Face De-Identification: A Domain-Centric Survey from Capture to Processing

Hui Wei, Hao Yu, and Guoying Zhao, Fellow, IEEE

Abstract—Face de-identification (De-ID) aims to remove or conceal personally identifiable facial features in images or videos to prevent identity recognition while preserving utility for downstream tasks. With the rising emphasis on data privacy and responsible AI, face De-ID has emerged as an active research area spanning computer vision and privacy-preserving communities. Early approaches, and many contemporary ones, operate in the **digital** domain by modifying pixel-level or appearance-level features through post-capture processing. Recent advances extend face De-ID beyond post-processing by integrating privacy mechanisms directly into **sensors** during image acquisition, bridging sensing systems and downstream vision algorithms. In parallel, **physical**-domain methods explore wearable accessories and materials that conceal identity information in real-world environments prior to capture. In this survey, we present the first unified overview that spans the full data acquisition pipeline, encompassing the physical, sensor, and digital domains. Through this domain-centric lens, we systematically analyze current methodologies, technical progress, and the distinct challenges inherent to each stage. We then review and organize existing evaluation protocols, examining current practices and highlighting the critical need for standardized, comprehensive benchmarks. Finally, we identify key open problems and outline emerging research directions to guide future work in this rapidly evolving field. To support ongoing research, we maintain a project page that organizes relevant literature with collected datasets and open source code: https://github.com/CV-AC/Awesome-FaceDe-ID.

Index Terms—Face De-identification, Utility Preservation, Evaluation Protocol, Privacy Protection, Literature Survey.

1 Introduction

Human face contains abundant identity-related information, making it a cornerstone of biometric recognition systems in modern society [1], [2]. However, the widespread collection and utilization of facial data pose significant privacy, security, and ethical concerns [3], [4]. Recent regulatory frameworks, most notably the EU AI Act (Regulation (EU) 2024/1689), underscore that the misuse of facial data constitutes a severe infringement of fundamental rights, including the right to privacy [5]. To mitigate these risks, the research community has developed **face de-identification (De-ID)**, which is defined as the class of image- or video-domain transformations that conceal personally identifiable facial information such that the depicted subject cannot be recognized, by either an automated face recognizer or a human observer, while non-identity attributes required by downstream applications (*e.g.*, medical analysis [6] and attribute recognition [7]) remain usable.

In recent years, face De-ID has been widely studied and has expanded from purely digital treatments to encompass sensor-integrated and physical-world interventions (see Fig. 1). Early face De-ID efforts predominantly operated in the **digital domain**, manipulating recorded imagery via hand-crafted filters (e.g., blurring, pixelation) [8] or statistical anonymization mechanisms such as k-Same family [9]. Subsequent advances leveraged adversarial perturbations [10], [11] and generative models [12], [13] to synthesize photorealistic yet privacy-safe faces. While these digital approaches achieved remarkable fidelity and controllability, they remain post hoc interventions, exposing raw sensitive data prior to De-ID [14]. To mitigate this vulnerability, recent research has shifted privacy protection into the imaging process. **Sensor domain** face De-ID paradigms integrate optical and computational imaging designs that suppress identity cues at capture time, embedding privacy filters in optics or sensors to eliminate identifiable cues before digital conversion (*e.g.*, PrivacyOptics [15]). In parallel, **physical domain** strategies extend face De-ID into the real world by modifying subjects' appearance through adversarial wearables [16], cosmetics [17], or projected light patterns [18] that confuse recognition systems under unconstrained environments. Collectively, these directions represent a continuum from post-capture face De-ID to sensor-level and physical-level protection, each offering distinct advantages and challenges regarding effectiveness, utility, and realism.

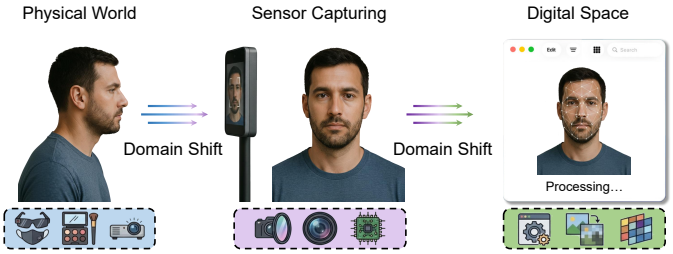


Fig. 1: **Overall perspective of this face De-ID survey paper**. The process begins in **physical** world, where identity cues are modified directly on the subject through adversarial wearables, cosmetics, or projection-based interventions. It then transitions through the **sensor**-capturing stage, where privacy-preserving optics and computational imaging encode De-ID at the sensor level. Finally, in **digital** space, post-capture algorithms apply pixel- or feature-level transformations to conceal identity while retaining utility.

- *Hui Wei, Hao Yu, and Guoying Zhao are with the ELLIS Institute Finland, Espoo, 02150, Finland, and also with the Center for Machine Vision and Signal Analysis, University of Oulu, 90570 Oulu, Finland.*
- *Guoying Zhao is the corresponding author. E-mail: guoying.zhao@oulu.fi*

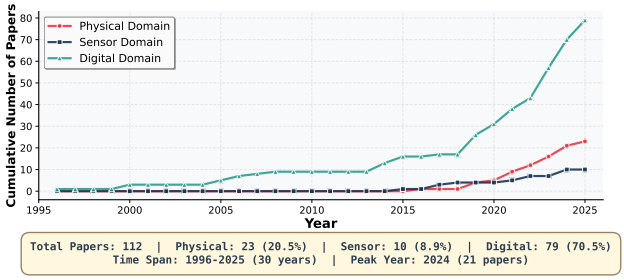


Fig. 2: **Cumulative growth of face De-ID methods across domains**. Data collected via Google Scholar using keywords: "face de-identification", "face privacy preserving", and "face adversarial attack".

This paper provides a unique and up-to-date survey of the rapidly growing area of face De-ID (Fig. 2 illustrates the growth trajectory and domain-specific research trends). Although several survey papers have reviewed aspects of face De-ID, existing works remain fragmented and narrow in scope (see TABLE 1). Specifically, most prior surveys concentrate only on digital domain techniques [19], [20], overlooking the rapid evolution of sensor-based and physical face De-ID methods that increasingly shape real-world privacy protection. Moreover, current surveys rarely provide evaluation protocol analysis [21], [22], resulting in insufficient overview of the datasets and metrics that are critical for advancing this field.

To address these gaps, this survey presents a comprehensive review and critical analysis of face De-ID techniques from **physical**, **sensor**, and **digital** perspectives, spanning more than three decades of research. Our goal is to bridge the fragmented literature across these domains, establish a unified taxonomy, and highlight both the fragmentation of evaluation protocols in this field and the emerging convergence between optical design, adversarial learning, and generative synthesis for privacy preservation. Specifically, this paper makes the following contributions:

- We provide the first **domain-centric taxonomy** of face De-ID methods, systematically categorizing approaches according to where privacy transformation occurs: before, during, or after image capture, and summarizing representative methodologies within each domain.
- We perform an **in-depth analysis of methodologies and trends** across physical, sensor-based, and digital face De-ID, highlighting their design principles, advantages, and limitations.
- We review and summarize **evaluation protocols**, emphasizing the need for unified assessment frameworks that jointly measure privacy protection, utility preservation, and visual quality.
- We identify **open challenges and future directions**, including cross-modal privacy preservation, foundation model level face De-ID, verifiable and reversible De-ID, and fairness-aware privacy guarantees.

The remainder of this paper is organized as follows. Sec. 2 establishes conceptual foundations. Sec. 3 reviews methodologies across the three domains. Sec. 4 systematizes evaluation protocols. Sec. 5 outlines open problems and Sec. 6 concludes the survey.

TABLE 1: **Summary of existing survey papers related to face De-ID**. Prior works predominantly focus on digital-domain methods, with limited exploration of sensor-based or physical de-identification. Our survey provides the first unified, domain-centric overview that spans all three levels and includes evaluation protocol analysis.

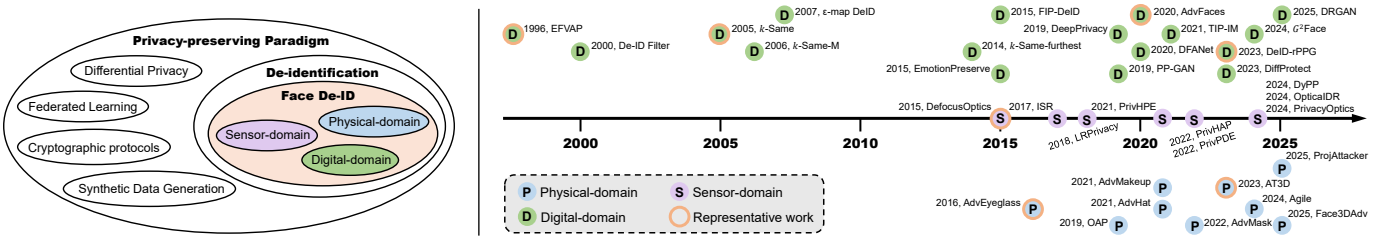
Paper	Year	Physical	Sensor	Digital	Protocol	Perspective
[23]	2015	✗	✓	✓	✗	Camera
[24]	2015	✗	✓	✓	✗	Image/video
[25]	2016	✗	✓	✓	✗	Multimedia
[26]	2021	✗	✓	✓	✗	Pipeline-level
[19]	2022	✗	✗	✓	✗	Social network
[20]	2022	✗	✗	✓	✗	3D face De-ID
[21]	2023	✓	✗	✗	✗	Presentation-level
[22]	2024	✗	✗	✓	✗	Processing level
[27]	2024	✗	✗	✓	✗	Person De-ID
[28]	2024	✗	✗	✓	✗	Digital era
[29]	2025	✗	✗	✓	✗	Generative
[30]	2025	✗	✗	✓	✗	Deepfake
[4]	2025	✗	✓	✓	✗	Leakage/solution
Ours	—	✓	✓	✓	✓	Domain-centric

2 PRELIMINARIES

Research on face De-ID traces back to the early era of computational face analysis. The first successful demonstration of automatic face recognition, Eigenfaces [31] by Turk and Pentland in 1991, highlighted both the feasibility of machine-based identification and the accompanying privacy risks. Shortly thereafter, the first dedicated De-ID technique, EFVAP [8], was introduced in 1996, marking the beginning of systematic efforts to suppress identifiable facial information. Since then, the field has expanded into a broad research direction. Fig. 3(b) presents a curated chronological trajectory of representative face De-ID methods, revealing a clear shift in research emphasis over time. A work appears in the timeline if it satisfies at least two of the following criteria: *(i)* it introduced a methodological primitive subsequently adopted by other works (*e.g.*, k-Same [9], AdvEyeglass [16], DefocusOptics [14]); *(ii)* it is the first reported method in its sub-category (*e.g.*, EFVAP [8]); *(iii)* it represents the current state of the art or a distinct subsequent generation within a sub-category (*e.g.*, AT3D [32], G^2Face [33]); or *(iv)* it has high scholarly impact (citations $\geq$ 200 for works published before 2022 and $\geq$ 50 from 2022 onwards). Exhaustive coverage of all 112 surveyed methods, including those not selected for the timeline, is provided in TABLE 3.

2.1 Face De-ID *vs.* Privacy-Preserving

As shown in Fig. 3(a), face De-ID represents one branch within a broader family of privacy-preserving methodologies. More broadly, privacy-preserving methodologies comprise complementary paradigms, including differential privacy [34], federated learning [35], cryptographic protocols [36], synthetic data generation [37], and de-identification [28], [38], each of which addresses distinct dimensions of data protection. Face De-ID is a vision-centric technique that complements but does not replace other privacy paradigms. Recognition in video and surveillance further exploits complementary modalities including body shape, gait, and other behavioral biometrics (*e.g.*, voice, eye gaze, hand motions). Face De-ID is thus one component of a broader *person* De-ID pipeline.



(a) Relationship between face De-ID and other privacy-preserving paradigms.

(b) Chronological overview of representative face De-ID works. The timeline highlights the shift from early digital methods to sensor-based and physical solutions.

Fig. 3: **Conceptual** (left) and **historical** (right) overview of face De-ID, clarifying the paper's domain-centric taxonomy.

2.2 Face De-ID *vs.* Face Identification

Face identification maximizes the discriminability of identity-bearing cues, whereas face De-ID suppresses those cues while preserving non-identity information. Most face De-ID approaches therefore assume a strong face recognition (FR) adversary as their primary threat model. Modern FR systems extract discriminative embeddings optimized for intra-class compactness and inter-class separability through metric learning objectives. Representative architectures include FaceNet [39], which pioneered triplet-loss deep metric learning; SphereFace [40], which introduced angular softmax for hypersphere embedding; ArcFace [41], the most prevalent in De-ID research owing to its additive angular margin loss; and AdaFace [42], which adapts margins to image quality. These systems typically operate on faces localized by detectors such as RetinaFace [43].

This adversarial relationship runs throughout the De-ID pipeline: physical methods optimize perturbations to fool FR models in the real world [32]; sensor-based approaches design optics that suppress identity-discriminative frequencies [44]; and digital methods incorporate pre-trained FR models as verifiers in their training objectives [45]. With FR systems defining both the attack surface and the evaluation criteria, face identification is foundational to designing and benchmarking De-ID techniques.

2.3 Problem Formulation

Let $\mathbf{I} \in \mathbb{R}^{H\times W\times 3}$ denote an input face image (or video frame) containing personally identifiable information about an individual. A face De-ID method seeks a transformation $\mathcal{F}(\cdot)$ that produces a de-identified output $\tilde{\mathbf{I}} = \mathcal{F}(\mathbf{I})$ in which identity-related cues are concealed while non-identity, task-relevant information is preserved. For applications that additionally require authorized recovery (*e.g.*, forensic investigation, medical audit, legal disclosure), $\mathcal{F}$ is paired with a key-conditioned inverse $\mathcal{R}(\cdot\,;\,k)$ such that $\mathcal{R}(\tilde{\mathbf{I}};\,k) \approx \mathbf{I}$ only when k is the legitimate recovery key.

Formally, we cast face De-ID as the joint *maximization* of privacy, utility, visual-quality, and reversibility *scores*:

$$\max_{\mathcal{F},\,\mathcal{R}} \; \lambda_p \mathcal{P} + \lambda_u \mathcal{U} + \lambda_q \mathcal{Q} + \delta_{\text{rev}}\, \lambda_r \mathcal{S}, \tag{1}$$

where $\mathcal{P}(\mathbf{I},\tilde{\mathbf{I}})$, $\mathcal{U}(\mathbf{I},\tilde{\mathbf{I}})$, $\mathcal{Q}(\tilde{\mathbf{I}})$, and $\mathcal{S}(\mathbf{I},\tilde{\mathbf{I}},\mathcal{R};\,k)$ denote privacy, utility, visual-quality, and key-conditioned reversibility *scores* (each to be maximized), $\lambda_p, \lambda_u, \lambda_q, \lambda_r \geq 0$ are trade-off weights, and $\delta_{\text{rev}} \in \{0,1\}$ is an application-level switch that enables the reversibility term when authorized recovery is required and disables it otherwise. The expected solution should satisfy the following criteria:

1) **Privacy Protection:** The de-identified $\tilde{\mathbf{I}}$ should not match any other identity. Formally, for any recognizer f_{id}, $\Pr\Big(f_{\text{id}}(\tilde{\mathbf{I}}) = f_{\text{id}}(\mathbf{I}_i)\Big) \approx 0,\ \forall i$, where $\mathbf{I}_i$ denotes any image of identity i. This prevents both re-identification to the original and misattribution to others.
2) **Utility Preservation:** For a set of downstream tasks $\mathcal{T} = \{t_1, \ldots, t_n\}$, the task-specific predictors f_{t_j} should maintain consistent outputs before and after face De-ID, *i.e.*, $f_{t_j}(\tilde{\mathbf{I}}) \approx f_{t_j}(\mathbf{I})$.
3) **Visual Fidelity and Naturalness:** The de-identified image should remain photorealistic and free from perceptible artifacts, ensuring usability and social acceptability in real applications.
4) **Reversibility:** When $\delta_{\text{rev}} = 1$, there should exist a key-conditioned inverse $\mathcal{R}(\cdot\,;\,k)$ satisfying $\mathcal{R}(\tilde{\mathbf{I}};\,k) \approx \mathbf{I}$ for the legitimate key while $\Pr\Big(\mathcal{R}(\tilde{\mathbf{I}};\,k') \approx \mathbf{I}\Big) \to 0$ for any $k' \neq k$. Reversibility is orthogonal to privacy: a method may be strongly de-identifying yet reversible (*e.g.*, password-conditioned generative inversion), strongly de-identifying and irreversible, or neither.

2.4 Taxonomy: A Domain-Centric Perspective

Face De-ID methods are fundamentally characterized by *where* in the imaging pipeline the privacy transformation occurs. This survey adopts a domain-centric taxonomy with three classes, **physical**, **sensor**, and **digital**, each embodying distinct design constraints and trade-offs.

Physical Domain. A transformation $\mathcal{F}_p$ alters the subject's real-world appearance before capture (*e.g.*, adversarial accessories, makeup, light projection, 3D mask) so that the captured image $\tilde{\mathbf{I}}$ inherently exhibits suppressed identity. These methods strongly protect against raw-data exposure but offer limited controllability and may be conspicuous.

Sensor Domain. Imaging components $\mathcal{F}_s$ embed privacy during acquisition through optical coding, sensor-level transformations, or computational imaging (*e.g.*, learned phase masks, PSF manipulation), encoding $\tilde{\mathbf{I}} = \mathcal{F}_s(\mathbf{I})$ in the raw measurement space so that no identifiable face is produced. They provide strong trust-boundary guarantees but face stringent hardware constraints (latency, resolution, power) and deployment challenges.

Digital Domain. Post-capture methods operate on recorded imagery, modifying identity through filtering, mapping, adversarial perturbations, or generative synthesis (*e.g.*, GANs [46], diffusion models [47]) to obtain $\tilde{\mathbf{I}} = \mathcal{F}_d(\mathbf{I})$. They are the most flexible and widely adopted, enabling fine-grained control and high fidelity, but expose raw data before De-ID and remain most vulnerable to re-identification.

2.5 Cross-Domain Interactions

A genuine unification of physical, sensor, and digital De-ID must articulate not only what each domain does in isolation, but how interventions in one domain interact with the others along the acquisition pipeline.

