DIFFCLEAN: Diffusion-based Makeup Removal for Accurate Age Estimation

Anonymous CVPR submission

Paper ID *****

Abstract

Accurate age verification can protect underage users from unauthorized access to online platforms and e-commerce sites that provide age-restricted services. However, accurate age estimation can be confounded by several factors, including facial makeup that can induce changes to alter perceived identity and age to fool both humans and machines. In this work, we propose DIFFCLEAN *which erases makeup traces using a text-guided diffusion model to defend against makeup attacks.* DIFFCLEAN *improves age estimation (minor vs. adult accuracy by 5.8%) and face verification (TMR by 5.1% at FMR=0.01%) compared to images with makeup. Our method is robust across digitally simulated and real-world makeup styles, and outperforms multiple baselines in terms of biometric and perceptual quality.*

1. Introduction

Motivation. Our constant need to stay digitally connected has given rise to a wide range of online services, including social media platforms, chat rooms, and dating websites. Several online services rely on *digital age verification*, typically, age verification from selfie-styled facial scans (*e.g.*, Yoti, Ondato, TrustStamp). But there have been reports of minors bypassing age verification checks exposing them to cyberbullying and online dangers. [1,2] In such scenarios, precise age estimation is critical in protecting underage users. Recently, government agencies and commercial content providers are actively pushing for online facial age verification for protecting teenagers (US's Identifying Minors Online, UK's Online Safety Act, YouTube) [6, 11] from unauthorized access to adult-themed websites.

Age estimation is intrinsically challenging due to complex genetic, environmental, demographic, and subjective factors; this is compounded when an individual puts on makeup. Facial makeup significantly alters the perception of facial features affecting reliable age prediction [13, 18, 46]. The perceived age in younger people with makeup is generally overestimated, while in older people the predicted age is underestimated [58, 61]. Makeup can be used to bypass automated age estimation [8], the primary issue addressed in this work, but we also look at its impact on face recognition [57]. Intuitively, we can design a makeup-invariant age estimator and face matcher, specifically designed for minors by using "more data". However, this immediately raises privacy concerns: we will need data on minors wearing makeup. The next strategy might be to generate synthetic makeup images for training new models. But makeup can vary in terms of intensity, texture, color, region of application, and subjective preferences. So, synthetic data may not capture the full diversity of facial makeup.

Our goal. We propose to defend against makeup attacks by designing a *makeup removal framework*, known as DIFFCLEAN. It will be deployed as a plug-in module that will filter the selfie-styled facial input image to erase makeup traces, which will primarily benefit accurate age estimation, and secondary tasks like identity verification.

Our approach. We formulate makeup removal as an image translation task, leveraging generative models known for seamless and controllable digital makeup transfer [9, 29, 34, 39, 49, 54, 63]. Unlike prior works that rely on attribute editing or reference-based adversarial transfer, our method DIFFCLEAN is a *reference-free*, text-guided diffusion model that transforms a face image *with* makeup to a face image *without* makeup using a combination of CLIP loss (for makeup detection), perceptual loss (for fidelity retention), biometric loss (for identity restoration), and age loss (for age restoration); see Fig. 1. Our method primarily enhances age estimation accuracy by removing makeup traces from face images, when present. It can also improve facial identity verification.

Contributions. Our main contributions are as follows:

1. We propose DIFFCLEAN, a novel *reference-free* text-guided diffusion-based makeup removal framework that digitally erases the traces of makeup present in face images, and improves age estimation (minor vs. adult accuracy by 5.8%) and face verification (TMR by 5.1% at FMR=0.01%) compared to images with makeup, while retaining visual quality (SSIM=0.98, PSNR=35.64).

[1] Children bypass age-checking measures on social media sites
[2] Underage teens bypassing age restrictions on dating apps.

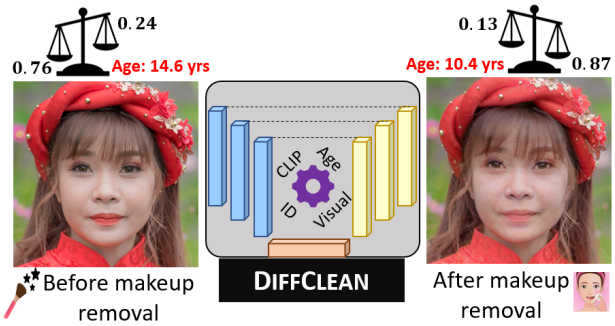


Figure 1. Overview of our method, DIFFCLEAN, that removes makeup on an example FFHQ image using text-guided diffusion model with a combination of CLIP loss, identity (ID) loss, Age loss and Perceptual (Visual) losses. Successful makeup removal results in shifting the softmax probabilities: $[p(\text{with makeup}), p(\text{without makeup})]$ from $[0.76, 0.24]$ to $[0.13, 0.87]$. Here, we use ViT-B/32 as CLIP-based classifier. Makeup removal reduces age overestimation from 14.6 to 10.4 yrs.

2. We utilize text-guided CLIP loss for detecting makeup, perceptual loss for minimizing visual artifacts, and a combination of age and identity losses for restoring the facial cues affected by makeup. The combined losses ensure that our model works on both makeup and non-makeup images without producing artifacts.
3. We empirically demonstrate that DIFFCLEAN works on both digitally simulated and real makeup images across variations in pose, illumination, facial accessories, and different makeup styles, while exhibiting demographic fairness. Our method can potentially assist with digital age verification to protect underage users from cyber harms. We will release the codes for reproducible research.

Responsible Usage. We recognize that our application focus is a sensitive topic, particularly since any modeling or algorithmic errors may adversely impact a vulnerable demographic (minors/teenagers). We strongly advocate for the ethical use of DIFFCLEAN *only* to assist with facial analytics (not for the purposes of malicious image editing).

2. Related Work

Facial Age Prediction. [32] used statistical face models generated by applying Principal Component Analysis (PCA) on a set of face images for age estimation. Texture features such as Local Binary Pattern (LBP) and its variants have been used for reliable age estimation [21, 50, 66]. Biologically-inspired features have been used in [22, 24]. Deep-learning -based approaches rely on convolutional neural networks (CNNs) and transformers for improved robustness. A multi-stream CNN was used to learn high-dimensional structured information from face patches in [65]. Deep CNN was used for age estimation in unconstrained settings in [33, 43, 60]. In [41], each age was treated as a rank to apply ordinal regression problem to age estimation using deep CNN. A multi-task learning approach for heterogeneous attribute estimation was adopted in [25]. Deep Expectation (DEX) [44] formulated age estimation by converting it into a classification-regression problem. SSRNet [64] adopted a compact soft stagewise regression model to perform multi-class classification. Auxiliary facial cues were used to boost reliable age estimation in [37, 67]. Deep random forests were used for age estimation in [51]. FP-age [36] used face parsing attention module to incorporate facial semantics into age estimation. Swin Transformer with attention-based convolution (ABC) for facial age estimation was proposed in [52]. In [31], a multi-input transformer-based model, MiVOLO, predicted age and gender using facial and full body information.

