, and translations, normally based on face landmarks like eyes [31], [32]. Misalignment continues to be one of the main sources of errors in practice, especially when dealing with lower quality surveillance videos or mobile use cases.

Face representation learning stage involves learning of face embedding from aligned face image. In deep metric learning tasks, we aim to minimize intra-class distance and maximize inter-class distance so that embeddings representing the same identity would be close to each other, whereas embeddings representing different identities would be further away. Finally, face verification/identification task is solved by measuring cosine/Euclidean distance between embeddings and thresholding (verification) or selecting the nearest neighbor (identification). The question about what information should be stored/transmitted: images, cropped faces, or encrypted/embedded faces is important for the amount of privacy exposed, bandwidth consumed, and latency incurred.

From a communications standpoint, a face recognition pipeline can be implemented according to different partitioning schemes. Centralized architecture sends images to the cloud where face detection, embedding, and verification occur. Edge-Assisted architecture performs detection and representation computation at the edge, sending only embeddings to the cloud. Fully On-device architecture performs everything at the edge with no transmission needed. Federated architecture allows for local training of the model without centralized storage of images.

Table 2. Pipeline stages, network implications and typical risks.

Stage	Main function	Network implication	Dominant risk
Capture	Acquire face image or video frame.	Raw data volume can be high, especially for multi-camera video streams.	Consent violation, low quality, sensor noise.
Detection	Locate face region and landmarks.	Running at edge reduces cloud traffic; missed detection causes failure.	Small faces, pose, background clutter.
Alignment	Normalize scale, rotation and pose.	Can be edge-side preprocessing before transmission.	Landmark failure, affine transformation error.
Representation	Generate embedding or template.	Sending embeddings reduces bandwidth but still may leak identity information.	Template inversion, bias, low-quality embeddings.
Matching	Compare probe against enrolled template(s).	Cloud matching scales to large galleries; edge matching reduces latency.	False accepts/rejects, threshold drift.
Decision	Grant/deny access or identify candidate.	Decision must be auditable and robust to network failure.	Spoofing, adversarial manipulation, fairness gap.
Storage/transmission	Store images/templates and transmit between nodes.	Requires encryption, access control, logging and revocation.	Data breach, linkability, non-renewable biometrics.

4. ALGORITHMIC FOUNDATIONS

4.1 From hand-crafted features to deep embeddings

The traditional FR methods are based on geometric parameters, eigenfaces, Fisherfaces, local binary pattern descriptors, Gabor filters, and shallow classifiers. Although they work well in laboratory settings, their scalability and robustness against the variability of facial appearance remain rather poor. DeepFace, FaceNet, and VGGFace have promoted the use of deep embeddings by learning discriminative features using extensive facial image collections [1]-[3]. The current FR paradigm consists of learning a function that transforms the input face into a compact feature representation for comparison through distance thresholding.

Margin losses contribute to increasing the discriminability of deep FR models by modifying the angular/cosine space. The representative margin losses include the additive angular margin loss called ArcFace, large-margin cosine loss known as CosFace, and multiplicative angular margin loss named SphereFace [4]-[6]. They decrease the intraclass variance while maximizing interclass differences, which become more critical for extensive face galleries and tight security requirements. Yet, there are margin losses that implicitly assume the informative nature of training samples; otherwise, they may harm model training when used with surveillance datasets.

4.2 Quality-adaptive and alignment-robust learning

AdaFace solves the problem of poor facial images through adaptive margins based on a quality metric generated from the feature norms [7]. The key idea revolves around the idea that difficult but high-quality images could be magnified based on the information they provide about identities, whereas impossible low-quality images would not contribute to the optimization process. This becomes even more relevant in the realm of networked FR, as images taken from surveillance cameras, drones, mobile phones, and IoT cameras are likely to suffer from low resolutions, blur, and poor illumination.

Vision transformers represent another approach to FR models as opposed to CNNs through modeling long-range dependencies using self-attention. Still, Vision Transformers can demonstrate vulnerability to issues with positional encoding and alignment failure. To overcome the problem, the authors propose an extension of the relative position encoding approach, adding the keypoints into the picture, which makes the model robust in terms of scale, translation, and rotation invariance [8]. Such research becomes particularly relevant in edge and surveillance systems where face crops are not always well-aligned.

