Makeup Like a Superstar: Deep Localized Makeup Transfer Network

Si Liu[1], **Xinyu Ou**[1,2,3], **Ruihe Qian**[1,4], **Wei Wang**[1] **and Xiaochun Cao**[1] *
[1]State Key Laboratory of Information Security,
Institute of Information Engineering, Chinese Academy of Sciences
[2]School of Computer Science and Technology, Huazhong University of Science and Technology
[3]YNGBZX, Yunnan Open University
[4]University of Electronic Science and Technology of China, Yingcai Experimental School
{liusi, caoxiaochun}@iie.ac.cn, ouxinyu@hust.edu.cn, 406618818@qq.com, wang_wei.buaa@163.com

Abstract

In this paper, we propose a novel Deep Localized Makeup Transfer Network to automatically recommend the most suitable makeup for a female and synthesis the makeup on her face. Given a before-makeup face, her most suitable makeup is determined automatically. Then, both the before-makeup and the reference faces are fed into the proposed Deep Transfer Network to generate the after-makeup face. Our end-to-end makeup transfer network have several nice properties including: **(1)** with complete functions: including foundation, lip gloss, and eye shadow transfer; **(2)** cosmetic specific: different cosmetics are transferred in different manners; **(3)** localized: different cosmetics are applied on different facial regions; **(4)** producing naturally looking results without obvious artifacts; **(5)** controllable makeup lightness: various results from light makeup to heavy makeup can be generated. Qualitative and quantitative experiments show that our network performs much better than the methods of [Guo and Sim, 2009] and two variants of NerualStyle [Gatys *et al.*, 2015a].

1 Introduction

Makeup makes the people more attractive, and there are more and more commercial facial makeup systems in the market. Virtual Hairstyle[1] provides manual hairstyle try-on. Virtual Makeover TAAZ[2] offers to try some pre-prepared cosmetic elements, such as, lipsticks and eye liners. However, all these softwares rely on the pre-determined cosmetics which cannot meet up with users' individual needs.

Different from the existing work, our goal is to design a real application system to automatically recommend the most suitable makeup for a female and synthesis the makeup on her face. As shown in Figure 1, we simulate an applicable makeup process with two functions. **I:** The first function is *makeup recommendation*, where personalization is taken special cares. More specifically, females with similar face, eye or mouth shapes are suitable for similar makeups [Liu *et al.*, 2014]. To this end, given a before-makeup face, we find the visually similar face from a reference dataset. The similarity is measured by the Euclidean distance between deep features produced by an off-the-shelf deep face recognition network [Parkhi *et al.*, 2015]. To sum up, the recommendation is personalized, data-driven and easy to implement. **II:** The second function is *makeup transfer* from the reference face to the before-makeup face. The makeup transfer function should satisfy five criteria. 1) *With complete functions*: we consider three kinds of most commonly used cosmetics, i.e., foundation, eye shadow and lip gloss. Note that our model is quite generic and easily extended to other types of cosmetics. 2) *Cosmetic specific*: different cosmetics are transferred in their own ways. For example, maintaining the shape is important

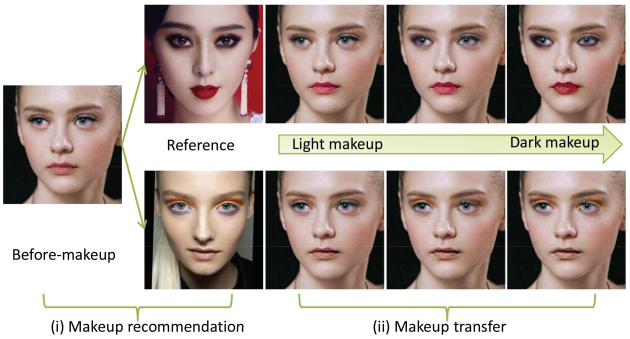

Figure 1: Our system has two functions. **I:** recommend the most suitable makeup for each before-makeup face **II:** transfer the foundation, eye shadow and lip gloss from the reference to the before-makeup face. The lightness of the makeup can be tuned. For better view of all figures in this paper, please refer to the original color pdf file and pay special attention to eye shadow, lip gloss and foundation transfer.

*corresponding author
[1]http://www.hairstyles.knowage.info/
[2]http://www.taaz.com/

for the eye shadow rather than for the foundation. 3) *Localized*: all cosmetics are applied locally on their corresponding facial parts. For example, lip gloss is put on the lip while eye shadow is worn the eyelid. 4) *Naturally looking*: the cosmetics need to be seamlessly fused with the before-makeup face. In other words, the after-makeup face should look natural. 5) *Controllable makeup lightness*: we can change the lightness of each type of cosmetic.