Age Estimation in Minors. Limited works have focused on improving age estimation performance in minors. In [5], a finetuned VGG-based 'children' network was used for apparent age estimation of children of age from 0 to 12 years. [4] trained ResNet50 on the underage facial dataset 'VisAGe' to improve the age estimation for minors. The need for fast, accurate, and robust age estimation and verification has gained traction recently. NIST Face Analysis Technology Evaluation (FATE) Age Estimation & Verification reports that subjects aged between 8 and 12 years tend to be overestimated to be 13 to 16 years old with false positives as high as 0.76 [3]; higher errors are reported for females than for males, specifically, in 17-year-olds [26]. The performance of facial age estimation algorithms can be affected by several factors like gender, ethnicity, image quality, facial expression, and makeup [3, 12].

Impact of Cosmetic Makeup on Facial Analytics. [57] demonstrated the impact of facial makeup on automatic face recognition. [18] observed that makeup applied to different facial regions had varying effects on age prediction quantified as Young Index (YI). [17] observed that the Identification Rate (IDR) decreased with increasing amount of makeup. [7, 23, 27] focused on making face recognition more robust to makeup. Authors in [8] observed that facial cosmetics (both real makeup and synthetically retouched) affect automated age estimation, causing a mean absolute error of up to 5.84 years. They further indicated that lip makeup had the least impact on automated age estimation, followed by eye makeup, followed by full face makeup (eye + lip + foundation). Makeup produces differential effects on apparent (manual) age estimation. 20-year-old women appeared older when wearing full face makeup, while 40- and 50-year-old women appeared younger when wearing makeup [12–14, 46]. Few works discuss the influence of makeup on facial age estimation in minors [3, 4].

Facial Makeup Transfer and Removal. Approaches for makeup transfer relied on GANs for style-transfer net-

[3] NIST FATE Report

works [9, 29, 34, 39, 56, 63]. Later, diffusion models were used [45, 55, 69]. Makeup removal, on the other hand, poses a more challenging task where the goal is to reveal the face under makeup without adversely affecting the biometric and perceptual features. Anti-makeup [35] integrated two adversarial networks, both at pixel and feature levels, to generate identity-preserving synthetic non-makeup faces. LADN [20] used local disentanglement with adversarial network for makeup transfer and removal. [68] replaced convolution layers in GAN with resnet blocks to design makeup removal network. PSGAN++ [38] used pose and expression robust spatial-aware GAN for detail-preserving makeup transfer and removal. CLIP2Protect [49] and DiffAM [54] use adversarial makeup transfer for privacy with capability to generate 'no makeup' faces. [45] used domain translation for a wide array of makeup tasks including makeup transfer and removal as a one-for-all diffusion-based framework; but suffers from lack of generalizability, and is limited to only training data and makeup styles. DeBeauty [59], is the *only* reference-free approach to the best of our knowledge that used adversarial de-makeup flow and relocation deformation flow to remove makeup.

3. Proposed Method

Our proposed method specifically targets mitigating the impact of makeup on automated facial age estimation and, additionally, facial identity verification. Defending against makeup attacks can prevent minors from bypassing age verification intended for protection against cyber abuse. One strategy may involve collecting data of minors wearing makeup for improving robustness of age estimators; this poses strong privacy concerns. So, we propose a generative model DIFFCLEAN, which is designed to mitigate age variations arising due to makeup. There are two challenges in designing this network: (1) we need a robust *makeup style generator* to synthesize data needed for training the makeup removal model and thus circumvent data collection; (2) we need a robust *age estimator* to serve as an auxiliary classifier to guide DIFFCLEAN to restore age-specific cues. For the *makeup style generator*, we use EleGANt [63] for locally-editable makeup transfer; see Fig. 2. For the *age estimator*, we use SSRNet [64] fine-tuned on UTKFace [70] dataset that has images from different age groups.

DIFFCLEAN uses the ablated diffusion model (ADM) with guided diffusion from [16] with improved U-Net architecture containing increased attention heads at multiple resolutions, BigGAN residual stacks for upsampling and downsampling, deeper networks and rescaled residual connections; refer to [16] for details. We used text prompts *face with makeup* and *face without makeup* for text-guided CLIP directional loss following [54] that detects if makeup is present or not. Makeup removal relies on the combined effect of CLIP-based makeup detection, perceptual utility

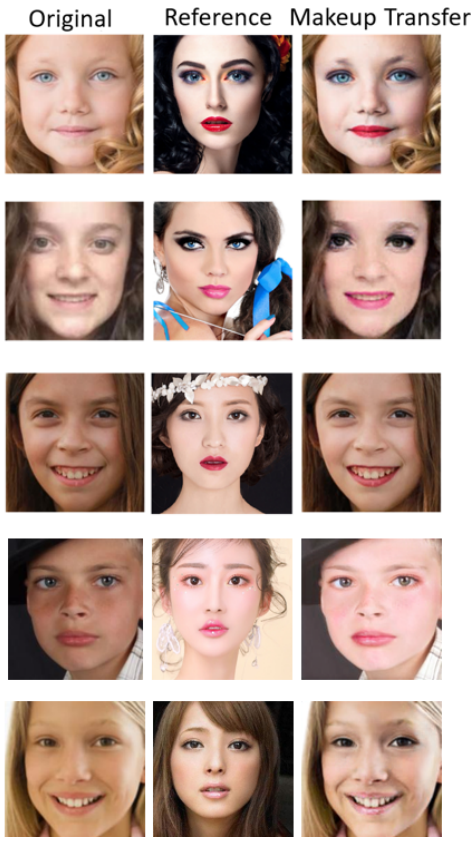


Figure 2. Synthetic makeup transfer results produced using EleGANt [63] on example images from UTKFace. Original image (first column), Makeup style reference (second column), and Makeup transferred image (third column).

retention, and restoration of identity and age-specific cues.

Methodology. We describe the terminology and notation here. Original/Clean image (without makeup), $\mathbf{I}_o$, Makeup image (with real/synthetic makeup), $\mathbf{I}_m$, and Makeup removed image, $\mathbf{I}_g$. Consider $\mathcal{M}$ as a real or synthetic makeup application operator, while $\mathcal{M}^{-1}$ represents the makeup removal operator. Thus, $\mathbf{I}_o \xrightarrow{\mathcal{M}} \mathbf{I}_m$, while $\mathbf{I}_m \xrightarrow{\mathcal{M}^{-1}} \mathbf{I}_g$. The goal is to model $\mathcal{M}^{-1}$ so that we can recover the clean image, $\mathbf{I}_g \approx \mathbf{I}_o$. Our proposed framework DIFFCLEAN approximates the role of $\mathcal{M}^{-1}$.

Architecture: The makeup removal framework ($\mathcal{M}^{-1}$) used in this work is a generative model which is a pre-trained ADM [16] that adds noise to the input image (with makeup), $\mathbf{I}_m$, to generate the noisy latent image during forward diffusion. The deterministic reverse diffusion through DDIM produces the makeup removed output, $\mathbf{I}_g$ supervised by CLIP, biometric identity, age and perceptual losses.

Loss functions: We present the loss functions used for supervising the age estimator and the diffusion model.

Age Estimator: SSRNet is fine-tuned using $\mathcal{L}_{WSL}$, a weighted adaptation of self-adjusting smoothed-$\mathcal{L}_1$ loss.