4.3 Multimodal and multispectral recognition

Multimodal biometrics combine evidence from several modalities in order to improve robustness. A recent multimodal face recognition architecture using visible and infrared facial images utilized VGG16 features, principal components analysis for dimensionality reduction, sequence-based neural networks, and weighted score level fusion, achieving enhanced robustness in various scenarios [17]. FR based on multispectral images is relevant to networked security applications like border crossing, night time surveillance, and critical infrastructure access, wherein the use of visible images might be insufficient under low-lighting conditions.

Nevertheless, multimodal systems necessitate additional networking and design challenges. Multiple sensors increase data volumes and synchronization requirements, whereas fusion may be carried out at the sensor, feature, score, or

decision levels. For bandwidth-limited communication channels, it would be more desirable to perform score level or embedding level fusion rather than transferring raw multispectral images, provided that privacy and security issues are adequately dealt with.

Table 3. Representative algorithmic directions and deployment relevance.

Direction	Representative methods	Strengths	Limitations in networked deployment
Deep embeddings	FaceNet, VGGFace, DeepFace [1]-[3].	Strong representation learning; simple matching using distances.	Requires large training data and robust preprocessing.
Margin-based losses	ArcFace, CosFace, SphereFace [4]-[6].	High discriminability for verification and identification.	Sensitive to noisy labels and low-quality samples.
Quality-adaptive loss	AdaFace [7].	Improves training with low-quality images; relevant to surveillance.	Requires careful interpretation of quality indicators.
Transformer-based FR	ViT-based models, KP-RPE [8].	Captures global context; can improve alignment robustness.	May be computationally heavy without compression.
Lightweight CNNs	MobileFaceNet, GhostFaceNets [9], [34].	Efficient inference for mobile/edge devices.	May sacrifice accuracy at strict FAR levels.
Multimodal fusion	VIS/IR score-level fusion [17].	Improves robustness under low light and environmental variability.	Increases sensor, synchronization and transmission complexity.
Synthetic-data training	DigiFace-1M, FRCSyn, Stylized-Face [14]-[16].	Reduces dependence on consent-sensitive web data; supports controlled variation.	Synthetic-to-real domain gap and identity realism remain challenging.

5. DEPLOYMENT ARCHITECTURES: CLOUD, EDGE, MOBILE AND FEDERATED FR

The deployment architecture describes the locus of computation as well as the data that flows through the network. In a cloud-based deployment scenario, raw images or video frames are sent to the server for detection, embedding and matching purposes. Cloud-based systems utilize huge amounts of computing resources along with centralized management, but have greater bandwidth and higher latency requirements, and pose potential privacy risks and vulnerability to network failures and data breaches.

In an edge-assisted deployment, computations are carried out at the edge nodes or camera gateways while only embeddings, scores or templates are sent to the cloud servers. Such systems reduce bandwidth usage, but may have some latency and require efficient models to be run on edge nodes, and require template security to be handled locally. Edge-assisted face recognition is especially promising for applications in smart campuses, factory access control, smart homes, or surveillance networks of entire cities.

In on-device face recognition, everything is done locally without sending any information to the network. An example of on-device face recognition is when many smartphones or smart locks do face verification locally. While minimizing exposure to network threats, this system requires very little network usage but consumes memory, battery power and CPU usage. Lightweight architectures such as MobileFaceNet and GhostFaceNets have emerged as important methods, where GhostFaceNets uses Ghost Modules and cheap operations to achieve efficient verification with a computational complexity of tens or hundreds of MFLOPs [9].

Federated face recognition is a new method of learning in a distributed manner. Federated learning trains or updates its model locally and sends out encrypted or aggregated model updates instead of raw facial image information. This eliminates raw data exposure to the network, but there are challenges associated with federated learning like non-IID data, poisoning attack vulnerabilities, communication costs, secure aggregation, and model inversion.

Table 4. Deployment architectures and trade-offs.

