

AI Tool Strips Away Makeup to Stop Minors Bypassing Age Checks

Originally published at <https://www.unite.ai/ai-tool-strips-away-makeup-to-stop-minors-bypassing-age-checks/>



The appearance of facial cosmetics is letting underage users, mostly girls, slip past selfie-based age checks on platforms such as dating apps and e-commerce sites. A new AI tool addresses this loophole, using a discriminative model trained to erase makeup while preserving identity, making it harder for minors to trick automated systems.

The use of third-party, selfie-based age verification services is on the rise, not least because of a [general global impetus](#) towards online age-based verification.

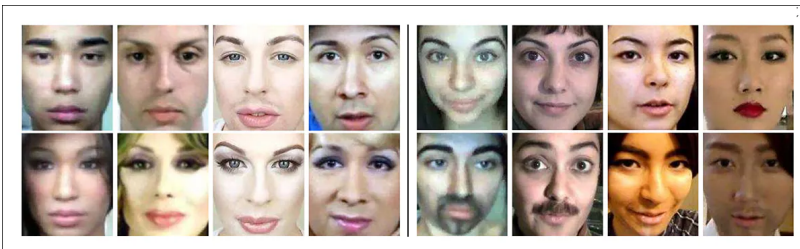
For instance, in the new enforcement regime that the UK's Online Safety Act now [mandates](#), age verification can be conducted by a variety of third-party [services](#), using various possible methods, [including visual age verification](#), where AI is used to visually predict the age of the user (usually from live mobile camera footage). Services that use approaches of this kind include [Ondato](#), [TrustStamp](#), and [Yoti](#).

However, age estimation is not infallible, and the traditional determination of teenagers to anticipate the rights of adulthood mean that young people have developed [a variety of effective methods](#) to enter dating sites, forums, and other environments that ban their age group.

One of these methods, most commonly used by females*, is by wearing facial make-up – a tactic [known to fool](#) automated age-estimation systems, which generally overestimate the age of young people and [underestimate the age](#) of older people.

Not Just the Girls

Before protest arises at considering makeup as 'female-focused', we must note that the presence of facial cosmetics on *anyone* is a [very unreliable indicator](#) of gender:



In the paper 'Impact of Facial Cosmetics on Automatic Gender and Age Estimation Algorithms' US researchers found that gender verification systems were foxed | Source: https://cse.msu.edu/~rossarun/pubs/ChenCosmeticsGenderAge_VISAPP2014.pdf

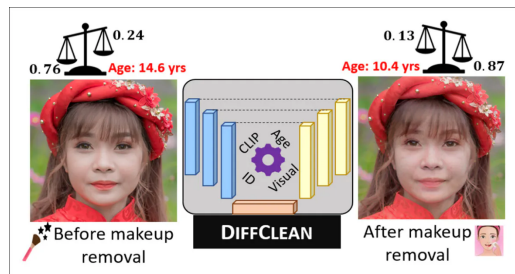
In 2024, 72% of US male consumers between the age of 18-24 were [estimated](#) to incorporate makeup into their grooming routine – though most use cosmetic products to boost the appearance of healthy skin, rather than indulging in the kind of performative[†] mascara/lipstick combinations [more associated](#) with women's visual aesthetics.

So we cannot help but treat the material studied in this article along the lines of the most common scenario explored in new research – that of female minors using makeup to subvert automated visual age-verification systems.

Effective Makeup Removal – The AI Way

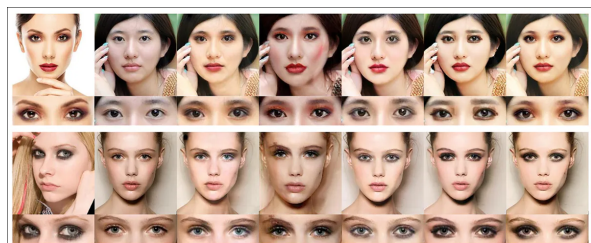
The research mentioned above comes from three contributors at New York University, in the form of the [new paper](#) *DiffClean: Diffusion-based Makeup Removal for Accurate Age Estimation*.

The objective of the project is to achieve an AI-driven method of removing the appearance of makeup from imagery (potentially including video imagery), in order to obtain a better idea of the true age of the person behind the makeup.



