Visibility-Aware Diffusion-Based Face Anonymization for Real-World Deployment*

Mohamed Jaouad Lahgazi[1][0000−0001−6619−3712] and Jean-Philippe Tarel[1][0000−0002−9241−5347]

COSYS - PICS-L laboratory, Gustave Eiffel University, Marne-la-Vallée, France
{mohamed-jaouad.lahgazi,jean-philippe.tarel}@univ-eiffel.fr

Abstract. Privacy regulations such as GDPR mandate face anonymization in visual data, yet existing methods optimize for benchmark performance while ignoring deployment challenges. State-of-the-art approaches achieve strong results on curated datasets of pre-cropped, aligned faces but lack infrastructure for real-world images containing multiple people at varying scales, poses, and occlusions. More critically, these methods process all detected faces indiscriminately, wasting computation on low-quality detections and risking artifacts on barely-visible faces. We address these deployment gaps with a production-ready pipeline combining diffusion-based inpainting with intelligent face filtering and robust preprocessing. Our visibility scoring system determines which faces warrant anonymization using geometric and recognition cues, preventing wasted computation and artifacts. The end-to-end architecture handles extreme scale variations, multiple overlapping people, and diverse conditions through adaptive mask generation with multi-fallback strategies. While achieving competitive performance on standard benchmarks, our approach demonstrates superior generalization to real-world scenarios including group photos, full-body images, and challenging outdoor scenes. The source codes are publicly available at : github.com/MJLahgazi/vis-aware-diffusion-anonymization.

Keywords: Face anonymization · Diffusion inpainting · Privacy preservation · Visibility scoring · Real-world deployment.

1 Introduction

Visual surveillance and facial recognition create urgent need for effective face anonymization. Privacy regulations including GDPR [6] and CCPA [34] mandate protection of personally identifiable information in collected visual data, with facial images explicitly classified as biometric data requiring stringent safeguards [1]. Face anonymization must remove identity information while preserving image utility for downstream tasks such as demographic analysis, pose estimation, and behavior understanding, presenting fundamental challenges in

* Supported by the ATUVVU project, funded by the French Road Safety Delegation (*Délégation à la Sécurité Routière*) under the 2022 annual call for projects.

computer vision. Early methods relied on blurring, pixelation, and masking, which severely degrade image quality and do not necessarily guarantee complete anonymization [25]. Deep generative models enabled realistic anonymization where synthetic faces replace original identities while preserving attributes and context. GANs initially dominated with DeepPrivacy [13] and DeepPrivacy2 [12]. Recent diffusion methods including FAMS [19], Diff-Privacy [8], and CAMOUFLaGE [28] achieve state-of-the-art results through high quality synthesis and sophisticated conditioning, reflecting the broader shift from adversarial to diffusion generation in computer vision.

However, a critical gap exists between benchmark optimization and practical deployment. State-of-the-art methods achieve strong results on curated datasets but lack completeness for a real-world scenario. FAMS operates exclusively on pre-cropped 512×512 facial images, requiring practitioners to implement detection, masking, and compositing pipelines. DeepPrivacy2 addresses full scenes but requires three separate generators, while CAMOUFLaGE's full-image regeneration incurs significant computational cost. More critically, no existing method addresses the filtering question: *which faces warrant processing?* Without principled quality assessment, systems waste computation on low-confidence detections and risk artifacts on marginal faces. This gap between curated benchmarks and production environments, where images contain multiple people at varying scales, poses, and occlusions, has been documented across computer vision [11,36]. Yet face anonymization research has largely overlooked these deployment challenges.

We address these deployment gaps with a production-ready pipeline combining YOLO detection [30] and Stable Diffusion XL inpainting [31], demonstrating that general-purpose inpainting models achieve competitive performance to task-specific networks when properly integrated. Our key innovations target practical requirements: a visibility scoring system combining facial keypoints and recognition confidence to filter non-recognizable faces, and scale-aware preprocessing with adaptive mask generation enabling robust handling of multiple people and extreme scale variations.

Our main contributions are:

- **Visibility filtering framework:** A scoring system combining facial keypoints and detection confidence to determine which faces warrant anonymization, preventing wasted computation and artifacts on marginal detections.
- **Production ready pipeline:** A complete architecture integrating detection, segmentation, and diffusion inpainting with adaptive preprocessing and multi fallback masking to handle multiple people and extreme scale variations.
- **Demonstrating general purpose inpainting viability:** We show that properly integrated diffusion inpainting models achieve competitive performance to specialized anonymization networks, challenging claims that task specific training is necessary while enabling superior generalization to real-world scenarios.