First, a physical-domain perturbation must survive lens distortion, sensor noise, demosaicing, white balance, and downstream JPEG/H.264 compression before reaching a digital recognizer. Empirically, perturbations optimized without explicit modeling of the ISP pipeline (*e.g.*, early adversarial patches) lose substantial attack success after realistic capture. Methods that explicitly model the capture chain (ProjAttacker [18], AT3D [32]) close this gap, illustrating the necessity of co-design between physical perturbations and the sensing stage they must traverse. Second, multi-stage interventions can compound or conflict. A sensor-domain phase mask combined with a downstream digital generative refiner (PrivacyOptics [15]) leverages the strengths of both: the optics provide hardware-level identity suppression while the generator restores task utility. Conversely, a physical adversarial patch followed by a digital denoiser may be neutralized, since the patch is treated as removable noise. Third, an intervention at an earlier stage can obviate the need for one at a later stage. Extreme low-resolution capture (ISR [48]) makes downstream digital pixelation redundant; conversely, strong digital generative De-ID (Swapping-DeID [13]) makes physical perturbations unnecessary when full digital pipeline control exists. Fourth, the three domains differ fundamentally in where raw identity-bearing signal first exists in digital form. Physical-domain De-ID prevents identifiable signal from ever being produced; sensor-domain De-ID ensures that identifiable signal never leaves the camera; digital-domain De-ID accepts a fully identifiable signal and removes identity in software layer. The choice of domain is therefore primarily a choice of trust model rather than a choice of algorithm.

3 Face De-ID Methodologies

Building upon the domain-centric taxonomy established in Sec. 2.4, we now present a comprehensive analysis of face De-ID methodologies across the physical, sensor, and digital domains. Fig. 4 provides a visual overview of the face De-ID methods discussed in this section.

3.1 Physical-Domain Face De-ID

Physical-domain face De-ID manipulates the physical appearance of individuals or their surrounding environment to prevent FR systems from correctly identifying them. Following Sec. 2.5, these methods intervene at the earliest stage of the acquisition pipeline: identity cues are suppressed on the subject before any digital signal exists, but the perturbation must subsequently survive optics, sensor, and ISP processing to remain effective at the recognizer. Based on their mechanism, existing methods fall into three classes: **wearable adversarial accessories** (printed or fabricated items such as eyeglasses, hats, stickers, masks, and bandages), **projected perturbations** (external devices that cast adversarial patterns onto faces using visible or infrared light), and **adversarial illumination** (manipulated environmental lighting that creates naturally-appearing yet recognition-disrupting effects). Fig. 5 illustrates representative examples.

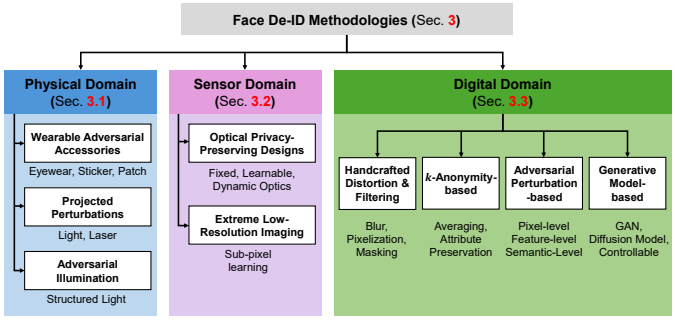


Fig. 4: **Hierarchical taxonomy of face De-ID methodologies**. Each domain is further categorized into specific subclasses based on their technical approaches.

Unlike digital or sensor methods, physical-domain face De-ID is governed by four dimensions that shape every design choice in this section. *(1) Manufacturability and configurability*: the perturbation must be physically realizable using printing, fabrication, or projection hardware with bounded resolution, color gamut, and material reflectance. *(2) Real-world robustness*: because the perturbation cannot be re-tuned at deployment time, it must remain effective across viewpoint, distance, motion, illumination, weather, background, and partial occlusion. *(3) Conspicuity and social acceptability*: the perturbation is worn or projected in public, so its visual impact and social cost (legality, suspicion, attention) directly determine deployability, independently of attack success rate. *(4) Propagation through sensor capture*: the perturbation passes through optics, sensor noise, ISP processing, and compression before reaching the recognizer, all of which can attenuate or destroy the adversarial signal. The following subsections examine the three classes of methods along these four axes.

3.1.1 Wearable Adversarial Accessories

Eyeglass-Based Approaches. Sharif *et al.* [16] (AdvEyeglass) established the foundational paradigm for physical attacks against FR, introducing eyeglass frames with optimized textures that enable evasion or impersonation. Building on this, AGNs [54] introduced a GAN-based framework that learns to generate adversarial eyeglasses from real eyeglass datasets, addressing vaguely specified objectives such as inconspicuousness that resist mathematical modeling.

Patch and Sticker-Based Methods. The small perturbation area imposed by L_0-norm restrictions on eyeglasses motivates larger patches and stickers. AdvHat [49] affixes rectangular stickers to hats for larger perturbation areas while preserving natural appearance. OAP [55] proposes grayscale patches for facial regions or eyeglasses, with a reproducible pipeline accounting for printing and camera color discrepancies. Physical AX [56] contributes a threshold-based smoothness loss with delayed total-variation penalties that improves convergence, and a patch-noise combo attack combining localized patches with imperceptible full-face noise for higher white-box and black-box success.

A central challenge is transferability in query-free black-box settings against unknown commercial systems. GenAP [57] regularizes patches on low-dimensional manifolds learned by pre-trained generative models (Pro-

(a) AdvEyeglass [16] 2016 (b) AdvHat [49] 2020 (c) ALPA [50] 2020 (d) AdvMakeup [17] 2021 (e) AdvMask [51] 2022 (f) AdvSticker [52] 2022 (g) AT3D [32] 2023 (h) Optical De-ID [53] 2023 (i) ProjAttacker [18] 2025

Fig. 5: **Sample figures of representative physical-domain face De-ID methods**. Wearable adversarial accessories: (a), (b), (d), (e), (f), (g). Projected perturbations: (c), (i). Adversarial illumination: (h).

GAN [58], StyleGAN [59], [60]), constraining them toward natural face features and narrowing the substitute-target response gap. PadvFace [61] models physical-world variations including chromatic aberration, sticker deformation, and illumination changes. EAP [62] targets impersonation via random similarity transformations, image pyramids, and meta-ensemble attacks for robustness across scales, resolutions, and models. Other work moves beyond optimized patterns: AdvSticker [52] uses real-life meaningful stickers (*e.g.*, decorative designs); PPAttack [63] jointly optimizes patch perturbation and position via reinforcement learning; and FaceAdv [64] uses Grad-CAM [65] to locate critical regions, with a generator-converter architecture simulating physical capture under limited sticker area.

Makeup and Mask-Based Approaches. Cosmetics offer a vehicle for inconspicuous perturbations. AdvMakeup [17] pioneered this with a GAN-based approach synthesizing natural adversarial eye shadow over the orbital region, combining gradient constraints with VGG16-based style/content losses [66] for imperceptibility. The COVID-19 normalization of face masks enabled mask-based methods. Adversarial Mask [51] embeds Universal Adversarial Perturbations into fabric masks through differentiable digital masking that places perturbations precisely along facial contours. SASMask [67] instead optimizes adversarial mask styles via continuous relaxation, maximizing impersonation success while staying visually realistic across digital, physical, and commercial platforms. AdvBandage [68] uses perturbed medical bandages as discreet patches, optimizing bandage size and placement with iterative FGSM [69] for both dodging and impersonation, with bandages providing natural cover without arousing suspicion.

3D Adversarial Meshes. 2D accessories are limited by detectability under anti-spoofing defenses and reduced effectiveness against commercial systems. 3D adversarial meshes address this by modeling the full geometry of facial perturbations for robust attacks across viewing conditions. AT3D [32] pioneered this with Adversarial Textured 3D Meshes that are 3D-printed and applied directly to faces; its key innovation is optimization in the low-dimensional coefficient space of 3D Morphable Models (3DMM) [70] rather than high-dimensional mesh space, which accelerates optimization, avoids local optima, and improves black-box transferability. Physical tests showed evasion of four commercial anti-spoofing APIs, two mobile systems, and two access control systems. Face3DAdv [71] extends this with a controllable simulation framework that reconstructs full 3D face information (texture, shape, viewpoint, lighting) via a differentiable renderer, using importance sampling to prioritize critical physical transformations.

3.1.2 *Projected Perturbations*

Projection-based methods shift from accessories to artifact-free De-ID by manipulating light patterns cast onto faces, offering real-time adaptability and multi-target flexibility.

ALPA [50] pioneered this direction with real-time adversarial light projections using an off-the-shelf camera-projector setup, incorporating landmark-based position calibration and Lab-space color calibration to accurately reproduce adversarial patterns, while generating transformation-invariant patterns through representative face averages. ProjAttacker [18] addressed digital-physical gap challenges through 3DMM [70] initialization for geometry alignment, Light Reflection Function modeling for interactions among projected light, skin reflectance, and ambient illumination, and a differentiable camera ISP proxy network simulating real-world imaging variations. Agile [72] pushes stealthiness further with adjustable, invisible infrared laser emissions directed into camera CMOS sensors rather than visible-light projection, enabling Denial-of-Service, dodging, and impersonation attacks.

3.1.3 *Adversarial Illumination*

Unlike projection-based methods that actively cast patterns onto faces, adversarial illumination exploits FR sensitivity to lighting by manipulating environmental lighting or the illumination used in specialized 3D acquisition systems, achieving De-ID through naturally-appearing lighting variations that avoid the conspicuousness of projected patterns.

Optical De-ID [53] targets structured-light-based 3D FR by corrupting the 3D data acquisition process itself, integrating 3D reconstruction and skin reflectance models into optimization through two strategies: phase shifting attacks that modify projected structured light patterns, and phase superposition attacks that project external adversarial noise. The dual approach enables adversarial point placement anywhere on the face with robustness to head movements. ARA [73] exploits illumination vulnerabilities in 2D FR by producing naturally relighted face images, combining physical model-based optimization (which guides adversarial lighting using feedback from FR systems) with efficient neural-network-based prediction (instantaneous prediction of adversarial light settings), with physical-world verification using precise relighting hardware.

3.1.4 *Discussion*

Physical-domain face De-ID has progressed from handcrafted, accessory-bounded textures to 3D- and optics-aware pipelines that explicitly model capture physics and sensor behavior. Reviewing this progress along the four physical-domain dimensions, several gaps persist. On *manufacturability*, low-dimensional priors such as GAN manifolds [46], 3DMM coefficients [70], and style spaces [57], [67] have systematically improved transfer and realism by constraining

perturbations near the face manifold, but standardized fabrication tolerances and material-aware optimization remain underdeveloped, particularly for 3D-printed meshes [32], [71] where print resolution and material reflectance directly affect deployment fidelity. On *real-world robustness*, non-contact photonic channels (visible/IR projection [18], [50], [72], structured-light interference [53]) widen the attack surface while alleviating occlusion-induced liveness failures, yet standardized benchmarks encompassing multi-view video, multi-sensor (RGB-IR-depth) fusion, and liveness under varied ambient conditions are still lacking, as are cross-system generalization studies under policy and threshold changes (including open-set verification) and principled evaluations of long-term stability such as washability, wear, and battery/thermal constraints for emitters [50], [72]. On *conspicuity and social acceptability*, quantitative stealth metrics aligned with human perception and social norms are needed; current proxies based on area or total-variation [56] under-capture social cost, and even methods designed for inconspicuousness [17], [51], [68] rely largely on qualitative argument rather than measurement. On *propagation through sensor capture*, methods that explicitly model the capture chain [18], [32] demonstrate the value of physical–sensor co-design, but co-design with defenses that jointly perform spoof detection, quality assessment, and identity inference remains largely unexplored. Addressing these gaps will require reproducible, physics-grounded testbeds, shared calibration protocols, and reporting standards that elevate claims from single-device demonstrations to robust, deployable face De-ID guarantees.

3.2 Sensor-Domain Face De-ID

Sensor-domain face De-ID methods intervene during image acquisition, preventing identifiable facial information from ever being recorded. Following the cross-domain analysis of Sec. 2.5, sensor-domain methods intervene inside the camera: identity-bearing signal exists optically but is encoded before digitization, and these methods frequently combine with downstream digital generators that restore task utility from the privacy-encoded measurement. These techniques fall into two paradigms: **optical privacy-preserving designs** that manipulate incident light through defocus, aberrations, or phase masks before it reaches the sensor, and **extreme low-resolution imaging** that captures faces below the recognition threshold. Fig. 6 contrasts the two paradigms: (a) shows the optical pipeline, in which a point source is encoded by the camera optics into a privacy-preserving point spread function (PSF) before reaching the sensor, across its fixed, learnable, and dynamic generations; (b) shows the low-resolution pipeline, in which a tiny camera records faces too coarse for recognition while a downstream network still recovers task utility.

3.2.1 Optical Privacy-Preserving Designs

Optical designs encode privacy in the imaging optics themselves. As illustrated in Fig. 6(a), they have progressed through three generations, namely *fixed* designs with hand-specified degradation, *learnable* designs that jointly optimize optics and downstream networks, and *dynamic* designs that vary the optical configuration at capture time.

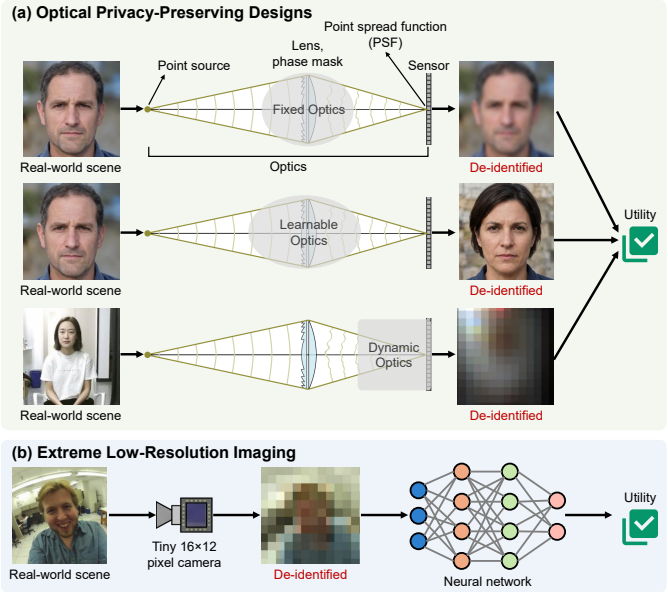


Fig. 6: **Taxonomy of sensor-domain face De-ID paradigms. (a) Optical privacy-preserving designs** encode privacy in the camera optics: a point source is shaped by a lens and phase mask into a point spread function (PSF) recorded by the sensor, yielding a de-identified image while downstream utility is retained. Three generations are shown, namely *fixed* optics with hand-specified degradation (DefocusOptics [14], DefocusOptics$^+$ [75]), *learnable* optics jointly optimized with downstream networks (PrivHPE [76], PrivHAR [77], PrivPDE [78], OpticalDR [44], PrivacyOptics [15]), and *dynamic* PSF that randomize the configuration per capture (DyPP [79]). **(b) Extreme low-resolution imaging** uses a tiny camera (e.g., 16×12 pixels) to capture faces below the recognition threshold, with a downstream neural network recovering task utility (ISR [48], LRPrivacy [80]).

Fixed Optics. Early optical privacy methods employed fixed designs to attenuate facial features. DefocusOptics [14] pioneered k-anonymity [74] preserving optical designs using intentional defocus blur, providing foundations for miniaturizing privacy optics within small sensor volumes. DefocusOptics$^+$ [75] extended this with programmable optics enabling pre-capture mask-based techniques. However, these hand-specified defocus mechanisms proved vulnerable to deconvolution attacks, motivating learnable optical designs with complex, irreversible degradations.

Learnable Optics. Learnable optical approaches jointly optimize hardware and software components for task-specific privacy-utility trade-offs. PrivHPE [76] introduced end-to-end optimization of optical encoders (parametrized via Zernike coefficients) and CNN decoders for pose estimation, using privacy-preserving loss functions to minimize face keypoint detection while maintaining task performance. PrivHAR [77] advanced this through adversarial training of phase masks for action recognition, achieving near-random classification performance on FR while preserving action recognition accuracy. Further refinements emerged through adversarial learning frameworks: PrivPDE [78] balanced FR prevention with depth estimation via aperture-plane phase masks, OpticalDR [44] achieved near-random FR while preserving depression-related features, and PrivacyOptics [15]

combined optical encoding with GAN-based face synthesis to close the security gap between capture and De-ID.

Dynamic Optics. A critical vulnerability of fixed and learnable-but-static designs is susceptibility to PSF inversion attacks, where adversaries recover the point spread function to reconstruct faces. DyPP [79] addressed this by introducing time-varying PSFs sampled from a learned privacy manifold, implemented via spatial light modulators. This dynamic design randomizes the optical configuration for each capture, significantly enhancing robustness against inversion attacks while maintaining task utility for object detection and pose estimation.

3.2.2 Extreme Low-Resolution Imaging

Rather than encoding privacy in the optics, extreme low-resolution imaging achieves face De-ID by capturing at resolutions (e.g., 16×12 pixels) where facial detail is inherently insufficient for reliable recognition (faces are reduced to 2×2 or smaller representations). As shown in Fig. 6(b), the coarse capture is fed directly to a downstream network that recovers task utility without ever recording an identifiable image. ISR [48] pioneered this paradigm by proposing inverse super resolution, generating multiple informative low-resolution training videos from high-resolution sources via optimized sub-pixel transformations. This enables learning robust decision boundaries in low-resolution feature space while ensuring hardware-level privacy, achieving comparative activity recognition accuracy. LRPrivacy [80] advanced this through a two-stream multi-Siamese CNN that learns transformation-invariant embeddings, enabling real-time activity classification on mobile GPUs without recording identifiable high-resolution videos.