Architecture	Data transmitted	Advantages	Challenges
Cloud-centric	Raw images/video or face crops.	Centralized compute; simple model updates; scalable gallery search.	High bandwidth, privacy exposure, latency, single breach target.
Edge-assisted	Embeddings, scores or protected templates.	Lower bandwidth and latency; less raw-image leakage.	Edge security, model compression, local template protection.
On-device/mobile	Usually no biometric data leaves device.	Offline operation; strong privacy if storage is secure.	Limited compute, energy and memory; difficult large-gallery search.
Hybrid edge-cloud	Adaptive split based on quality, load or risk.	Balances accuracy, latency and resource usage.	Requires orchestration, QoS and dynamic policy control.
Federated/distributed	Model updates or gradients.	Reduces raw-data centralization; supports distributed learning.	Communication overhead, poisoning, leakage from updates, non-IID data.

6. RESOURCE-EFFICIENT FACE RECOGNITION FOR EDGE AND IOT

Resource efficiency is critical for networked face recognition (FR). Edge nodes and IoT cameras can often run with limited compute/graphics processing unit (CPU/GPU), memory, power, and cooling resources. The operational goal goes beyond benchmark accuracy towards high accuracy with fast speed and low energy. A model with high accuracy in LFW testing may not be feasible in an edge cloud environment if it involves a large number of floating point operations per second (FLOPS) or needs to communicate with the cloud all the time.

Lightweight FR includes lightweight CNNs, depth-wise convolution, ghost modules, pruning, quantization, knowledge distillation, and efficient similarity calculation. MobileFaceNet adopts mobile CNN techniques for face recognition and verification tasks. Similarly, GhostFaceNets are built on GhostNetV1 and GhostNetV2 by removing redundancies and lowering the computational burden of features [9], [34]. Such models make it easier to implement on smartphones, locks, robots, and embedded security cameras.

The evaluation of efficient design should include both recognition and systems measures. It is necessary to assess the processing latency per frame, throughput when handling multiple video streams, energy usage per prediction, memory requirements, storage capacity of the model, communication volume, and the effect under false acceptance rate restrictions. For edge–cloud systems, the optimal point will change depending on network traffic, camera quality, device configuration, and data privacy concerns. System design can adaptively switch from local recognition to cloud-based identification according to image quality, risks, and computing loads.

Table 5. Efficiency techniques for edge and IoT face recognition.

Technique	Purpose	Deployment benefit	Risk/limitation
Lightweight backbone	Reduce parameters and FLOPs.	Enables real-time on-device inference.	Lower capacity can reduce accuracy on hard cases.
Knowledge distillation	Transfer knowledge from large teacher to small student.	Improves compact models without full complexity.	Teacher bias may be inherited.
Quantization	Represent weights/activations with lower precision.	Reduces memory and accelerates inference.	Accuracy may drop at high-security thresholds.
Pruning	Remove redundant weights/channels.	Smaller model and faster execution.	Requires careful retraining and validation.
Edge-cloud partitioning	Split pipeline between edge and server.	Balances local privacy and cloud scalability.	Needs secure communication and orchestration.
Efficient matching	Optimize cosine/distance computation.	Useful for large galleries and low-power devices.	Approximation can affect matching stability.

7. PRIVACY AND SECURITY IN NETWORKED FACE RECOGNITION

Since biometrics (FR) cannot easily change like passwords, privacy-preserving FR (PPFR) must ensure that leaks do not occur throughout the FR lifecycle, including acquisition, preprocessing, feature extraction, inference, storage, transmission, matching, and auditing [11], [12]. The primary privacy threats in PPFR surveys are leakage of raw image, template inversion, linking across different databases, illegal matching, demographic inference, and model leakage. Template inversion is very important because the embedding or template might leak information on the identity even without raw images being sent. Thus, the scheme to protect templates needs to meet irreversibility, unlinkability, and renewability criteria – the protected template should not reveal the biometric value, templates across different systems are not linkable, and leaked templates can be renewed.

The cryptographic approaches are homomorphic encryption, secure multi-party computation, and secure enclaves. While homomorphic encryption allows operations over encrypted data, it imposes high computational latency and bandwidth cost. The secure multi-party computation can be used for private matching but involves considerable communication overhead. To address the threat of leakage in training data or model updates, differential privacy can be employed, although too much noise can reduce accuracy. Federated learning avoids raw data collection at the server side but cannot completely prevent leakage through gradients and model updates, as well as poisoned clients.

In addition to the above threats, networked FR systems are also subjected to presentation and adversarial attacks. Presentation attacks utilize printing, replaying video, mask, or face swapping to impersonate enrolled subjects. Adversarial attacks involve input manipulations to cause false acceptances/rejections. Moreover, there are more attack vectors in the network setting: compromised cameras, man-in-the-middle, replayed embedding, API abuse in cloud-based FR service, poisoned federated clients, and internal threats.