From the new paper, an example of how makeup removal can notably alter an age prediction. Source: <https://arxiv.org/pdf/2507.13292>

One of the challenges of developing such a system is the potential sensitivity around gathering or curating imagery of minor-aged girls wearing adult makeup. In the end, the researchers used a third-party Generative Adversarial Network-based system called [EleGANT](#) to artificially impose makeup styles, a technique which proved very effective:



Tsinghua University's 2022 EleGANT system uses a Generative Adversarial Network (GAN) to superimpose cosmetics authentically onto source photos. Source: <https://arxiv.org/pdf/2207.09840>

With the aid of synthetic data obtained in this way, and with the help of a diverse range of ancillary projects and datasets, the authors were able to exceed state-of-the-art methods in age-estimation when confronted by performative or 'evident' makeup.

The paper states:

'DiffClean [erases] makeup traces using a text-guided diffusion model to defend against makeup attacks. [It] improves age estimation (minor vs. adult accuracy by 4.8%) and face verification (TMR by 8.9% at FMR=0.01%) over competing baselines on digitally simulated and real makeup images.'

Let's take a look at how they went about the task.

Method

To avoid sourcing real images of minors in makeup, the authors used EleGANT to apply synthetic cosmetics to images sourced from the [UTKFace](#) dataset, producing before-and-after pairs for training.

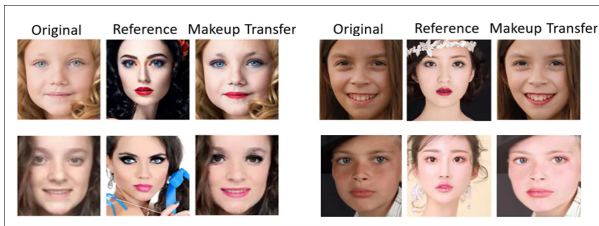


Examples from the UTKFace dataset. Source: <https://susanqq.github.io/UTKFace/>

DiffClean was then trained to reverse this transformation. Since age-estimation algorithms err most when dealing with younger age groups, the researchers found it necessary to develop a proxy age classifier [fine-tuned](#) on the target ages (10-19 years). To this end they used the [SSRNet](#) architecture trained on UTKFace, with a weighted [L1 loss](#).

A stripped-down version of the 2021 OpenAI [diffusion model](#) provided the backbone for the transformation, with the authors retaining the core architecture, but modifying it with extra [attention heads](#) at diverse resolutions, deeper layers, and [BigGAN](#)-style blocks to improve the upsampling and downsampling stages.

Directional control was introduced using [CLIP](#) prompts: specifically, *face with makeup* and *face without makeup*, so that the model learned to move in the desired semantic direction, allowing makeup to be removed without compromising facial detail, age cues, or identity.



Synthetic makeup applied using EleGANt. Each triplet shows the original UTKFace image (left), the reference makeup style (center), and the result after style transfer (right). Makeup transfer of this kind is rife in computer vision literature, and this facility is also available in the neural filters of Adobe Photoshop, which can similarly impose makeup from a reference image onto a target image.

Four key [loss functions](#) guided makeup removal without affecting facial identity or age cues. Besides the above-noted CLIP-based loss, identity was preserved using a weighted pair of [ArcFace](#) losses drawn from the [InsightFace](#) library – losses which measured the similarity between the generated face and both the original clean image and the ‘made-up’ version, ensuring that the subject remained visually consistent before and after makeup removal.

Thirdly, the perceptual loss [Learned Perceptual Similarity Metrics](#) (LPIPS) used L1 distance to enforce pixel-level realism, and retain the overall look of the original image after the makeup was excised.

Finally, age was supervised using a fine-tuned SSRNet trained on the UTKFace dataset, with the model using a smoothed L1 loss (with heavier penalties for errors in the 10–29 age range, where misclassification is most common). A variant of the model replaced this with a CLIP-based age prompt, prompting the model to match the appearance of a specific age.

For age estimation at inference time (as opposed to the use of SSRNet at training time), the 2023 [MiVOLO](#) framework was used.

Data and Tests

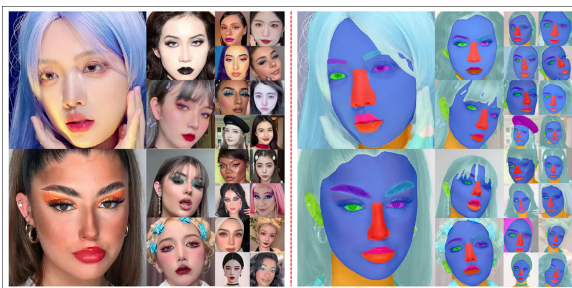
The SSRNet fine-tune of UTKFace employed a training set of 15,364 images, against a [test-set](#) of 6,701 images. The original 20,000 images were filtered to remove anyone aged over 70, and then likewise split 70:30.