2 Related Work

Traditional anonymization methods including blurring, pixelation, and masking severely degrade image quality and remain vulnerable to de-obfuscation attacks [25]. Generative Adversarial Networks enabled realistic synthesis where synthetic faces replace original identities while preserving attributes and context. DeepPrivacy [13] pioneered conditional face generation, introducing the Facial De-identification Dataset with diverse poses and occlusions. DeepPrivacy2 [12] extended this to full body scenarios through a cascade of three GAN networks (face, body, and full scene generators) trained jointly on 1.53 million images, though this multi-network architecture introduces significant training complexity. CIAGAN [24] enables controllable identity swapping from predefined sets.

Diffusion models have emerged as the dominant paradigm for high quality synthesis. FAMS [19] achieves state-of-the-art performance through dual ReferenceNet architecture built on Stable Diffusion v2.1, using only reconstruction loss to eliminate dependency on face recognition models during training. DiffPrivacy [8] unifies anonymization and visual identity hiding through Multi-Scale Image Inversion with different denoising strategies. CAMOUFLaGE [28] regenerates entire images via ControlNet [38] conditioning on depth, pose, and semantic information for stronger privacy guarantees at significant computational cost. Text-guided approaches [26,17] combine attribute extraction via vision-language models with BrushNet [15] inpainting.

An important question is whether general-purpose inpainting models can match networks trained specifically for anonymization. Stable Diffusion inpainting variants [31,33] extend latent diffusion models with additional masked-region channels, fine-tuned on inpainting tasks [31], while training-free methods [23,7] guide pretrained models through iterative or gradient-based denoising. Proponents of task-specific training argue that face-specific losses and conditioning are necessary for optimal identity removal [19], while LDFA [18] demonstrates that pretrained Stable Diffusion with proper integration achieves reasonable anonymization. This debate remains unresolved with limited empirical comparison between properly integrated inpainting pipelines and specialized networks.

While these methods optimize for benchmark performance on pre-detected, cropped, and aligned faces, practical deployment requires addressing overlooked challenges. Modern YOLO detectors [37,21] achieve strong multi-scale detection through Receptive Field Enhancement modules, but detection alone does not answer: *should a detected face be processed?* Distant, occluded, or low-resolution faces may be detected but lack sufficient information to warrant anonymization. Processing such faces wastes computation and risks artifacts.

Face Image Quality Assessment (FIQA) methods [9,35,27] filter images unsuitable for biometric matching but optimize for stricter algorithmic requirements than human identifiability demands. Anonymization requires a different criterion: *is this face clear enough for a human to identify the person?* This deployment gap, documented across computer vision [11,22], is particularly acute for face anonymization, where benchmark datasets like CelebA-HQ [16] feature pre-cropped, aligned images under controlled conditions rarely met in real-world

contexts [3]. Our work addresses these overlooked challenges by introducing visibility filtering and demonstrating that general-purpose diffusion inpainting, when properly integrated, can achieve competitive performance to specialized networks while providing the complete infrastructure needed for deployment.

3 Method

We present a production-ready face anonymization pipeline addressing these deployment challenges. Our approach integrates YOLO detection [30], visibility filtering, and Stable Diffusion XL inpainting [31] into a complete system for real-world scenarios with multiple people, extreme scale variations, and diverse image conditions. Figure 1 illustrates the pipeline architecture.

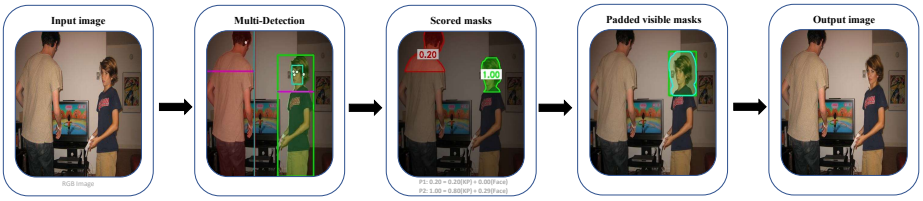


Fig. 1: Complete anonymization pipeline. Input images undergo person detection, face detection with visibility scoring, and filtering. high visibility faces are processed through scale-aware preprocessing, adaptive mask generation, SDXL inpainting, and seamless compositing.