Table 6. Privacy and security mechanisms for networked FR.

Mechanism	Protects against	Advantages	Limitations
Secure transport and storage	Eavesdropping, database breach.	Essential baseline for networked systems.	Does not prevent misuse by authorized services.
Template protection	Inversion, linkability, template theft.	Supports revocability and cross-system separation.	May affect matching accuracy or interoperability.
Homomorphic encryption	Untrusted server inference/matching.	Computes on encrypted data.	High computational and

			communication overhead.
Secure multi-party computation	Private matching across parties.	Reduces exposure between organizations.	Complex protocols; latency can be high.
Differential privacy	Training-data leakage.	Formal privacy guarantee.	Noise-accuracy trade-off.
Federated learning	Central raw-data collection.	Supports distributed training.	Gradient leakage, poisoning, non-IID data.
Liveness/PAD	Spoofing and replay attacks.	Improves access-control security.	False rejections and new attack types remain.
Audit and governance	Unauthorized use and accountability gaps.	Supports compliance and incident response.	Requires organizational discipline and policy enforcement.

8. SYNTHETIC DATA, STYLIZED FACES AND PRIVACY-FRIENDLY TRAINING

Previous success of FR has always been achieved with the help of large face datasets built on crawled web content. Despite the obvious advantage of such datasets, they bring about ethical, legal, and technical issues related to consent, demographic bias, label noise, and accessibility. Notable face datasets have become inaccessible or even retracted due to these reasons, thus driving the community towards the idea of using synthetic faces [13], [14]. The use of synthetic data allows for creating a variety of poses, expressions, lighting conditions, accessories, demographics, and even identity changes without collecting any real biometric information.

A notable example of a synthetic face dataset is DigiFace-1M, created via computer graphics-based pipeline. It has been constructed with the purpose of addressing consent, bias, and label noise problems typical for web-crawled face datasets, and it has proven that proper data augmentation and limited real-world fine-tuning can help close the synthetic-to-real domain gap [14]. Similarly, FRCSyn-OnGoing evaluates synthetic data from the perspective of the challenge, studying aspects such as performance, demographic bias, domain adaptation, and various challenging conditions like age, pose, and occlusions [15]. Altogether, the mentioned works show that synthetic data can be considered an advanced tool for evaluation, augmentation, and training of FR systems while maintaining privacy.

Stylization and generation of faces bring their specific challenges as well. A personalized generative model could create portraits in an artistic, anime, painting, or cyberpunk style of a real identity. Stylized-Face introduces a million-styled-faces dataset and the benchmark to recognize the same identity across different stylistic domains [16]. From the standpoint of governance of generative technologies, portrait right protection, and identity misuse detection, this research direction is extremely important. However, there is a number of complex questions here as well – when does a generated image start keeping an identity, when should we interpret consent, and how do we avoid violating artistry?

Synthetic-data approaches require rigorous evaluation, too. For instance, a model trained purely on synthetic examples would perform worse on real-world faces when distributions differ regarding geometry, texture, lighting, or demographics. On the contrary, using both synthetic and real data can improve robustness or fairness by filling out the gaps in the distribution of real images. The most valuable benchmarks in the future should include both performance scores and fairness gaps, privacy risks, domain gaps, and data-generation specifics.

Table 7. Synthetic and generated data in FR.

Data direction	**Representative examples**	**Primary value**	**Open concern**
Synthetic training data	DigiFace-1M, synthetic-data surveys [13], [14].	Consent-friendly large-scale training; controlled variation.	Synthetic-to-real domain gap.
Challenge benchmarking	FRCSyn-onGoing [15].	Compares synthetic strategies for performance and fairness.	Benchmark tasks may not cover all deployment domains.
Stylized/generated faces	Stylized-Face [16].	Recognizes identities across generative styles; supports media-governance use cases.	Identity preservation and consent boundaries.
Privacy enhancement	Synthetic replacements for sensitive data.	Reduces exposure of real biometric images.	Synthetic data may still encode demographic or identity-like patterns.
Attack simulation	Deepfake, morphing and presentation-attack datasets.	Improves stress testing and robustness.	Arms race with generative attack methods.

