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'Oral Research and Innovation, France
[2] LTCI, Tlcom Paris, Institut Polytechnique de Paris, France

Abstract. While existing makeup style transfer models perform an image synthesis whose results cannot be explicitly controlled, the ability to modify makeup color continuously is a desirable property for virtual try-on applications. We propose a new formulation for the makeup style transfer task, with the objective to learn a color controllable makeup style synthesis. We introduce CA-GAN, a generative model that learns to modify the color of specific objects (e.g. lips or eyes) in the image to an arbitrary target color while preserving background. Since color labels are rare and costly to acquire, our method leverages weakly supervised learning for conditional GANs. This enables to learn a controllable synthesis of complex objects, and only requires a weak proxy of the image attribute that we desire to modify. Finally, we present for the first time a quantitative analysis of makeup style transfer and color control performance.

Keywords: Image Synthesis, GANs, Weakly Supervised Learning, Makeup Style Transfer

1 Introduction

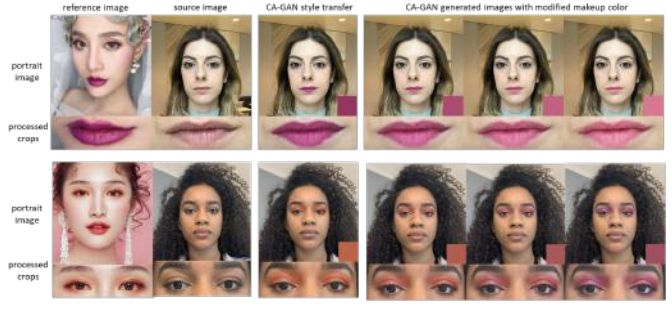

Fig. 1. Our CA-GAN model performs a color controllable makeup style transfer. The makeup color is explicitly estimated from the reference image and passed to the generator. Represented at the bottom right corner of each image, the makeup color can be modified to explore makeup style and reach the desired result.

The development of online cosmetic purchase has led to a growing interest in makeup virtual try-on technologies. Based on image filtering [32] or physical modeling of skin and makeup optical properties [25], makeup can be virtually applied to a source portrait image. Furthermore, thanks to the development of real-time facial landmark tracking [19], consumers can now try new cosmetics directly from their smartphone using Augmented Reality (AR) applications [29,30]. However, conventional makeup rendering models often fail to take into account complex appearance effects such as specular highlights. In addition, makeup is applied on face pixels according to an estimated segmentation mask. This can lead to large failures for images with extreme facial poses, which are common when trying cosmetics such as lipsticks.

More recently, the development of style transfer and image-to-image translation based on neural networks has led to new advances in the domain of makeup synthesis. The task of makeup style transfer, which consists in extracting makeup style from a reference portrait image, and applying it to the target image of a different person, has been widely studied [3,4,12,26,27,33]. In contrast to standard augmented reality, such methods can implicitly model and transfer more complex makeup in a realist manner. Yet, makeup style transfer models suffer from a lack of control as the generated makeup style cannot be modified by the user to explore different cosmetic shades. Consequently, the obtained rendering cannot be transformed to simulate another close makeup shade or a target cosmetic product. Furthermore, the ability to try various shades is an indispensable characteristic expected by consumers in virtual try-on applications.

In this paper, we propose to develop a makeup style transfer method in which the user can have fine control over the color of the synthesized makeup. Our main contributions can be summarized as follows:

- We propose CA-GAN, a color aware conditional gan that can modify the color of specific objects in the image to an arbitrary target color. This model is based on the use of a color regression loss combined with a novel background consistency loss that preserves the color attributes of non-targeted objects.
- To remove the need for costly color labeled data, we introduce weakly supervised learning for GAN based controllable synthesis. This method enables to learn a controllable synthesis of complex objects, and only requires a weak proxy of the image attribute that we desire to modify.
- We share a novel makeup dataset, the *social media* dataset [3] of 9K images, with largely increased variability in skin tones, facial poses, and makeup color.
- For the first time, we introduce a quantitative analysis of color accuracy and makeup style transfer performance for lipsticks cosmetics using ground-truth images and demonstrate that our model outperforms state of the art.

[3] available upon demand at *contact.ia@rd.loreal.com*

2 Related Work

In this section, we review related work on image synthesis and makeup style transfer. We first review GAN based methods for image-to-image translation that is the starting point of our approach. Then, we describe recent advances in controllable image-to-image synthesis using GANs. Finally, we present existing popular approaches for makeup style transfer.