The adaptive weight ensures stricter penalty for misclassification in vulnerable age groups [10-29] years as follows:

$$\mathcal{L}_{WSL} = \begin{cases} 3.0 * \mathcal{L}_{SL}, \text{if age } \{10-29\} \text{ yrs.} \\ 1.0 * \mathcal{L}_{SL}, \text{otherwise.} \end{cases} \quad (1)$$

Refer to [19] for details about $\mathcal{L}_{SL}$. Note we ablated with different losses such as $\mathcal{L}_1$, $\mathcal{L}_2$ and Huber loss functions but empirically observed $\mathcal{L}_{WSL}$ loss to perform the best.

DIFFCLEAN: During *training*, we used four losses to fine-tune the diffusion model.

CLIP-based Makeup Loss: The transition from makeup domain to non-makeup domain is guided via textual prompts 'face without makeup' and 'face with makeup' in CLIP space using ViT-B/16 encoder following [54]. We define the loss as one minus the cosine similarity between the $\mathcal{L}_2$-normalized image and text embeddings as follows:

$$\mathcal{L}_{clip} = 1 - \frac{\Delta\mathbf{I}.\Delta\mathbf{T}}{\|\Delta\mathbf{I}\| \, \|\Delta\mathbf{T}\|} \quad (2)$$

where $\Delta\mathbf{I} = \mathcal{E}_\mathbf{I}(\mathbf{I}_g) - \mathcal{E}_\mathbf{I}(\mathbf{I}_m)$ and $\Delta\mathbf{T} = \mathcal{E}_\mathbf{T}(t_{\text{face w/o makeup}}) - \mathcal{E}_\mathbf{T}(t_{\text{face w/ makeup}})$. Here, $t_{\text{face w/o makeup}}$ is textual prompt after makeup removal and $t_{\text{face w/ makeup}}$ is textual prompt before makeup removal. Here, $\mathcal{E}_\mathbf{I}$ and $\mathcal{E}_\mathbf{T}$ are CLIP image and text encoders, respectively.

Biometric identity Loss: Our goal involves identity restoration of the makeup removed image with respect to the original image (without makeup). During training with synthetic makeup pairs, we have access to $(\mathbf{I}_o, \mathbf{I}_m)$ to generate $\mathbf{I}_g$. So, we compute weighted pairwise biometric identity losses as $\mathcal{L}_{og}$ and $\mathcal{L}_{mg}$. This is different than [54] that uses a single identity loss between input and output. The final identity loss is as follows:

$$\mathcal{L}_{id} = \lambda_1 \mathcal{L}_{og} + \lambda_2 \mathcal{L}_{mg} \quad (3)$$

where $\mathcal{L}_{.,.}$ represents the cosine distance between the extracted features using the ResNet ArcFace from the InsightFace library, with weights $\lambda_1 = 0.75$ and $\lambda_2 = 0.25$.

Perceptual Loss: To preserve the perceptual features, we use the perceptual loss $\mathcal{L}_{pips}$ with VGG [53] network and $\mathcal{L}_1$ loss between generated makeup-removed image $\mathbf{I}_g$ and input image with makeup $\mathbf{I}_m$.

Age Loss: The age loss is computed as the smoothed-$\mathcal{L}_1$ loss (introduced in RetinaMask [19]) between the the ground truth age, a_i of the original image, $\mathbf{I}_o$, and the predicted age, $\hat{a}_i$ of the makeup removed image, $\mathbf{I}_g$. We used the fine-tuned SSRNet model to predict $\hat{a}_i$. We used the age labels provided by [15] for FFHQ. We introduce the SSRNet age loss as follows:

$$\mathcal{L}_{\text{SSRNet-age}} = \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}_{SL}(a_i, \hat{a}_i) \quad (4)$$

We further tested with CLIP-based age loss in lieu of SSRNet, where we probed the CLIP model with the generated image, $\mathbf{I}_g$, using the prompt $\mathbf{t}_{age}$ = *face of* $\{a_i\}$*-year old*, where a_i indicates the ground truth age label as follows:

$$\mathcal{L}_{\text{CLIP-age}} = 1 - \frac{\mathbf{I}_g.\mathbf{t}_{age}}{\|\mathbf{I}_g\| \, \|\mathbf{t}_{age}\|} \quad (5)$$

Total loss for makeup removal: Combining all the above losses, the final loss for DIFFCLEAN is as follows:

$$\mathcal{L}_{total} = \lambda_{clip}\mathcal{L}_{clip} + \lambda_{id}\mathcal{L}_{id} + \lambda_{lpips}\mathcal{L}_{lpips} + \lambda_{L1}\mathcal{L}_{L1} + \lambda_{age}\mathcal{L}_{age} \quad (6)$$

where loss weights are tuned as $\lambda_{clip} = 5.0$, $\lambda_{id} = 1.0$, $\lambda_{lpips} = 5.0$, $\lambda_{L1} = 2.0$, and $\lambda_{age} = 0.5$ for SSRNet age loss and $\lambda_{age} = 5.0$ for CLIP age loss. Hyperparameters were selected using grid search; see sensitivity analysis in Sec. 5.6. Thus, we design two models: DIFFCLEAN with SSRNet age loss ($\mathcal{L}_{age} = \mathcal{L}_{\text{SSRNet-age}}$) and DIFFCLEAN with CLIP age loss ($\mathcal{L}_{age} = \mathcal{L}_{\text{CLIP-age}}$).

During *testing*, our model accepts face images captured under unconstrained conditions, either without makeup or with real or synthetic makeup as input.

4. Experiments and Analysis

4.1. Datasets

Training/Fine-tuning: We used the UTKFace [70] dataset to fine-tune SSRNet for age estimation with a training set of 15,364 images and a testing set of 6,701 images. The original set of 20K images was split into training and testing splits (70:30) after filtering out images with ages ≥70 yrs. We adopted a two-step approach for fine-tuning DIFFCLEAN for makeup removal. First, we used 300 images (200 training, 100 validation) from MT dataset [34] following the guidelines in [54]. We further refined the model by additionally fine-tuning it on 300 images (200 training, 100 validation) from the UTKFace dataset after applying makeup using EleGANt [63] with the combined loss. Thus, our training set D_{tr} consisted of a total of 600 images.

Evaluation: We tested our proposed DIFFCLEAN on both *synthetic* makeup (on FFHQ) and *real-world* makeup (LADN, BeautyFace, Makeup-Wild) images.

Synthetic makeup data generation: We sampled 2,556 images from the Flickr-Faces-HQ Dataset (FFHQ) [30] by sampling 284 images from each of the 9 age groups (excluding ≥ 70 yrs.) using the age bins provided by FFHQ-Aging [42]. Next, we applied EleGANt [63] that needs reference images for makeup transfer. We used five reference images [34] to simulate varied makeup styles, representing lip, face and eye makeup; see Fig. 2. We note that there can be other makeup styles but the representative reference styles selected for data creation mimic real-world samples.

