Supplementary Material for CA-GAN: Weakly Supervised Color Aware GAN for Controllable Makeup Transfer

Robin Kips[1,2], Pietro Gori[2], Matthieu Perrot[1], and Isabelle Bloch[2]

[1] L'Oréal Research and Innovation, France
[2] LTCI, Télécom Paris, Institut Polytechnique de Paris, France

Examples of makeup style transfer on videos are also available from the project page: https://robinkips.github.io/CA-GAN/

Table 1. The detailed architecture of our generator and discriminator models.

Generator

Inputs / outputs	Layer	Activation	Normalization	Output shape
image input	-	-	-	128 x 128 x 3
color input	-	-	-	1 x 1 x 3
-	tile + concatenation	-	-	128 x 128 x 6
-	conv 7 x 7, 64 filters	relu	instance norm	128 x 128 x 64
-	conv 4 x 4, 128 filters	relu	instance norm	64 x 64 x 128
-	conv 4 x 4, 256 filters	relu	instance norm	32 x 32 x 256
-	residual block	-	instance norm	32 x 32 x 256
-	residual block	-	instance norm	32 x 32 x 256
-	residual block	-	instance norm	32 x 32 x 256
-	residual block	-	instance norm	32 x 32 x 256
-	residual block	-	instance norm	32 x 32 x 256
-	residual block	-	instance norm	32 x 32 x 256
-	transposed conv 4 x 4, 128 filters	relu	instance norm	64 x 64 x 128
-	transposed conv 4 x 4, 64 filters	relu	instance norm	128 x 128 x 64
-	conv 7x7, 3 filters	tanh	-	128 x 128 x 3
image output	add input image	-	-	128 x 128 x 3

Discriminant

Inputs / outputs	Layer	Activation	Normalization	Output shape
image input			-	128 x 128 x 3
-	conv 4 x 4, 64 filters	relu	-	648 x 64 x 64
-	conv 4 x 4, 64 filters	relu	-	32x 32 x 128
-	conv 4 x 4, 64 filters	relu	-	16 x 16 x 256
-	conv 4 x 4, 64filters	relu	-	8 x 8 x 512
-	conv 4 x 4, 64 filters	relu	-	4 x 4 x 1024
-	conv 4 x 4, 64 filters	relu	-	2 x 2 x 2048
real/fake classif output	conv 2x2, 1 filter	linear	-	2 x 2 x 1
color regression output	conv 2x2, 1 filter	linear	-	3 x 1
background color regression output	conv 2x2, 1 filter	linear	-	3 x 1