GANs for Image-to-Image Translation. GAN based methods are at the origin of a large variety of recent success in image synthesis and image-to-image-translation tasks. The idea of adversarial training of a discriminator and a generator model was first introduced in [9]. Then, this method was extended in [16] to image-to-image translation with conditional GANs. However, this method requires the use of pixel aligned image pairs for training, which is rare in practice. To overcome this limitation, the cycle consistency loss was introduced in [37], allowing to train GAN for image-to-image translation from unpaired images. The use of GAN for solving image-to-image translation problems has later been extended to many different applications such as image completion [15], super-resolution [24] or video frame interpolation [17].

Controllable Image Synthesis with GANs. In the field of GANs, efforts have recently been made to develop methods that can control one or more *attributes* of the generated images. A first research direction gathers works that attempt to implicitly control the model outputs in an unsupervised manner, through operations in the latent space. Among them, InfoGAN [5] aims to learn interpretable representations in the latent space based on information regularization. Furthermore, StyleGAN [18] is an architecture that leverages AdaIn layers [14] to implicitly diversify and control the style of generated images at different scales. More recently, an unsupervised method was proposed in [34] to identify directions in a GAN model latent space that are semantically meaningful. However, while these methods introduce a meaningful modification of the generated images, they have no control over which attributes are edited. Hence, the meaning of each modified attribute is described *a posteriori* by the researchers while observing empirically the induced modification ("zoom", "orientation", "gender", etc.) On the other hand, other studies attempt to provide explicit control of the generated images through supervised methods that leverage image labels. For instance, the method in [23] achieves continuous control along a specific class attribute by using adversarial training in the latent space. Besides, the StarGAN architecture in [6] extends image-to-image translation to multiple class domains. This provides control over multiple attributes simultaneously, each being encoded as a discrete class. Later, in [7], inspired by the success of StyleGAN [18], the StarGAN architecture was improved using a style vector to enforce diversity of generated images within each target classes domain. However, it cannot be directly extended to continuous attributes. While some studies attempt to modify color attributes, they only provide control on discrete color categories [6] (e.g. "blond hair", "dark hair"), or on the intensity of a discrete color class [23].

Other synthesis methods are based on sketch conditions, that might contain color information as in [31]. However, in practice, such conditions can be complex for non-artist users and are not adapted to consumer-level applications. To the best of our knowledge, there is no existing method that enables high-level color control to an arbitrary shade in the continuous color space.

Makeup style transfer. The task of makeup style transfer has drawn interest throughout the evolution of computer vision methods. Traditional image processing methods such as image analogy [13] were first applied to this problem in [33]. Other early methods, such as in [12], propose to decompose an image into face structure, skin and color layers and transfer information between corresponding layers of different images. Later, neural networks based style transfer [8] were used for makeup images [27]. However, such a method requires aligned faces and similar skin tones in source and target images. Inspired by recent successes in GANs, makeup style transfer was formulated in [3] as an asymmetric domain translation problem. The authors of this work jointly trained a makeup transfer and makeup removal network using a conditional GAN approach. In a later work, BeautyGAN [26] improved this GAN based approach by introducing a makeup instance-level transfer in addition to the makeup domain transfer. This is ensured through makeup segmentation and histogram matching between the source and the reference image. Furthermore, in [10], makeup style transfer models were extended from processing local lips and eyes patches to the entire region of the face by using multiple overlapping discriminators. Such an improvement allows accurately transferring extreme makeup styles.

However, existing methods suffer from several limitations. First, the makeup extracted from the reference image is represented implicitly. It is therefore impossible to associate the synthesized makeup style with an existing cosmetic product that could be recommended to obtain that look. Furthermore, once the makeup style has been transferred, the generated image cannot be modified to explore other makeup shades. Prior studies [4,27,36] attempted to propose makeup style transfer methods that are controllable, but only in terms of transfer intensity.

3 Problem Formulation

We propose a new formulation for the makeup style transfer problem, where the objective is to learn a color controllable makeup style synthesis. Hence, we propose to train a generator G to generate a makeup style of an arbitrary target color c from source image x. Furthermore, in order to perform makeup style transfer from reference image y to source image x we also need to train a discriminator D_{color} to estimate the makeup color c^y from y. Equation 1 describes the objective of color controllable makeup style transfer, where c^y belongs to a continuous three-dimensional color space:

$$G(x, c^y) = G(x, D_{color}(y)) \tag{1}$$

With this new objective, the makeup color is transferred from the reference to the source image, and at the same time, explicitly controlled to reach the

desired result. Furthermore, the estimated makeup color can be used to compute a correspondence with existing cosmetics products that can be recommended. In contrast to other studies, we do not decompose between before and after makeup image domains. Indeed, in practice consumers desire to virtually try new shades without removing their current makeup. For this reason, it is desirable to train a model that can generate makeup style from portrait images with or without makeup. Finally, we desire to train our model from unlabeled unpaired images. Indeed, while massively available, makeup images are rarely qualified with labels on which cosmetics were used. Furthermore, there is currently no large database containing image pairs before and after makeup using the same makeup style, and collecting one would be particularly costly.