3.1 Detection and visibility scoring

We detect people using YOLO, providing bounding boxes (bbox) for subsequent face and pose analysis. For each detected person, we perform face detection using DeepFace [32] with RetinaFace backend, obtaining face bounding boxes and confidence scores. Faces smaller than 256 pixels are upscaled to 512 pixels during detection to improve accuracy, with scale factors tracked for coordinate transformation. Simultaneously, we detect pose keypoints using YOLO pose, focusing on the five facial keypoints (nose, left eye, right eye, left ear, right ear) with confidence threshold 0.9 to identify visible landmarks. The visibility scoring system determines which detected faces require anonymization. Not all detected faces pose privacy risks: faces that are heavily occluded, too small, turned away from the camera, or poorly detected may not be identifiable by humans and can be safely filtered to reduce computation and avoid artifacts. Our visibility score V combines geometric visibility through keypoints with face detection confidence:

$$V = \min(\frac{k}{5} \cdot w_k + c_f \cdot w_c, 1) \tag{1}$$

where k is the count of visible facial keypoints, c_f is the DeepFace face confidence, $w_k = 1.0$ is the keypoint weight, and $w_c = 0.3$ is the confidence

weight. Faces with $V > 0.4$ are processed, others are filtered to avoid wasted computation and potential artifacts.

This formulation treats keypoint visibility as the primary indicator while validating recognizability through detection confidence. Profile views or distant faces have few visible keypoints regardless of resolution, providing robust visibility assessment independent of image quality. Figure 2 illustrates scoring behavior across different viewing conditions.

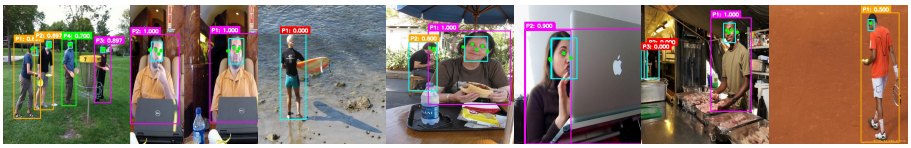


Fig. 2: Visibility scoring across occlusion levels. Examples show no visibility (score ≤ 0.2), medium visibility (score ≈ 0.3–0.6), and high visibility (score > 0.6), illustrating adaptive threshold filtering.

3.2 Adaptive mask generation

For faces passing the visibility threshold, we generate head masks through segmentation-based extraction with synthetic fallback for robustness. We apply YOLO seg with 10% padded crop around person bounding boxes (bbox) to ensure complete capture. All masks undergo morphological dilation with elliptical kernel (typically 10–15% of mask size) followed by Gaussian blur (kernel 21) for smooth boundaries. This provides inpainting context while maintaining natural transitions. Figure 3a shows the expansion strategy.

When segmentation succeeds, we separate head from body using adaptive cutting strategies based on available detection information. With high confidence face detection ($c_f \geq 0.4$), we extend 40% of face height below the face boundary. With lower confidence but sufficient keypoints ($k \geq 2$), we extend 50% from bbox top to the lowest keypoint. Otherwise, we cut at 30% from person bbox top. When segmentation fails, we generate elliptical masks using three fallback tiers. Reliable face detection enables head dimension estimation (assuming face occupies 60% of head height). Keypoint-based construction derives bbox from visible features with compensating multipliers. Otherwise, geometric placement positions the ellipse at 20% from bbox top with 70% bbox width and 30% bbox height. Figure 3b illustrates these strategies.

3.3 Inpainting and compositing

We inpaint tight crops around the expanded mask rather than full images, reducing computational cost and improving quality by focusing the diffusion model's attention. Each crop is processed independently through Stable Diffusion XL inpainting [31]. We use SDXL [29] for its quality at resolutions up to 1024×1024



(a) Adaptive mask expansion and cropping. Padded head masks (light blue) with crop regions (boxes) across scales and scenarios.

(b) Mask generation fallback rules. Left to right: segmentation mask, face detection ellipse, keypoint ellipse, and geometric fallback.

Fig. 3: Mask processing pipeline visualization.

and proven capabilities on diverse image types. The pipeline is independent of the generative model, thus allowing its substitution by an alternative diffusion model.

Prompts emphasize photorealistic qualities (natural skin texture, realistic features, authentic lighting) while suppressing artificial artifacts and smooth textures. We use guidance scale 7.5 and 60 inference steps. The inpainting strength parameter $(0.0 - 1.0)$ controls anonymization degree: lower values (≤ 0.4) produce subtle modifications preserving demographic attributes, while higher values (≥ 0.7) enable dramatic transformations. Even at maximum strength, the pipeline maintains visual consistency with surrounding context. Figure 4 demonstrates adjustable anonymization granularity (strength $0.3 - 0.9$), multi-scale processing (close-up portrait, mid-range upper body, distant full body), and multi-person handling.