3.2.3 Discussion

Sensor-based face De-ID offers strong security-by-design advantages by eliminating the raw identifiable face from the acquisition path, but several challenges persist. First, threat models and evaluation protocols require standardization: claims of irreversibility should be stress-tested against optical inversion, deconvolution/deblurring, super-resolution, face restoration, and ISP-level side channels across diverse sensors and color pipelines. Second, dynamic optics (e.g., DyPP) improve resilience but introduce calibration, latency, and energy costs; quantifying these trade-offs for mobile and embedded platforms is essential. Third, privacy–utility objectives remain task- and population-dependent: preserving low-frequency geometry can leak soft-biometrics, and fairness across demographics under optical distortions is underexplored. Fourth, co-design with downstream ISPs and multi-camera systems is largely unaddressed, as is long-term robustness to hardware drift and manufacturing tolerances. Finally, reproducible cross-device benchmarks coupling privacy, task, and perception metrics are needed for on-sensor De-ID at scale.

3.3 Digital-Domain Face De-ID

Following the cross-domain analysis of Sec. 2.5, digital-domain methods accept a fully identifiable digital signal and suppress identity in software. Their flexibility comes at the cost of accepting raw exposure and, in some regimes, replacing rather than complementing upstream physical or sensor interventions.

3.3.1 Handcrafted Distortion and Filtering

Handcrafted distortion methods are the earliest face De-ID approaches, employing simple image processing operations. Fig. 7(a) shows naive distortion (Gaussian blur, pixelization, masking), while Fig. 7(b) depicts the k-Same framework, marking the shift from ad-hoc distortion to statistically grounded face De-ID.

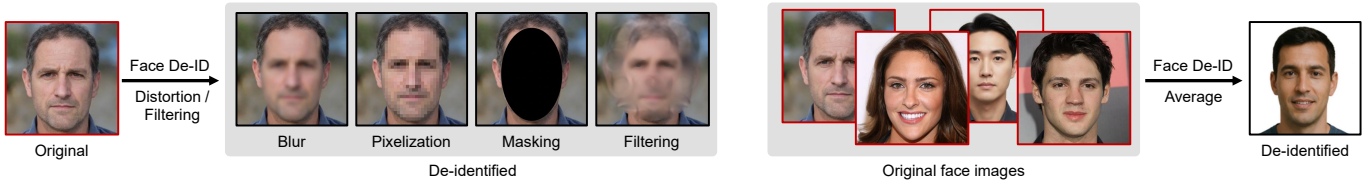
Early work evaluated these operations for their privacy-utility trade-offs. EFVAP [8] and De-ID Filter [82] studied blur and pixelization in workplace video, finding that moderate blur conceals identity while preserving awareness cues such as posture, activity, and people count. Blur outperformed pixelization, which showed abrupt jumps in information disclosure across levels. However, BFPP [83] exposed a limitation in home telecommuting: blur weak enough to preserve awareness cues failed to de-identify, since motion and context still enabled recognition. This tension motivated spatially adaptive techniques that selectively preserve task-relevant information. PUI [84] introduced the Eigen-space filter, projecting live images onto a PCA basis learned from "socially correct" reference images so that reconstruction retains only reference-set information while suppressing non-conforming elements, though this raised ethical concerns over misuse for face animation and masquerading. EmotionPreserve [81] instead preserved expressions via variational adaptive filtering with Total Variation regularization, strongly blurring identity-contributing structures (nose, wrinkles) while sparing expression-related regions (eyes, eyebrows, mouth).

Discussion. Handcrafted distortion established the foundational framework for digital face De-ID but suffers from critical limitations. Uniform operations cannot achieve strong De-ID while preserving utility, owing to the coupling between identity and task-relevant features: privacy-preserving settings eliminate awareness cues, while utility-preserving settings leave enough signal for recognition through motion and context. These methods also lack formal privacy guarantees, relying on user studies or model-specific recognition accuracy rather than rigorous metrics. Their need for empirical, context-dependent tuning shows that handcrafted distortion cannot universally balance privacy and utility, motivating learning-based approaches that adaptively optimize this trade-off.

3.3.2 k-Anonymity-Based Methods

Formal privacy notions migrated into face De-ID via k-anonymity [74], which bounds re-identification risk by ensuring each released record is indistinguishable from at least $k-1$ others. The foundational k-Same algorithm [9] translates this principle to faces by clustering images and replacing each with the cluster average, guaranteeing recognition accuracy $\leq 1/k$. Early refinements addressed utility and quality: k-Same-Select [85] imposed utility constraints to preserve gender and expression during clustering, while k-Same-M [86] leveraged active appearance models for improved alignment and photometric fidelity. These works systematically exposed weaknesses in ad-hoc pixelation, including susceptibility to parrot recognition and resolution enhancement attacks [86].

Subsequent developments expanded beyond closed-set scenarios. The ϵ-map framework [87] formalized three pri-



(a) Handcrafted methods: Blur, pixelization, masking, and filtering. (b) k-Same method: Cluster k faces, substitute with aggregate.

Fig. 7: **Traditional digital-domain face De-ID approaches.** (a) Handcrafted filters, including blur, pixelization, and masking, uniformly reduce the recognizability of the facial region, alongside spatially adaptive filtering (EmotionPreserve [81]) that selectively blurs identity-bearing structures while preserving utility-bearing regions. (b) The k-Same framework aggregates a cluster of k face images to synthesize a mean face that replaces the original, theoretically limiting the recognition probability to $1/k$. Together they illustrate the shift from ad-hoc distortion to statistically grounded face De-ID.

vacy targets: ϵ-map ("like no one"), wrong-map ("like someone else"), and (ϵ, k)-map ("like everyone"), via adaptive pixelation and likelihood equalization under reference models, supporting multiple images per subject with explicit privacy bounds. Multi-Factor DeID [88] disentangled identity from nuisance factors using unified linear/bilinear/quadratic models with semi-supervised fitting. An alternative emerged with k-Same-furthest [89], which averaged over maximally dissimilar clusters rather than similar ones, theoretically driving recognition toward zero while maintaining k-anonymity.

Attribute preservation became central in later methods. GARP-Face [90] blended faces with attribute-matched gallery images to preserve demographics while removing identity. APFD [91] formulated this as joint optimization over AAM parameters, computing optimal fusion weights for k attribute-similar images. Deep learning enabled new approaches: FIP-DeID [92] exploited pose-invariant features to generate averaged faces; k-Same-Net [93] replaced averaging with a generator trained on a disjoint proxy identity set, eliminating pixel-averaging artifacts while preserving the $1/k$ bound; ELEGANT-DeID [94] combined k-Same with generative attribute transfer, achieving complete preservation with minimal gallery requirements.

Discussion. k-Anonymity methods established the first formal privacy framework for face de-identification, providing provable guarantees that recognition accuracy cannot exceed $1/k$. However, this guarantee entails fundamental limitations: clustering constraints reduce scalability and introduce bias with skewed distributions, while pixel averaging produces perceptual artifacts. More critically, k-anonymity addresses only closed-world threat models, offering no protection against adversaries with auxiliary information. The utility–privacy trade-off remains coarse-grained (reducing k to preserve visual detail discontinuously increases re-identification risk) and attribute-preservation extensions [91], [94] require extensive labeled data with limited control over attribute retention. Despite these constraints, k-anonymity methods laid essential conceptual groundwork for subsequent developments.

3.3.3 *Adversarial Perturbation-based Methods*

Early adversarial perturbation approaches generated minimal, imperceptible noise to fool FR systems (Fig. 8, left). EvolutionaryAttack [95] showed that perturbations as small as 10^{-5} MSE enable dodging attacks, while P-FGVM [96] added realism constraints to spatial-domain perturbations, and AdvFaces [97] used discriminators for high perceptual quality under both dodging and impersonation. These pixel-level approaches, though effective, struggled with visual naturalness and cross-system transferability.

To address this, researchers turned to feature-space manipulation in learned representations rather than pixel space. FSAP [98] optimizes latent vectors via bi-loss alternation between identity dissimilarity and attribute preservation; TIP-IM [99] combines relative identification loss with Maximum Mean Discrepancy constraints for open-set naturalness; FE-DeID [100] concentrates perturbations on critical regions (eyes, nose, mouth) via feature embedding; and Adv-Inversion [101] optimizes latent codes through reconstruction-fidelity losses, improving transferability over pixel-level perturbations.

Other methods use natural visual modifications as perturbation carriers for stealthiness. AMT-GAN [102] reconciles adversarial noise with cycle consistency in makeup transfer through joint training, while Makeup De-ID [103] embeds perturbations in eye-makeup regions for more realistic dodging and impersonation. CLIP2Protect [104] searches latent manifolds guided by textual makeup prompts, and DiffProtect [105] produces target-agnostic examples that avoid noise-based artifacts, while Weaken-Diff [11] targets purification effects via learned unconditional embeddings to raise impersonation success.

Semantic-level manipulation has gained attention for its natural appearance and transferability. Adv-Attribute [10] perturbs high-level attributes in disentangled latent spaces through importance-aware selection and multi-objective optimization, excelling against robust black-box models. Adv-CPG [106] integrates customized portrait generation with progressive two-layer identity encryption, enabling multimodal control while achieving strong black-box success.

Black-box transferability remains a central challenge. DFANet [107] uses dropout to diversify examples and prevent surrogate overfitting, establishing the TALFW robustness benchmark; Sibling-Attack [108] adds attribute recognition as an auxiliary task to improve generalization; SMAP [109] jointly optimizes patch texture, position, and shape via gradient-guided location selection; and ADA [110] and DPA [111] respectively survey transferability enhancement across FR architectures and advance unrestricted-attack embedding strategies.

A further group targets specialized scenarios. GMAA [112] produces expression-robust impersonation by training across facial action units; UAXs [113] exposes

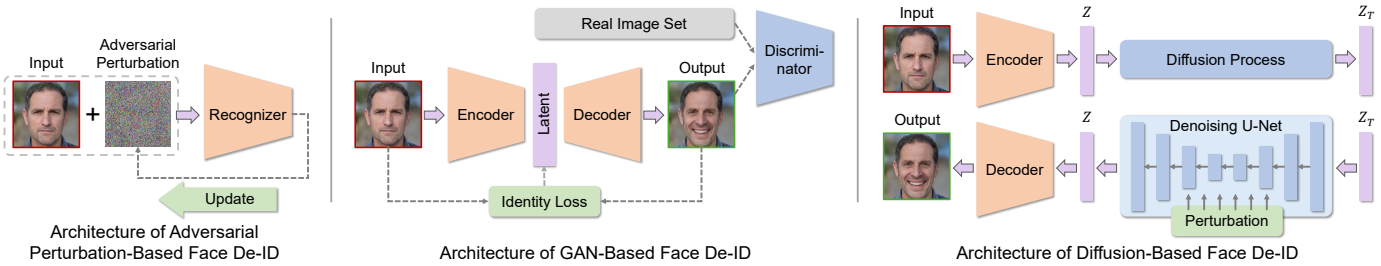


Fig. 8: **Architectural paradigms of contemporary digital-domain face De-ID**. (Left) Adversarial perturbation-based methods inject imperceptible noise at the pixel level to mislead face recognizers. (Middle) GAN-based generative approaches synthesize identity-removed faces through encoder–decoder frameworks guided by discriminators and identity losses. (Right) Diffusion-based methods employ iterative denoising processes to achieve high-fidelity, photorealistic anonymization within structured latent spaces. Together, these architectures illustrate the evolution from explicit perturbation to implicit generative synthesis for controllable, realistic, and utility-preserving face De-ID.

vulnerabilities via identity-agnostic universal perturbations that spoof multiple identities at once; VLA [114] extends attacks to physical settings using projected visible light with alternating perturbation and concealing frames; Chameleon [115] offers instant protection through identity-preserving targeted perturbations; and Veil Privacy [116] generates veiled data that conceals identity from humans while preserving, or enhancing, DNN utility.

Discussion. Adversarial perturbation methods have evolved from pixel-level noise injection [95], [96] through feature-space methods [98], [99] to semantic manipulations [10], [101], a clear trend toward more natural-looking and transferable perturbations. Yet four challenges persist. First, the *transferability-imperceptibility trade-off*: perturbations effective on surrogates often fail on black-box systems, while larger magnitudes compromise naturalness [107], [108]. Second, limited *robustness to preprocessing* (JPEG compression, resizing, filtering), which attenuates adversarial effects before reaching the target. Third, variable *computational efficiency*, with iterative optimization [98], [99] limiting real-time use. Fourth, *evaluation inconsistency* across FR models, datasets, and metrics, which hinders systematic comparison. Future work should prioritize unified protocols assessing not only attack success but also visual quality, transferability, and efficiency.

3.3.4 Generative Model-based Methods

GANs have emerged as a dominant paradigm for face De-ID, enabling the synthesis of photorealistic faces while preserving non-identity attributes (Fig. 8, middle). Early approaches established foundational architectures through style transfer and conditional generation. FATM [117] pioneered facial attribute transfer through encoder-decoder architectures mapping non-identity attributes to donor identities. EPD-Net [118] introduced dual auxiliary networks with identity and emotion verificators to maximize emotion preservation while minimizing identity similarity. LiveDeID [119] extended these principles to video streams through attractor-repeller mechanisms where low-to-mid level features enforce similarity to input frames while high-level representations enforce distance from target identities.

Several methods have addressed privacy-utility trade-offs through specialized architectures and loss formulations. PP-GAN [120] combined Siamese verificators for identity removal with SSIM-based regulators for utility preservation. Generative-DeID [121] extended De-ID to full-body scenarios through segmentation-conditioned synthesis. AutoDeID [122] integrated k-anonymity principles through Anonymous Semantic Masks and identity-adversarial discriminators ensuring distance from original and sensitive identities. DeID-GAN [123] utilized StyleGAN-generated proxy faces as anonymous templates, fusing non-biometric attributes based on emotion and pose similarity.

Attribute-aware and controllable generation has become increasingly prominent. A^{3}GAN [12] formulated face De-ID as joint semantic suppression and controllable attribute injection through suppressive convolutional units and attribute-aware injective networks. FaceSwap [124] combined StyleGAN-based attribute disentanglement with adversarial vector mapping in latent space to generate recognition-resistant images. SF-GAN [125] distinguished between shallow attributes (hairstyle, glasses) and deep attributes (expression, gender) through specialized processing networks, employing uniqueness loss to ensure distinctness. AnonymousNet [126] achieved measurable privacy through four-stage frameworks integrating attribute selection compliant with k-anonymity, l-diversity, t-closeness, and ε-differential privacy. Disguise [7] grounded De-ID in differential privacy and ensemble learning with mixture-of-experts utility networks for attribute disentanglement.

Reversibility represents critical advances for practical deployment. IdentityMask [127] introduced reversible video De-ID with Protection and Recovery Modules using user-specific keys and motion flow guidance for temporal consistency. PIDIM [128] enabled personalized invertible De-ID through password-controlled hyperspherical transformations. UU-Net [129] proposed landmarks-free reversible solutions using sequential encoder-decoder models for public anonymization and private reconstruction. IDeudemon [130] adopted divide-and-conquer strategies obfuscating 3D disentangled identity codes from NeRF models while preserving utility through visual similarity assistance.

Contemporary methods leverage advanced architectures including StyleGAN variants and diffusion models. StyleGAN-based approaches demonstrate particular promise: StyleGAN-DeID [131] achieved effective De-ID through style mixing preserving utility attributes; FALCO [132] operated in StyleGAN2 latent space with margin-based identity obfuscation in ArcFace space and feature-matching attribute preservation in FaRL's ViT space,

eliminating background-based re-identification risks; and CPP-DeID [133] enabled customizable privacy-utility trade-offs through GAN inversion with privacy parameter-controlled identity suppression in W^+ latent space.

Diffusion-based De-ID (Fig. 8, right) has emerged as a powerful alternative to GANs. DiffAM [134] introduced diffusion-based adversarial makeup with superior black-box transferability. Diff-Privacy [135] unified anonymization via Multi-Scale image Inversion with embedding scheduling. Synthetic-DeID [136] reformulated De-ID as training-free identity editing in frozen diffusion models, and Swapping-DeID [13] simplified the pipeline through reconstruction loss within Stable Diffusion, controlling De-ID via a single parameter.

Multi-spectral and 3D-aware approaches extend De-ID beyond RGB still imagery. Thermal-F De-ID [137] combined RGB and thermal imagery with ensemble learning, showing that thermal features prevent photo-based spoofing. MVC-DeID [138] leveraged 3D-aware StyleNeRF for viewpoint-consistent identity disentanglement, while G^{2}Face [33] integrated 3D face models and StyleGAN decoders for password-based reversibility.

Context-specific applications are defined by application-driven utility constraints: DeID-rPPG [139] preserves remote photoplethysmography signals; FaceMotionPreserve [140] maintains facial motion for medical diagnosis; Egocentric-DeID [141] targets wearable-camera footage; Veil Privacy [116] balances human-imperceptible obfuscation with DNN utility; and DRGAN [142] pioneered 2D/3D De-ID preserving expression, gender, and ethnicity through disentanglement-reconstruction.

3.3.5 Discussion

Generative methods represent the most sophisticated paradigm for face De-ID. The field has progressed from early conditional GANs prioritizing photorealism to frameworks incorporating formal privacy guarantees (k-anonymity [122], differential privacy [126]), reversible mechanisms [127], [128], and latent space manipulation [132], [133]. Despite superior visual quality, critical challenges remain: (1) the privacy-utility trade-off constitutes a fundamental limitation: perfect identity removal conflicts with perfect attribute preservation given representation entanglement; (2) evaluation fragmentation persists across disparate datasets and metrics, precluding meaningful comparison; (3) computational costs prohibit real-time deployment, particularly for diffusion approaches [13]; and (4) temporal consistency for video receives insufficient attention, with most methods designed for images. Recent diffusion models [135], [136] and geometric priors [33], [138] suggest promising directions, yet the field lacks frameworks characterizing achievable privacy-utility frontiers. Future priorities include: (a) standardized evaluation protocols; (b) information-theoretic analysis of privacy-utility trade-offs; and (c) video extensions with temporal consistency guarantees via 3D-aware or NeRF approaches [130], [138].

4 Evaluation Protocols

Face De-ID lacks the standardized benchmarks and metrics characteristic of mature vision tasks. Instead, evaluation practice has evolved heterogeneously, shaped by diverse threat models and application contexts across the three domains. TABLE 3 summarizes the evaluation protocols of the 112 surveyed methods, revealing both remarkable breadth and concerning fragmentation. We first review the underlying datasets, then analyze the metrics used to assess privacy, utility, and visual quality.

TABLE 2: **Overview of primary face De-ID evaluation datasets**, grouped by modality. Env.: C = Controlled, S = Semi-controlled, W = In-the-wild.