4 CA-GAN: Color Aware GAN

To solve this problem, we introduce the novel CA-GAN architecture, a color aware generative adversarial network that learns to modify the color of specific objects to an arbitrary target color. Our proposed model is not specific to makeup images and could be trained on any object category that can be described by a single color. Furthermore, the CA-GAN model does not require images with color labels since it can be trained in a weakly supervised manner. While the architecture of our model is close to existing popular methods, we introduce new losses for both generator and discriminator that are critical for accurate color control (see Section 4.3).

4.1 Weakly supervised color features

Since our objective is to learn to modify the color of an object in the image to an arbitrary color, we need color values to support the training of our generator. However, most available datasets do not contain labels on objects color. Furthermore, labeling the apparent color value of an object in an image is a tedious task that is highly subjective. On the other hand, GAN based models require a large amount of data to be trained. To overcome this difficulty, we introduce a method to train our model in a weakly supervised manner. Instead of using manually annotated color labels, we propose to use a weak proxy for the target object color attributes that can be obtained without supervision. In particular, in the case of makeup, we build on the assumption that makeup is generally localized on specific regions of the face, which can be approximately estimated for each image using traditional face processing methods. We denote by $C_m(x)$ our weak makeup colors feature extractor, illustrated in Figure 2. This weak estimator consists in first estimating the position of facial landmarks using the popular *dlib* library [20] and then computing the median pixel in a fixed region defined from landmarks position, for lips and eye shadow. Similarly, we also use $C_s(x)$, a weak skin color model to compute the skin color in each image, using the inverse makeup segmentation mask. Skin color will be used to ensure background color consistency when processing local crops.

Fig. 2. Example on test images of $C_m(x)$ the weak makeup extractor versus $D_{color}(x)$ our learned color discriminator module. Estimated facial landmarks are represented as red dots. While the two models agree on many images (top row), our learnt model seems superior in case of disagreement (bottom row).

This color feature extractor is *weak* in the sense that it produces a noisy estimate of a makeup color. The landmarks estimation often fails for complex poses, and the median color estimation does not take into account shading effects nor occlusion, as illustrated in Figure 2. Furthermore, the spatial information on which $C_m(x)$ relies only captures a simplified information of the makeup style, in particular for eye makeup. For this reason, we avoid to use $C_m(x)$ to directly control the generator output, and instead use it as a weak supervisor for $D_{color}(x)$ learned color discriminant module. By leveraging the noisy signal of $C_m(x)$ over a large amount of data, $D_{color}(x)$ learns a better representation for the attribute of interest, and outperforms $C_m(x)$ as discussed in Section 5.3.

4.2 CA-GAN architecture

Our CA-GAN model consists of two different networks, a generator and a discriminator, that are jointly trained. To achieve a higher resolution in the generated images, the model only processes crops of the region of interest. Besides, we train two independent CA-GAN models to process lips and eyes images.

Generator Our generator takes as input a source image together with a target color and outputs an estimated image. Its architecture is described in detail in Table 1 of the supplementary material. As in StarGAN [6] the input condition is concatenated as an additional channel of the source image. The residual blocks that we use consist of two convolutional layers with 4×4 kernels and a skip connection. Similarly to [3], the generator outputs a pixel difference that is added to the source image in order to obtain the generated image.

Discriminator As described in Table 1 of the supplementary material, our discriminator network is a fully convolutional neural network, similar to PatchGAN [16], with multi-task output branches. The discriminator network simultaneously estimates makeup color, skin color, and classifies the image as real or fake, as illustrated in Figure 3.

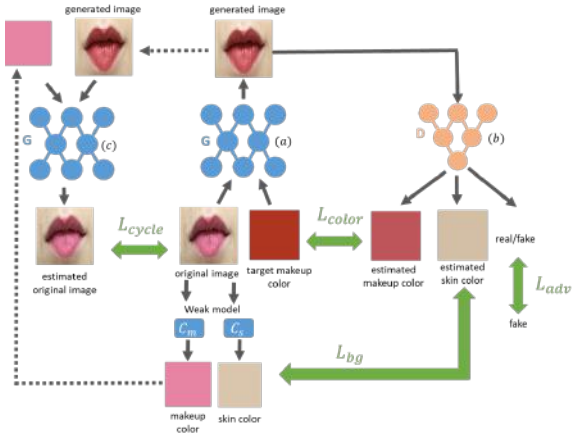

Fig. 3. The training procedure of our CA-GAN model. First (a) the generator G estimates an image from a source image and a target makeup color. Secondly (b) the discriminator D estimates the makeup color, skin color and a real/fake classification from the generated image, used to compute the color regression loss L_{color}, background consistency loss L_{bg} and adversarial loss L_{adv}, respectively. Thirdly (c), the source image is reconstructed from the generated one using the makeup color as target. The reconstruction is used to compute the cycle consistency loss L_{cycle}.