Fig. 4: Core capabilities demonstration. Each row shows (left to right): original image and anonymization outputs at strength $0.3 - 0.9$ (increments of 0.1). Three scenarios: close up portrait (top), mid-range upper body (middle) and distant two person scene (bottom), demonstrating adjustable anonymization granularity, multi scale processing, and multi person handling.

The inpainted crop is composited back through alpha blending using a tighter mask than used for generation, both derived from the orginal segmentation mask with reduced padding for compositing to ensure precise blending. For multiple high visibility faces, we process each sequentially, maintaining independence be-

tween anonymization operations while handling overlapping or closely spaced faces.

4 Experiments

We evaluate our approach through qualitative and quantitative analysis on both benchmark and real-world data. We compare against DeepPrivacy2 [12] and FAMS [19] using official implementations with default parameters. Experiments use 2000 images: 1000 CelebA-HQ portraits [16] (aligned, high quality faces) and 1000 COCO scenes [20] (occlusions, scale variation, complex backgrounds). Our evaluation examines utility preservation, privacy protection, and perceptual quality.

4.1 Implementation details

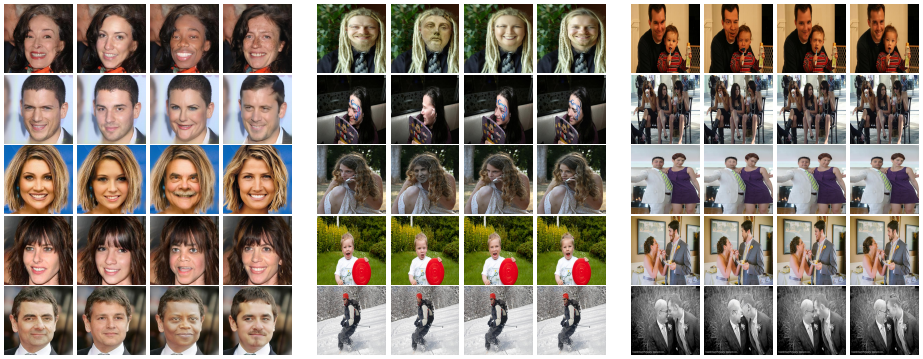
The pipeline is implemented in Python using PyTorch and the Diffusers library. We employ YOLO11n variants [14] for detection (person, segmentation, pose), DeepFace [32] with RetinaFace [4] backend for face detection, and SDXL inpainting [33]. Inference runs on an NVIDIA RTX 5000 Ada Generation with 16GB VRAM using float16 precision. The modular architecture allows component replacement without pipeline redesign. All thresholds and parameters were determined empirically.

4.2 Qualitative comparison

We evaluate visual quality across three scenarios: aligned portraits (CelebA-HQ), multi-scale scenes (COCO), and multi-person groups. For FAMS, we use the standard pipeline on aligned faces and the alignment-augmented pipeline with SFD face detector [40] for unaligned faces, as specified in the official repository. DeepPrivacy2 handles alignment internally through its hierarchical architecture.

Close up portrait anonymization On aligned CelebA-HQ portraits, all methods successfully anonymize identity while preserving image quality. FAMS occasionally produces dramatic transformations including gender or ethnicity changes, prioritizing strong anonymization over attribute preservation. DeepPrivacy2 maintains better demographic consistency. Our approach achieves competitive visual quality to specialized methods despite using general-purpose SDXL inpainting, as shown in Figure 5a.

Multi scale real-world scenarios COCO scenes containing people at varying distances expose limitations of methods trained on curated datasets. DeepPrivacy2's GAN architecture produces visible artifacts on small or distant faces, while FAMS generates inconsistent results across scales. Our pipeline maintains quality through scale-aware preprocessing: small faces are upscaled before inpainting, while adaptive cropping preserves context for larger faces. Figure 5b illustrates these differences.



(a) Close up portraits (b) Multi scale scenes (c) Multi person groups

Fig. 5: Qualitative comparison across three scenarios. Columns show (left to right): Original, DeepPrivacy2 [12], FAMS [19], and our method. (a) Close-up portraits (CelebA-HQ), (b) Multi-scale scenes (COCO), (c) Multi-person groups (COCO).

Group scenes with multiple people Group photographs with multiple people at various poses and occlusions require adaptive masking that automatically adjusts boundaries based on detection confidence and keypoint visibility for each individual. While specialized methods degrade on occluded or non-frontal faces, our approach maintains synthesis quality across these variations (Figure 5c).

4.3 Quantitative evaluation

Effective anonymization must balance utility preservation, identity protection, and visual realism. We evaluate performance under both ideal (aligned, high quality faces) and challenging (occlusions, scale variation, complex backgrounds) conditions. Our method uses inpainting strength 0.6, balancing utility and privacy, while baselines use default parameters. The following subsections examine utility (detection and attribute preservation), privacy (identity removal), and quality (perceptual metrics and sharpness).