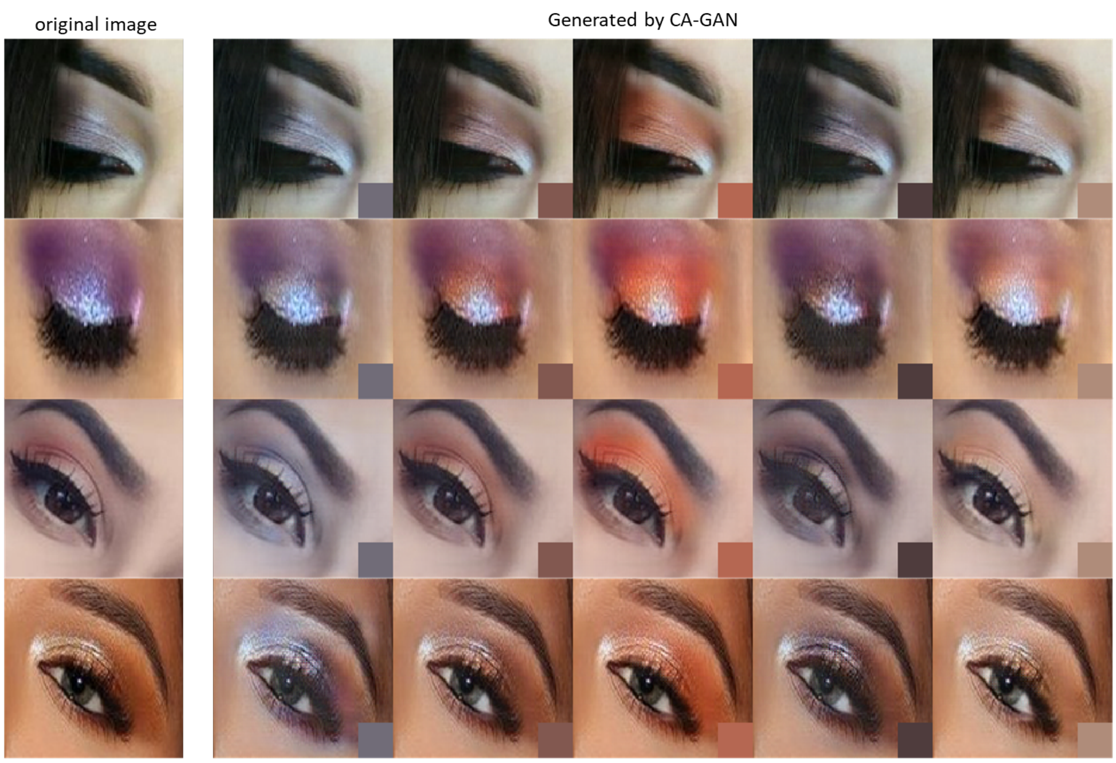

Fig. 1. Eyes pose variation: examples of eye shadow rendering using our CA-GAN model on images with various eyes poses. The segmentation accuracy and image realism is consistent across various poses.

Fig. 2. Lips pose variation: examples of lipstick rendering using our CA-GAN model on images with various lips poses. The segmentation accuracy and image realism is consistent across various poses.

Fig. 3. Skin color variation: examples of lipstick rendering using our CA-GAN model on subjects with different skin tones. The color accuracy and image realism is consistent across various skin tones.

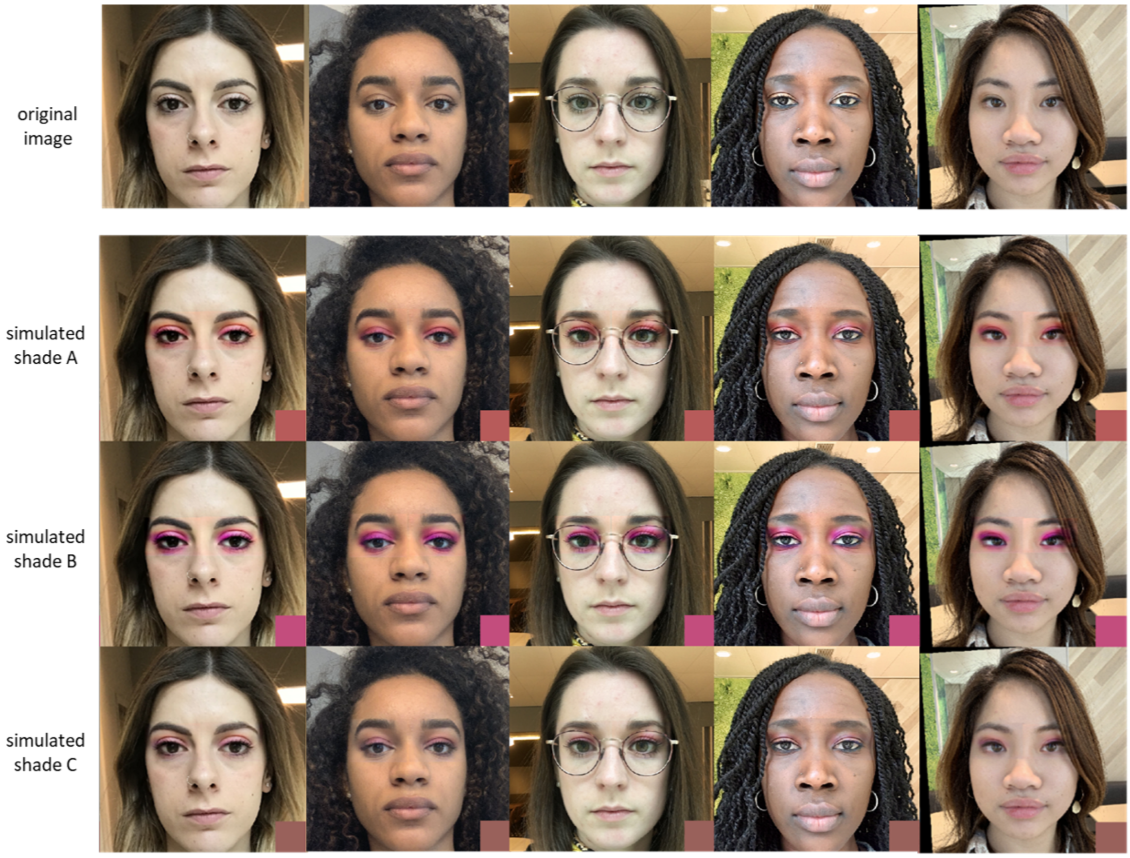

Fig. 4. Skin color variation: examples of eye shadow rendering using our CA-GAN model on subjects with different skin tones. The color accuracy and image realism is consistent across various skin tones.