4.3 CA-GAN objective function

In this section, we introduce the loss functions that are the key components of our novel CA-GAN model. The training procedure is summarized in Figure 3.

Color regression loss The color regression loss ensures that the makeup color in the generated image is close to the target color condition passed to the generator. During training, for each image x_i among the n training examples, a target color c_i is randomly sampled at each epoch among existing colors in the training set. The color regression loss computes a color distance between a target color c_i and $D_{color}(G(x_i, c_i))$, the color of the generated image as estimated by the makeup color branch of the discriminator. As a color regression loss, we propose to use $mse - lab$, the mean squared error in the CIE $L^*a^*b^*$ space. Introduced for neural networks in [22], the $mse - lab$ loss inherits from the perceptual properties of the color distance *CIE ΔE* 1976* [28] which is key for color estimation problems. The color regression loss is described in Equations 2 and 3 for the discriminator and the generator respectively, where D_{color} is the makeup color regression output of the discriminator, and $c_i^{x_i} = C_m(x_i)$ the color label for image x_i obtained using our weak model:

$$L_{color}^{D} = \frac{1}{n}\sum_{i=1}^{n} \|c_i^{x_i} - D_{color}(x_i)\|^2 \tag{2}$$

$$L_{color}^{G} = \frac{1}{n}\sum_{i=1}^{n} \|c_i - D_{color}(G(x_i, c_i))\|^2 \tag{3}$$

Adversarial loss As in any GAN problem, we use an adversarial loss whose objective is to make generated images indistinguishable from real images. In particular, we use the Wasserstein GAN loss [2] and more specifically the one from [11] with gradient penalty. Our used adversarial loss is described in Equation 4 for the discriminator and Equation 5 for the generator, where D_{proba} is the realism classification output of the discriminator and $\lambda_{gp}\, gp(D)$ the weighted gradient penalty term computed on D:

$$L_{adv}^{D} = \frac{1}{n}\sum_{i=1}^{n} D_{proba}(G(x_i, c_i)) - \frac{1}{n}\sum_{i=1}^{n} D_{proba}(x_i) + \lambda_{gp}\, gp(D) \tag{4}$$

$$L_{adv}^{G} = -\frac{1}{n}\sum_{i=1}^{n} D_{proba}(G(x_i, c_i)) \tag{5}$$

Cycle consistency loss Since we are learning image-to-image translation from unpaired images, we need an additional loss to ensure that we will not modify undesired content in the source image. Consequently, we employ a cycle consistency loss described in Equation 6, where we compute a perceptual distance between x_i and its reconstruction $\hat{x_i} = G(G(x_i, c_i), c_i^{x_i}))$. As a perceptual distance, we choose $MSSIM$, the multiscale structural similarity loss introduced by [35], leading to:

$$L_{cycle} = 1 - MSSIM(x_i,\, \hat{x_i}) \tag{6}$$

Background consistency loss Since the generator is only processing local crops of the image, we need to ensure that the background color stays consistent with the rest of the image. Besides, if the background color is modified by the generator, the adversarial loss and the cycle consistency loss will not be able to penalize this change as it might lead to a realistic image and modify color in the same direction as the target color. Thus, we propose a background consistency loss that penalizes the color modification of the background. In the case of makeup color, the background color is represented by the skin color on the source image. Equations 7 and 8 describe background consistency for the discriminator and the generator, respectively, where D_{bg} is the background color estimation output of the discriminator and $b_i^{x_i} = C_s(x_i)$ the extracted background color of the image x_i:

$$L_{bg}^{D} = \frac{1}{n}\sum_{i=1}^{n} \|b_i^{x_i} - D_{bg}(x_i)\|^2 \tag{7}$$

$$L_{bg}^{G} = \frac{1}{n}\sum_{i=1}^{n} \|D_{bg}(x_i) - D_{bg}(G(x_i, c_i))\|^2 \tag{8}$$

Fig. 4. Modification of makeup color along each dimension of the $CIEL^*a^*b^*$ color space, using images from our social media dataset. The color patch on the bottom-right of each image illustrates the target color passed to the model. Our approach generalizes to lips and eyes images with various makeup textures and facial poses.