Visibility filtering impact: Table 1 validates our visibility scoring system. On curated CelebA-HQ portraits, all faces exceed the visibility threshold. On real-world COCO scenes, 36.9% of detected faces are filtered as non-identifiable, demonstrating the practical value of our visibility-aware approach for deployment.

Table 1: Visibility filtering statistics.

Dataset	CelebA-HQ	COCO
Total faces	1000	2650
Filtered	0	978
Processed	1000	1672
Filter rate	0.0%	36.9%

Utility preservation Anonymized images must remain usable for downstream computer vision tasks. We measure utility through three hierarchical levels: person detection (YOLOv11), face detection (DeepFace with RetinaFace backend),

and demographic attribute classification (gender and age via DeepFace). Person detection validates that anonymization preserves overall scene structure, face detection confirms synthetic faces remain recognizable as faces, and attribute preservation ensures demographic information persists despite identity modification. Evaluation uses 1000 images from each dataset.

Table 2: Utility preservation across benchmark portraits (CelebA-HQ) and real-world scenes (COCO). All methods maintain comparable performance on aligned portraits, while specialized methods show degradation on complex scenes.

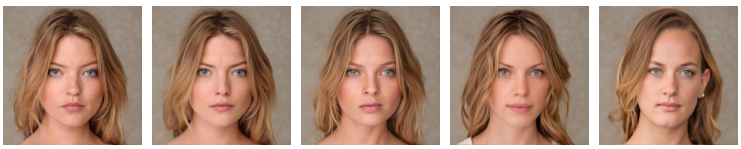
	Original		**DeepPrivacy2**		**FAMS**		**Ours**	
Metric	*CelebA*	*COCO*	*CelebA*	*COCO*	*CelebA*	*COCO*	*CelebA*	*COCO*
YOLO det rate (%) (↑)	100%	100%	**100%**	**100%**	**100%**	99.3%	**100%**	**100%**
YOLO conf. (↑)	0.81	0.84	0.81	0.84	**0.82**	0.83	**0.82**	**0.85**
DeepFace det rate(%) (↑)	100%	99.0%	**100%**	94.2%	**100%**	84.0%	**100%**	**98.0%**
Face Count Preserv. (↑)	100%	100%	**100%**	94.4%	**100%**	90.6%	**100%**	**99.7%**
Gender accuracy (%) (↑)	100%	100%	**94.6%**	74.8%	59.2%	72.8%	94.0%	**87.2%**
Age MAE (years) (↓)	0.0	0.0	5.3	4.8	7.7	**4.7**	**3.4**	**4.7**

Table 2 reveals a critical divergence between benchmark and real-world performance. On CelebA-HQ's aligned portraits, all methods maintain perfect detection rates with reasonable attribute preservation, reflecting the dataset's controlled conditions. However, COCO's complex scenes expose fundamental robustness gaps in specialized methods. Our approach achieves substantial improvements across all utility metrics: +4 to +14 percentage points on face detection (98.0% vs 84.0–94.2%), face count preservation (99.7% vs 90.6–94.4%), and gender accuracy (87.2% vs 72.8–74.8%). These gains stem from our visibility-aware filtering and adaptive preprocessing, which handle scale variations, occlusions, and pose diversity that cause specialized methods to produce undetectable or failed outputs. Notably, FAMS exhibits severe distribution mismatch on CelebA-HQ gender classification (59.2%), while our method maintains consistent performance across both datasets.

Privacy protection Identity protection is measured via ArcFace [5] cosine similarity between original and anonymized face embeddings, where lower similarity indicates stronger anonymization. We define strong anonymization as similarity ≤ 0.3 and moderate as ≤ 0.5, following established evaluation protocols.

Table 3 shows competitive anonymization across both datasets. On CelebA-HQ, all methods achieve strong identity removal with over 81% reaching strong anonymization thresholds. Our method's slightly higher similarity (0.217) reflects the moderate inpainting strength (0.6) balancing privacy and utility. On COCO, all methods exhibit substantially weaker performance, with mean similarities above 0.82 and only 11–14% achieving strong anonymization. This uniform degradation likely reflects known limitations of face recognition metrics on challenging conditions including low resolution, extreme poses, occlusions,

and poor lighting [2]. Such conditions degrade recognition model performance and potentially yield high similarity scores even when visual identity differs substantially. This underscores why utility preservation and qualitative assessment remain critical for real-world deployment evaluation. Moreover, the degraded imaging conditions that impair recognition metrics such as low resolution, extreme poses, and occlusions are precisely those where re-identification risk is inherently lowest in practice. Our architecture's parametrizable nature enables deployment-specific tuning: Figure 6 demonstrates how increasing inpainting strength progressively reduces similarity from 0.74 to 0.18, allowing users to navigate the privacy-quality tradeoff based on regulatory requirements.