Dataset	#Img/Vid	#IDs	Resolution	Env.	Annotations
RGB Images					
LFW [143]	13,233	5,749	250×250	W	identity
CelebA [144]	202,599	10,177	178×218	W	identity, 40 attr., 5 lmk
CelebA-HQ [58]	30,000	—	1024×1024	W	40 attr., 5 lmk
FFHQ [59]	70,000	—	1024×1024	W	—
VGGFace2 [145]	3.31M	9,131	varies	W	identity, pose, age
CASIA-WebFace [146]	494,414	10,575	varies	W	identity
MS-Celeb-1M [147]	10M	100,000	varies	W	identity
MegaFace [148]	1M	690,572	varies	W	identity
PubFig [149]	58,797	200	varies	W	73 attr.
AgeDB [150]	16,488	568	varies	W	identity, age, gender
CFP [151]	7,000	500	varies	W	frontal/profile
FERET [152]	14,126	1,199	256×384	C	pose, expre.
Multi-PIE [153]	750,000	337	varies	C	15 poses, 19 illu., 6 expre.
MORPH [154]	55,134	13,000	varies	S	age, gender, race
RaFD [155]	8,040	67	681×1024	C	8 expre., gaze, 5 poses
CK+ [156]	593 seq.	123	640×490	C	expre., AU
AffectNet [157]	450,000	—	varies	W	valence-arousal, 8 expre.
LADN [158]	635	—	varies	W	makeup style
RGB Video					
YTF [159]	3,425 vid.	1,595	varies	W	identity
VoxCeleb [160]	150K vid.	6,112	varies	W	speaker, audio
VidTIMIT [161]	430 vid.	43	512×384	C	speaker, audio
Ego4D [162]	3,670 hrs	—	varies	W	narration, activity
3D Face					
Bosphorus [163]	4,666 scans	105	3D mesh	C	35 expre., AU, occlusion
BU-3DFE [164]	2,500 scans	100	3D mesh	C	6 expre. × 4 levels
Eurecom Kinect [165]	936	52	RGB-D	C	pose, expre., occlusion
SIAT-3DFE [166]	8,000 scans	500	3D mesh	C	expre.
Physiological (rPPG / Affective)					
PURE [167]	60 vid.	10	640×480	C	heart rate ground truth
OBF [168]	200 vid.	100	varies	C	heart rate, atrial fibrillation
AVEC [169], [170]	—	84+	varies	S	depression score

4.1 Evaluation Datasets

TABLE 2 summarizes the principal characteristics of the most widely adopted face De-ID datasets, grouped by modality and including capture environment, scale, resolution, and annotation richness. Across the 112 surveyed methods, a clear concentration emerges: LFW [143], CelebA / CelebA-HQ [58], [144], VGGFace2 [145], and FFHQ [59] together account for the majority of evaluations. Task-specific benchmarks address narrower utility goals, *e.g.*, RaFD [155] and AffectNet [157] for expression, Bosphorus [163] for 3D-aware methods, and PURE [167] for rPPG-preserving De-ID, while a smaller subset evaluates privacy–utility trade-offs in activity recognition (HMDB [171], DogCentric [172]).

Beyond adoption statistics, these datasets exhibit structural properties that determine their suitability. *(i) Web-scraped, celebrity-dominated composition*: the dominant benchmarks (LFW, CelebA(-HQ), VGGFace2, MS-Celeb-1M, MegaFace) contain adult, quasi-frontal, well-lit faces with demographic skew, so results on them overstate robustness for children, older adults, and under-represented groups. *(ii) Identity as the only universal annotation*: utility labels (expression, AUs, gaze, physiological signals) exist only in separate, smaller, often lab-controlled corpora; this asymmetry is a root cause of the fragmentation in Fig. 9, since

TABLE 3: **Overview of the face De-ID landscape highlighting the heterogeneity in evaluation protocols**. Methods are organized by domain (Physical, Sensor, Digital) and publication year, summarizing the specific datasets, performance metrics (Privacy[a], Utility[b], and Quality[c]), and key innovations.

Method	Year	Dataset	Privacy[a]	Utility[b]	Quality[c]	Key Innovation
Physical-Domain Face De-ID: Real-world Adversarial Modifications (Sec. 3.1)						
AdvEyeglass [16]	2016	PubFig [149]	Success Rate (↑): 31.0%	✗	✗	Eyeglass-frame perturbation
AGNs [54]	2019	PubFig [149]	Success Rate (↑): 70.0%	✗	✗	GAN-based natural eyeglasses
OAP [55]	2019	CASIA [146]	Similarity (↑)	✗	✗	Grayscale patches for ArcFace
VLA [114]	2019	CusFace, LFW [143]	Success Rate (↑): 85.6%	✗	Similarity, Distance, SelfDis	Visible light-based physical attack
ALPA [50]	2020	—	Success Rate (↑): 92.0%	✗	✗	Adversarial light projection
AdvHat [49]	2021	CASIA [146]	Similarity (↑)	✗	✗	Hat-based surface attack
AdvMakeup [17]	2021	LFW [143], LADN [158]	Success Rate (↑): 63.74%	✗	✗	Physical eye-shadow makeup attack
FaceAdv [64]	2021	LFW [143], VolFace [64]	Success Rate (↑): 100%	✗	✗	Multi-shape sticker GAN
GenAP [57]	2021	LFW [143], CelebA-HQ [58]	Success Rate (↑): ≤99%	✗	✗	GAN manifold regularization
AdvSticker [52]	2022	LFW [143], CelebA [144]	Fooling Rate (↑)	✗	✗	Attack using real-life stickers
Physical AX [56]	2022	VGGFace2 [145]	Success Rate (↑): ≤83%	✗	✗	Threshold-based smoothness loss
AdvMask [51]	2022	CASIA [146], CelebA [144], MS-Celeb-1M [147]	Recognition Rate (↓)	✗	✗	Universal fabric mask patterns
PPAttack [63]	2023	LFW [143], CelebA [144]	Fooling Rate (↑)	✗	✗	Simultaneously optimization
AT3D [32]	2023	LFW [143], CelebA-HQ [58]	Success Rate (↑): ≤100%	✗	✗	Low-dim 3-DMM coefficient
Optical De-ID [53]	2023	Bosphorus [163], Eurecom [165], SIAT-3DFE [166]	Success Rate (↑): ≤99%	✗	RMSE	Adversarial-illumination 3D attack
PadvFace [61]	2023	LFW [143]	Success Rate (↑): 31%	✗	✗	Curriculum-optimized sticker
SASMask [67]	2024	LFW [143], VGGFace2 [145], AgeDB [150], CFP [151]	Success Rate (↑): 43.4%	✗	SSIM	Adversarial style optimization
Agile [72]	2024	LFW [143], YTF [159]	Success Rate (↑): 80%	✗	✗	Invisible infrared laser perturbations
EAP [62]	2024	LFW [143], CelebA-HQ [58]	Success Rate (↑): 91.2%	✗	✗	Meta-ensemble gradient extraction
AdvBandage [68]	2024	LFW [143]	Accuracy (↓)	✗	✗	Physical attack using bandages
ARA [73]	2024	VGGFace2 [145], CelebA [144]	Success Rate (↑): ≤99%	✗	BRISQUE, NIQE	Physical natural adversarial lighting
Face3DAdv [71]	2025	LFW [143], CelebA-HQ [58]	Success Rate (↑): ≤99%	✗	✗	Robust 3D adversarial patches
ProjAttacker [18]	2025	LFW [143], CelebA-HQ [58]	Success Rate (↑): ≤98%	✗	✗	Configurable physical attack
Sensor-Domain Face De-ID: Acquisition-Level Privacy Protection (Sec. 3.2)						
DefocusOptics [14]	2015	—	k-anonymity	Depth, Tracking	✗	Optical PSF-based k-anonymity
DefocusOptics+ [75]	2016	FERET [152]	k-anonymity	Depth, Counting, Detection	✗	Pre-capture privacy-preserving optics
ISR [48]	2017	HMDB [171], DogCentric [172], JPL-Interaction [173]	Subjective	Action	✗	Extreme low resolution
LRPrivacy [80]	2018	HMDB [171], DogCentric [172]	Subjective	Action	✗	Multi-view embedding
PrivHPE [76]	2021	LFW [143], MS-Celeb-1M [147], AgeDB [150], CFP [151]	AUC (↓)	Pose	PSNR, SSIM	End-to-end optimize optical encoder
PrivHAR [77]	2022	HMDB [171], VISPR [174], PA-HMDB [175]	C-MAP (↓), AUC (↓)	Action	✗	Adversarial optical encoding
PrivPDE [78]	2022	VGGFace2 [145], NYUv2 [176]	AUC (↓)	Depth, Action	✗	Optimize phase mask
OpticalDR [44]	2024	CelebA [144], CK+ [156], AVEC [169], [170]	AUC (↓)	Emotion, Depression	✗	Depression signal preservation
DyPP [79]	2024	PubFig [149], LFW [143], AgeDB [150]	Accuracy (↓), AUROC (↓)	Counting, Pose, Detection	PSNR, SSIM	Time-varying optical privacy
PrivacyOptics [15]	2024	CelebA-HQ [58], FFHQ [59]	L_2 Distance (↑)	Attribute	FID	De-ID via learned optical encoder
Digital-Domain Face De-ID: Post-Capture Processing Methods (Sec. 3.3)						
Handcrafted Distortion & Filtering (Sec. 3.3.1)						
EFVAP [8]	1996	–	Subjective	✗	✗	Gaussian blur and pixelation
De-ID Filter [82]	2000	–	Protection Rate (↑)	✗	✗	Evaluation of filters
PUI [84]	2000	–	Subjective	✗	✗	Eigen-space filtering
BFPP [83]	2006	–	Awareness Rate (↓)	✗	✗	Home telepresence study
EmotionPreserve [81]	2015	LFW [143]	Verification Rate (↓)	Expression, Gaze	✗	Variational adaptive filtering
k-Anonymity & Statistical Methods (Sec. 3.3.2)						
k-Same [9]	2005	FERET [152]	k-anonymity	✗	✗	First formal framework
k-Same-Select [85]	2005	FERET [152]	Recognition Rate (↓)	Gender, Expression	✗	Incorporate utility constraints
k-Same-M [86]	2006	Multi-PIE [153], FERET [152]	Recognition Rate (↓)	Expression	✗	Model-based face De-ID
ϵ-map DeID [87]	2007	–	Rank-1 Accuracy (↓)	Gender, Expression	✗	k-anonymity guarantees
Multi-Factor DeID [88]	2008	IMM Face Dataset [177]	Recognition Rate (↓)	Expression	✗	Multi-factor separation
k-Same-furthest [89]	2014	IMM Face Dataset [177]	Recognition Rate (↓)	Gender, Age, Expression	✗	Cluster furthest away
GARP-Face [90]	2014	MORPH Dataset [154]	Recognition Accuracy (↓)	Race, Gender, Age	✗	Structured utility hierarchy
DiffPose-DeID [178]	2014	IMM Face Dataset [177]	Recognition Rate (↓)	Pose, Emotion	✗	De-ID across different poses
Photorealistic-DeID [179]	2014	Multi-PIE [153], MUCT [180], PUT [181]	ROC Analysis	Gender, Expression, Pose	PSNR	Component-based face synthesis
APFD [91]	2015	FaceTracer [182], FaceScrub [183], ND1 [184], FERET [152], CAS-PEAL [185], BioID [186]	Recognition Rate (↓)	Attribute	PSNR	Optimal weighted fusion
FIP-DeID [92]	2015	Multi-PIE [153]	Recognition Rate (↓)	Gender, Race, Age, Expression	✗	Deep learning-based method
k-Same-Net [93]	2018	RaFD [155], XM2VTS [187], CK+ [156]	Recognition Rate (↓)	Expression	✗	Generative realization of k-anonymity
ELEGANT-DeID [94]	2019	CelebA [144]	k-anonymity	Age, Gender, Expression	✗	ELEGANT-based attribute transfer
Adversarial Perturbation Methods (Sec. 3.3.3)						
EvolutionaryAttack [95]	2019	LFW [143], MegaFace [148]	MSE (↑)	✗	✗	Decision-based attack on recognition
P-FGVM [96]	2019	CelebA [144]	Misclassification Rate (↑)	✗	MSSIM	Realism-penalized gradient
Makeup De-ID [103]	2019	Custom Dataset	Success Rate (↑)	✗	✗	Makeup-based adversarial attack
AdvFaces [97]	2020	CASIA [146], LFW [143]	Success Rate (↑): 99.67%	✗	SSIM	GAN-based face synthesis
DFANet [107]	2020	LFW [143], MS-Celeb-1M [147], VGGFace2 [145], CASIA [146], IMDb-Face [188]	Success Rate (↑)	✗	SSIM	Dropout-based method
UAXs [113]	2021	LFW [143], VGGFace2 [145]	FMR (↑)	✗	✗	Universal multi-identity spoofing
TIP-IM [99]	2021	LFW [143], MS-Celeb-1M [147], MegaFace [148]	Rank-N-T (↓)	✗	PSNR, SSIM, MMD	Principled distribution matching
Adv-Attribute [10]	2022	FFHQ [59], CelebA-HQ [58]	Success Rate (↑)	✗	FID, MSE	Compositional attribute editing
AMT-GAN [102]	2022	MT [189], CelebA-HQ [58], LADN [158]	Success Rate (↑): 52.8%	✗	FID, PSNR, SSIM	Makeup transfer (cycle-consistent)
FSAP [98]	2023	FFHQ [59], CelebA [144]	SPR (↑)	Detection	MSE	StyleGAN W-space perturbation
CLIP2Protect [104]	2023	CelebA-HQ [58], LADN [158], LFW [143]	Success Rate (↑)	✗	FID, PSNR, SSIM	Text-guided face De-ID
DiffProtect [105]	2023	CelebA-HQ [58], FFHQ [59]	Success Rate (↑)	✗	FID	Semantic code optimization
Sibling-Attack [108]	2023	CelebA-HQ [58], LFW [143]	Success Rate (↑)	✗	SSIM, MSE	Multi-task transferable attack
SMAP [109]	2023	CelebA-HQ [58], LFW [143], VGGFace2 [145]	Success Rate (↑)	✗	FID, LPIPS, SSIM	Spatial mutable adversarial patch
GMAA [112]	2023	CelebA-HQ [58], LFW [143]	Success Rate (↑)	✗	✗	Generalized manifold (AU-aware)
Chameleon [115]	2024	FaceScrub [183], LFW [143]	Success Rate (↑)	Pose, Expression, Age, Gender	FID, LPIPS, SSIM, User Study	Universal privacy protection mask
ADA [110]	2024	CelebA-HQ [58], LFW [143]	Success Rate (↑)	✗	SSIM, PSNR, FID	Injection and denoising
FE-DeID [100]	2024	LFW [143], CelebA [144], GFSG2 [60]	Success Rate (↑)	✗	PSNR, SSIM, LPIPS	Feature embedding via CNN
WeakenDiff [11]	2025	CelebA-HQ [58], LADN [158]	Success Rate (↑): 79.2%	✗	FID, PSNR, SSIM	Counter-purification targeting
DPA [111]	2025	CelebA-HQ [58], LFW [143]	Success Rate (↑): ≤98.2%	✗	✗	Diverse parameters augmentation
Adv-Inversion [101]	2025	FFHQ [59], CelebA-HQ [58]	Success Rate (↑), Rank-N-T (↓)	✗	FID, SSIM, PSNR	GAN inversion framework
Adv-CPG [106]	2025	FGID [190], FFHQ [59], CelebA-HQ [58], LFW [143]	Success Rate (↑)	✗	FID, SSIM, PSNR	Double-layer encryption
Generative Model-Based Methods (Sec. 3.3.4)						
Generative-DeID [121]	2017	CCP [191], Human3.6M [192]	Accuracy (↓)	Garment Shape, Silhouette	Naturalness Score	GAN-based full-body De-ID
PP-GAN [120]	2019	MORPH Dataset [154]	De-ID Rate (↑)	Detection, Age	SSIM	Verificator+regulator paradigm