Total objective functions Finally, to combine all the loss functions, we propose to use weighting factors for each loss of the generator. Indeed, some factors such as the cycle consistency loss and the reconstruction loss must be balanced as they penalize opposite transformations. Equations 9 and 10 describe the total objective functions of the discriminator and the generator, where λ_{color}, λ_{bg} and λ_{cycle} are weighting factors for each generator loss that are set experimentally:

$$L_D = L_{adv}^D + L_{color}^D + L_{bg}^D \tag{9}$$

$$L_G = L_{adv}^G + \lambda_{color}\, L_{color}^G + \lambda_{bg}\, L_{bg}^G + \lambda_{cycle}\, L_{cycle} \tag{10}$$

5 Experiments

5.1 Data

Since our model does not require images before and after makeup to be trained, we are not restricted to the conventional makeup style transfer datasets such as the MT dataset [26]. Instead, we collected a database of 5000 social media images from makeup influencers. Compared to MT, this dataset contains a larger variety of skin tones, facial poses, and makeup color, with 1591 shades of 294 different cosmetics products. Since these images are unpaired and unlabeled, they are used to train our model using our proposed weakly supervised approach. We will refer to this database as the social media dataset. Furthermore, for model evaluation purposes, we collected a more controlled database focusing on the lipstick category. Therefore, we gathered images of 100 panelists with a range of 80 different lipsticks with various shades and finish. For each panelist, we collected images without makeup, and with three different lipsticks drawn from the 80 possible shades. This dataset will be referred to as the lipstick dataset.

Fig. 5. Our background consistency loss improves the preservation of the skin color in the modified image, which is essential at the portrait scale.

5.2 Implementation

Our CA-GAN model is implemented using the Tensorflow [1] deep learning framework. The generator and discriminator are jointly trained on 90 percent of our social media dataset, with lips and eyes crops of size 128 by 128 pixels. The weighting factors of the generator loss are set to $\lambda_{gp} = 10$, $\lambda_{color} = 10$, $\lambda_{bkg} = 5$, $\lambda_{cycle} = 200$. We train our model over 200 epochs using the adam optimizer [21] with a learning rate of 10^{-3} for the discriminator and 3.10^{-3} for the generator. Finally, we train separated CA-GAN models for processing lips and eyes images, as well as a joint model trained on both categories. As illustrated in Section 5.4 and the supplementary material, separated models slightly overperform the joint model, and are thus used for the image results presented in this study.

5.3 Qualitative evaluation

Color controllable makeup synthesis First, we use images from our social media dataset that are unseen during training and modify their makeup color independently in each dimension of the $CIE\ L^*a^*b^*$ color space. This experiment intends to illustrate the performance of our model with complex poses and makeup textures, and the results are displayed in Figure 4. In addition, we generate portrait images typically encountered in augmented reality tasks using our lipstick dataset, visible in Figure 1. For both experiments, it can be observed that the synthesized images reach well the target color while preserving their realistic appearance. Our approach generalizes well for both lips and eye images, with various skin colors, makeup colors, and textures. In particular, for images of eyes without makeup, eye shadow seemed to be synthesized on an average position around the eye, as visible in Figure 1. Furthermore, our model implicitly learns to only modify the makeup color attributes, preserving other dimensions such as shine or eye color, as it can be observed in Figure 4. Such results are usually obtained through complex image filtering techniques and would need a specific treatment depending on each object category. Additional generated images and videos [4] are presented in the supplementary material, illustrating performance on various skin tones and illuminants.

[4] also accessible at `https://robinkips.github.io/CA-GAN/`

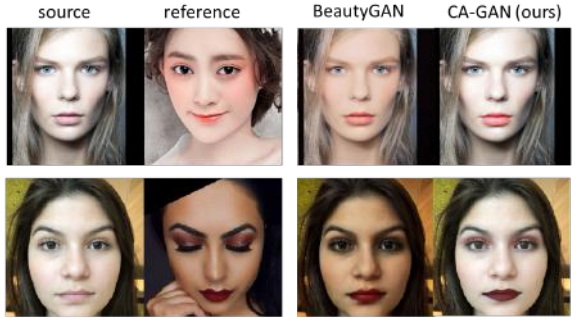

Fig. 6. Our model shows makeup style transfer performances that are equivalent to state of the art models, while obtaining better preservation of the skin color of the source subject. More results are presented in the supplementary material.

Skin color preservation Even though our model only processes a local crop of the image, the color of skin pixels is preserved, and the crop modification is not easily perceivable at the portrait scale, as seen in Figure 1. For this reason, we do not need to use Poisson blending to insert the processed crop in the final image as used in [3,4], which speeds up computations and avoids using a segmentation of the lips or eyes region. As an ablation study, we train a CA-GAN model without using the proposed background consistency loss. As observed in Figure 5, skin pixels are also modified in the generated image. Even though these changes might look realistic at the patch level and thus are not penalized by the adversarial loss, they are not acceptable at the portrait image level. Using our background consistency loss however, skin color modification is penalized by the discriminator, which leads to significantly improved results.