(0.0, 1.00) (0.3, 0.74) (0.5, 0.51) (0.7, 0.303) (0.9, 0.18)

Fig. 6: Effect of inpainting strength on identity modification. Tuples show (strength, similarity). Higher strength produces greater visual change and lower similarity to the original face.

Perceptual quality Visual quality is assessed through perceptual similarity (LPIPS [39]), Fréchet Inception Distance (FID [10]), and relative sharpness (blur ratio normalized to 100 for the original image). LPIPS measures perceptual distance using deep features, FID quantifies distribution-level realism, and blur ratio measures sharpness degradation relative to the input. Lower values indicate better performance for all metrics.

Table 3: Identity removal performance via ArcFace similarity across face quality conditions. Lower mean similarity and higher percentages of low similarity pairs (< 0.3, < 0.5) indicate stronger anonymization.

Method	**Mean** (↓)	**Std**	**<0.3**(↑)	**<0.5**(↑)
CelebA-HQ				
DP2	**0.161**	0.159	**88.9%**	**97.6%**
FAMS	0.172	0.170	88.8%	96.0%
Ours	0.217	0.151	81.7%	94.1%
COCO				
DP2	**0.828**	0.308	**14.0%**	**19.6%**
FAMS	0.856	0.282	11.3%	15.1%
Ours	0.843	0.283	11.3%	17.8%

Table 4: Perceptual quality and blur metrics. LPIPS measures perceptual similarity, blur ratio indicates sharpness relative to baseline, FID measures realism. Lower values are better.

	LPIPS (↓)		**Blur (%)**		
Method	**Face**	**Whole**	**Face**	**Whole**	**FID**(↓)
CelebA-HQ					
DP2	**0.1691**	**0.1412**	**79.6**	**80.5**	**21.14**
FAMS	0.1903	0.1731	141.5	139.9	56.64
Ours	0.2124	0.1804	120.8	125.7	39.73
COCO					
DP2	0.0997	**0.0157**	78.5	96.8	**42.42**
FAMS	**0.0623**	0.0307	78.1	100.3	50.65
Ours	0.1076	0.0357	**63.1**	**92.2**	49.31

Table 4 reveals trade-offs between masking strategies. Baseline methods achieve lower LPIPS by restricting inpainting to tightly cropped central regions, minimizing pixel modifications. Our padded masks extend over original face boundaries for seamless context blending, necessarily modifying more pixels and incurring higher LPIPS. On CelebA, DeepPrivacy2 achieves the best overall metrics on controlled portraits. However, on COCO's challenging conditions, our method produces substantially sharper outputs (blur: 63.1 face, 92.2 whole vs 78+ for baselines), demonstrating that higher LPIPS reflects generous masking for coherence rather than quality loss. Our approach maintains superior sharpness on real-world scenes where baseline methods struggle with complex backgrounds, occlusions, and pose variations.

4.4 Flexibility of general purpose models

Building on general-purpose diffusion models provides two key deployment advantages. First, our modular design enables full body anonymization through a trivial modification of the cutoff parameter to keep the entire person mask (Figure 7). This adaptability allows the same pipeline to address different anonymization scenarios without retraining, a flexibility that specialized networks cannot match.



Fig. 7: Full body anonymization via person masks. Each pair shows input (left) and output (right), demonstrating seamless extension beyond face only processing.

Second, Stable Diffusion's inherent text conditioning enables fine-grained attribute control through natural language prompts. Figure 8 demonstrates controllable generation across ages, ethnicities, and emotions by simply modifying prompt text. This capability allows practitioners to tailor anonymization to specific requirements, preserving demographic distributions or generating particular emotional states, without model retraining.