Continued on next page

FATM [117]	2019	LFW [143], PIPA [193], VidTIMIT [161]	Accuracy Drop (↑)	Expression	SSIM	Encoder-decoder architecture
DeepPrivacy [194]	2019	FDF, WIDER FACE [195]	Subjective	Detection	FID	Conditional GAN with U-Net
AnonymousNet [126]	2019	CelebA [144]	k-anonymity	Attribute	PSNR, SSIM, MS-SSIM	StarGAN + formal privacy
Live-DeID [119]	2019	LFW [143], CelebA [144], PubFig [149], CelebA-HQ [58]	Accuracy (↓)	Attribute	SSIM	First video De-ID
CIAGAN [45]	2020	CelebA [144], MOTS [196], LFW [143]	Recall (↑)	Detection	FID	Conditional GAN-Based De-ID
EPD-Net [118]	2020	RaFD [155]	De-ID Rate (↑)	Emotion	SSIM	pix2pix-based face De-ID
Thermal-F De-ID [137]	2020	Custom dataset	Accuracy (↓)	Detection	✗	Novel feature extraction
FaceSwap [124]	2021	FFHQ [59]	ID Similarity (↓)	Detection, Expression	FID, LPIPS	Attribute disentanglement
PIDIM [128]	2021	CelebA-HQ [58], FFHQ [59], CASIA [146]	De-ID Rate (↑)	Detection	✗	Password-based reversibility
DeID-GAN [123]	2021	RaFD [155], XM2VTS [187], CelebA [144], FFHQ [59]	DMOS (↑)	Emotion, Attribute	FMOS	Three-stage pipeline with StyleGAN
AutoDeID [122]	2021	LFW [143], VGGFace2 [145], CelebA-HQ [58]	Re-ID Recall (↑)	Detection, Attribute	FID	Combining k-anonymity with GAN
SF-GAN [125]	2021	CelebA-HQ [58], FFHQ [59]	SSIM (↓)	Expression, Gender, Hairstyle, Glasses	✗	Multi-attribute preservation
A^{3}GAN [12]	2022	CelebA [144], WIDER FACE [195], ExpW [197], NTHU-DDD [198]	Accuracy (↓)	Detection, Expression	FID	Attribute editing-based De-ID
IdentityMask [127]	2022	VoxCeleb [160]	CSIM (↓)	Pose, Expression, Landmark	FID	Reversible face video De-ID
UU-Net [129]	2022	YTF [159], P-DESTRE [199], BIODI, MARS [200]	AUC (↓)	Pose, Lighting, Background, Expression	✗	ROI steganography (surveillance)
MVC-DeID [138]	2023	–	ID Distance (↑)	Detection	PSNR, SSIM, LPIPS	3D-aware generators
CPP-DeID [133]	2023	CelebA-HQ [58], RaFD [155], XM2VTS [187], AffectNet [157]	AUC (↓), De-ID score (↑)	Gender, Expression	FID, LPIPS, SSIM, MSE	Controllable privacy protection
DartBlur [201]	2023	WIDER FACE [195], FDDB [202], CrowdHuman [203]	Subjective	Face Detection	PSNR, SSIM	Learnable U-Net-based blur model
DeID rPPG [139]	2023	PURE [167], OBF [168]	Accuracy (↓)	rPPG signal	PSNR, SSIM	De-ID preserving rPPG signals
IDeudemon [130]	2023	CelebA-HQ [58], FFHQ [59]	L_2 Distance (↑)	Pose, Expression	PSNR, SSIM, FID	Parametric NeRF De-ID
FALCO [132]	2023	CelebA-HQ [58], LFW [143]	Re-ID Rate (↓)	Detection	FID	Latent code optimization
Egocentric DeID [141]	2023	Ego4D [162]	Similarity (↓), Accuracy (↓)	✗	FID, LPIPS, SSIM	Consistent identity replacement
StyleGAN-DeID [131]	2023	CelebA [144], LFW [143]	Success Rate (↑)	Detection, Gender, Pose, Expression, Age	✗	StyleGAN's style-mixing
Verifiable DeID [204]	2024	LFW [143], FFHQ [59]	Selection Rate (↓)	✗	FID, SSIM, BRISQUE	Verifiable proof framework
Disguise [7]	2024	VGGFace2 [145], LFW [143], CelebA-HQ [58]	TPR (↓)	Landmark, Gaze, Emotion	SER-FIQ	DP-inspired pseudo-identities
RBGAN [205]	2024	CelebA-HQ [58], FFHQ [59], CASIA [146], LFW [143]	CSIM (↓): 23.0%	Face Detection, Attribute	SSIM, PSNR, FID	Symmetric-consistency guidance
DiffAM [134]	2024	CelebA-HQ [58], LADN [158]	Success Rate (↑)	✗	FID, PSNR, SSIM	Diffusion makeup adversarial
FaceCaricature [206]	2024	CelebA-HQ [58]	Recognition Rate (↓)	Attribute	LPIPS, SSIM, FID	Face De-ID using face caricatures
Veil Privacy [116]	2024	BioID [186], ORL [207], LFW [143], CelebA [144]	SSIM (↓), User Study	Classification	SSIM	Random pixel flipping
FaceMotionPreserve [140]	2024	PD Dataset	Similarity (↓)	Medical Info	✗	GAN-based face swapping
Diff-Privacy [135]	2024	CelebA-HQ [58], LFW [143]	Re-ID Rate (↓), mAP (↓), Accuracy (↓)	Pose, Expression, Age, Gender	FID, KID, SSIM, LPIPS	Unified diffusion framework
G^{2}Face [33]	2024	CelebA-HQ [58], LFW [143], FFHQ [59]	CSIM (↓), TAR (↓)	Detection, 3D Shape, Expression, Pose	FID	Reversible face De-ID
DRGAN [142]	2025	BU-3DFE [164], BU-4DFE [208], Bosphorus [163]	ACC (↓), EER (↑)	Biometric, Expression, Gender, Ethnicity	✗	2D/3D face De-ID
Face-DeID-Net [209]	2025	CelebA-HQ [58], MyStyle [210]	Re-ID Rate (↓)	Attribute	IQA	Identity extraction and removal
Synthetic DeID [136]	2025	CelebA-HQ [58], FFHQ [59]	SID (↑), TID (↓)	Landmark, Emotion	LPIPS	Training-free identity editing
VCA-DeID [211]	2025	CelebAMask-HQ [212], FFHQ [59], LFW [143]	IS (↓), VR (↓), EER (↓)	Attribute, Gender, Race	FID, FIQ	Image-text aligned attributes
Swapping De-ID [13]	2025	CelebA-HQ [58], FFHQ [59]	Re-ID Rate (↓), Face Distance (↑)	Pose, Gaze, Expression	IQA	Face swapping for De-ID

privacy scores are comparable across papers while utility must be evaluated on disjoint datasets. Resolution and labeling further partition roles: LFW supports recognition-based privacy evaluation but not generative-quality assessment. CelebA(-HQ) and FFHQ provide the resolution and attribute labels needed by generative methods, though FFHQ lacks identity labels and thus cannot support formal privacy measurement. FERET and Multi-PIE remain the testbeds for factor-isolating k-anonymity analyses. *(iii) Static RGB stills as the default modality*: video, 3D, and physiological resources (YTF, Ego4D; Bosphorus, BU-3DFE; PURE, OBF) are scarce and small-scale, explaining why temporal consistency and micro-level utility signals remain under-evaluated.

In total, 66 distinct datasets appear across the method-level protocols catalogued in TABLE 3, of which TABLE 2 distills the 29 primary, recurrently adopted ones. Despite this apparent abundance, three critical gaps remain. First, **evaluation fragmentation across domains** (TABLE 3, Fig. 9): physical-domain methods (23 techniques) report privacy exclusively, sensor-domain methods (10 techniques) universally report both privacy and utility, and digital-domain methods are the most heterogeneous. Second, **lack of multi-modal semantic annotations**: as TABLE 2 shows, no single dataset jointly provides identity, fine-grained expression, demographics, and physiological labels, forcing researchers to evaluate utility preservation across heterogeneous datasets (*e.g.*, expression on RaFD, attributes on CelebA), which impedes fair benchmarking. Third, **paucity of child-centric resources**: although datasets such as FG-NET [213] (1,002 images, 82 subjects), ITWCC [214] (1,705 images, 304 subjects), and CLF [215] exist, they lack the scale and diversity required for training and analyzing, leaving vulnerable populations underserved.

4.2 Evaluation Metrics

Face De-ID evaluation balances three competing objectives: privacy protection, utility preservation, and visual quality. Fig. 9 summarizes the evaluation landscape: the left reports how often each aspect is measured per domain, and the right ranks the 69 distinct metrics extracted from existing papers by adoption. Two forms of fragmentation are evident: the set of reported aspects differs across domains, and the choice of metric within each aspect is largely method-specific. The metric families below are organized along the three axes; per-metric adoption is deferred to Fig. 9.

4.2.1 *Privacy Protection Evaluation Metrics*

Identity Verification and Recognition Metrics. The most direct privacy assessment employs FR systems as adversaries. True Acceptance Rate (TAR) at fixed False Acceptance Rate (FAR) thresholds (commonly FAR=0.1% or 0.01%) measures genuine pair acceptance, with lower TAR indicating stronger De-ID. Equal Error Rate (EER), where FAR equals False Rejection Rate, provides a threshold-independent metric pushed toward 50% (random chance) by effective De-ID. Rank-N accuracy assesses whether the true identity appears in the top-N retrievals. Attack Success Rate (ASR), or fooling rate, quantifies adversarial effectiveness: for dodging it measures the proportion of pairs failing verification, and for impersonation the rate of successful targeted misclassification. As Fig. 9 shows, ASR is the single most widely reported metric (∼36%) and dominates the physical domain.

Embedding Space Metrics. Latent-space measurements provide fine-grained privacy quantification beyond binary

verification. Cosine Similarity (CSIM) and L_2 Distance between original and De-ID embeddings measure feature-space proximity, with lower similarity and larger distance indicating stronger privacy. Source/Target Identity Distance (SID/TID) extends these to targeted scenarios, with SID maximized against the original identity and TID minimized toward the target. Maximum Mean Discrepancy (MMD) quantifies distributional divergence between embedding sets, enabling dataset-level assessment valuable for evaluating transferability.

Formal Privacy Guarantees. k-anonymity-based methods enforce group indistinguishability, upper-bounding recognition accuracy at $1/k$; De-ID rate measures the proportion of faces meeting this bound or failing verification at a given threshold. Near-perfect De-ID (TPR@FPR) reports the re-identification rate at extremely low False Positive Rates (e.g., FPR=0.02%), relevant for high-security scenarios where even rare breaches are unacceptable.

Human Perceptual Metrics. Algorithmic metrics may diverge from human recognition. Human recognition rate through user studies provides ground-truth privacy assessment, and Mean Opinion Scores (MOS) capture subjective De-ID and naturalness. Such evaluations are critical for social acceptability (used by ∼15% of surveyed methods).

4.2.2 Utility Preservation Evaluation Metrics

Utility metrics assess the retention of task-relevant, non-identity signals, and are coupled to the target application.

Facial Attribute Preservation. Attribute classifiers measure retention of soft biometrics, with classification accuracy for gender, age, and race quantifying utility degradation.Landmark localization error (normalized mean error over 68 fiducial points) assesses geometric structure preservation, critical for gaze estimation and medical diagnostics. Detection metrics (mAP, recall) evaluate whether De-ID faces remain detectable, essential for surveillance utility.

Expression and Affect Recognition. Emotion-preserving methods report accuracy on expression datasets (e.g., CK+ [156], AVEC [169], [170]), with strong methods retaining $>85\%$ of original performance. Action Unit (AU) detection [216] provides fine-grained assessment of facial movement preservation.

Pose and Gaze Estimation. Head pose error (mean angular error in yaw/pitch/roll) and gaze estimation error quantify geometric utility, the latter critical for attention analysis and driver monitoring. Depth estimation metrics (RMSE, absolute relative error) assess 3D utility, particularly for the sensor-domain methods in TABLE 3 that prioritize depth.

Remote Physiological Signal Preservation. Remote photoplethysmography (rPPG) enables contactless extraction of cardiovascular signals from facial videos. Evaluation metrics include pulse-waveform SNR, heart-rate MAE, and Pearson correlation of inter-beat interval sequences, with PURE [167] and OBF [168] providing synchronized ground truth. DeID rPPG [139] showed that conventional face De-ID severely degrades rPPG quality, motivating approaches that explicitly preserve frequency-domain characteristics.

4.2.3 Visual Quality Metrics

Visual quality metrics assess photorealism, naturalness, and artifact presence of the de-identified outputs.

Pixel-Level Distortion Metrics. Peak Signal-to-Noise Ratio (PSNR) measures pixel-wise fidelity but correlates poorly with perceptual quality for synthesized content. Structural Similarity Index (SSIM) [217] captures structural preservation (0-1 scale), and Multi-Scale SSIM extends it across spatial scales. SSIM is the most widely reported quality metric (∼30%, Fig. 9), yet aligns poorly with human perception for GAN-synthesized or heavily perturbed images.

Perceptual / Distribution Metrics. Fréchet Inception Distance (FID) [218] measures distributional divergence in Inception-v3 [219] feature space; Kernel Inception Distance (KID) is an unbiased alternative with better small-sample properties [135]; and LPIPS [220] correlates strongly with human judgment. FID (∼25%) is the third most reported metric overall; generative methods adopt FID and LPIPS almost universally, whereas adversarial methods rarely use them due to minimal pixel changes.

Image Quality Assessment. No-reference metrics assess naturalness without a reference image. BRISQUE and NIQE detect artifact-indicative statistical anomalies [73], [204], with lower scores indicating higher naturalness, and Face Image Quality (FIQ) scores (e.g., SER-FIQ) evaluate recognizability for face recognition [7], providing coupled privacy-utility insight.

Human Perceptual Studies. User studies provide ground-truth quality assessment where algorithmic proxies fail, via MOS for photorealism, the De-ID-specific FMOS and DMOS, and pairwise preference tests that reduce absolute-rating bias. As Fig. 9 confirms, such studies sit in the long tail of rarely adopted metrics despite their deployment importance, limited by cost and scalability.

4.3 Performance Analysis

Despite the evaluation fragmentation, cross-domain analysis of TABLE 3 reveals discernible performance patterns.

Physical-domain methods achieve remarkably high attack success rates: 3D adversarial meshes (AT3D [32]) and projection-based approaches (ProjAttacker [18]) reach near-perfect rates (98–100%) against FR models, while eyeglass-based methods (AdvEyeglass [16], AGNs [54]) trail at 31–70% due to limited perturbation area. However, all 23 physical-domain methods report privacy metrics exclusively, reflecting their attack-centric origins.

Sensor-domain methods exhibit the most balanced privacy-utility trade-offs, with all 10 techniques reporting both privacy and utility. PrivHAR [77] and PrivPDE [78] reduce FR to near-random levels (AUC approaching 0.5) while keeping depth estimation within acceptable bounds, OpticalDR [44] preserves depression-related diagnostic signals, and DyPP [79] adds robustness to PSF inversion. This balance reflects the inherent advantage of optical encoding, which attenuates identity-discriminative frequencies while preserving task-relevant information.

Digital-domain methods exhibit the greatest heterogeneity across the ∼80 techniques catalogued. Handcrafted distortion and k-anonymity methods (k-Same [9], k-Same-Select [85]) provide formal guarantees but limited utility and quality. Adversarial perturbation methods achieve high success rates but transfer poorly to black-box systems. Generative approaches deliver the best holistic performance: GAN-based (FALCO [132], CPP-DeID [133], Disguise [7]) and diffusion-based methods (Diff-Privacy [135],

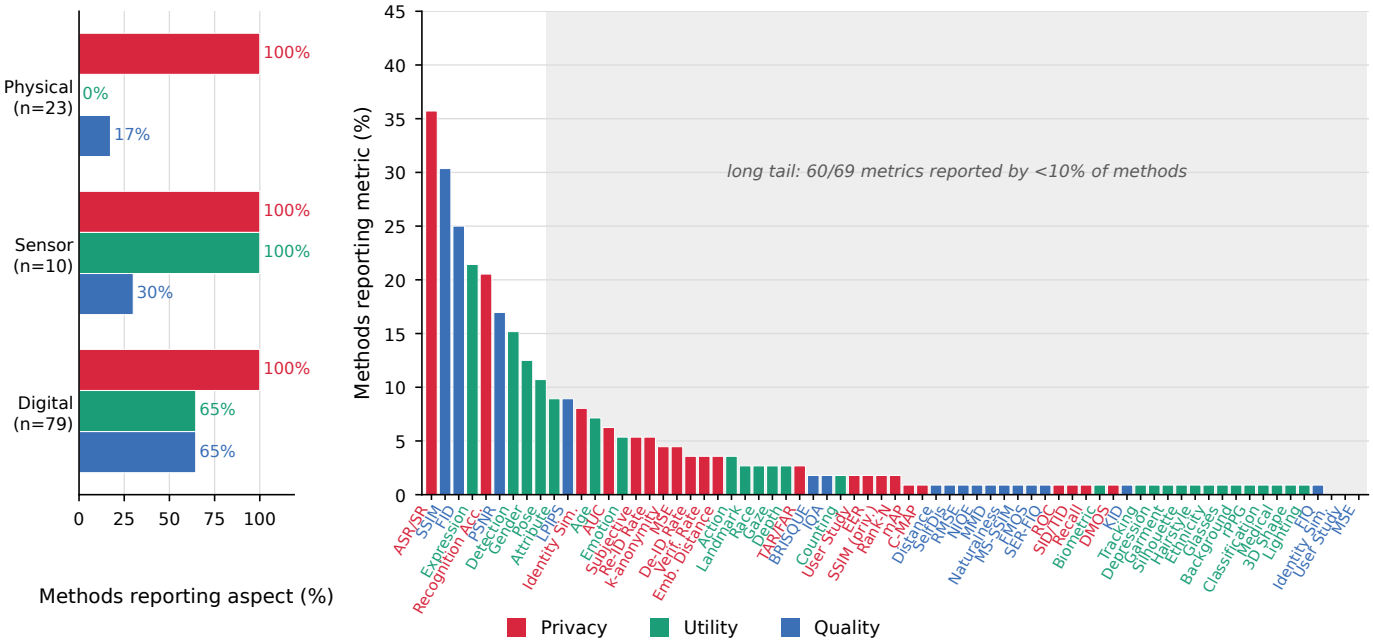


Fig. 9: **Evaluation-protocol fragmentation across the surveyed face De-ID methods.** *Left*: per-domain reporting rates of the three evaluation aspects. Physical-domain methods (n=23) report privacy universally but utility never and quality rarely; sensor-domain methods report privacy and utility universally and quality in 30% of cases; digital-domain methods report privacy universally but utility and quality in only 65% of cases each. *Right*: adoption of the 69 distinct metrics used by existing papers, ranked by reporting share and colored by aspect. Only ASR (∼36%), SSIM (∼30%), and FID (∼25%) exceed a quarter of methods, while 60 metrics (grey) fall below 10%.

Synthetic DeID [136]) retain attribute preservation above 90% while achieving effective De-ID and competitive quality, and hybrid adversarial-generative methods (DiffProtect [105], WeakenDiff [11]) point toward unified privacy-utility-quality optimization.

In summary, physical-domain methods achieve the strongest privacy but neglect utility; sensor-domain methods offer principled privacy-utility optimization through differentiable optical design; and digital-domain methods provide maximum flexibility, with generative approaches delivering the best holistic performance. The field shows a clear progression from ad-hoc distortion toward learned representations that jointly optimize competing objectives, yet the absence of standardized benchmarks remains a critical barrier to definitive cross-method comparison.

5 Future Directions and Open Challenges

While substantial progress has been achieved across the physical, sensor, and digital domains, critical challenges and emerging opportunities warrant focused investigation.

Physical Domain: Bridging the Utility Gap. Sec. 4 shows that physical-domain methods universally report privacy metrics without evaluating **utility preservation**, reflecting adversarial origins where recognition evasion supersedes downstream task enablement. Yet real-world deployment (*e.g.*, healthcare monitoring, emotion-aware interaction) demands physical De-ID that preserves task-relevant attributes. Closing this gap requires two innovations: (1) multi-objective optimization that jointly maximizes privacy and minimizes utility degradation across tasks (expression recognition, gaze estimation, age perception) under manufacturability constraints (printability, material reflectance, durability) absent in digital methods; and (2) attribute-disentangled adversarial patterns that suppress identity-discriminative features while sparing utility-related regions. Social acceptability is a further concern: bandages and medical masks offer plausible cover, whereas adversarial eyeglasses and stickers may attract unwanted attention, undermining covert protection.