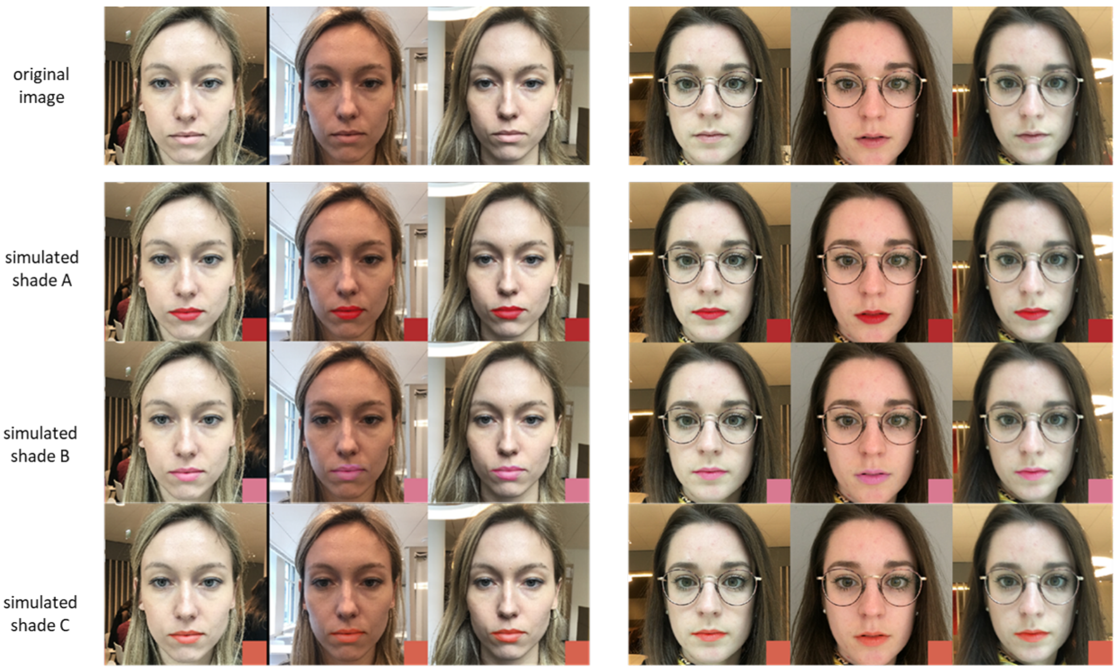

Fig. 5. Lighting variation: Example of lipstick rendering using our CA-GAN model under various illuminants