In accordance with the prior method established by the 2023 [DiffAM](#) project, training then proceeded in two stages, with the initial session using 300 real-world makeup images (this time a 200/100 split between training and validation) from BeautyGAN’s [MT dataset](#).

The model was then refined further using 300 additional UTKFace images, augmented with synthetic makeup via EleGANt. This created a final training set of 600 examples, paired across five reference styles from BeautyGAN. Because makeup removal involves mapping many makeup styles to a single clean face, the training focused on broad [generalization](#) rather than covering every possible cosmetic variation.

Performance was evaluated on both synthetic and real-world images. Synthetic testing used 2,556 [Flickr-Faces-HQ](#) Dataset (FFHQ) images, evenly sampled across nine age groups below 70, and modified with EleGANt.

Generalization was assessed using 3,000 images from [BeautyFace](#) and 355 from [LADN](#), both containing authentic makeup.



Examples from the BeautyFace dataset, exemplifying the semantic segmentation that defines various areas of affected face surface.

Source: https://li-chongyi.github.io/BeautyREC_files/

Metrics and Implementation

For metrics, the authors used [Mean Absolute Error](#) (MAE) between the ground truth (real images with factual ages established) and the predicted age values, where lower results are better; *age group accuracy* was used to assess if predicted ages ended up in the correct groupings (in which case, lower results are better); minor/adult accuracy was used to evaluate correct identification of 18+ people (wherein a higher result is better).

Additionally, though it does not center on the particular topic at hand, the authors also report identity verification metrics in the form of True Match Rate (TMR) and False Match Rate (FMR), with further reporting of related [Receiver Operating Characteristic](#) (ROC) values.

SSRNet was fine-tuned on 64×64px images using a [batch size](#) of 50 under the [Adam](#) optimizer with a [weight decay](#) of 1e−4, as well as a [cosine annealing scheduler](#), and a learning rate of 1e−3 over 200 [epochs](#), with [early stopping](#).

By contrast, the DiffClean module received 256×256px input images, and was fine-tuned for five epochs using Adam, at a coarser learning rate of $4e-3$. Sampling used 40 [DDIM inversion](#) steps, and 6 DDIM forward steps. All training was performed on a single NVIDIA A100 GPU (whether with 40GB or 80GB of VRAM was not specified).

Rival systems tested were [CLIP2Protect](#) and the earlier-mentioned DiffAM. The authors used ‘matte’ makeup styles in the workflow, as this has been noted in CLIP2Protect as achieving a higher success rate (presumably allowing an avenue of opportunity for those seeking to defeat this approach – but that is a matter for another time).

To replicate DiffAM as a baseline, the pretrained model from BeautyGAN was fine-tuned on the MT dataset. For adversarial makeup transfer, the checkpoint from DiffAM was used with default parameters for the target model, reference image, and identity.

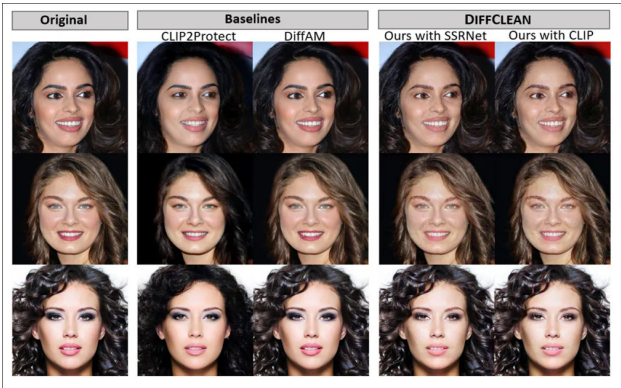
Metrics	Makeup images	CLIP2Protect ('no makeup') prompt	DiffAM (MR)	DiffClean (Ours) SSRNet age loss	DiffClean (Ours) with CLIP age loss
Minor/Adult Acc (%)	82.8	83.8	85.5	87.8	88.6
Age group Acc (%)	34.4	33.3	34.5	36.8	37.0
MAE	6.56	6.54	6.32	5.76	5.71

Performance of DiffClean compared to baselines on age estimation tasks, using MiVOLO. Metrics reported are Minor/Adult classification accuracy, age group accuracy, and mean absolute error (MAE). DiffClean with CLIP age loss achieves the best results across all metrics.