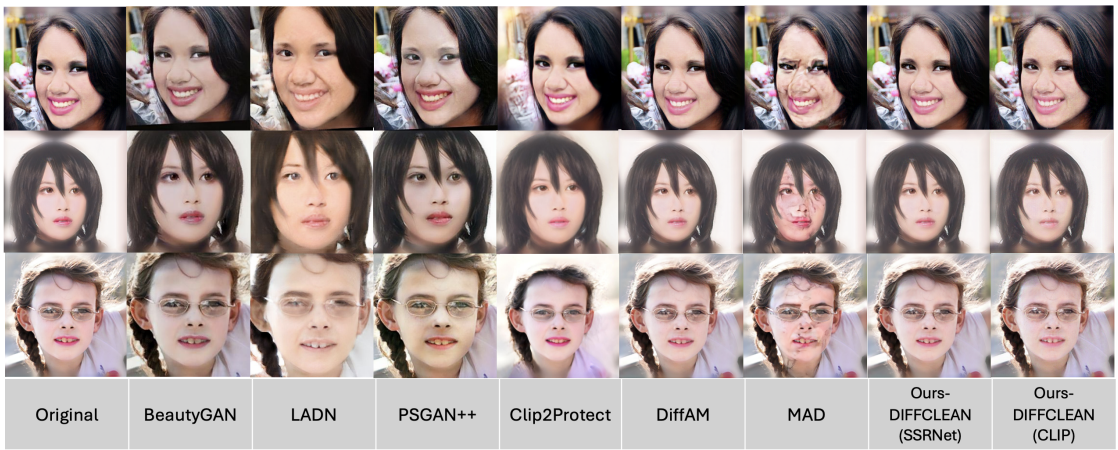


Figure 3. Comparison of makeup removal results generated by six baselines and our proposed DIFFCLEAN (last two columns) on three examples images from FFHQ dataset. GAN-based baselines (BeautyGAN, LADN, PSGAN++) introduce visual artifacts, while CLIP2Protect alters hair color and style, DiffAM does not effectively remove makeup, and MAD produces distortions on unseen data.

Real-world makeup data curation: We used 3,000 images from the BeautyFace [62] dataset, 355 images from the LADN dataset [20], and 384 images from the Makeup-Wild dataset [29] for testing generalizability of our method across variations in pose, illumination, accessories and occlusions.

4.2. Metrics

We evaluated DIFFCLEAN for (1) **age restoration** using Mean Absolute Error (MAE) between ground truth (if available) and predicted age values (*lower is better*); Age group accuracy to assess if the predicted age is correctly grouped into age bins (*higher is better*); Minor/Adult Accuracy to evaluate if the predicted age ≥ 18 to classify as adult, otherwise, minor (*higher is better*). We evaluated DIFFCLEAN for (2) **identity verification** by reporting the True Match Rate (TMR) at a False Match Rate (FMR) of 0.01% and illustrated the ROC curves. Finally, we computed (3) **image quality** of generated images using SSIM and PSNR.

4.3. Implementation Details

The SSRNet model was fine-tuned with batch size 50 with Adam optimizer with weight decay as 1e-4, cosine annealing scheduler and learning rate set to 1e-3 for 200 epochs with early stopping. The input image size is 64×64. The makeup removal module takes the makeup image, $\mathbf{I}_m$, of size 256×256 as input, and uses 40 DDIM inversion steps with 6 DDIM sampling steps and (total time steps=80), and noise scale for latent image generation is controlled by cosine beta scheduler [40]. We used the Adam optimizer with a learning rate of 4e-3. The model is fine-tuned for 5 epochs. All experiments were carried out on a single NVIDIA A100 GPU. Hyperparameters in Eqn. 6 were obtained using grid search on the validation set (see sensitivity analysis in Sec. 5.6). We perform age estimation using MiVOLO [31] (while training uses SSRNet) and MobileFace [10] and FaceNet [47] for identity verification (while training uses ArcFace IRSE50 [28]).

4.4. Baselines

We compared DIFFCLEAN with 6 baselines (4 GAN-based and 2 diffusion-based models) : BeautyGAN [34], LADN [20], PSGAN++ [38] CLIP2Protect [49], DiffAM [54], and MAD [45]. We randomly selected six non-makeup images as references for baselines following [59].

5. Results and Discussion

We present qualitative results for subjective evaluation (Figs. 3, 4, 6, and 7). We compare with the baselines in terms of age estimation (Tables 1 and 3), identity verification (Fig. 5) and visual quality (Table 2).

5.1. Age estimation

We report the results of the predicted age estimated by MiVOLO after makeup removal on FFHQ dataset in Table 1. We compute the Original performance by comparing the clean FFHQ face images (without makeup) with images simulated using EleGANt (with makeup), and compute the minor vs adult accuracy (%), age group accuracy (%), and mean absolute error (MAE). In Table 1, our method DIFFCLEAN outperforms all baselines and can successfully restore age cues disrupted due to makeup by lowering the MAE (by 0.85) and improving the minor/adult prediction accuracy (by 5.1%). Our objective is focused on minor age groups and the results indicate that we achieve a superior minor vs. adult age classification of 88.6%.

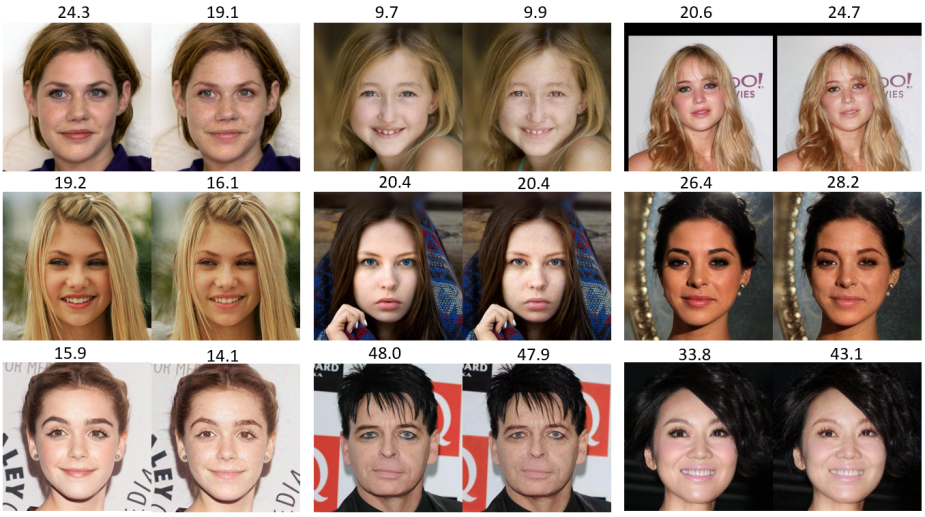


Figure 4. Examples of differential effects of age estimation before and after makeup removal. $\mathbf{1}^{st}$ **column:** DIFFCLEAN causes *lower* predicted ages than with makeup, thus reducing overestimation errors; $\mathbf{2}^{nd}$ **column:** DIFFCLEAN *preserves* original age when there is minimal or no makeup; $\mathbf{3}^{rd}$ **column:** DIFFCLEAN causes *higher* predicted ages than with makeup, thus reducing underestimation errors.