9. ROBUSTNESS, FAIRNESS AND ETHICAL DEPLOYMENT

Robustness is a core necessity in networked FR. Natural cameras produce faces in settings with occlusions, masking, illumination changes, pose, motion blur, sensor noise, and low resolution. Masked-face recognition became an important research field after the emergence of COVID-19, employing occlusion-robust feature extraction, occlusion-aware recognition, and occlusion recovery [18]. This task remains difficult since masks deprive faces of discriminative information in the lower region, and most of the existing face databases are artificially masked.

Low-quality FR is critical for surveillance networks and drones. AdaFace explicitly trains on low-quality samples during training [7]. On the other hand, KP-RPE reduces the adverse effect of misalignment in transformer-based face recognizers [8]. In general, this trend indicates that robust FR must incorporate quality modeling and uncertainty management; the system can use quality predictions to decide, e.g., to execute local matching, request a better-quality image, or proceed with cloud computing.

Fairness is yet another critical aspect in practice. Public reports and academic studies have demonstrated that FR and facial analysis algorithms can demonstrate disparate results among demographic groups if their training sets are unbalanced. Synthetic data and benchmarking datasets for fairness and bias mitigation are growing [14], [15], [37], [38]. Thus, a networked FR system must evaluate not only overall accuracy but also conditional FAR and FRR among different demographics. If such technology is used in law enforcement, immigration, and financial inclusion scenarios, then demographic disparities may lead to harmful consequences.

Beyond technical accuracy, responsible usage requires consent, purpose limitation, data minimization, transparency, auditability, and human review. From a legal perspective, FR is a technology with privacy and accountability implications, especially in critical infrastructures and sensitive environments [43]. In addition to technical considerations, ethical deployment calls for questions regarding necessity, availability of alternatives, storage policies, permitted entities, retention period, and rights of individuals against system decisions.

Table 8. Robustness and fairness challenges.

Challenge	Typical cause	System-level mitigation
Masks and occlusion	Covered facial regions; accessories; hands.	Occlusion-aware training, periocular features, quality-aware thresholds.
Low resolution/blur	Distance from camera, motion, compression.	Quality-adaptive learning, frame selection, edge enhancement, cloud escalation.
Misalignment	Landmark failure, pose, poor lighting.	Robust detectors, KP-RPE, multi-frame alignment.
Illumination variation	Night scenes, shadows, backlighting.	VIS/IR fusion, illumination normalization, sensor quality control.
Synthetic-to-real gap	Unrealistic generated data distributions.	Domain adaptation, mixed real/synthetic training, benchmark audits.
Demographic bias	Imbalanced datasets and deployment populations.	Balanced data, fairness metrics, group-specific error reporting, governance.
Spoofing/deepfakes	Presentation attacks and generative media.	PAD, liveness, challenge-response, forensic detectors.

10. EVALUATION METRICS FOR NETWORKED FR

However, accuracy alone could be a misguide when reporting verification system performance. The suite of metrics that need to be presented includes FAR, FRR, EER, ROC curves, AUC, and true acceptance rate at extreme FARs (e.g., TAR@FAR=10^-4). In the case of identification systems, the suite of metrics should include rank-1, rank-k, and CMC curves. In open-set identification, particular emphasis needs to be placed on whether the correct identity exists in the gallery.

In networked FR, the following metrics become particularly important. End-to-end latency should account for face detection, image pre-processing, inference, transmission, feature extraction, feature matching, and decision time. Throughput should be measured using multiple concurrent camera streams. Communication cost should take into account the amount of traffic involved in sending images, embeddings, templates, and model updates. Energy consumption and memory usage should be stated for edge and mobile platforms. Privacy should be evaluated based on template leakage, inversion robustness, linkability, and data retention. Fairness should be assessed in terms of demographic error gaps in addition to the average recognition error.

The most critical demand going forward will be the creation of a standard benchmark testbed for FR that will simultaneously measure recognition, system, and privacy metrics. Current benchmarks tend to focus exclusively on one aspect of FR, such as recognition accuracy, synthetic data performance, face mask handling, or model efficiency. These narrow benchmarks do not adequately characterize the risk posed by deploying biometric services over a network.

Table 9. Recommended metric set for networked FR evaluation.