5 Discussion and conclusion

We presented a production-ready face anonymization pipeline addressing the gap between benchmark optimization and practical deployment. Our approach

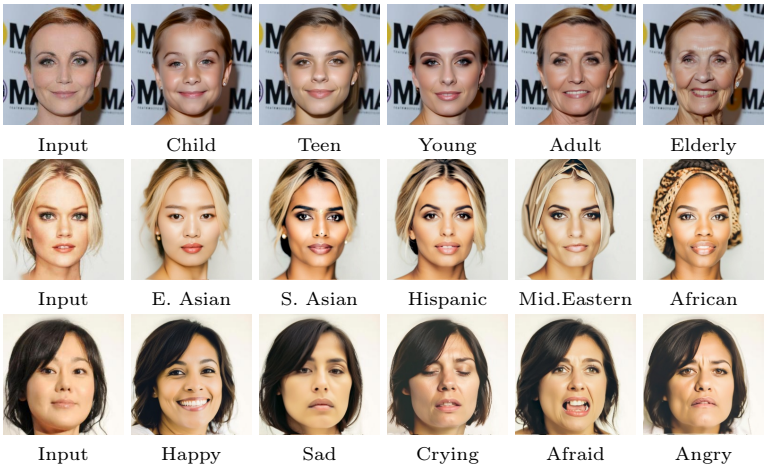


Fig. 8: Text guided attribute control demonstrating age variations (row 1), ethnic diversity (row 2), and emotional expressions (row 3) via prompt modification.

demonstrates that general-purpose diffusion inpainting, integrated with visibility filtering and adaptive preprocessing, achieves competitive benchmark performance while providing superior generalization to real-world scenarios.

The visibility scoring system successfully filters non-recognizable faces, preventing wasted computation and artifacts on marginal detections. On real-world COCO images, our method demonstrates substantial utility advantages over specialized approaches across face detection, count preservation, and gender accuracy (Table 2). These gains stem from visibility-aware filtering and adaptive preprocessing that handle scale variations, occlusions, and pose diversity. These utility improvements come with anonymization performance reflecting an intentional utility-privacy trade-off (Table 3) and superior output sharpness on challenging data (Table 4). Our padded masking strategy prioritizes seamless context blending over similarity metrics, accepting higher LPIPS as the cost of artifact-free integration. The parametrizable architecture enables users to tune the privacy-quality tradeoff based on deployment constraints, with higher inpainting strengths providing stronger anonymization when required (Figure 6). The general-purpose foundation further enables straightforward adaptation to diverse deployment scenarios without retraining, a flexibility specialized networks inherently lack. Computational efficiency is accordingly a function of the chosen backbone rather than the pipeline itself, enabling substitution of lighter generative models as deployment constraints demand.

Limitations include reduced privacy at moderate strengths and challenges on heavily degraded faces where all methods struggle. The reliance on face recognition metrics for privacy evaluation remains problematic given their known degradation on challenging conditions. Future work could explore lightweight generative models for improved efficiency, adaptive strength selection per face, video extension through temporal consistency by conditioning on face attributes

from previous frames of the same instance, and integration with multimodal anonymization techniques. Our modular architecture facilitates such extensions while maintaining deployment viability.

References

1. Almeida, D., Shmarko, K., Lomas, E.: The ethics of facial recognition technologies, surveillance, and accountability in an age of artificial intelligence: A comparative analysis of us, eu, and uk regulatory frameworks. AI and Ethics **2**(3), 377–387 (2021). https://doi.org/10.1007/s43681-021-00077-w
2. Burgos-Artizzu, X.P., Perona, P., Dollár, P.: Robust face landmark estimation under occlusion. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 1513–1520 (2013)
3. Cao, J., Chen, X., Liu, B., Ding, M., Xie, R., Song, L., Li, Z., Zhang, W.: Face de-identification: State-of-the-art methods and comparative studies. arXiv preprint arXiv:2411.09863 (2024)
4. Deng, J., Guo, J., Ververas, E., Kotsia, I., Zafeiriou, S.: Retinaface: Single-shot multi-level face localisation in the wild. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
5. Deng, J., Guo, J., Xue, N., Zafeiriou, S.: Arcface: Additive angular margin for deep face recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4690–4699. IEEE (2019)
6. European Parliament and Council: Regulation (eu) 2016/679 of the european parliament and of the council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (general data protection regulation). Official Journal of the European Union (2016), l 119/1
7. Grechka, A., Couairon, G., Cord, M.: Gradpaint: Gradient-guided inpainting with diffusion models. Computer Vision and Image Understanding **240**, 103928 (2024)
8. He, X., Zhu, M., Chen, D., Wang, N., Gao, X.: Diff-privacy: Diffusion-based face privacy protection. IEEE Transactions on Circuits and Systems for Video Technology (2024)
9. Hernandez-Ortega, J., Fierrez, J., Morales, A., Vera-Rodriguez, R.: Face image quality assessment: A literature survey. ACM Computing Surveys **55**(2), 1–41 (2022). https://doi.org/10.1145/3494793
10. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans are too slow to catch up. In: Advances in Neural Information Processing Systems (NIPS). vol. 30, pp. 548–557 (2017)
11. Hönel, S., Nordqvist, J.: From lab to factory: Pitfalls and guidelines for self-/unsupervised defect detection on low-quality industrial images. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. pp. 511–527. Springer (2025)
12. Hukkelås, H., Lindseth, F.: Deepprivacy2: Towards realistic full-body anonymization. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 1329–1338 (January 2023)
13. Hukkelås, H., Mester, R., Lindseth, F.: Deepprivacy: A generative adversarial network for face anonymization. In: Advances in Visual Computing: 14th International Symposium on Visual Computing (ISVC). pp. 565–578. Springer (2019)
14. Jocher, G., Qiu, J.: Ultralytics yolo11 (2024), https://github.com/ultralytics/ultralytics