Makeup exhibits a differential effect on facial age. It may result in a decrease or increase in apparent age and, in some cases, may not affect perceived age. **Our method does not arbitrarily reduce the predicted age in all cases**. Instead, it strategically erases makeup traces to restore the original age-specific cues. Fig. 4 clearly depicts the differential age restoration to minimize overestimation and underestimation errors while retaining the original age in presence of light to no makeup.

5.2. Identity verification

We present the results of biometric identity verification in Fig. 5 using FaceNet and MobileFace face recognition (FR) models at FMR=0.01%. We compute the Original performance by comparing the clean FFHQ face images (without makeup) with images simulated using EleGANt (with makeup). Baseline performances are computed by comparing the original FFHQ face images with outputs of respective baselines (makeup removed images). We observe that makeup removal with our approach, improves the True Match Rate (TMR) by 1.3% with FaceNet and 5.1% with MobileFace compared to 'Original' performance. MAD does not perform well on unseen data (trained on MT; tested on FFHQ). This manifests as visual artifacts that lowers overall biometric performance.

5.3. Image quality evaluation.

We compared the images before and after makeup removal in terms of image quality metrics: SSIM and PSNR and reported our findings in Table 2. We compute the Original performance by comparing the clean FFHQ face images (without makeup) with images simulated using EleGANt (with makeup). Our algorithm is at par with DiffAM in terms of SSIM=0.98 while being second best in terms of PSNR=35.64, showing high retention of visual fidelity.

5.4. Real-world makeup generalizability

Real-world use cases, such as, online age verification *only* include a facial image with or without makeup, it does not include the individual's actual age label. So, in this experiment, we applied DIFFCLEAN to real-world makeup images from the **BeautyFace** [62], **LADN** [20], and **MakeupWild** [29] datasets, and used MiVOLO to predict age before and after makeup removal. See examples of real-world makeup removal in Figs. 6 and 7. Since we do not have ground truth age labels, instead, we report both overestimation and underestimation error statistics in Table 3. Further, we analyzed the confidence interval (CI) using a statistically significant sample size of 3K images from BeautyFace using the student's t-distribution at a 95% confidence level. We obtain CI = [2.88, 3.09] with margin of error = 0.10 on the underestimation errors, and CI = [2.75, 2.89] with margin of error of 0.07 on the overestimation errors. Results indicate that DIFFCLEAN generalizes across various real-world makeup styles, pose and illumination variations, and can be readily used in practical settings.

5.5. Ablation study

Training Dataset Diversity. We fine-tuned on 600 images (300 from MT dataset with smoky-eyes, flashy, retro, Korean and Japanese makeup styles + 300 from UTKFace with EleGANt-based makeup transfer with five reference styles and adaptive interpolated shading). To further assess training data diversity, we augmented our training set, D_{tr}, and

Table 1. Results of age estimation on FFHQ in terms of minor vs adult accuracy (%) ↑, age group accuracy (%) ↑ and mean absolute error (MAE) ↓. **Bolded** values indicate best results while underlined values indicate second-best results.

Metrics	Original	BeautyGAN	LADN	PSGAN++	Clip2Protect	DiffAM	MAD	Ours-DIFFCLEAN (SSRNet)	Ours-DIFFCLEAN (CLIP)
Minor/Adult Acc (%)	82.8	86.12	84.34	82.7	83.8	85.5	72.53	87.8	**88.6**
Age group Acc (%)	34.4	35.57	30.9	32.54	33.3	34.5	28.03	36.8	**37.0**
MAE	6.56	6.09	6.98	6.95	6.54	6.32	9.51	5.76	**5.71**

Table 2. Results of makeup removal in terms of image quality metrics (SSIM↑ and PSNR↑) on FFHQ. **Bolded** values indicate best results while underlined indicates second-best results.

	SSIM	PSNR
Original	0.58	14.94
BeautyGAN	0.89	21.48
LADN	0.43	13.53
PSGAN++	0.85	20.65
CLIP2Protect	0.46	15.74
DiffAM	**0.98**	**39.52**
MAD	0.92	26.46
Ours-DIFFCLEAN (SSRNet)	**0.98**	35.64
Ours-DIFFCLEAN (CLIP)	**0.98**	35.29

Table 3. Statistics of predicted ages on the real-world makeup images from the BeautyFace, LADN and Makeup-Wild datasets before and after makeup removal, and associated overestimation and under estimation errors in predicted age.

Before makeup removal $\mu \pm \sigma$		**After makeup removal** $\mu \pm \sigma$	Overestimation $\mu \pm \sigma$	Underestimation $\mu \pm \sigma$
22.38 ± 3.36 [BeautyFace]	DIFFCLEAN (SSRNet)	21.28 ± 4.58	2.99 ± 2.92	2.79 ± 1.96
	DIFFCLEAN (CLIP)	21.34 ± 4.66	3.14 ± 2.97	2.82 ± 1.95
24.64 ± 3.88 [LADN]	DIFFCLEAN (SSRNet)	22.59 ± 4.47	1.60 ± 1.11	3.02 ± 1.90
	DIFFCLEAN (CLIP)	22.55 ± 4.51	1.62 ± 1.11	3.11 ± 1.97
25.56 ± 4.12 [Makeup-Wild]	DIFFCLEAN (SSRNet)	24.16 ± 4.38	2.45 ± 2.31	3.34 ± 3.06
	DIFFCLEAN (CLIP)	24.25 ± 4.40	2.98 ± 2.80	3.20 ± 3.07

report MAE on validation set, as follows: 4.42 ($|D_{tr}| = 600$), 4.55 ($|D_{tr}| = 1,200$) and 4.68 ($|D_{tr}| = 1,600$). We observe consistent MAE across varying training data scale.

Influence of each loss component. We report MAE (↓) on the validation set by fine-tuning separate models with loss components added incrementally, as follows: 9.28 ($\mathcal{L}_{clip}$), 7.43 ($\mathcal{L}_{clip} + \mathcal{L}_{id}$), 4.70 ($\mathcal{L}_{clip} + \mathcal{L}_{id} + \mathcal{L}_{pips}$), and **4.42** ($\mathcal{L}_{clip} + \mathcal{L}_{id} + \mathcal{L}_{pips} + \mathcal{L}_{SSRNet-age}$). Evidently, the combined losses improve the overall performance.

Demographic fairness analysis. Table 4 reports demographic fairness analysis on FFHQ using DeepFace library [48], indicating DIFFCLEAN does not exhibit demographic disparity. FFHQ has more White images than Black and Indian images, but a balanced gender distribution.

Table 4. Demographic fairness analysis on FFHQ using our DIFFCLEAN SSRNet/CLIP-age loss models in terms of MAE. Values in parentheses indicate count after being processed by DeepFace.