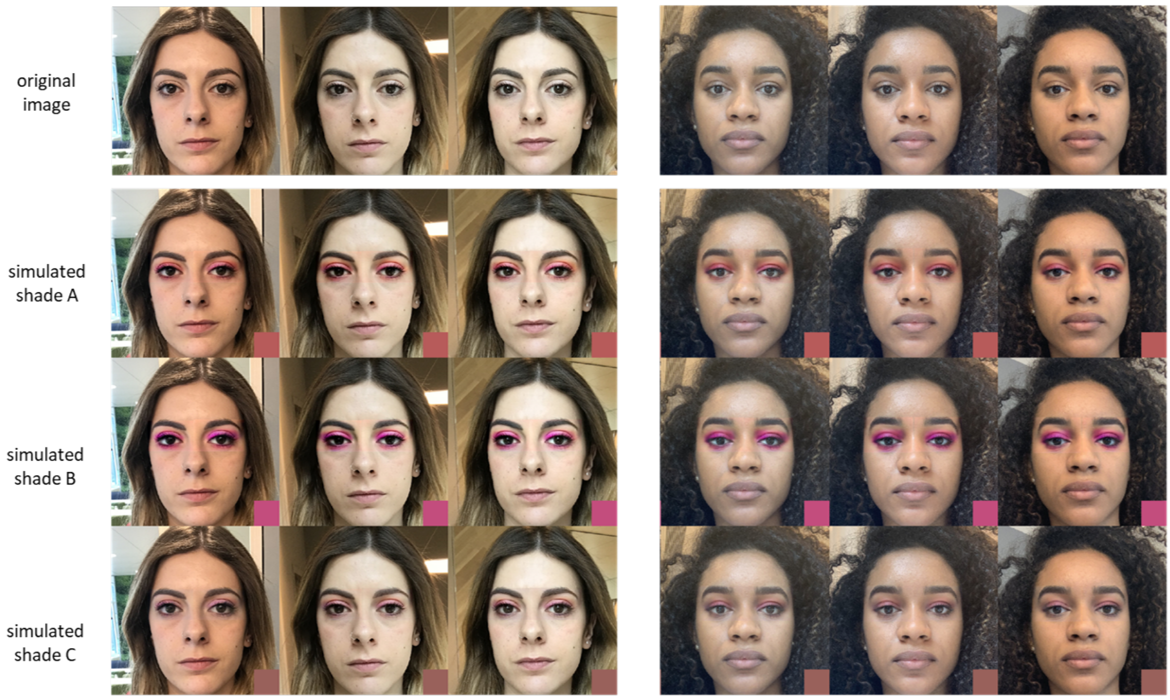

Fig. 6. Lighting variation: Example of eye shadow rendering using our CA-GAN model under various illuminants.