Of these results, the authors state:

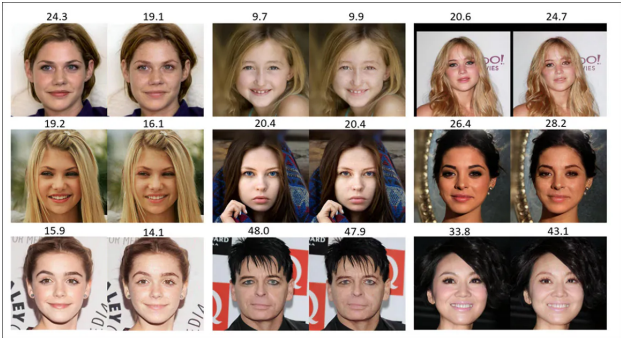
‘[Our] method DIFFCLEAN outperforms both baselines, CLIP2Protect and DiffAM, and can successfully restore the age cues disrupted due to makeup by lowering the MAE (to 5.71) and improving the overall age group prediction accuracy (to 37%).

‘Our objective focused on minor age groups, and results indicate that we achieve superior minor vs adult age classification of 88.6%.’



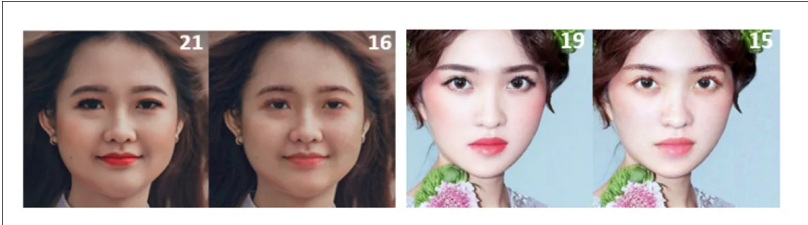
Makeup removal results from baseline and proposed methods. The left-most column shows source images, the next outputs from CLIP2Protect and DiffAM. The third column shows results from DiffClean via SSRNet and CLIP-based age loss. The authors contend that DiffClean removes makeup more effectively, avoiding the feature distortion seen in CLIP2Protect, and the residual cosmetics missed by DiffAM.

The authors further note that makeup does not have a uniform effect on perceived age, but rather can increase, decrease, or leave unchanged the apparent age of a face. Therefore DiffClean does not apply a ‘blanket reduction’ in predicted age, but instead attempts to recover the original age indicators by removing cosmetic traces:



Makeup removal examples from the CelebA-HQ and CACD datasets. Each column shows a pair of images before (left) and after (right) makeup removal. In the first column, predicted age decreases after makeup is removed; in the second, it remains unchanged; and in the third, it increases.

To test how well DiffClean performed on novel data, it was run on the BeautyFace and LADN datasets, which contain authentic makeup, but no paired images of the same subjects without cosmetics. Age predictions made before and after makeup removal were compared, to assess how effectively DiffClean reduced the distortion introduced by makeup:



Makeup removal results on real-world images from the LADN (left pair) and BeautyFace (right pair) datasets. DiffClean reduces the predicted ages by removing cc between apparent and actual age. White numbers show estimated ages before and after processing.

Results showed that DiffClean consistently narrowed the gap between apparent and actual age. Across both datasets, it lowered the overestimation and underestimation errors by about three years on average, suggesting that the system generalizes well to real-world cosmetic styles.

Conclusion

It is interesting, and perhaps inevitable, that performative cosmetic makeup would be used in an adversarial manner. Given that girls mature at different rates, but [consistently mature faster](#) as a group, the task of identifying the cusp between minor and adult female status may be one of the most ambitious that the research scene has yet set itself.

Nonetheless, time and data may eventually determine consistent age-related signs that can be used to anchor visual age-verification systems.

** Since this subject invites charged language, and since 'girls' is exclusionary (while 'women and girls', the currently acceptable term for female-gendered people, is not an accurate description in this case), I have defaulted to 'females' as the best compromise that I could devise – though it does not capture all demographic subtleties, for which I apologize.*

† In this article I use 'performative' to indicate makeup that is intended to be seen and recognized as makeup, such as mascara, eyeliner, blusher and foundation, as opposed to concealing creams and other 'surreptitious' kinds of cosmetic applications.

First published Friday, July 18, 2025

Writer on machine learning, domain specialist in human image synthesis. Former head of research content at Metaphysic.ai, until its dissolution into DNEG's Brahma.ai.

Portfolio site: martinanderson.ai

Contact: martin@martinanderson.ai



Discover More