Race	White (1,420)	Black (23)	Asian (507)	Hispanic (254)	Indian (23)	Middle Eastern (129)
MAE	5.9/5.9	4.2/4.3	5.3/5.3	5.5/5.4	6.9/6.8	5.3/5.2
Gender	Male (1,188)			Female (1,168)		
MAE	5.6/5.5			5.9/5.8		

5.6. Practical Considerations and Further Analysis.

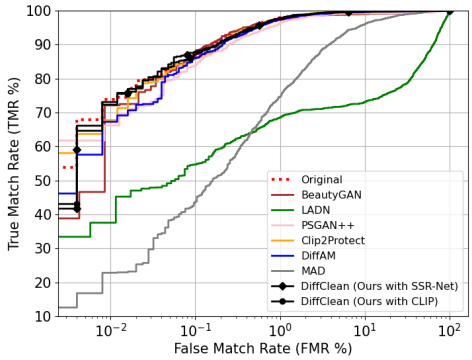
Deployment scenario. For real-world deployment, the model needs to accept an unconstrained facial image, and remove makeup efficiently and effectively for facial analytics. DIFFCLEAN can handle pose, lighting variations, etc., and is faster (needs 2 secs.) than CLIP2Protect (30 secs.), MAD (3 mins.), and comparable with PSGAN++ (2.35 secs.) and DiffAM (2 secs.).

Hyperparameter sensitivity. We used *optuna* [2], an automated hyperparameter optimization framework to perform multivariate sensitivity analysis to observe the interaction between the hyperparameters in Eqn. 6. Hyperparameter importance plot after 30 trials shows that λ_{age} has the highest importance=0.89 in the objective function; see Fig. 8.

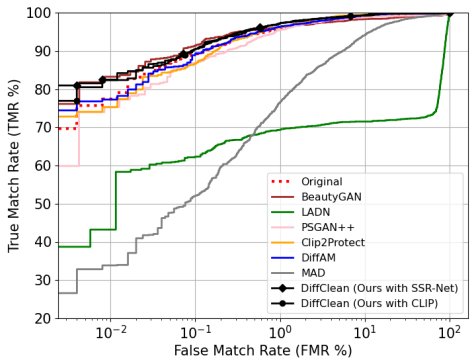
Effect of hallucinations. We computed the hallucination metric, Hal(x), proposed in [1] to detect hallucinations in DIFFCLEAN generated makeup-removed images. We used 200 images from FFHQ test set over 80 timesteps, and computed: $\text{Hal}(x) = \frac{1}{|T_2 - T_1|} \sum_{i=T_1}^{T_2} \left(\hat{x}_0^{(i)} - \overline{\hat{x}_0^{(t)}} \right)^2$ where $x_0^{(t)}$ represents the predicted values of the DIFFCLEAN outputs at (t) time steps, and $\hat{x}_0^{(t)}$ is the mean predicted value from $T_1 = 0$ to $T_2 = 80$. Fig. 9 shows overlapping histogram plots of generated samples both for in-distribution data, *i.e.*, no-makeup input (blue bars), and makeup-removed data, *i.e.*, with-makeup input (orange bars), with mean values of Hal(x) =0.04 and 0.06, respectively. Refer to Supplementary Materials for additional analysis.

6. Conclusion

In this work, we investigate the inadequacies of automated facial age estimators in presence of makeup. Makeup can be used to bypass age verification exposing minors to cyber



FaceNet



MobileFace

Methods	**FR model**	
	FaceNet	MobileFace
Original	74.5	79.1
BeautyGAN	72.1 (↓)	<u>83.2</u> (↑)
LADN	45.2 (↓)	58.3 (↓)
PSGAN++	66.2 (↓)	75.3 (↓)
CLIP2Protect	67.8 (↓)	75.3 (↓)
DiffAM	69.1 (↓)	78.2 (↓)
MAD	22.8 (↓)	33.9 (↓)
Ours-DIFFCLEAN (SSRNet)	**75.8** (↑)	**84.2** (↑)
Ours-DIFFCLEAN (CLIP)	<u>75.5</u> (↑)	82.4 (↑)

Figure 5. (Top): ROC curve with FaceNet. (Middle): ROC curve with MobileFace. (Bottom): Biometric matching in terms of TMR (%) @FMR = 0.01% (*higher is better*) with FaceNet and MobileFace matchers on FFHQ. **Bolded** values indicate best results while <u>underlined</u> values indicate second-best results.

abuse. To address these issues, we designed DIFFCLEAN, a *reference-free* diffusion-based makeup removal framework that can be deployed as a ready-to-plugin pre-processing module in a real-world application of online age verification for protecting minors. We combined text-guided CLIP loss with identity preservation and visual quality retention for makeup removal while restoring age-relevant features. Evaluations show DIFFCLEAN outperforms competing baselines in terms of age estimation (by 5.8%) and identity verification (by 5.1%) while exhibiting robustness across real-world makeup styles.

Figure 6. Results of DIFFCLEAN on real-world makeup images from LADN [20] (left), BeautyFace [62] (center), and Makeup-Wild [29] (right) datasets.

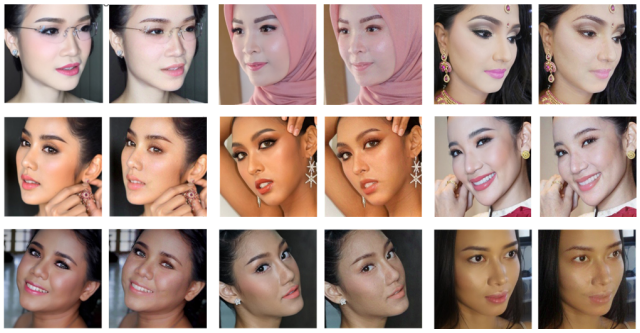


Figure 7. Examples from Makeup-Wild dataset with variations in pose, illumination, background and facial accessories. DIFFCLEAN preserves details such as fingers, jewelry, and expressions.

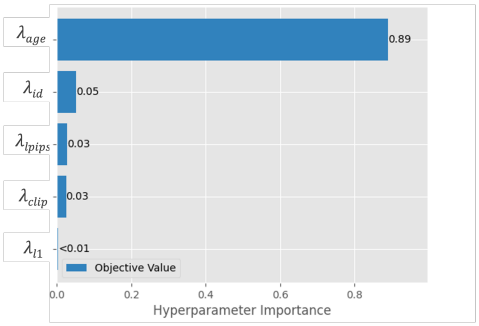


Figure 8. Importance of hyperparameters used in Eqn. 6.

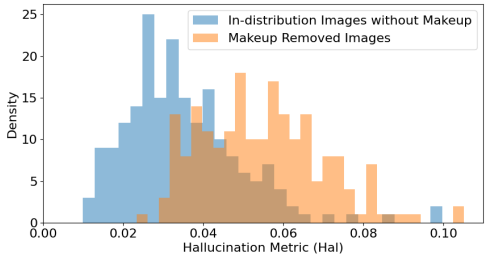


Figure 9. Histogram plot of hallucination metric [1] for DIFFCLEAN generated images.