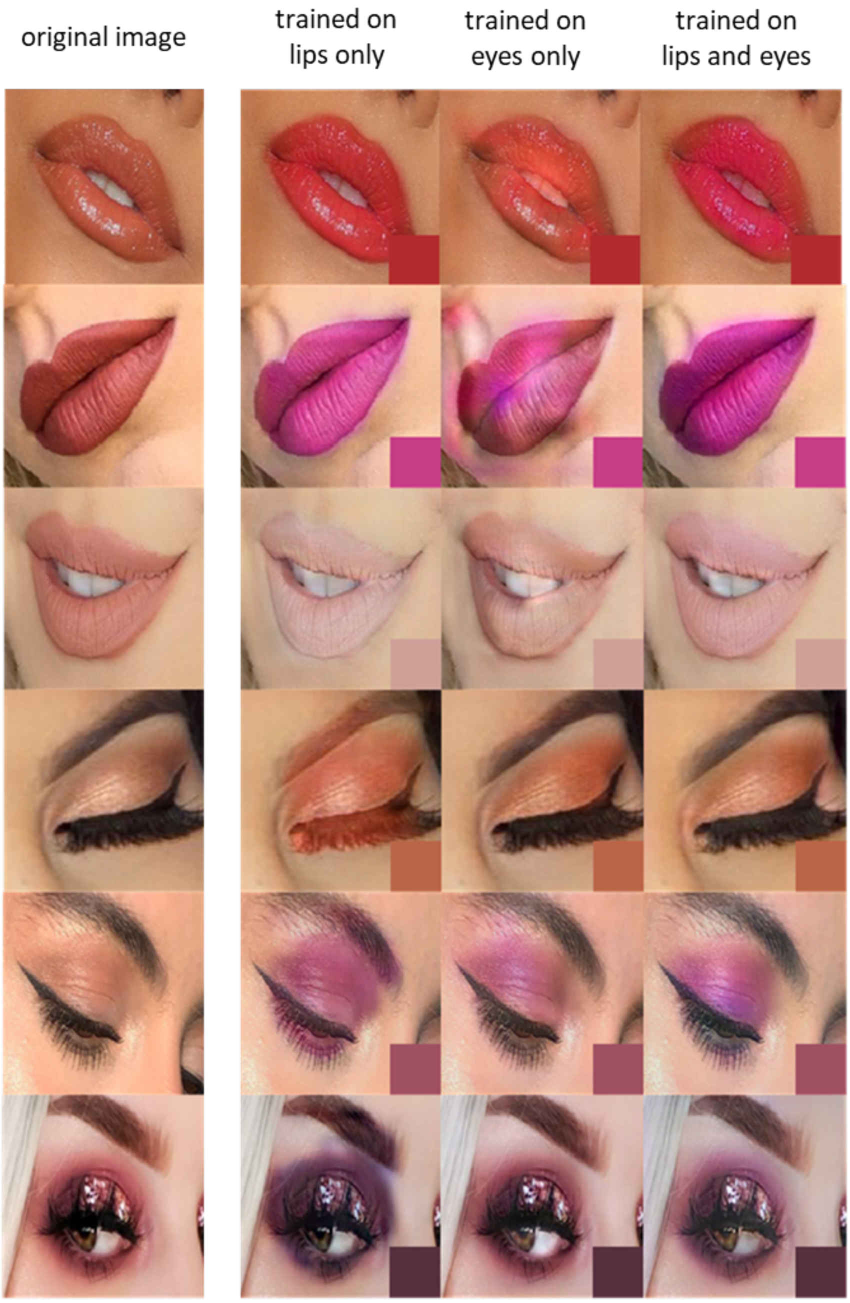

Fig. 7. Specialized models, trained on lips or eyes images only, do not generalize well to other unseen image categories. Training a joint model on both lips and eyes makeup synthesis leads to qualitatively similar results for both categories. However, qualitative analysis established that specialized models lead to higher color accuracy and style transfer performance, as described in Table 1 and 2 of the paper.

Fig. 8. Example of lipstick synthesis on portrait images for various colors. The skin consistency enables our model to generate images that preserve skin color and are consistent at the portrait scale.

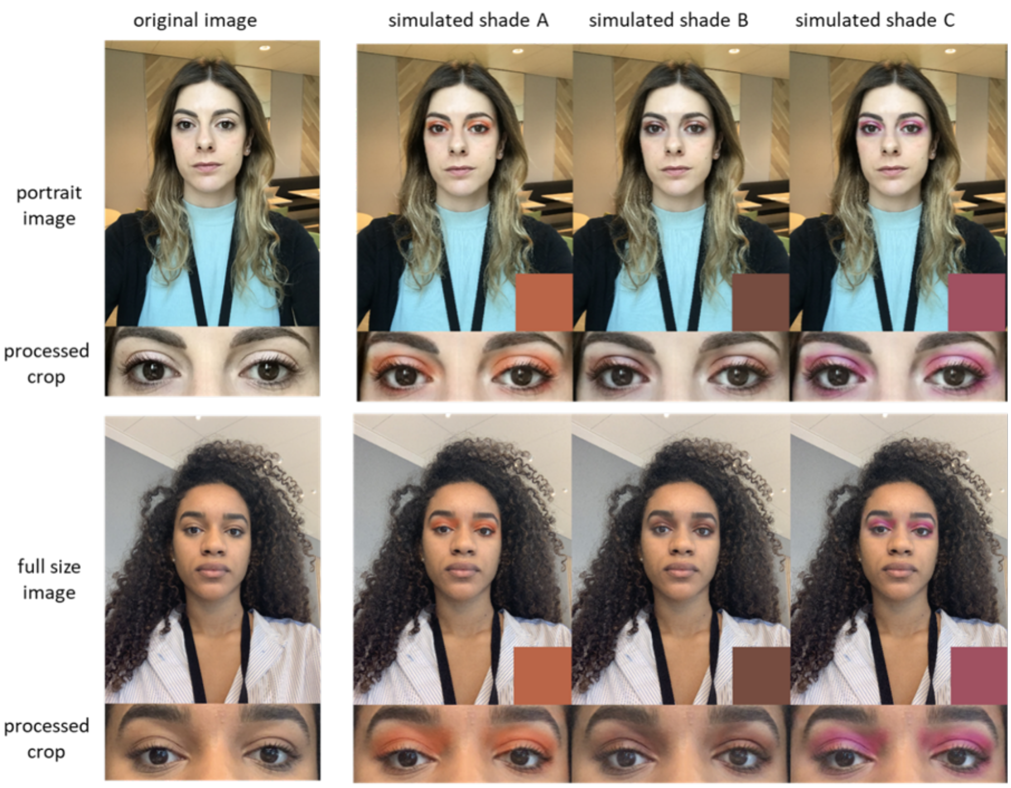

Fig. 9. Example of eye shadow synthesis on portrait images for various colors. The skin consistency enables our model to generate images that preserve skin color and are consistent at the portrait scale.