Sensor Domain: Advancing Programmable Optics. Sensor-based face De-ID offers strong theoretical guarantees by preventing identity capture before digitization, yet current implementations rely on static or learned-but-fixed optics (Fig. 6), lacking **adaptive systems** that adjust privacy-utility trade-offs to scene context or task. Three directions warrant priority: (1) context-aware programmable optics that pair scene understanding (medical consultation versus public space) with real-time optical reconfiguration; (2) multi-modal sensor fusion combining privacy-encoded RGB with unencoded depth or infrared to preserve geometric utility (pose, tracking) while suppressing identity-rich texture; and (3) co-design of optical encoders and downstream networks through differentiable imaging pipelines. Programmable optics, however, incur latency, power, and cost penalties that limit deployment, so standardized protocols must assess privacy-utility trade-offs alongside computational efficiency, hardware complexity, and long-term robustness.

Digital Domain: Verifiable and Reversible Face De-ID. Two coupled challenges define the next phase of digital-domain face De-ID: *verifiability*, the ability to prove that a de-identified output meets a stated privacy criterion without revealing the original $\mathbf{I}$, and *reversibility* (criterion (4) of Sec. 2.3), required by forensic, medical, and legal workflows but difficult to combine with strong irreversible guaran-

tees. Existing reversible methods rely predominantly on password- or key-conditioned generative inversion [33], [127], which is vulnerable to brute-force, key-leakage, and auxiliary-information attacks because the recovery mapping is learned rather than cryptographically protected. A promising direction is to integrate cryptographic primitives, particularly homomorphic encryption and secure multi-party computation, with generative models, so that the recovery oracle $\mathcal{R}(\cdot\,; k)$ is realized through provably secure protocols rather than learned mappings alone. Verifiability can be addressed in parallel through zero-knowledge proofs or audit logs that attest, without disclosing $\mathbf{I}$, that $\tilde{\mathbf{I}}$ was produced by a sanctioned $\mathcal{F}$ meeting a privacy threshold.

Addressing Micro-Level Utility Signals. Face De-ID research predominantly targets macro-level attributes (age, gender, expression), yet emerging healthcare, affective computing, and behavioral analysis applications demand preservation of **micro-level** signals far more sensitive to identity suppression. Three categories stand out. (1) *Physiological signals* [139] (rPPG, heart rate, respiration) require signal-aware methods with frequency-domain constraints or correlation losses validated across skin tones, lighting, and De-ID paradigms. (2) *Micro-expressions* (fleeting movements revealing genuine emotion) require temporal robustness through optical-flow consistency and micro-expression-aware losses, supporting deception detection, mental health assessment, and human-robot interaction. (3) *Neurological diagnostic features* (stroke asymmetries, Parkinsonian tremor, autism-related expression abnormalities) require interdisciplinary collaboration to extend the OpticalDR paradigm [44], validate clinical accuracy, and earn medical community acceptance. Underlying all three is the fine-grained coupling between identity and utility, which calls for information-theoretic bounds on achievable privacy-utility trade-offs and principled, application-driven navigation of those bounds.

Unified Evaluation Protocols and Benchmarks. The **evaluation fragmentation** identified in Sec. 4 represents a primary bottleneck. Addressing these issues requires community-wide coordination: (1) curated multi-task datasets with comprehensive annotations (identity, age, gender, expression, landmarks); (2) standardized metrics spanning privacy (TAR@FAR, embedding distance), utility (attribute classification, landmark localization, expression recognition), and quality (FID, LPIPS, BRISQUE, NIQE); (3) reference implementations across all three domains with reproducible scripts; and (4) continuously updated leaderboards tracking SOTA performance. Beyond technical infrastructure, benchmark design must address fairness by stratifying performance across race, gender, and age, while adversarial robustness evaluation (JPEG compression, resizing, denoising) remains critical but underexplored.

Cross-Domain Co-Design for Privacy-First Sensing. The convergence of optical and algorithmic privacy mechanisms invites a rethinking of the sensing-to-analytics pipeline, with next-generation systems integrating the **physical-sensor-digital** stack to enforce privacy at the earliest stages of acquisition. Programmable sensors can implement De-ID in analog or early digital domains through neuromorphic architectures or event-based cameras that inherently suppress identity-rich spatial detail while retaining motion and scene dynamics. Differentiable imaging pipelines complement this by jointly optimizing optical elements (metasurfaces, diffractive optics), image signal processing, and downstream task networks under unified privacy-utility objectives. Realizing this vision requires differentiable models of the full image-formation process and optimization strategies that account for physical realizability, manufacturing tolerances, and deployment conditions, marking a shift from post-capture De-ID toward architectures where protection is intrinsic to the imaging system.

Emerging Application Domains. Beyond traditional surveillance, access control, and social media, face De-ID must address **emerging domains** where conventional approaches fall short. Three stand out. (1) *Autonomous vehicles and smart cities* deploy ubiquitous sensing and require perception systems that balance public safety against individual privacy. (2) *Medical and telehealth* applications need domain-specific methods that preserve clinically relevant cues (expressions for neurological assessment, dermatological conditions for diagnosis) while preventing re-identification. (3) *Metaverse and extended reality* introduce new attack surfaces, where avatars, behavioral patterns, and physiological signals (eye tracking, gestures, locomotion) enable identity inference beyond facial biometrics. Meeting these challenges demands collaboration across computer vision, cryptography, hardware design, and policy.

Joint De-ID Across Face, Body, and Behavioral Channels. Face De-ID alone is insufficient when other identity-bearing channels remain intact: body shape, gait, voice, eye gaze, and hand motions are each independently sufficient for re-identification, and recent surveys document active research on each [27], [38]. Face De-ID nonetheless retains substantial standalone value. The face is the most discriminative and most widely operationalized biometric cue [1], targeted by commercial recognition systems, watchlists, and web-scale search [4], so its suppression removes the adversary with the greatest practical reach. In the data-release scenarios that most often motivate De-ID (photographs, medical and telehealth recordings [6], social-media imagery [19]), the face is often the only channel captured at recognition-grade quality; gait requires temporal sequences [221] and body-appearance matching degrades under clothing and viewpoint variation [193]. Regulatory instruments such as the EU AI Act [5] and GDPR [222] further attach specific obligations to facial data. Face De-ID thus mitigates the dominant re-identification risk, while joint multi-channel anonymization remains necessary against adversaries with high-quality dynamic footage. Three open problems emerge at the intersection of face and these complementary channels: (1) *joint anonymization frameworks* that suppress identity across face, body, and motion simultaneously without re-introducing leakage in any single channel; (2) *temporal-consistency guarantees* for dynamic traits (*e.g.*, gait, micro-expressions), where frame-wise face De-ID may still leak identity through motion signatures; and (3) *cross-modal evaluation protocols* that quantify residual leakage when one channel is anonymized but others are not. Addressing these problems requires bridging the face De-ID literature reviewed in this survey with the soft biometric anonymization

literature, an essential direction for future research.

6 Conclusion

This survey presents the first unified treatment of face de-identification (De-ID) spanning the full acquisition pipeline, from the physical world through the sensor interface to the digital domain. Our domain-centric analysis exposes a clear trade-off: digital methods offer mature generative fidelity but expose raw identity signals; physical methods push the trust boundary outward yet remain limited by robustness and social acceptability; and sensor-domain methods embed privacy at capture, the closest realization of a privacy-by-design paradigm. Our review of evaluation protocols further reveals substantial fragmentation across datasets, thresholds, and metrics. Progress will require standardized benchmarks jointly measuring privacy, utility, and quality, together with cross-domain co-design that advances physical materials, programmable optics, and verifiable generative models in tandem. Meeting the emerging challenges of foundation-model-scale recognition and multimodal surveillance will be central to building vision systems that are trustworthy and socially responsible.

Acknowledgement

This work was supported by the Research Council of Finland (former Academy of Finland) Academy Professor project EmotionAI (grants 336116, 359894), the University of Oulu & Research Council of Finland Profi 7 (grant 352788), EU HORIZON-MSCA-SE-2022 project ACMod (grant 101130271), and the Finnish Doctoral Program Network in Artificial Intelligence, AI-DOC (decision number VN/3137/2024-OKM-6). As well, the authors wish to acknowledge CSC – IT Center for Science, Finland, for computational resources.

References

[1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *CSUR*, vol. 35, no. 4, pp. 399–458, 2003. 1, 15

[2] Y. Guo, H. Wang, L. Wang, Y. Lei, L. Liu, and M. Bennamoun, "3d face recognition: Two decades of progress and prospects," *CSUR*, vol. 56, no. 3, pp. 1–39, 2023. 1

[3] Y. Mi, Z. Zhong, Y. Huang, J. Ji, J. Xu, J. Wang, S. Wang, S. Ding, and S. Zhou, "Privacy-preserving face recognition using trainable feature subtraction," in *CVPR*, 2024, pp. 297–307. 1

[4] L. Laishram, M. Shaheryar, J. T. Lee, and S. K. Jung, "Toward a privacy-preserving face recognition system: A survey of leakages and solutions," *CSUR*, vol. 57, no. 6, pp. 1–38, 2025. 1, 2, 15

[5] European Union, "Regulation (eu) 2024/1689 of the european parliament and of the council of 13 june 2024 laying down harmonised rules on artificial intelligence," *Official Journal of the European Union*, vol. L, no. 2024/1689, 2024, entered into force: 2 August 2024. [Online]. Available: http://data.europa.eu/eli/reg/2024/1689/oj 1, 15

[6] Y. Yang, J. Lyu, R. Wang, Q. Wen, L. Zhao, W. Chen, S. Bi, J. Meng, K. Mao, Y. Xiao *et al.*, "A digital mask to safeguard patient privacy," *Nature Medicine*, vol. 28, no. 9, pp. 1883–1892, 2022. 1, 15

[7] Z. Cai, Z. Gao, B. Planche, M. Zheng, T. Chen, M. S. Asif, and Z. Wu, "Disguise without disruption: Utility-preserving face de-identification," in *AAAI*, vol. 38, no. 2, 2024, pp. 918–926. 1, 9, 12, 13

[8] S. E. Hudson and I. Smith, "Techniques for addressing fundamental privacy and disruption tradeoffs in awareness support systems," in *ACM CSCW*, 1996, pp. 248–257. 1, 2, 7, 11

[9] E. M. Newton, L. Sweeney, and B. Malin, "Preserving privacy by de-identifying face images," *IEEE TKDE*, vol. 17, no. 2, pp. 232–243, 2005. 1, 2, 7, 11, 13

[10] S. Jia, B. Yin, T. Yao, S. Ding, C. Shen, X. Yang, and C. Ma, "Adv-attribute: Inconspicuous and transferable adversarial attack on face recognition," in *NIPS*, 2022, pp. 34 136–34 147. 1, 8, 9, 11

[11] A. Salar, Q. Liu, Y. Tian, and G. Zhao, "Enhancing facial privacy protection via weakening diffusion purification," in *CVPR*, 2025, pp. 8235–8244. 1, 8, 11, 14

[12] L. Zhai, Q. Guo, X. Xie, L. Ma, Y. E. Wang, and Y. Liu, "A^{3}gan: Attribute-aware anonymization networks for face de-identification," in *ACM MM*, 2022, pp. 5303–5313. 1, 9, 12

[13] H.-W. Kung, T. Varanka, S. Saha, T. Sim, and N. Sebe, "Face anonymization made simple," in *WACV*. IEEE, 2025, pp. 1040–1050. 1, 4, 10, 12

[14] F. Pittaluga and S. J. Koppal, "Privacy preserving optics for miniature vision sensors," in *CVPR*, 2015, pp. 314–324. 1, 2, 6, 11

[15] J. Lopez, C. Hinojosa, H. Arguello, and B. Ghanem, "Privacy-preserving optics for enhancing protection in face de-identification," in *CVPR*, 2024, pp. 12 120–12 129. 1, 4, 6, 11

[16] M. Sharif, S. Bhagavatula, L. Bauer, and M. K. Reiter, "Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition," in *ACM SIGSAC*, 2016, pp. 1528–1540. 1, 2, 4, 5, 11, 13

[17] B. Yin, W. Wang, T. Yao, J. Guo, Z. Kong, S. Ding, J. Li, and C. Liu, "Adv-makeup: A new imperceptible and transferable attack on face recognition," in *IJCAI*, 2021, pp. 1–7. 1, 5, 6, 11

[18] Y. Liu, H. Wei, C. Jia, R. Xiao, W. Ruan, X. Wei, J. T. Zhou, and Z. Wang, "Projattacker: A configurable physical adversarial attack for face recognition via projector," in *CVPR*, 2025, pp. 21 248–21 257. 1, 4, 5, 6, 11, 13

[19] C. Liu, T. Zhu, J. Zhang, and W. Zhou, "Privacy intelligence: A survey on image privacy in online social networks," *CSUR*, vol. 55, no. 8, pp. 1–35, 2022. 2, 15

[20] K. H. Cheng, Z. Yu, H. Chen, and G. Zhao, "Benchmarking 3d face de-identification with preserving facial attributes," in *ICIP*. IEEE, 2022, pp. 656–660. 2

[21] M. R. Hasan, R. Guest, and F. Deravi, "Presentation-level privacy protection techniques for automated face recognition—a survey," *CSUR*, vol. 55, no. 13s, pp. 1–27, 2023. 2

[22] J. Cao, X. Chen, B. Liu, M. Ding, R. Xie, L. Song, Z. Li, and W. Zhang, "Face de-identification: State-of-the-art methods and comparative studies," *arXiv preprint arXiv:2411.09863*, 2024. 2

[23] J. R. Padilla-López, A. A. Chaaraoui, and F. Flórez-Revuelta, "Visual privacy protection methods: A survey," *Expert Systems with Applications*, vol. 42, no. 9, pp. 4177–4195, 2015. 2

[24] S. Ribaric and N. Pavesic, "An overview of face de-identification in still images and videos," in *FG*, vol. 4. IEEE, 2015, pp. 1–6. 2

[25] S. Ribaric, A. Ariyaeeinia, and N. Pavesic, "De-identification for privacy protection in multimedia content: A survey," *Signal Processing: Image Communication*, vol. 47, pp. 131–151, 2016. 2

[26] B. Meden, P. Rot, P. Terhörst, N. Damer, A. Kuijper, W. J. Scheirer, A. Ross, P. Peer, and V. Štruc, "Privacy–enhancing face biometrics: A comprehensive survey," *TIFS*, vol. 16, pp. 4147–4183, 2021. 2

[27] W. Khan, L. Topham, U. Khayam, S. Ortega-Martorell, P. Heather, D. Ansell, D. Al-Jumeily, and A. Hussain, "Person de-identification: A comprehensive review of methods, datasets, applications, and ethical aspects along with new dimensions," *IEEE TBBIS*, pp. 293–312, 2024. 2, 15

[28] Y. Wen, B. Liu, L. Song, J. Cao, and R. Xie, *Face De-identification: Safeguarding Identities in the Digital Era*. Springer, 2024. 2

[29] S. Goswami, S. K. Paul, S. Jawlia, and A. Goel, "Faces in the fog: A deep dive into face de-identification techniques and their comparative analysis," in *DICCT*. IEEE, 2025, pp. 603–608. 2

[30] S. Park, H. Kim, S.-K. Choi, T. Kim, and E. Park, "Privacy-driven faces: A survey on generative facial de-identification," in *WSIDC*, 2025, pp. 27–32. 2

[31] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991. 2

[32] X. Yang, C. Liu, L. Xu, Y. Wang, Y. Dong, N. Chen, H. Su, and J. Zhu, "Towards effective adversarial textured 3d meshes on physical face recognition," in *CVPR*, 2023, pp. 4119–4128. 2, 3, 4, 5, 6, 11, 13