Metric group	Examples	Why it matters
Recognition	Accuracy, TAR@FAR, EER, rank-1/rank-k, ROC/AUC.	Measures biometric decision quality.
Robustness	Performance under masks, blur, pose, occlusion, low light.	Captures real-world deployment variation.
Network/system	Latency, throughput, bandwidth, packet loss sensitivity, availability.	Measures service feasibility over networks.
Edge efficiency	FLOPs, model size, memory, energy per inference.	Determines mobile/IoT deployability.
Security	Attack success rate, spoof detection, adversarial robustness.	Measures resilience to intentional misuse.
Privacy	Template inversion risk, unlinkability, revocability, leakage from updates.	Measures biometric data protection.
Fairness	Group-conditioned FAR/FRR, gap, standard deviation across groups.	Measures demographic reliability and accountability.

11. APPLICATIONS IN NETWORKED ENVIRONMENTS

Biometrics is also used in door access, office entry, campus security, ATM transactions, and financial services. In these applications, the issues of delay and FAR become very important, as the system makes immediate decisions on the user's identity. The research that merges MTCNN and DeepFace highlights the importance of offline or smart-door biometric solutions, but they have to be reviewed under more general threat models, privacy aspects, and working conditions before going into production [45].

Public surveillance and safety projects rely on multi-camera streams, person detection, tracking, and identification. The review of sensors shows the interactions between person detection, tracking, identification, and FR technologies within live video systems [19]. In these scenarios, there is a high demand for bandwidth and throughput, as well as serious privacy concerns, as people do not give their consent actively. The network architecture should avoid transmitting raw video content and confine identification to legitimate reasons only.

Mobile identification and online banking require secure local storage, liveness detection, and fast local inference. Although cloud-based identification could be useful for account recovery or fraudulent activity detection, the transfer of raw biometric data needs to be limited. Multimodal or multispectral sensors could be valuable in healthcare and border control applications, but they need to be audited and followed for compliance. Smart campuses and attendance management provide examples of the attractiveness of FR technology for automation, but these systems should be developed with the prevention of surveillance creep and the option for users not using biometrics.

12. OPEN CHALLENGES AND FUTURE DIRECTIONS

Optimized accuracy-latency-privacy trade-offs. Future FR systems should adaptively choose between local, edge, and cloud-based processing according to image quality, network congestion, privacy requirements, and security thresholds. This calls for policies that consider biometric identification as a network service with a quantifiable risk level.

Integrated, network-aware performance benchmarks. The benchmarks should measure the impact of recognition accuracy, latency, bandwidth utilization, power consumption, template security, and fairness on each other. Without these criteria, algorithms trained on static image databases might not be as successful in practice.

Edge-cloud matching with end-to-end encryption. Practical secure matching is highly resource-intensive. Further research is needed to minimize the latency and bandwidth overheads introduced by homomorphic encryption and secure multi-party computation on biometric templates.

Federated and continual learning of FR under non-IID data. Distributed FR systems will inevitably face changes in the target population, varying illumination conditions, local biases, and potentially even malicious clients. Secure aggregation, poisoning, and privacy audit techniques need to be implemented.

Reliability and accountability of synthetic data generation. Although synthetic data offers a promising way to reduce the need for real, consent-obtaining data, it must still be rigorously tested for domain adaptation, identity plausibility, fairness, and possible malicious uses. Transparency and auditing standards for generation pipelines are required.

Resilience to generative media and stylization. Various transformations, including deepfakes, face morphing, and stylized identity photos, present difficulties to existing matching algorithms. It is crucial to test FR systems' resilience to these transformations and identity-preserving style changes.

Renewable biometric templates. Biometric templates must be revocable and updatable in case of leakage without compromising on recognition accuracy and inter-vendor compatibility. However, this remains challenging in practical, large-scale, networked systems.

Equity and responsibility of FR use. FR services should support group-level performance monitoring, purpose limitation, user notifications, human review, and appeal mechanisms, especially for critical deployments.

13. CONCLUSION

Facial recognition has evolved to become a highly reliable biometric technology; yet, beyond the mere efficacy of a facial recognition model, its utility in Internet of Things (IoT), edge, mobile, cloud, and surveillance contexts hinges upon other considerations such as latency, bandwidth, energy costs, template safety, privacy, fairness, and robustness. This review presents a taxonomy for current research trends by associating them with pipeline components, model types, architecture options, resource limitations, privacy and security schemes, data augmentation methods, and assessment frameworks.