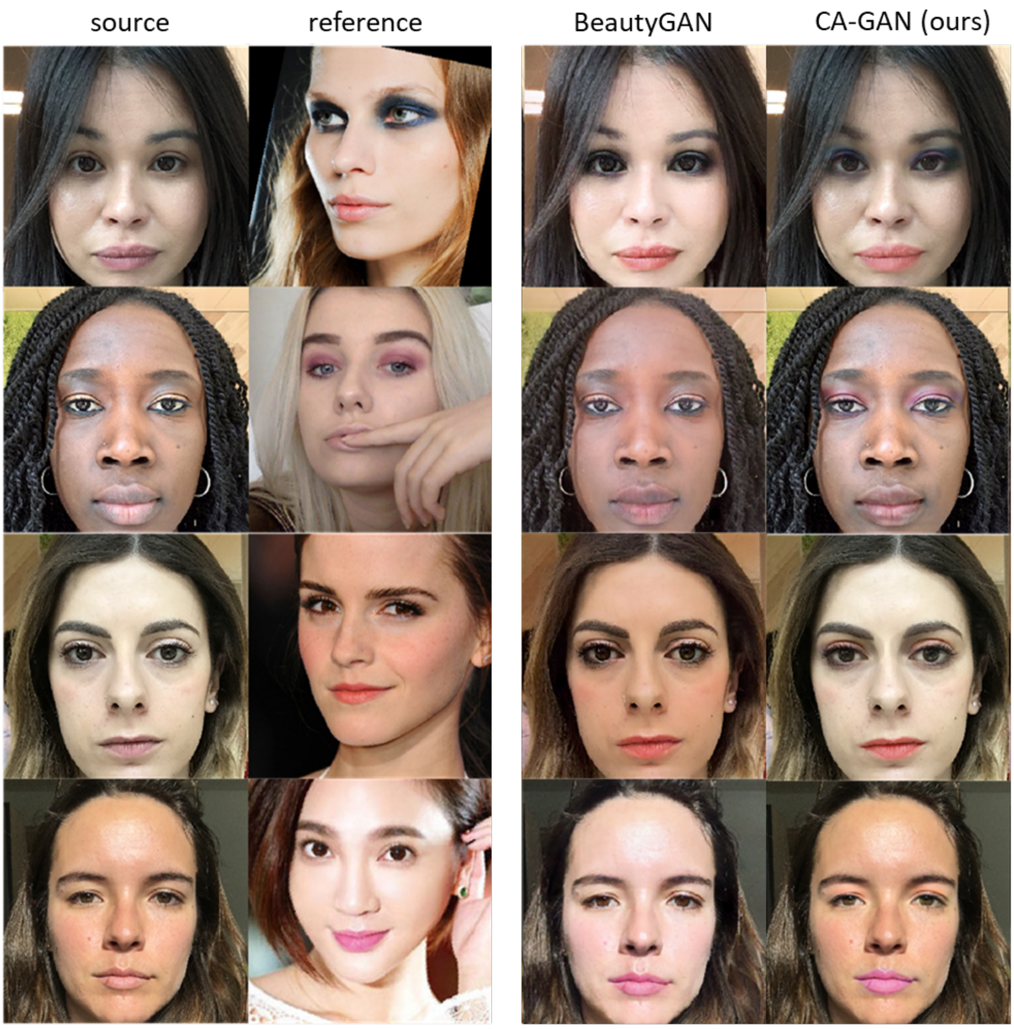

Fig. 10. Additional examples of makeup style transfer. Compared to other methods, our model learns to preserve the skin color of the subject in the source image. The disentangling of makeup style and skin color is an essential property for virtual try-on applications, as consumers expect their skin tones to be preserved.