15. Ju, X., Liu, X., Wang, X., Bian, Y., Shan, Y., Xu, Q.: Brushnet: A plug-and-play image inpainting model with decomposed dual-branch diffusion. In: European Conference on Computer Vision (ECCV). pp. 150–168. Springer (2024)
16. Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of gans for improved quality, stability, and variation. In: International Conference on Learning Representations (ICLR) (2018)
17. Kim, H., Shim, J., Park, S., Hwang, E.: Visual context-aware attribute-preserving face de-identification. Neurocomputing **638** (2025)
18. Klemp, M., et al.: Ldfa: Latent diffusion face anonymization for self-driving applications. arXiv preprint arXiv:2302.08931 (2023)
19. Kung, H.W., Varanka, T., Saha, S., Sim, T., Sebe, N.: Face anonymization made simple. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 1040–1050 (February 2025)
20. Lin, T.Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: European conference on computer vision (ECCV). pp. 740–755 (2014)
21. Liu, L., Wang, G., Miao, Q.: ADYOLOv5-Face: An enhanced YOLO-based face detector for small target faces. Electronics **13**(21), 4184 (2024). https://doi.org/10.3390/electronics13214184
22. Lucas, T.: Overcoming challenges of deploying computer vision at scale (2024)
23. Lugmayr, A., et al.: Repaint: Inpainting using denoising diffusion probabilistic models. In: CVPR. pp. 11451–11461 (2022)
24. Maximov, M., Elezi, I., Leal-Taixé, L.: Ciagan: Conditional identity anonymization generative adversarial networks. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2020)
25. McPherson, R., Shokri, R., Shmatikov, V.: Defeating image obfuscation with deep learning. In: arXiv preprint arXiv:1609.00408 (2016)
26. Muştu, M.İ.: Facial attribute based text guided face anonymization. arXiv preprint arXiv:2505.21002 (2025)
27. Ou, F.Z., et al.: Clib-fiqa: Face image quality assessment with confidence calibration. In: CVPR (2024)
28. Piano, L., Basci, P., Lamberti, F., Morra, L.: Latent diffusion models for attribute-preserving image anonymization. arXiv preprint arXiv:2403.14790 (2024)
29. Podell, D., English, Z., Lacey, K., Blattmann, A., Dockhorn, T., Müller, J., Penna, J., Rombach, R.: Sdxl: Improving latent diffusion models for high-resolution image synthesis. In: arXiv preprint arXiv:2307.01952 (2023)
30. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
31. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10684–10695 (2022)
32. Serengil, S.I., Ozpinar, A.: Lightface: A hybrid deep face recognition framework. In: Innovations in Intelligent Systems and Applications Conference (ASYU). pp. 23–27. IEEE (2020)
33. Stability AI: Stable diffusion xl 1.0 inpainting (2023)
34. State of California: California consumer privacy act. California Civil Code Section 1798.100 et seq. (2018), california statute on consumer data privacy rights
35. Terhörst, P., et al.: Ser-fiq: Unsupervised estimation of face image quality. In: CVPR. pp. 5650–5659 (2020)

36. Yao, S., Ardabili, B.R., Pazho, A.D., Noghre, G.A., Neff, C., Bourque, L., Tabkhi, H.: From lab to field: Real-world evaluation of an ai-driven smart video solution to enhance community safety. arXiv preprint arXiv:2312.02078 (2024)
37. Yu, Z., et al.: Yolo-facev2: A scale and occlusion aware face detector. Pattern Recognition **155**, 110714 (2024)
38. Zhang, L., Rao, A., Agrawala, M.: Adding conditional control to text-to-image diffusion models. In: IEEE/CVF International Conference on Computer Vision (ICCV) (2023)
39. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 586–595 (2018)
40. Zhang, S., Zhu, X., Lei, Z., Shi, H., Wang, X., Li, S.Z.: S^3fd: Single shot scale-invariant face detector. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 192–201 (2017)