From the collected literature, it is clear that there does not exist a solution that meets all deployment criteria. Margin losses and adaptive losses improve discriminatory capacity and low-quality tolerance; light models enable edge computing; generated data may mitigate dependence on consent-based Internet datasets; privacy-preserving techniques minimize data leakage but increase complexity; and fair evaluations are required before deploying the system. Thus, future research will need to focus on the development of hybrid solutions that dynamically choose computation and defense strategies based on input quality and network status. A holistic view of face recognition as a networked, secure, and efficient service is vital to biometric authentication in modern systems.

REFERENCES

[1] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2014, pp. 1701-1708.
[2] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2015, pp. 815-823.
[3] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in Proc. British Machine Vision Conference (BMVC), 2015.
[4] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), 2019, pp. 4690-4699.
[5] H. Wang et al., "CosFace: Large margin cosine loss for deep face recognition," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), 2018, pp. 5265-5274.
[6] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "SphereFace: Deep hypersphere embedding for face recognition," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2017, pp. 212-220.
[7] M. Kim, A. K. Jain, and X. Liu, "AdaFace: Quality adaptive margin for face recognition," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), 2022, pp. 18750-18759.
[8] M. Kim, Y. Su, F. Liu, A. Jain, and X. Liu, "KeyPoint relative position encoding for face recognition," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), 2024.
[9] M. Alansari, O. Abdul Hay, S. Javed, A. Shoufan, Y. Zweiri, and N. Werghi, "GhostFaceNets: Lightweight face recognition model from cheap operations," IEEE Access, vol. 11, pp. 35429-35446, 2023.
[10] S. I. Serengil and A. Ozpinar, "A benchmark of facial recognition pipelines and co-usability performances of modules," Bilisim Teknolojileri Dergisi, vol. 17, no. 2, pp. 95-107, 2024.
[11] L. Laishram, M. Shaheryar, J. T. Lee, and S. K. Jung, "Toward a privacy-preserving face recognition system: A survey of leakages and solutions," ACM Computing Surveys, vol. 57, no. 6, Article 147, 2025.
[12] Z. Sun and Z. Liu, "Ensuring privacy in face recognition: A survey on data generation, inference and storage," Discover Applied Sciences, vol. 7, Article 441, 2025.
[13] F. Boutros, V. Struc, J. Fierrez, and N. Damer, "Synthetic data for face recognition: Current state and future prospects," arXiv:2305.01021, 2023.
[14] G. Bae et al., "DigiFace-1M: 1 million digital face images for face recognition," in Proc. IEEE/CVF Winter Conf. Applications of Computer Vision (WACV), 2023, pp. 3526-3535.
[15] I. DeAndres-Tame et al., "Second FRCSyn-onGoing: Winning solutions and post-challenge analysis to improve face recognition with synthetic data," Information Fusion, vol. 120, Article 103099, 2025.
[16] Z. Peng et al., "Stylized-Face: A million-level stylized face dataset for face recognition," in Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV), 2025.
[17] M. Abdul-Al, G. K. Kyeremeh, R. Qahwaji, N. T. Ali, and R. A. Abd-Alhameed, "Fusion-enhanced hybrid multimodal biometric system: Integrating visible and infrared facial recognition for robust authentication," IEEE Access, vol. 14, pp. 6006-6028, 2026.
[18] S. Sharma, M. A. Khan, H. M. Mir, and S. Sharma, "A comprehensive survey on masked face recognition techniques using deep learning: Motivations, research progress, and future challenges," ICT Express, vol. 12, pp. 55-75, 2026.