Makeup style transfer We use our lipstick dataset as typical source images, and perform makeup style transfer from reference images drawn from the MT dataset, as illustrated in Figure 1. In addition to obtaining a realistic generated image, the makeup style can be edited to explore other makeup styles in a continuous color space. Furthermore, our model also estimates the makeup color which can be used to recommend existing cosmetics that can be used in practice to achieve a similar result. Moreover, we compared our results on the style transfer tasks against other popular models for which the code is available. To perform style transfer with our CA-GAN model, we estimate the makeup color in the reference image using the color regression branch of the discriminator, and generate a synthetic makeup image using the generator. The obtained results can be observed in Figure 6. We compared our model against BeautyGAN [26] which is a state of the art method for conventional makeup style transfer. Our model transfers makeup color with equivalent performance. Furthermore, while BeautyGAN tends to transfer the skin color together with the makeup style, our

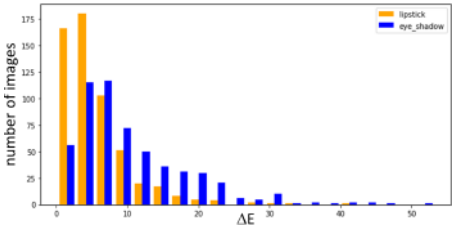

Fig. 7. Color difference between weak color features and learnt discriminant. Large differences between the two models are generally due to failure of the weak feature extractor.

model obtains better preservation of the original skin tone of the source subject, which is a desirable property for virtual try-on applications.

Weak vs learned color estimator While the weak color estimator $C_m(x)$ used for weak supervision is fixed, the learnt color extractor in the discriminant $D_{color}(x)$ leverages a large dataset. Hence, even if $C_m(x)$ has high variance and largely fails for some images, $D_{color}(x)$ learns a more robust color estimator. To illustrate this idea, we computed on test images the color difference between estimates of the weak model and the corresponding learned discriminant, as illustrated in Figure 7. Even if the two models agree for most images, large differences occur in some cases. In practice, we found that in most large difference cases, the weak estimator was failing due to poor facial landmark localization, occlusion, or complex appearance with shading and specularities (see Figure 2 and supplementary material). The difference is even larger for the eye shadow region in which appearance is more complex due to hair and eyelash occlusion. This reinforces the interest of weakly supervised learning for GAN based model, since improved color estimation will improve the generator control and in turn the style transfer accuracy.

5.4 Quantitative evaluation

In this section, we focus on the evaluation of the model on lips images. Indeed, while there is no existing approach for eye makeup segmentation, that might be on a larger region than the eyelid, it is possible to segment the lips makeup region for our experiments using face parsing models.

Color accuracy evaluation First, we evaluate the ability of our CA-GAN model to generate makeup images that are close to the chosen color target. For this experiment, illustrated in the supplementary material, we use the 500 test images from our social media dataset. First, we choose a set of 50 representative lipstick shades by computing the centroids of a k-means clustering of the lipstick colors in our training data. Then, for each test sample, we generate an image with

Table 1. The ablation study demonstrates that our color regression loss and background consistency loss significantly increase the makeup color synthesis accuracy and skin color preservation.

Model	color loss	background consistency loss	training images	lips color accuracy (ΔE mean)	skin color preservation (ΔE mean)
CA-GAN	rgb-mse	no	lips	25.82	19.49
CA-GAN	lab-mse	no	lips	9.62	10.18
CA-GAN	lab-mse	yes	lips	**6.80**	**6.05**
CA-GAN	lab-mse	yes	eyes and lips	7.78	8.76

each representative lipstick color as target. Finally, using a lips segmentation algorithm we estimate the median color of lips to compute a color distance to the model target. We also estimate the difference between the color of the skin before and after image synthesis to control its preservation. The results of these experiments are reported in Table 1. The ablation study confirms that the use of the $lab - mse$ for color loss largely increases the color accuracy of our model. Furthermore, our novel background consistency loss helps the generator to disentangle skin and lips color, which leads to significantly improved lipstick color accuracy and skin color preservation.

Style transfer performance evaluation For the first time, we introduce a quantitative evaluation of model performance on the makeup style transfer task, as illustrated in Figure 8. We use our collected lipstick dataset that contains images of multiple panelists wearing the same lipstick shade. Thus, it is possible to construct ground-truth triplets with a reference portrait, a source portrait, and the associated ground-truth image with the reference makeup. The style transfer accuracy is then computed using the MSSIM similarity [35] as a measure of a perceptual distance. Furthermore, to avoid lighting bias, we select the ground-truth among several images of the same panelist, using the most similar skin color compared to the source image. We perform this experiment on 300 image triplets with 100 different panelists and 80 different lipstick shades. The results of this experiment are reported in Table 2. The ablation study confirms that our color regression loss and background consistency loss significantly improve the style transfer performance. Furthermore, we observe that our model outperforms BeautyGAN by a significant margin. This is expected given the ability of our model to preserve the skin color in the source image.