References

[1] Sumukh K Aithal, Pratyush Maini, Zachary C. Lipton, and J. Zico Kolter. Understanding hallucinations in diffusion models through mode interpolation. In *Advances in Neural Information Processing Systems*, pages 134614–134644. Curran Associates, Inc., 2024. 7, 8

[2] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework. In *The 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2623–2631, 2019. 7

[3] Felix Anda, Brett A Becker, David Lillis, Nhien-An Le-Khac, and Mark Scanlon. Assessing the influencing factors on the accuracy of underage facial age estimation. In *International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*, pages 1–8. IEEE, 2020. 2

[4] Felix Anda, Nhien-An Le-Khac, and Mark Scanlon. Deepuage: improving underage age estimation accuracy to aid csem investigation. *Forensic Science International: Digital Investigation*, 32:300921, 2020. 2

[5] Grigory Antipov, Moez Baccouche, Sid-Ahmed Berrani, and Jean-Luc Dugelay. Apparent age estimation from face images combining general and children-specialized deep learning models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 96–104, 2016. 2

[6] James Beser. Extending our built-in protections to more teens on youtube. Published on: July 29, 2025 https://blog.youtube/news-and-events/extending-our-built-in-protections-to-more-teens-on-youtube/. [Online accessed: August 29, 2025]. 1

[7] Cunjian Chen, Antitza Dantcheva, and Arun Ross. Automatic facial makeup detection with application in face recognition. In *International Conference on Biometrics (ICB)*, pages 1–8. IEEE, 2013. 2

[8] Cunjian Chen, Antitza Dantcheva, and Arun Ross. Impact of facial cosmetics on automatic gender and age estimation algorithms. *International Conference on Computer Vision Theory and Applications (VISAPP)*, 2:182–190, 2014. 1, 2

[9] Hung-Jen Chen, Ka-Ming Hui, Szu-Yu Wang, Li-Wu Tsao, Hong-Han Shuai, and Wen-Huang Cheng. Beautyglow: On-demand makeup transfer framework with reversible generative network. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, pages 10042–10050, 2019. 1, 3

[10] Sheng Chen, Yang Liu, Xiang Gao, and Zhen Han. Mobilefacenets: Efficient cnns for accurate real-time face verification on mobile devices. In *Chinese conference on biometric recognition*, pages 428–438. Springer, 2018. 5

[11] Clare Y Cho. Identifying Minors Online. *CRS Report R47884 - US Library of Congress*, June 10, 2024. 1

[12] Albert Clapés, Ozan Bilici, Dariia Temirova, Egils Avots, Gholamreza Anbarjafari, and Sergio Escalera. From apparent to real age: gender, age, ethnic, makeup, and expression bias analysis in real age estimation. In *proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 2373–2382, 2018. 2

[13] Hannah Davis and Janice Attard-Johnson. Your id, please? the effect of facemasks and makeup on perceptions of age of young adult female faces. *Applied Cognitive Psychology*, 36 (2):453–459, 2022. 1

[14] Steven H Dayan, Katherine Cho, Mary Siracusa, and Selika Gutierrez-Borst. Quantifying the impact cosmetic makeup has on age perception and the first impression projected. *Journal of drugs in dermatology : JDD*, 14 4:366–74, 2015. 2

[15] DCGM. Gender, age, and emotions for flickr-faces-hq dataset (ffhq). https://github.com/DCGM/ffhq-features-dataset. [Online accessed: 21 st February, 2025]. 4

[16] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat GANs on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021. 3

[17] Marie-Lena Eckert, Neslihan Kose, and Jean-Luc Dugelay. Facial cosmetics database and impact analysis on automatic face recognition. In *IEEE 15th international workshop on multimedia signal processing (MMSP)*, pages 434–439. IEEE, 2013. 2

[18] Ranran Feng and Balakrishnan Prabhakaran. Quantifying the makeup effect in female faces and its applications for age estimation. In *IEEE International Symposium on Multimedia*, pages 108–115. IEEE, 2012. 1, 2

[19] Cheng-Yang Fu, Mykhailo Shvets, and Alexander C Berg. Retinamask: Learning to predict masks improves state-of-the-art single-shot detection for free. *arXiv preprint arXiv:1901.03353*, 2019. 4

[20] Qiao Gu, Guanzhi Wang, Mang Tik Chiu, Yu-Wing Tai, and Chi-Keung Tang. LADN: Local adversarial disentangling network for facial makeup and de-makeup. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 10481–10490, 2019. 3, 5, 6, 8

[21] Asuman Gunay and Vasif V. Nabiyev. Automatic age classification with lbp. In *23rd International Symposium on Computer and Information Sciences*, pages 1–4, 2008. 2

[22] Guodong Guo, Guowang Mu, Yun Fu, and Thomas S. Huang. Human age estimation using bio-inspired features. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 112–119, 2009. 2

[23] Guodong Guo, Lingyun Wen, and Shuicheng Yan. Face authentication with makeup changes. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(5):814–825, 2013. 2

[24] Hu Han, Charles Otto, and Anil K. Jain. Age estimation from face images: Human vs. machine performance. In *International Conference on Biometrics (ICB)*, pages 1–8, 2013. 2

[25] Hu Han, Anil K Jain, Fang Wang, Shiguang Shan, and Xilin Chen. Heterogeneous face attribute estimation: A deep multi-task learning approach. *IEEE transactions on pattern analysis and machine intelligence*, 40(11):2597–2609, 2017. 2

[26] Kayee Hanaoka, Mei Ngan, Joyce Yang, George W. Quinn, Austin Hom, and Patrick Grother. Face Analysis Technology Evaluation (FATE) Age Estimation & Verification. *NIST Interagency Report NIST IR 8525* https://nvlpubs.

nist.gov/nistpubs/ir/2014/NIST.IR.7995.pdf, 2014. 2

[27] Junlin Hu, Yongxin Ge, Jiwen Lu, and Xin Feng. Makeup-robust face verification. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2342–2346. IEEE, 2013. 2

[28] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018. 5

[29] Wentao Jiang, Si Liu, Chen Gao, Jie Cao, Ran He, Jiashi Feng, and Shuicheng Yan. PSGAN: Pose and expression robust spatial-aware GAN for customizable makeup transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5194–5202, 2020. 1, 3, 5, 6, 8

[30] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019. 4

[31] Maksim Kuprashevich and Irina Tolstykh. Mivolo: Multi-input transformer for age and gender estimation. In *International Conference on Analysis of Images, Social Networks and Texts*, pages 212–226. Springer, 2023. 2, 5

[32] Andreas Lanitis, Chrisina Draganova, and Chris Christodoulou. Comparing different classifiers for automatic age estimation. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 34(1):621–628, 2004. 2

[33] Gil Levi and Tal Hassner. Age and gender classification using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 34–42, 2015. 2

[34] Tingting Li, Ruihe Qian, Chao Dong, Si Liu, Qiong Yan, Wenwu Zhu, and Liang Lin. Beautygan: Instance-level facial makeup transfer with deep generative adversarial network. In *Proceedings of the 26th ACM international conference on Multimedia*, pages 645–653, 2018. 1, 3, 4, 5

[35] Yi Li, Lingxiao Song, Xiang Wu, Ran He, and Tieniu Tan. Anti-makeup: Learning a bi-level adversarial network for makeup-invariant face verification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018. 3

[36] Yiming Lin, Jie Shen, Yujiang Wang, and Maja Pantic. FP-age: Leveraging face parsing attention for facial age estimation in the wild. *IEEE Transactions on Image Processing*, 2022. 2

[37] Na Liu, Fan Zhang, and Fuqing Duan. Facial age estimation using a multi-task network combining classification and regression. *IEEE Access*, 8:92441–92451, 2020. 2