[33] H. Yang, X. Xu, C. Xu, H. Zhang, J. Qin, Y. Wang, P.-A. Heng, and S. He, "G²face: High-fidelity reversible face anonymization via

generative and geometric priors," *TIFS*, vol. 19, pp. 8773–8785, 2024. 2, 10, 12, 15
[34] Y. Zhao and J. Chen, "A survey on differential privacy for unstructured data content," *CSUR*, vol. 54, no. 10s, pp. 1–28, 2022. 2
[35] W. Huang, M. Ye, Z. Shi, G. Wan, H. Li, B. Du, and Q. Yang, "Federated learning for generalization, robustness, fairness: A survey and benchmark," *IEEE TPAMI*, vol. 46, no. 12, pp. 9387–9406, 2024. 2
[36] D. Basin, C. Cremers, J. Dreier, and R. Sasse, "Tamarin: Verification of large-scale, real-world, cryptographic protocols," *IEEE Security & Privacy*, vol. 20, no. 3, pp. 24–32, 2022. 2
[37] X. Jia, J. Du, H. Wei, R. Xue, Z. Wang, H. Zhu, and J. Chen, "Balancing privacy and performance: A many-in-one approach for image anonymization," in *AAAI*, vol. 39, no. 17, 2025, pp. 17 608–17 616. 2
[38] S. Hanisch, P. Arias-Cabarcos, J. Parra-Arnau, and T. Strufe, "Anonymization techniques for behavioral biometric data: A survey," *CSUR*, vol. 57, no. 11, pp. 1–54, 2025. 2, 15
[39] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *CVPR*, 2015, pp. 815–823. 3
[40] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "Sphereface: Deep hypersphere embedding for face recognition," in *CVPR*, 2017, pp. 212–220. 3
[41] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *CVPR*, 2019, pp. 4690–4699. 3
[42] M. Kim, A. K. Jain, and X. Liu, "Adaface: Quality adaptive margin for face recognition," in *CVPR*, 2022, pp. 18 750–18 759. 3
[43] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou, "Retinaface: Single-shot multi-level face localisation in the wild," in *CVPR*, 2020, pp. 5203–5212. 3
[44] Y. Pan, J. Jiang, K. Jiang, Z. Wu, K. Yu, and X. Liu, "Opticaldr: A deep optical imaging model for privacy-protective depression recognition," in *CVPR*, 2024, pp. 1303–1312. 3, 6, 11, 13, 15
[45] M. Maximov, I. Elezi, and L. Leal-Taixé, "Ciagan: Conditional identity anonymization generative adversarial networks," in *CVPR*, 2020, pp. 5447–5456. 3, 12
[46] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *NIPS*, 2014, pp. 1–9. 3, 5
[47] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *NIPS*, 2020, pp. 6840–6851. 3
[48] M. Ryoo, B. Rothrock, C. Fleming, and H. J. Yang, "Privacy-preserving human activity recognition from extreme low resolution," in *AAAI*, vol. 31, no. 1, 2017, pp. 1–8. 4, 6, 7, 11
[49] S. Komkov and A. Petiushko, "Advhat: Real-world adversarial attack on arcface face id system," in *ICPR*. IEEE, 2021, pp. 819–826. 4, 5, 11
[50] D.-L. Nguyen, S. S. Arora, Y. Wu, and H. Yang, "Adversarial light projection attacks on face recognition systems: A feasibility study," in *CVPRW*, 2020, pp. 814–815. 5, 6, 11
[51] A. Zolfi, S. Avidan, Y. Elovici, and A. Shabtai, "Adversarial mask: Real-world universal adversarial attack on face recognition models," in *MLKDD*. Springer, 2022, pp. 304–320. 5, 6, 11
[52] X. Wei, Y. Guo, and J. Yu, "Adversarial sticker: A stealthy attack method in the physical world," *IEEE TPAMI*, vol. 45, no. 3, pp. 2711–2725, 2022. 5, 11
[53] Y. Li, Y. Li, X. Dai, S. Guo, and B. Xiao, "Physical-world optical adversarial attacks on 3d face recognition," in *CVPR*, 2023, pp. 24 699–24 708. 5, 6, 11
[54] M. Sharif, S. Bhagavatula, L. Bauer, and M. K. Reiter, "A general framework for adversarial examples with objectives," *ACM TPS*, vol. 22, no. 3, pp. 1–30, 2019. 4, 11, 13
[55] M. Pautov, G. Melnikov, E. Kaziakhmedov, K. Kireev, and A. Petiushko, "On adversarial patches: Real-world attack on arcface-100 face recognition system," in *SIBIRCON*. IEEE, 2019, pp. 0391–0396. 4, 11
[56] I. Singh, T. Araki, and K. Kakizaki, "Powerful physical adversarial examples against practical face recognition systems," in *WACV*, 2022, pp. 301–310. 4, 6, 11
[57] Z. Xiao, X. Gao, C. Fu, Y. Dong, W. Gao, X. Zhang, J. Zhou, and J. Zhu, "Improving transferability of adversarial patches on face recognition with generative models," in *CVPR*, 2021, pp. 11 845–11 854. 4, 5, 11
[58] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," in *ICLR*, 2018, pp. 1–12. 5, 10, 11, 12
[59] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *CVPR*, 2019, pp. 4401–4410. 5, 10, 11, 12
[60] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of stylegan," in *CVPR*, 2020, pp. 8110–8119. 5, 11
[61] X. Zheng, Y. Fan, B. Wu, Y. Zhang, J. Wang, and S. Pan, "Robust physical-world attacks on face recognition," *PR*, vol. 133, p. 109009, 2023. 5, 11
[62] X. Liu, F. Shen, J. Zhao, and C. Nie, "Eap: An effective black-box impersonation adversarial patch attack method on face recognition in the physical world," *Neurocomputing*, vol. 580, p. 127517, 2024. 5, 11
[63] X. Wei, Y. Guo, J. Yu, and B. Zhang, "Simultaneously optimizing perturbations and positions for black-box adversarial patch attacks," *IEEE TPAMI*, vol. 45, no. 7, pp. 9041–9054, 2023. 5, 11
[64] M. Shen, H. Yu, L. Zhu, K. Xu, Q. Li, and J. Hu, "Effective and robust physical-world attacks on deep learning face recognition systems," *TIFS*, vol. 16, pp. 4063–4077, 2021. 5, 11
[65] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *ICCV*, 2017, pp. 618–626. 5
[66] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *ICLR*, 2015. 5
[67] H. Gong, M. Dong, S. Ma, S. Camtepe, S. Nepal, and C. Xu, "Stealthy physical masked face recognition attack via adversarial style optimization," *TMM*, vol. 26, pp. 5014–5025, 2023. 5, 11
[68] S. Bhilare and A. Hati, "Fooling face recognition systems through physical adversarial attack," in *ICCVIP*. Springer, 2024, pp. 420–435. 5, 6, 11
[69] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," in *ICLR*, Y. Bengio and Y. LeCun, Eds., 2015, pp. 1–11. 5
[70] A. Tuan Tran, T. Hassner, I. Masi, and G. Medioni, "Regressing robust and discriminative 3d morphable models with a very deep neural network," in *CVPR*, 2017, pp. 5163–5172. 5
[71] X. Yang, L. Xu, T. Pang, Y. Dong, Y. Wang, H. Su, and J. Zhu, "Face3dadv: Exploiting robust adversarial 3d patches on physical face recognition," *IJCV*, vol. 133, no. 1, pp. 353–371, 2025. 5, 6, 11
[72] Y. Wang, Z. Liu, B. Luo, R. Hui, and F. Li, "The invisible polyjuice potion: an effective physical adversarial attack against face recognition," in *ACM SIGSAC*, 2024, pp. 3346–3360. 5, 6, 11
[73] Q. Zhang, Q. Guo, R. Gao, F. Juefei-Xu, H. Yu, and W. Feng, "Adversarial relighting against face recognition," *TIFS*, vol. 19, pp. 9145–9157, 2024. 5, 11, 13
[74] L. Sweeney, "k-anonymity: A model for protecting privacy," *IJUFKS*, vol. 10, no. 05, pp. 557–570, 2002. 6, 7
[75] F. Pittaluga and S. J. Koppal, "Pre-capture privacy for small vision sensors," *IEEE TPAMI*, vol. 39, no. 11, pp. 2215–2226, 2016. 6, 11
[76] C. Hinojosa, J. C. Niebles, and H. Arguello, "Learning privacy-preserving optics for human pose estimation," in *ICCV*, 2021, pp. 2573–2582. 6, 11
[77] C. Hinojosa, M. Marquez, H. Arguello, E. Adeli, L. Fei-Fei, and J. C. Niebles, "Privhar: Recognizing human actions from privacy-preserving lens," in *ECCV*, 2022, pp. 314–332. 6, 11, 13
[78] Z. Tasneem, G. Milione, Y.-H. Tsai, X. Yu, A. Veeraraghavan, M. Chandraker, and F. Pittaluga, "Learning phase mask for privacy-preserving passive depth estimation," in *ECCV*. Springer, 2022, pp. 504–521. 6, 11, 13
[79] J. Cheng, X. Dai, J. Wan, N. Antipa, and N. Vasconcelos, "Learning a dynamic privacy-preserving camera robust to inversion attacks," in *ECCV*. Springer, 2024, pp. 349–367. 6, 7, 11, 13
[80] M. Ryoo, K. Kim, and H. Yang, "Extreme low resolution activity recognition with multi-siamese embedding learning," in *AAAI*, vol. 32, no. 1, 2018. 6, 7, 11
[81] G. Letournel, A. Bugeau, V.-T. Ta, and J.-P. Domenger, "Face de-identification with expressions preservation," in *ICIP*. IEEE, 2015, pp. 4366–4370. 7, 8, 11
[82] M. Boyle, C. Edwards, and S. Greenberg, "The effects of filtered video on awareness and privacy," in *ACM CSCW*, 2000, pp. 1–10. 7, 11
[83] C. Neustaedter, S. Greenberg, and M. Boyle, "Blur filtration fails to preserve privacy for home-based video conferencing," *ACM TOCHI*, vol. 13, no. 1, pp. 1–36, 2006. 7, 11

[84] J. L. Crowley, J. Coutaz *et al.*, "Things that see," *Communications of the ACM*, pp. 1–10, 2000. 7, 11

[85] R. Gross, E. Airoldi, B. Malin, and L. Sweeney, "Integrating utility into face de-identification," in *PETW*. Springer, 2005, pp. 227–242. 7, 11, 13

[86] R. Gross, L. Sweeney, F. De la Torre, and S. Baker, "Model-based face de-identification," in *CVPRW*, 2006, pp. 161–161. 7, 11

[87] R. Gross and L. Sweeney, "Towards real-world face de-identification," in *ICB*. IEEE, 2007, pp. 1–8. 7, 11

[88] R. Gross, L. Sweeney, F. De La Torre, and S. Baker, "Semi-supervised learning of multi-factor models for face de-identification," in *CVPR*, 2008, pp. 1–8. 8, 11

[89] L. Meng and Z. Sun, "Face de-identification with perfect privacy protection," in *ICICTEM*. IEEE, 2014, pp. 1234–1239. 8, 11

[90] L. Du, M. Yi, E. Blasch, and H. Ling, "Garp-face: Balancing privacy protection and utility preservation in face de-identification," in *IJCB*. IEEE, 2014, pp. 1–8. 8, 11

[91] A. Jourabloo, X. Yin, and X. Liu, "Attribute preserved face de-identification," in *ICB*. IEEE, 2015, pp. 278–285. 8, 11

[92] H. Chi and Y. H. Hu, "Face de-identification using facial identity preserving features," in *GCSIP*. IEEE, 2015, pp. 586–590. 8, 11

[93] B. Meden, Ž. Emeršič, V. Štruc, and P. Peer, "k-same-net: k-anonymity with generative deep neural networks for face deidentification," *Entropy*, vol. 20, no. 1, p. 60, 2018. 8, 11

[94] B. Yan, M. Pei, and Z. Nie, "Attributes preserving face de-identification," in *ICCVW*, 2019, pp. 1217–1221. 8, 11

[95] Y. Dong, H. Su, B. Wu, Z. Li, W. Liu, T. Zhang, and J. Zhu, "Efficient decision-based black-box adversarial attacks on face recognition," in *CVPR*, 2019, pp. 7714–7722. 8, 9, 11

[96] E. Chatzikyriakidis, C. Papaioannidis, and I. Pitas, "Adversarial face de-identification," in *ICIP*. IEEE, 2019, pp. 684–688. 8, 9, 11

[97] D. Deb, J. Zhang, and A. K. Jain, "Advfaces: Adversarial face synthesis," in *IJCB*. IEEE, 2020, pp. 1–10. 8, 11

[98] H. Xue, B. Liu, X. Yuan, M. Ding, and T. Zhu, "Face image de-identification by feature space adversarial perturbation," *CCPE*, vol. 35, no. 5, p. e7554, 2023. 8, 9, 11

[99] X. Yang, Y. Dong, T. Pang, H. Su, J. Zhu, Y. Chen, and H. Xue, "Towards face encryption by generating adversarial identity masks," in *ICCV*, 2021, pp. 3897–3907. 8, 9, 11

[100] G. Hanawa, K. Ito, and T. Aoki, "Face image de-identification based on feature embedding," *EURASIP Journal on Image and Video Processing*, vol. 2024, no. 1, p. 25, 2024. 8, 11

[101] H. Wang, W. Luo, X. Xie, P. Zheng, W. Huang, and J. Huang, "Adv-inversion: Stealthy adversarial attacks via gan-inversion for facial privacy protection," *TIFS*, pp. 1–15, 2025. 8, 9, 11

[102] S. Hu, X. Liu, Y. Zhang, M. Li, L. Y. Zhang, H. Jin, and L. Wu, "Protecting facial privacy: Generating adversarial identity masks via style-robust makeup transfer," in *CVPR*, 2022, pp. 15 014–15 023. 8, 11

[103] Z.-A. Zhu, Y.-Z. Lu, and C.-K. Chiang, "Generating adversarial examples by makeup attacks on face recognition," in *ICIP*. IEEE, 2019, pp. 2516–2520. 8, 11

[104] F. Shamshad, M. Naseer, and K. Nandakumar, "Clip2protect: Protecting facial privacy using text-guided makeup via adversarial latent search," in *CVPR*, 2023, pp. 20 595–20 605. 8, 11

[105] J. Liu, C. P. Lau, and R. Chellappa, "Diffprotect: Generate adversarial examples with diffusion models for facial privacy protection," *arXiv preprint arXiv:2305.13625*, pp. 1–14, 2023. 8, 11, 14

[106] J. Wang, H. Zhang, and Y. Yuan, "Adv-cpg: A customized portrait generation framework with facial adversarial attacks," in *CVPR*, 2025, pp. 21 001–21 010. 8, 11

[107] Y. Zhong and W. Deng, "Towards transferable adversarial attack against deep face recognition," *TIFS*, vol. 16, pp. 1452–1466, 2020. 8, 9, 11

[108] Z. Li, B. Yin, T. Yao, J. Guo, S. Ding, S. Chen, and C. Liu, "Sibling-attack: Rethinking transferable adversarial attacks against face recognition," in *CVPR*, 2023, pp. 24 626–24 637. 8, 9, 11

[109] H. Ma, K. Xu, X. Jiang, Z. Zhao, and T. Sun, "Transferable black-box attack against face recognition with spatial mutable adversarial patch," *TIFS*, vol. 18, pp. 5636–5650, 2023. 8, 11

[110] C. Hu, Y. Li, Z. Feng, and X. Wu, "Toward transferable attack via adversarial diffusion in face recognition," *TIFS*, vol. 19, pp. 5506–5519, 2024. 8, 11

[111] F. Zhou, B. Yin, H. Ling, Q. Zhou, and W. Wang, "Improving the transferability of adversarial attacks on face recognition with diverse parameters augmentation," in *CVPR*, 2025, pp. 3516–3527. 8, 11

[112] Q. Li, Y. Hu, Y. Liu, D. Zhang, X. Jin, and Y. Chen, "Discrete point-wise attack is not enough: Generalized manifold adversarial attack for face recognition," in *CVPR*, 2023, pp. 20 575–20 584. 8, 11

[113] T. Amada, S. P. Liew, K. Kakizaki, and T. Araki, "Universal adversarial spoofing attacks against face recognition," in *IJCB*. IEEE, 2021, pp. 1–7. 8, 11

[114] M. Shen, Z. Liao, L. Zhu, K. Xu, and X. Du, "Vla: A practical visible light-based attack on face recognition systems in physical world," *ACM IMWUT*, vol. 3, no. 3, pp. 1–19, 2019. 9, 11

[115] K.-H. Chow, S. Hu, T. Huang, and L. Liu, "Personalized privacy protection mask against unauthorized facial recognition," in *ECCV*. Springer, 2024, pp. 434–450. 9, 11

[116] S. Pang, R. Ma, B. Li, Y. Zhou, and Y. Yao, "Veil privacy on visual data: Concealing privacy for humans, unveiling for dnns," in *ECCV*. Springer, 2024, pp. 280–297. 9, 10, 12

[117] Y. Li and S. Lyu, "De-identification without losing faces," in *IHMSW*, 2019, pp. 83–88. 9, 12

[118] A. Aggarwal, R. Rathore, P. Chattopadhyay, and L. Wang, "Epd-net: A gan-based architecture for face de-identification from images," in *IOTEM*. IEEE, 2020, pp. 1–7. 9, 12

[119] O. Gafni, L. Wolf, and Y. Taigman, "Live face de-identification in video," in *ICCV*, 2019, pp. 9378–9387. 9, 12

[120] Y. Wu, F. Yang, Y. Xu, and H. Ling, "Privacy-protective-gan for privacy preserving face de-identification," *JCST*, vol. 34, no. 1, pp. 47–60, 2019. 9, 11

[121] K. Brkic, I. Sikiric, T. Hrkac, and Z. Kalafatic, "I know that person: Generative full body and face de-identification of people in images," in *CVPRW*. IEEE, 2017, pp. 1319–1328. 9, 11

[122] Z. Kuang, H. Liu, J. Yu, A. Tian, L. Wang, J. Fan, and N. Babaguchi, "Effective de-identification generative adversarial network for face anonymization," in *ACM MM*, 2021, pp. 3182–3191. 9, 10, 12

[123] A. Agarwal, P. Chattopadhyay, and L. Wang, "Privacy preservation through facial de-identification with simultaneous emotion preservation," *SIVP*, vol. 15, no. 5, pp. 951–958, 2021. 9, 12

[124] S. Yang, W. Wang, Y. Cheng, and J. Dong, "A systematical solution for face de-identification," in *CCBR*. Springer, 2021, pp. 20–30. 9, 12

[125] Y. Li, Q. Lu, Q. Tao, X. Zhao, and Y. Yu, "Sf-gan: Face de-identification method without losing facial attribute information," *IEEE SPL*, vol. 28, pp. 1345–1349, 2021. 9, 12

[126] T. Li and L. Lin, "Anonymousnet: Natural face de-identification with measurable privacy," in *CVPRW*, 2019, pp. 0–10. 9, 10, 12

[127] Y. Wen, B. Liu, J. Cao, R. Xie, L. Song, and Z. Li, "Identitymask: Deep motion flow guided reversible face video de-identification," *IEEE TCSVT*, vol. 32, no. 12, pp. 8353–8367, 2022. 9, 10, 12, 15

[128] J. Cao, B. Liu, Y. Wen, R. Xie, and L. Song, "Personalized and invertible face de-identification by disentangled identity information manipulation," in *ICCV*, 2021, pp. 3334–3342. 9, 10, 12

[129] H. Proença, "The uu-net: Reversible face de-identification for visual surveillance video footage," *IEEE TCSVT*, vol. 32, no. 2, pp. 496–509, 2022. 9, 12

[130] Y. Wen, B. Liu, J. Cao, R. Xie, and L. Song, "Divide and conquer: a two-step method for high quality face de-identification with model explainability," in *ICCV*, 2023, pp. 5148–5157. 9, 10, 12

[131] S. M. S. M. Khorzooghi and S. Nilizadeh, "Examining stylegan as a utility-preserving face de-identification method," in *PET*, 2023, p. 341–358. 9, 12

[132] S. Barattin, C. Tzelepis, I. Patras, and N. Sebe, "Attribute-preserving face dataset anonymization via latent code optimization," in *CVPR*, 2023, pp. 8001–8010. 9, 10, 12, 13

[133] B. Meden, M. Gonzalez-Hernandez, P. Peer, and V. Štruc, "Face deidentification with controllable privacy protection," *IVC*, vol. 134, p. 104678, 2023. 10, 12, 13

[134] Y. Sun, L. Yu, H. Xie, J. Li, and Y. Zhang, "Diffam: Diffusion-based adversarial makeup transfer for facial privacy protection," in *CVPR*, 2024, pp. 24 584–24 594. 10, 12