Table 2. A quantitative evaluation of the style transfer performance using style transfer image triplets.

Model	color loss	background consistancy loss	training images	L1	1 - MSSIM
BeautyGAN [26]	-	-	-	0.124	0.371
CA-GAN	rgb-mse	no	lips	0.231	0.698
CA-GAN	lab-mse	no	lips	0.097	0.313
CA-GAN	lab-mse	yes	lips	**0.085**	**0.283**
CA-GAN	lab-mse	yes	eyes and lips	0.087	0.312

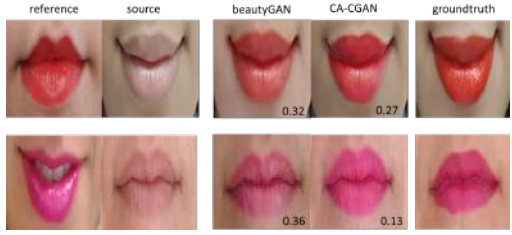

Fig. 8. The style transfer performance is evaluated using triplets of lips images. The makeup is extracted from the reference image and transferred to the source image of a different panelist. We use a ground-truth image of the source panelist with the same lipstick to compute a style transfer performance. The computed perceptual distance $1 - MSSIM$ is given at the bottom right of each generated image.

6 Conclusion and Future Work

In this paper, we introduced CA-GAN, a generative model that learns to modify the color of objects in an image to an arbitrary target color. This model is based on the combined use of a color regression loss with a novel background consistency loss that learns to preserve the color of non-target objects in the image. Furthermore, CA-GAN can be trained on unlabeled images using a weakly supervised approach based on a noisy proxy of the attribute of interest. Using this architecture on makeup images of eyes and lips we show that we can perform makeup synthesis and makeup style transfer that are controllable in a continuous color space. For the first time, we introduce a quantitative analysis of makeup style transfer and color control performance. Our results show that our model can accurately modify makeup color, while outperforming conventional models such as [26] in makeup style transfer realism. Since our CA-GAN model does not require labeled images, it could be directly applied to other object categories for which it is possible to compute pixel color statistics, such as hair, garments, cars, or animals.

Finally, we emphasize some perspectives for future work. First, we represent eyes and lips makeup by three-dimensional color coordinates. However, extreme makeup can be composed of multiple different cosmetics, in particular for the eye shadow category. To achieve color control on multiple cosmetics simultaneously, our model should be extended with a spatial information condition in addition to our current color condition. However, while our model can currently be trained in a weakly supervised manner, using segmentation masks to carry the spatial information would require to have annotated images. Moreover, the representation of cosmetics could also be completed using a shine representation. While the current model objective is to learn to modify color only, without affecting the other image attributes such as shine and specularities, using a shine score as an additional generator condition would make it possible to simulate mat and shine cosmetics with more accuracy.

References

1. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al.: Tensorflow: A system for large-scale machine learning. In: Operating Systems Design and Implementation. pp. 265–283 (2016)
2. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein generative adversarial networks. In: International Conference on Machine Learning. pp. 214–223 (2017)
3. Chang, H., Lu, J., Yu, F., Finkelstein, A.: Pairedcyclegan: Asymmetric style transfer for applying and removing makeup. In: Computer Vision and Pattern Recognition. pp. 40–48 (2018)
4. Chen, H.J., Hui, K.M., Wang, S.Y., Tsao, L.W., Shuai, H.H., Cheng, W.H.: Beautyglow: On-demand makeup transfer framework with reversible generative network. In: Computer Vision and Pattern Recognition. pp. 10042–10050 (2019)
5. Chen, X., Duan, Y., Houthooft, R., Schulman, J., Sutskever, I., Abbeel, P.: Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In: Advances in Neural Information Processing Systems. pp. 2172–2180 (2016)
6. Choi, Y., Choi, M., Kim, M., Ha, J.W., Kim, S., Choo, J.: Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In: Conference on Computer Vision and Pattern Recognition. pp. 8789–8797 (2018)
7. Choi, Y., Uh, Y., Yoo, J., Ha, J.W.: Stargan v2: Diverse image synthesis for multiple domains. In: Conference on Computer Vision and Pattern Recognition. pp. 8188–8197 (2020)
8. Gatys, L.A., Ecker, A.S., Bethge, M.: Image style transfer using convolutional neural networks. In: Computer Vision and Pattern Recognition. pp. 2414–2423 (2016)
9. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in Neural Information Processing Systems. pp. 2672–2680 (2014)
10. Gu, Q., Wang, G., Chiu, M.T., Tai, Y.W., Tang, C.K.: Ladn: Local adversarial disentangling network for facial makeup and de-makeup. In: International Conference on Computer Vision. pp. 10481–10490 (2019)
11. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C.: Improved training of Wasserstein GANs. In: Advances in Neural Information Processing Systems. pp. 5767–5777 (2017)
12. Guo, D., Sim, T.: Digital face makeup by example. In: Computer Vision and Pattern Recognition. pp. 73–79. IEEE (2009)
13. Hertzmann, A., Jacobs, C.E., Oliver, N., Curless, B., Salesin, D.H.: Image analogies. In: Computer Graphics And Interactive Techniques. pp. 327–340 (2001)
14. Huang, X., Belongie, S.: Arbitrary style transfer in real-time with adaptive instance normalization. In: International Conference on Computer Vision. pp. 1501–1510 (2017)
15. Iizuka, S., Simo-Serra, E., Ishikawa, H.: Globally and locally consistent image completion. ACM Transactions on Graphics **36**(4), 1–14 (2017)
16. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Computer Vision and Pattern Recognition. pp. 1125–1134 (2017)
17. Jiang, H., Sun, D., Jampani, V., Yang, M.H., Learned-Miller, E., Kautz, J.: Super slomo: High quality estimation of multiple intermediate frames for video interpolation. In: Conference on Computer Vision and Pattern Recognition. pp. 9000–9008 (2018)

18. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: Computer Vision and Pattern Recognition. pp. 4401–4410 (2019)
19. Kazemi, V., Sullivan, J.: One millisecond face alignment with an ensemble of regression trees. In: Conference on Computer Vision and Pattern Recognition. pp. 1867–1874 (2014)
20. King, D.E.: Dlib-ml: A machine learning toolkit. Journal of Machine Learning Research **10**(Jul), 1755–1758 (2009)
21. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. International Conference for Learning Representations (2015)
22. Kips, R., Tran, L., Malherbe, E., Perrot, M.: Beyond color correction: Skin color estimation in the wild through deep learning. Electronic Imaging (2020)
23. Lample, G., Zeghidour, N., Usunier, N., Bordes, A., Denoyer, L., Ranzato, M.: Fader networks: Manipulating images by sliding attributes. In: Advances in Neural Information Processing Systems. pp. 5967–5976 (2017)
24. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: Conference on Computer Vision and Pattern Recognition. pp. 4681–4690 (2017)
25. Li, C., Zhou, K., Lin, S.: Simulating makeup through physics-based manipulation of intrinsic image layers. In: Conference on Computer Vision and Pattern Recognition. pp. 4621–4629 (2015)
26. Li, T., Qian, R., Dong, C., Liu, S., Yan, Q., Zhu, W., Lin, L.: Beautygan: Instance-level facial makeup transfer with deep generative adversarial network. In: International Conference on Multimedia. pp. 645–653 (2018)
27. Liu, S., Ou, X., Qian, R., Wang, W., Cao, X.: Makeup like a superstar: Deep localized makeup transfer network. In: IJCAI (2016)
28. McLaren, K.: XIII-The development of the CIE 1976 (L* a* b*) uniform colour space and colour-difference formula. Journal of the Society of Dyers and Colourists **92**(9), 338–341 (1976)
29. Modiface, Inc.: Modiface - augmented reality, `http://modiface.com/`, Last accessed on 2020-02-24
30. Perfect Corp.: Perfect corp. - virtual makeup, `https://www.perfectcorp.com/business/products/virtual-makeup`, Last accessed on 2020-02-24
31. Portenier, T., Hu, Q., Szabo, A., Bigdeli, S.A., Favaro, P., Zwicker, M.: Faceshop: Deep sketch-based face image editing. ACM Transactions on Graphics **37**(4) (2018)
32. Sokal, K., Kazakou, S., Kibalchich, I., Zhdanovich, M.: High-quality AR lipstick simulation via image filtering techniques. In: CVPR Workshop on Computer Vision for Augmented and Virtual Reality (2019)
33. Tong, W.S., Tang, C.K., Brown, M.S., Xu, Y.Q.: Example-based cosmetic transfer. In: Pacific Conference on Computer Graphics and Applications. pp. 211–218 (2007)
34. Voynov, A., Babenko, A.: Unsupervised discovery of interpretable directions in the gan latent space. arXiv preprint arXiv:2002.03754 (2020)
35. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multiscale structural similarity for image quality assessment. In: Thirty-Seventh Asilomar Conference on Signals, Systems & Computers. vol. 2, pp. 1398–1402 (2003)
36. Zhang, H., Chen, W., He, H., Jin, Y.: Disentangled makeup transfer with generative adversarial network. arXiv preprint arXiv:1907.01144 (2019)
37. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: International Conference on Computer Vision (Oct 2017)