[19] B. Amirgaliyev, M. Mussabek, T. Rakhimzhanova, and A. Zhumadillayeva, "A review of machine learning and deep learning methods for person detection, tracking and identification, and face recognition with applications," Sensors, vol. 25, Article 1410, 2025.
[20] A. Nemavhola, C. Chibaya, and S. Viriri, "A systematic review of CNN architectures, databases, performance metrics, and applications in face recognition," Information, vol. 16, Article 107, 2025.
[21] H. L. Gururaj, B. C. Soundarya, S. Priya, J. Shreyas, and F. Flammini, "A comprehensive review of face recognition techniques, trends, and challenges," IEEE Access, vol. 12, pp. 107903-107926, 2024.
[22] N. El Fadel, "Facial recognition algorithms: A systematic literature review," Journal of Imaging, vol. 11, Article 58, 2025.
[23] J. Guo, J. Deng, N. Xue, and S. Zafeiriou, "Stacked dense U-nets with dual transformers for robust face alignment," in Proc. BMVC, 2018.
[24] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, "MS-Celeb-1M: A dataset and benchmark for large-scale face recognition," in Proc. European Conf. Computer Vision (ECCV) Workshops, 2016.
[25] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "VGGFace2: A dataset for recognising faces across pose and age," in Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition, 2018, pp. 67-74.
[26] 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," University of Massachusetts, Amherst, Technical Report 07-49, 2007.
[27] I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller, and E. Brossard, "The MegaFace benchmark: 1 million faces for recognition at scale," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016, pp. 4873-4882.
[28] B. Maze et al., "IARPA Janus Benchmark-C: Face dataset and protocol," in Proc. Int. Conf. Biometrics (ICB), 2018.
[29] B. Maze et al., "IARPA Janus Benchmark-S: Face recognition in video surveillance," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition Workshops, 2018.
[30] Z. Zhu et al., "WebFace260M: A benchmark unveiling the power of million-scale deep face recognition," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), 2021, pp. 10492-10502.
[31] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499-1503, 2016.
[32] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou, "RetinaFace: Single-shot multi-level face localisation in the wild," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), 2020, pp. 5203-5212.
[33] A. G. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," arXiv:1704.04861, 2017.
[34] S. Chen, Y. Liu, X. Gao, and Z. Han, "MobileFaceNets: Efficient CNNs for accurate real-time face verification on mobile devices," in Proc. Chinese Conf. Biometric Recognition, 2018, pp. 428-438.
[35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778.
[36] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in Proc. Int. Conf. Learning Representations (ICLR), 2021.
[37] J. Buolamwini and T. Gebru, "Gender shades: Intersectional accuracy disparities in commercial gender classification," in Proc. Conf. Fairness, Accountability and Transparency, 2018, pp. 77-91.
[38] K. Karkkainen and J. Joo, "FairFace: Face attribute dataset for balanced race, gender, and age for bias measurement and mitigation," in Proc. IEEE/CVF Winter Conf. Applications of Computer Vision (WACV), 2021, pp. 1548-1558.
[39] Y. Liu et al., "MagFace: A universal representation for face recognition and quality assessment," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), 2021, pp. 14225-14234.
[40] J. Deng et al., "Sub-center ArcFace: Boosting face recognition by large-scale noisy web faces," in Proc. European Conf. Computer Vision (ECCV), 2020, pp. 741-757.
[41] S. Sengupta, J.-C. Chen, C. Castillo, V. M. Patel, R. Chellappa, and D. W. Jacobs, "Frontal to profile face verification in the wild," in Proc. IEEE Winter Conf. Applications of Computer Vision (WACV), 2016, pp. 1-9.
[42] R. Kute, V. Vyas, and A. Anuse, "Transfer learning for face recognition using fingerprint biometrics," Egyptian Informatics Journal, 2021.
[43] S. Al Ali, I. Al Shawabkeh, and E. M. AlQodsi, "Facial recognition technology: Protecting biometric privacy in the digital age," Journal of Legal Affairs and Dispute Resolution in Engineering and Construction, 2026.
[44] Agustiyar, R. R. Isnanto, and C. E. Widodo, "Face recognition for attendance systems: A bibliometric review of research trends and opportunities," Jurnal SISFOKOM, vol. 15, no. 1, pp. 8-13, 2026.
[45] V. Godase, "Optimized algorithm for face recognition using Deepface and Multi-task Cascaded Convolutional Network (MTCNN)," Optimum Science Journal, vol. 3, pp. 66-74, 2025.
[46] D. Sanyal, G. Saha, A. Pakhira, A. Bhattacharya, and P. Sarkar, "Harnessing ancient Vedic mathematics with Karatsuba multiplication for efficient and secure face recognition in surveillance," Discover Artificial Intelligence, vol. 6, Article 146, 2026.