[135] X. He, M. Zhu, D. Chen, N. Wang, and X. Gao, "Diff-privacy: Diffusion-based face privacy protection," *IEEE TCSVT*, vol. 34, no. 12, pp. 13 164–13 176, 2024. 10, 12, 13

[136] J. Park, S. Lee, M. Shaheryar, and S. K. Jung, "Facial identity editing: Towards effective de-identification," in *ICIP*. IEEE, 2025, pp. 1792–1797. 10, 12, 14

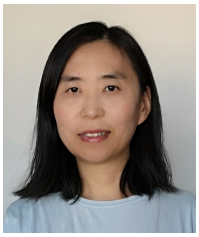
[137] C.-H. Lin, Z.-H. Wang, and G.-J. Jong, "A de-identification face recognition using extracted thermal features based on deep learn-

ing," *IEEE Sensors Journal*, vol. 20, no. 16, pp. 9510–9517, 2020. 10, 12
[138] J. Cao, B. Liu, Y. Wen, R. Xie, and L. Song, "Achieving privacy-preserving multi-view consistency with advanced 3d-aware face de-identification," in *ACM MM Asia*, 2023, pp. 1–7. 10, 12
[139] M. Savic and G. Zhao, "De-identification of facial videos while preserving remote physiological utility," in *BMVC*. BMVA Press, 2023, pp. 1–14. 10, 12, 13, 15
[140] B. Zhu, C. Zhang, Y. Sui, and L. Li, "Facemotionpreserve: A generative approach for facial de-identification and medical information preservation," *Scientific Reports*, vol. 14, no. 1, p. 17275, 2024. 10, 12
[141] B. Puangthamawathanakun, C. Arpnikanondt, W. Krathu, G. Healy, and C. Gurrin, "Towards face de-identification for wearable cameras," in *ICCMI*, 2023, pp. 210–216. 10, 12
[142] Y. Liu, K. H. Cheng, M. Savic, H. Chen, Z. Yu, and G. Zhao, "3d face de-identification with preserving multi-facial attributes: A benchmark," *IEEE TBIOM*, pp. 1–14, 2025. 10, 12
[143] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," in *Workshop on Faces in'Real-Life'Images*, 2008, pp. 1–11. 10, 11, 12
[144] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *ICCV*, 2015, pp. 3730–3738. 10, 11, 12
[145] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," in *FG*. IEEE, 2018, pp. 67–74. 10, 11, 12
[146] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," *arXiv preprint arXiv:1411.7923*, pp. 1–9, 2014. 10, 11, 12
[147] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, "Ms-celeb-1m: A dataset and benchmark for large-scale face recognition," in *ECCV*. Springer, 2016, pp. 87–102. 10, 11
[148] I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller, and E. Brossard, "The megaface benchmark: 1 million faces for recognition at scale," in *CVPR*, 2016, pp. 4873–4882. 10, 11
[149] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and simile classifiers for face verification," in *ICCV*. IEEE, 2009, pp. 365–372. 10, 11, 12
[150] S. Moschoglou, A. Papaioannou, C. Sagonas, J. Deng, I. Kotsia, and S. Zafeiriou, "Agedb: The first manually collected, in-the-wild age database," in *CVPRW*, 2017, pp. 51–59. 10, 11
[151] 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 *WACV*. IEEE, 2016, pp. 1–9. 10, 11
[152] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The feret database and evaluation procedure for face-recognition algorithms," *IVC*, vol. 16, no. 5, pp. 295–306, 1998. 10, 11
[153] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-pie," in *FG*, 2008, pp. 1–8. 10, 11
[154] K. Ricanek and T. Tesafaye, "Morph: A longitudinal image database of normal adult age-progression," in *FG*. IEEE, 2006, pp. 341–345. 10, 11
[155] O. Langner, R. Dotsch, G. Bijlstra, D. H. Wigboldus, S. T. Hawk, and A. Van Knippenberg, "Presentation and validation of the radboud faces database," *Cognition and Emotion*, vol. 24, no. 8, pp. 1377–1388, 2010. 10, 11, 12
[156] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *CVPRW*. IEEE, 2010, pp. 94–101. 10, 11, 13
[157] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "Affectnet: A database for facial expression, valence, and arousal computing in the wild," *IEEE TAC*, vol. 10, no. 1, pp. 18–31, 2017. 10, 12
[158] Q. Gu, G. Wang, M. T. Chiu, Y.-W. Tai, and C.-K. Tang, "Ladn: Local adversarial disentangling network for facial makeup and de-makeup," in *ICCV*, 2019, pp. 10 481–10 490. 10, 11, 12
[159] L. Wolf, T. Hassner, and I. Maoz, "Face recognition in unconstrained videos with matched background similarity," in *CVPR*. IEEE, 2011, pp. 529–534. 10, 11, 12
[160] A. Nagraniy, J. S. Chungy, and A. Zisserman, "Voxceleb: A large-scale speaker identification dataset," in *ISCA*, vol. 2017, 2017, pp. 2616–2620. 10, 12
[161] C. Sanderson and B. C. Lovell, "Multi-region probabilistic histograms for robust and scalable identity inference," in *ICB*. Springer, 2009, pp. 199–208. 10, 12
[162] K. Grauman, A. Westbury, E. Byrne, Z. Chavis, A. Furnari, R. Girdhar, J. Hamburger, H. Jiang, M. Liu, X. Liu *et al.*, "Ego4d: Around the world in 3,000 hours of egocentric video," in *CVPR*, 2022, pp. 18 995–19 012. 10, 12
[163] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun, "Bosphorus database for 3d face analysis," in *BIMW*. Springer, 2008, pp. 47–56. 10, 11, 12
[164] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3d facial expression database for facial behavior research," in *FG*. IEEE, 2006, pp. 211–216. 10, 12
[165] R. Min, N. Kose, and J.-L. Dugelay, "Kinectfacedb: A kinect database for face recognition," *TSMCS*, vol. 44, no. 11, pp. 1534–1548, 2014. 10, 11
[166] Y. Ye, Z. Song, J. Guo, and Y. Qiao, "Siat-3dfe: A high-resolution 3d facial expression dataset," *IEEE Access*, vol. 8, pp. 48 205–48 211, 2020. 10, 11
[167] R. Stricker, S. Müller, and H.-M. Gross, "Non-contact video-based pulse rate measurement on a mobile service robot," in *IEEE RHIC*. IEEE, 2014, pp. 1056–1062. 10, 12, 13
[168] X. Li, I. Alikhani, J. Shi, T. Seppanen, J. Junttila, K. Majamaa-Voltti, M. Tulppo, and G. Zhao, "The obf database: A large face video database for remote physiological signal measurement and atrial fibrillation detection," in *FG*. IEEE, 2018, pp. 242–249. 10, 12, 13
[169] M. Valstar, B. Schuller, K. Smith, F. Eyben, B. Jiang, S. Bilakhia, S. Schnieder, R. Cowie, and M. Pantic, "Avec 2013: The continuous audio/visual emotion and depression recognition challenge," in *ACM IWAVEC*, 2013, pp. 3–10. 10, 11, 13
[170] M. Valstar, B. Schuller, K. Smith, T. Almaev, F. Eyben, J. Krajewski, R. Cowie, and M. Pantic, "Avec 2014: 3d dimensional affect and depression recognition challenge," in *ACM IWAVEC*, 2014, pp. 3–10. 10, 11, 13
[171] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre, "Hmdb: A large video database for human motion recognition," in *ICCV*. IEEE, 2011, pp. 2556–2563. 10, 11
[172] Y. Iwashita, A. Takamine, R. Kurazume, and M. S. Ryoo, "First-person animal activity recognition from egocentric videos," in *ICPR*. IEEE, 2014, pp. 4310–4315. 10, 11
[173] M. S. Ryoo and L. Matthies, "First-person activity recognition: What are they doing to me?" in *CVPR*, 2013, pp. 2730–2737. 11
[174] T. Orekondy, B. Schiele, and M. Fritz, "Towards a visual privacy advisor: Understanding and predicting privacy risks in images," in *ICCV*, 2017, pp. 3686–3695. 11
[175] Z. Wu, H. Wang, Z. Wang, H. Jin, and Z. Wang, "Privacy-preserving deep action recognition: An adversarial learning framework and a new dataset," *IEEE TPAMI*, vol. 44, no. 4, pp. 2126–2139, 2020. 11
[176] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from rgbd images," in *ECCV*. Springer, 2012, pp. 746–760. 11
[177] M. B. Stegmann, B. K. Ersboll, and R. Larsen, "Fame-a flexible appearance modeling environment," *IEEE TMI*, vol. 22, no. 10, pp. 1319–1331, 2003. 11
[178] B. Samarzija and S. Ribaric, "An approach to the de-identification of faces in different poses," in *ICICTEM*. IEEE, 2014, pp. 1246–1251. 11
[179] S. Mosaddegh, L. Simon, and F. Jurie, "Photorealistic face de-identification by aggregating donors' face components," in *ACCV*. Springer, 2014, pp. 159–174. 11
[180] S. Milborrow, J. Morkel, and F. Nicolls, "The muct landmarked face database," *PRASA*, vol. 201, no. 0, p. 535, 2010. 11
[181] A. Kasinski, A. Florek, and A. Schmidt, "The put face database," *IPC*, vol. 13, no. 3-4, pp. 59–64, 2008. 11
[182] N. Kumar, P. Belhumeur, and S. Nayar, "Facetracer: A search engine for large collections of images with faces," in *ECCV*. Springer, 2008, pp. 340–353. 11
[183] H.-W. Ng and S. Winkler, "A data-driven approach to cleaning large face datasets," in *ICIP*. IEEE, 2014, pp. 343–347. 11
[184] K. Chang, K. Bowyer, and P. Flynn, "Face recognition using 2d and 3d facial data," in *ACM MUAW*, 2003, pp. 25–32. 11
[185] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, and D. Zhao, "The cas-peal large-scale chinese face database and baseline evaluations," *TSMC*, vol. 38, no. 1, pp. 149–161, 2007. 11
[186] O. Jesorsky, K. J. Kirchberg, and R. W. Frischholz, "Robust face detection using the hausdorff distance," in *ICAVBPA*. Springer, 2001, pp. 90–95. 11, 12

[187] K. Messer, J. Matas, J. Kittler, J. Luettin, G. Maitre *et al.*, "Xm2vtsdb: The extended m2vts database," in *ICAVBPA*, vol. 964. Washington, DC, 1999, pp. 965–966. 11, 12
[188] F. Wang, L. Chen, C. Li, S. Huang, Y. Chen, C. Qian, and C. C. Loy, "The devil of face recognition is in the noise," in *ECCV*, 2018, pp. 765–780. 11
[189] T. Li, R. Qian, C. Dong, S. Liu, Q. Yan, W. Zhu, and L. Lin, "Beautygan: Instance-level facial makeup transfer with deep generative adversarial network," in *ACM MM*, 2018, pp. 645–653. 11
[190] J. Huang, X. Dong, W. Song, Z. Chong, Z. Tang, J. Zhou, Y. Cheng, L. Chen, H. Li, Y. Yan *et al.*, "Consistentid: Portrait generation with multimodal fine-grained identity preserving," *arXiv preprint arXiv:2404.16771*, 2024. 11
[191] W. Yang, P. Luo, and L. Lin, "Clothing co-parsing by joint image segmentation and labeling," in *CVPR*, 2014, pp. 3182–3189. 11
[192] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, "Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments," *IEEE TPAMI*, vol. 36, no. 7, pp. 1325–1339, 2013. 11
[193] N. Zhang, M. Paluri, Y. Taigman, R. Fergus, and L. Bourdev, "Beyond frontal faces: Improving person recognition using multiple cues," in *CVPR*, 2015, pp. 4804–4813. 12, 15
[194] H. Hukkelås, R. Mester, and F. Lindseth, "Deepprivacy: A generative adversarial network for face anonymization," in *International Symposium on Visual Computing*. Springer, 2019, pp. 565–578. 12
[195] S. Yang, P. Luo, C.-C. Loy, and X. Tang, "Wider face: A face detection benchmark," in *CVPR*, 2016, pp. 5525–5533. 12
[196] P. Voigtlaender, M. Krause, A. Osep, J. Luiten, B. B. G. Sekar, A. Geiger, and B. Leibe, "Mots: Multi-object tracking and segmentation," in *CVPR*, 2019, pp. 7942–7951. 12
[197] Z. Zhang, P. Luo, C. C. Loy, and X. Tang, "From facial expression recognition to interpersonal relation prediction," *IJCV*, vol. 126, no. 5, pp. 550–569, 2018. 12
[198] C.-H. Weng, Y.-H. Lai, and S.-H. Lai, "Driver drowsiness detection via a hierarchical temporal deep belief network," in *ACCV*. Springer, 2016, pp. 117–133. 12
[199] S. A. Kumar, E. Yaghoubi, A. Das, B. Harish, and H. Proença, "The p-destre: A fully annotated dataset for pedestrian detection, tracking, and short/long-term re-identification from aerial devices," *TIFS*, vol. 16, pp. 1696–1708, 2020. 12
[200] L. Zheng, Z. Bie, Y. Sun, J. Wang, C. Su, S. Wang, and Q. Tian, "Mars: A video benchmark for large-scale person re-identification," in *ECCV*. Springer, 2016, pp. 868–884. 12
[201] B. Jiang, B. Bai, H. Lin, Y. Wang, Y. Guo, and L. Fang, "Dartblur: Privacy preservation with detection artifact suppression," in *CVPR*, 2023, pp. 16 479–16 488. 12
[202] V. Jain and E. Learned-Miller, "Fddb: A benchmark for face detection in unconstrained settings," UMass Amherst Technical Report, Tech. Rep., 2010. 12
[203] S. Shao, Z. Zhao, B. Li, T. Xiao, G. Yu, X. Zhang, and J. Sun, "Crowdhuman: A benchmark for detecting human in a crowd," *arXiv preprint arXiv:1805.00123*, pp. 1–9, 2018. 12
[204] S. Park, H. Na, and D. Choi, "Verifiable facial de-identification in video surveillance," *IEEE Access*, vol. 12, pp. 67 758–67 771, 2024. 12, 13
[205] Y. Zhang, Y. Fang, Y. Cao, and J. Wu, "Rbgan: Realistic-generation and balanced-utility gan for face de-identification," *IVC*, vol. 141, p. 104868, 2024. 12
[206] L. Laishram, J. T. Lee, and S. K. Jung, "Face de-identification using face caricature," *IEEE Access*, vol. 12, pp. 19 344–19 354, 2024. 12
[207] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *IEEE Workshop on Applications of Computer Vision*. IEEE, 1994, pp. 138–142. 12
[208] X. Zhang, L. Yin, J. F. Cohn, S. Canavan, M. Reale, A. Horowitz, and P. Liu, "A high-resolution spontaneous 3d dynamic facial expression database," in *FG Workshop*. IEEE, 2013, pp. 1–6. 12
[209] Y. Zeng, M. Zhang, and H. Xin, "Face-deid-net: Generative face de-identification with identity removal and attribute preservation for latent diffusion model training," in *ICSIP*. IEEE, 2025, pp. 489–495. 12
[210] Y. Nitzan, K. Aberman, Q. He, O. Liba, M. Yarom, Y. Gandelsman, I. Mosseri, Y. Pritch, and D. Cohen-Or, "Mystyle: A personalized generative prior," *TOG*, vol. 41, no. 6, pp. 1–10, 2022. 12
[211] H. Kim, J. Shim, S. Park, and E. Hwang, "Visual context-aware attribute-preserving face de-identification," *Neurocomputing*, vol. 638, p. 130205, 2025. 12
[212] C.-H. Lee, Z. Liu, L. Wu, and P. Luo, "Maskgan: Towards diverse and interactive facial image manipulation," in *CVPR*, 2020, pp. 5549–5558. 12
[213] G. Panis and A. Lanitis, "An overview of research activities in facial age estimation using the fg-net aging database," in *ECCV*. Springer, 2014, pp. 737–750. 12
[214] D. Deb, D. Aggarwal, and A. K. Jain, "Child face age-progression via deep feature aging," *arXiv preprint arXiv:2003.08788*, 2020. 12
[215] D. Deb, N. Nain, and A. K. Jain, "Longitudinal study of child face recognition," in *ICB*. IEEE, 2018, pp. 225–232. 12
[216] W. V. Friesen and P. Ekman, "Facial action coding system: A technique for the measurement of facial movement," *Palo Alto*, vol. 3, no. 2, p. 5, 1978. 13
[217] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE TIP*, vol. 13, no. 4, pp. 600–612, 2004. 13
[218] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," in *NIPS*, 2017, pp. 815–823. 13
[219] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *CVPR*, 2016, pp. 2818–2826. 13
[220] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *CVPR*, 2018, pp. 586–595. 13
[221] A. Sepas-Moghaddam and A. Etemad, "Deep gait recognition: A survey," *IEEE TPAMI*, vol. 45, no. 1, pp. 264–284, 2022. 15
[222] European Parliament and the Council of the European Union, "General data protection regulation," Official Journal of the European Union, pp. 1–88, May 2016, oJ L 119, 4.5.2016. [Online]. Available: http://data.europa.eu/eli/reg/2016/679/oj 15

Hui Wei received the PhD degree in computer science from Wuhan University, China. He is currently a postdoctoral researcher with the Center for Machine Vision and Signal Analysis, University of Oulu, Finland. He has authored or coauthored papers in mainstream conferences and journals, including CVPR, NeurIPS, AAAI, ACM Multimedia, and the IEEE TPAMI. His research interests include trustworthy AI, privacy protection, and machine learning.

Hao Yu is currently a PhD student at the Center for Machine Vision and Signal Analysis, University of Oulu, Finland. His research interests focus on visual representation learning, facial attribute analysis, and visual generative models.



Guoying Zhao (IEEE Fellow 2022) received the Ph.D. degree in computer science from the Chinese Academy of Sciences, Beijing, China, in 2005. She is currently an Academy Professor and full Professor (tenured in 2017) with University of Oulu, and a PI with ELLIS Institute Finland. She is a member of Academia Europaea, a member of Finnish Academy of Sciences and Letters, Fellow of IEEE, IAPR, ELLIS and AAIA. Her current research interests include image and video representation, facial-expression and micro-expression recognition, emotional gesture analysis, affective computing, and biometrics.