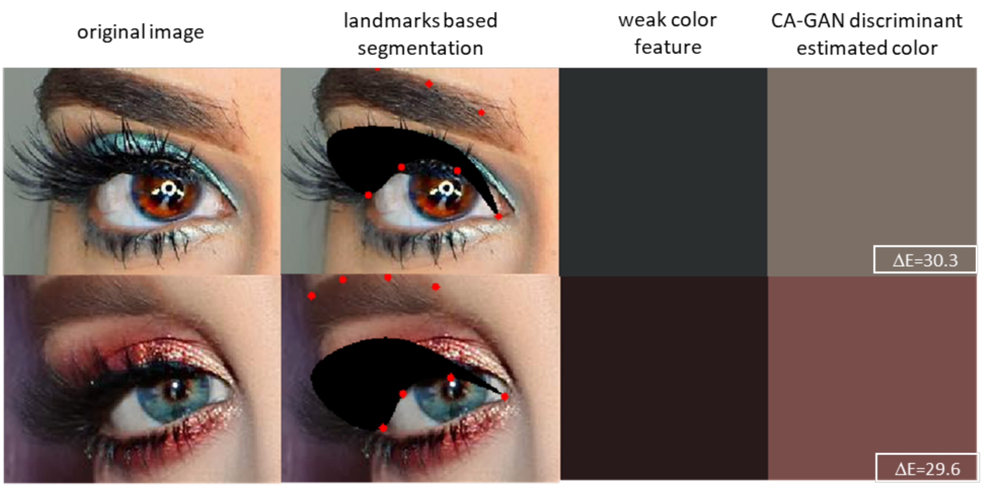

Fig. 11. The weak makeup color estimator often fails because of eyelash occlusion, which leads to too dark estimates. By contrast, our learnt model obtains results that are qualitatively superior. Displayed ΔE is the distance between the color estimated by weak and learnt model.

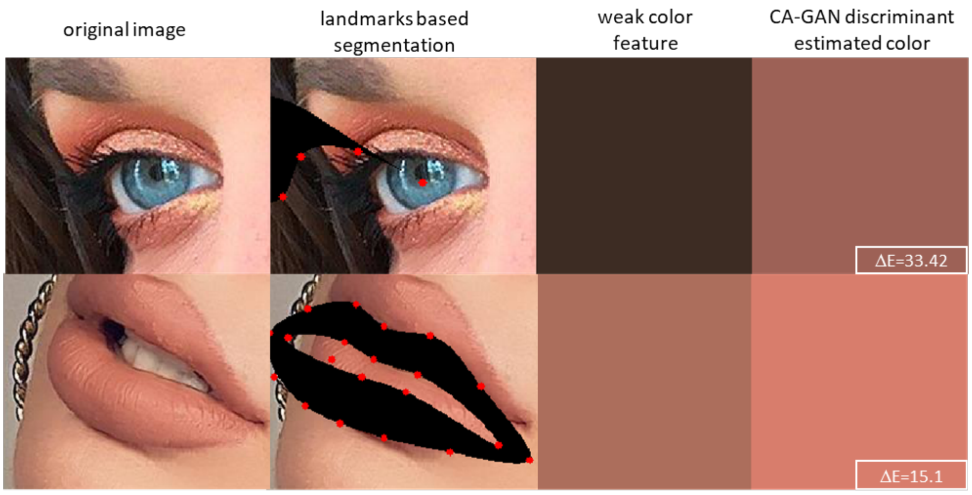

Fig. 12. Failure in facial landmarks estimation leads to large error in the weak estimator. Our learnt discriminant module seems to be more robust to extreme poses. Displayed ΔE is the distance between the color estimated by weak and learnt model

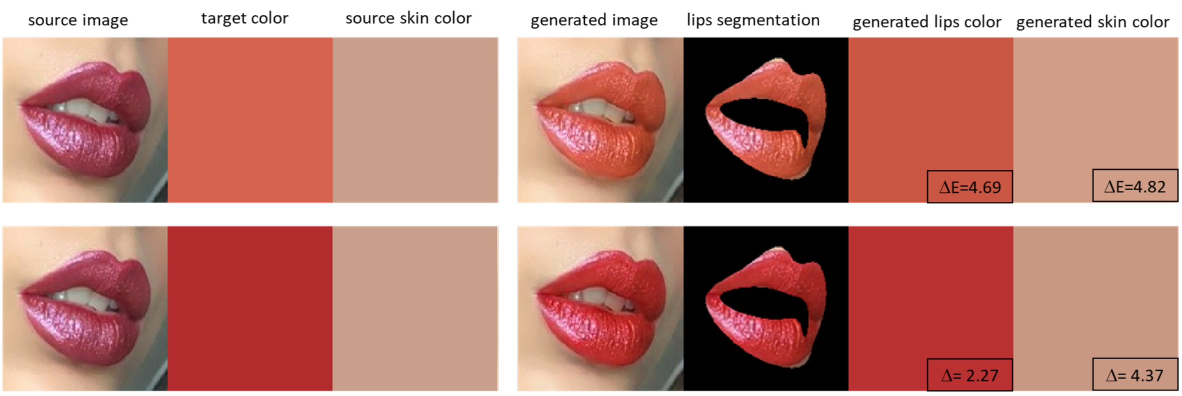

Fig. 13. Example of color accuracy evaluation. For each generated image we use a lips segmentation model and compute the pixel color median to measure the generated lips and skin color, and compute the color difference with the target lips color and source skin color.