[38] Si Liu, Wentao Jiang, Chen Gao, Ran He, Jiashi Feng, Bo Li, and Shuicheng Yan. PSGAN++: Robust Detail-Preserving Makeup Transfer and Removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44: 8538–8551, 2021. 3, 5

[39] Thao Nguyen, Anh Tuan Tran, and Minh Hoai. Lipstick ain't enough: beyond color matching for in-the-wild makeup transfer. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, pages 13305–13314, 2021. 1, 3

[40] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International conference on machine learning*, pages 8162–8171. PMLR, 2021. 5

[41] Zhenxing Niu, Mo Zhou, Le Wang, Xinbo Gao, and Gang Hua. Ordinal regression with multiple output cnn for age estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4920–4928, 2016. 2

[42] Roy Or-El, Soumyadip Sengupta, Ohad Fried, Eli Shechtman, and Ira Kemelmacher-Shlizerman. Lifespan age transformation synthesis. In *16th European Conference on computer Vision (ECCV)*, page 739–755, 2020. 4

[43] Rajeev Ranjan, Sabrina Zhou, Jun Cheng Chen, Amit Kumar, Azadeh Alavi, Vishal M Patel, and Rama Chellappa. Unconstrained age estimation with deep convolutional neural networks. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 109–117, 2015. 2

[44] Rasmus Rothe, Radu Timofte, and Luc Van Gool. Deep expectation of real and apparent age from a single image without facial landmarks. *International Journal of Computer Vision*, 126(2):144–157, 2018. 2

[45] Bo-Kai Ruan and Hong-Han Shuai. MAD: Makeup All-in-One with Cross-Domain Diffusion Model. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2025. 3, 5

[46] Richard Russell, Carlota Batres, Sandra Courrèges, Gwenaël Kaminski, Frédérique Soppelsa, Frédérique Morizot, and Aurélie Porcheron. Differential effects of makeup on perceived age. *British Journal of Psychology*, 110(1):87–100, 2019. 1, 2

[47] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015. 5

[48] Sefik Serengil and Alper Ozpinar. A benchmark of facial recognition pipelines and co-usability performances of modules. *Journal of Information Technologies*, 17(2):95–107, 2024. 7

[49] Fahad Shamshad, Muzammal Naseer, and Karthik Nandakumar. Clip2protect: Protecting facial privacy using text-guided makeup via adversarial latent search. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20595–20605, 2023. 1, 3, 5

[50] Caifeng Shan. Learning local features for age estimation on real-life faces. In *Proceedings of the 1st ACM International Workshop on Multimodal Pervasive Video Analysis*, page 23–28, New York, NY, USA, 2010. Association for Computing Machinery. 2

[51] Wei Shen, Yilu Guo, Yan Wang, Kai Zhao, Bo Wang, and Alan Yuille. Deep differentiable random forests for age estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(2):404–419, 2021. 2

[52] Chaojun Shi, Shiwei Zhao, Ke Zhang, Yibo Wang, and Longping Liang. Face-based age estimation using improved swin transformer with attention-based convolution. *Frontiers in Neuroscience*, 17, 2023. 2

[53] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 4

[54] Yuhao Sun, Lingyun Yu, Hongtao Xie, Jiaming Li, and Yongdong Zhang. Diffam: Diffusion-based adversarial makeup transfer for facial privacy protection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 24584–24594, 2024. 1, 3, 4, 5

[55] Zhaoyang Sun, Shengwu Xiong, Yaxiong Chen, Fei Du, Weihua Chen, Fan Wang, and Yi Rong. Shmt: Self-supervised hierarchical makeup transfer via latent diffusion models. *Advances in Neural Information Processing Systems*, 37:16016–16042, 2024. 3

[56] Zhaoyang Sun, Shengwu Xiong, Yaxiong Chen, and Yi Rong. Content-style decoupling for unsupervised makeup transfer without generating pseudo ground truth. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7601–7610, 2024. 3

[57] Sayako Ueda and Takamasa Koyama. Influence of make-up on facial recognition. *Perception*, 39(2):260–264, 2010. 1, 2

[58] Manuel C Voelkle, Natalie C Ebner, Ulman Lindenberger, and Michaela Riediger. Let me guess how old you are: effects of age, gender, and facial expression on perceptions of age. *Psychology and aging*, 27(2):265, 2012. 1

[59] Jinghang Wang, Yi Li, Zexing Zhang, Cunrui Zou, Zibo Liu, and Zhiguo Zhou. Debeauty: A joint framework for facial beautification removal based on spatial collaborative adaptation and hyperplane relocation. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5, 2025. 3, 5

[60] Xiaolong Wang, Rui Guo, and Chandra Kambhamettu. Deeply-learned feature for age estimation. In *IEEE Winter Conference on Applications of Computer Vision*, pages 534–541. IEEE, 2015. 2

[61] Tamara Lea Watson, Yumiko Otsuka, and Colin Walter Giles Clifford. Who are you expecting? biases in face perception reveal prior expectations for sex and age. *Journal of Vision*, 16(3):5–5, 2016. 1

[62] Qixin Yan, Chunle Guo, Jixin Zhao, Yuekun Dai, Chen Change Loy, and Chongyi Li. BeautyREC: Robust, Efficient, and Component-Specific Makeup Transfer . In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1102–1110, 2023. 5, 6, 8

[63] Chenyu Yang, Wanrong He, Yingqing Xu, and Yang Gao. Elegant: Exquisite and locally editable gan for makeup transfer. In *European Conference on Computer Vision*, pages 737–754. Springer, 2022. 1, 3, 4

[64] Tsun-Yi Yang, Yi-Hsuan Huang, Yen-Yu Lin, Pi-Cheng Hsiu, and Yung-Yu Chuang. SSR-Net: A compact soft stage-wise regression network for age estimation. In *IJCAI*, page 7, 2018. 2, 3

[65] Dong Yi, Zhen Lei, and Stan Z Li. Age estimation by multi-scale convolutional network. In *Asian conference on computer vision*, pages 144–158. Springer, 2014. 2

[66] Juha Ylioinas, Abdenour Hadid, and Matti Pietikäinen. Age classification in unconstrained conditions using lbp variants. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR)*, pages 1257–1260, 2012. 2

[67] Xusheng Zeng, Junyang Huang, and Changxing Ding. Soft-ranking label encoding for robust facial age estimation. *IEEE Access*, 8:134209–134218, 2020. 2

[68] Haoyu Zhang, Zhaohui Wang, and Jiawei Hou. Makeup removal for face verification based upon deep learning. In *2021 IEEE 6th International Conference on Signal and Image Processing (ICSIP)*, pages 446–450. IEEE, 2021. 3

[69] Yuxuan Zhang, Lifu Wei, Qing Zhang, Yiren Song, Jiaming Liu, Huaxia Li, Xu Tang, Yao Hu, and Haibo Zhao. Stable-makeup: When real-world makeup transfer meets diffusion model. *arXiv preprint arXiv:2403.07764*, 2024. 3

[70] Zhifei Zhang, Yang Song, and Hairong Qi. Age progression/regression by conditional adversarial autoencoder. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017. 3, 4