To meet the five afore-mentioned criteria, we propose a Deep Localized Makeup Transfer Network, whose flowchart is shown in Figure 2. The network transfers the makeup from the recommended reference face to a before-makeup face. Firstly, both before-makeup and reference faces are fed into a face parsing network to generate two corresponding labelmaps. The parsing network is based on the Fully Convolutional Networks [Long *et al.*, 2014] by (i) emphasizing the makeup relevant facial parts, such as eye shadow and (ii) considering the symmetry structure of the frontal faces. Based on the parsing results, the local region of the before-makeup face (e.g., mouth) corresponds to its counterpart (e.g., lip) in the reference face. Secondly, three most common cosmetics, i.e, eye shadow, lip gloss and foundation are transferred in their own manners. For example, keeping the shape is the most important for transferring eye shadow while the smoothing the skin's texture is the most important for foundation. So the eye shadow is transferred via directly altering the corresponding deep features [Mahendran and Vedaldi, 2015] while foundation is transferred via regularizing the inner product of the feature maps [Gatys *et al.*, 2015a]. The after-makeup face is initialized as the before-makeup face, then gradually updated via Stochastic Gradient Descent to produce naturally looking results. By tuning up the weight of each cosmetic, a series of after-makeup faces with increasingly heavier makeup can be generated. In this way, our system can produce various results with controllable makeup lightness.

Compared with traditional makeup transfer methods [Guo and Sim, 2009; Tong *et al.*, 2007; Scherbaum *et al.*, 2011; Liu *et al.*, 2014], which require complex data preprocessing or annotation and their results are inferior, our contributions are as follows. **I:** To the best of our knowledge, it is the first makeup transferring method based on deep learning framework and can produce very natural-looking results. Our system can transfer foundation, eye shadow and lip gloss, and therefore is with complete functions. Furthermore, the lightness of the makeup can be controlled to meet the needs of various users. **II:** we propose an end-to-end Deep Localized Makeup Transfer Network to first build part vs cosmetic correspondence and then transfer makeup. Compared with NeruralStyle [Gatys *et al.*, 2015a] which fuses two images globally, our method transfers makeup locally from the cosmetic regions to their corresponding facial parts. Therefore, a lot of unnatural results are avoided.

2 Related Work

We will introduce the related makeup transfer methods and the most representative deep image synthesis methods.

Facial Makeup Studies: The work of Guo *et al.* [Guo and Sim, 2009] is the first attempt for the makeup transfer task. It first decomposes the before-makeup and reference faces into three layers. Then, they transfer information between the corresponding layers. One major disadvantage is that it needs warp the reference face to the before-makeup face, which is very challenging. Scherbaum *et al.* [Scherbaum *et al.*, 2011] propose to use a 3D morphable face model to facilitate facial makeup. It also requires the before-after makeup face pairs for the same person, which are difficult to collect in real application. Tong *et al.* [Tong *et al.*, 2007] propose a "cosmetic-transfer" procedure to realistically transfer the cosmetic style captured in the example-pair to another person's face. The requirement of before-after makeup pairs limits the practicability of the system. Liu et al [Liu *et al.*, 2014] propose an automatic makeup recommendation and synthesis system called beauty e-expert. Their contribution is in the recommendation module. To sum up, our method greatly relaxes the requirements of traditional makeup methods, and generates more naturally-looking results.

Deep Image Synthesis Methods: Recently, deep learning has achieved great success in fashion analysis works [Liu *et al.*, 2015; Liang *et al.*, 2015]. Dosovitskiy *et al.* [Dosovitskiy *et al.*, 2015; Dosovitskiy and Brox, 2015] use a generative CNN to generate images of objects given object type, viewpoint, and color. Simonyan *et al.* [Simonyan *et al.*, 2013] generate an image, which visualizes the notion of the class captured by a net. Mahendran *et al.* [Mahendran and Vedaldi, 2015] contribute a general framework to invert both hand-crafted and deep representations to the images. Gatys *et al.* [Gatys *et al.*, 2015b] present a parametric texture model based on the CNN which can synthesise high-quality natural textures. Generative adversarial network [Goodfellow *et al.*, 2014] consists of two components; a generator and a discriminator. The generated image is very natural without obvious artifacts. Goodfellow *et al.* introduce [Goodfellow *et al.*, 2015] a simple and fast method of generating adversarial examples. Their main aim is to enhance the CNN training instead image synthesis. Kulkarni *et al.* [Kulkarni *et al.*, 2015] present the Deep Convolution Inverse Graphics Network, which learns an interpretable representation of images for 3D rendering. All existing deep methods can only generate one image. However, we mainly focus on how to generate a new image having the nature of the two input images. Recently, deep learning techniques have been applied in many image editing tasks, such as image colorization [Cheng *et al.*, 2015], photo adjustment [Yan *et al.*, 2014], filter learning [Xu *et al.*, 2015], image inpainting [Ren *et al.*, 2015], shadow removal [Shen *et al.*, 2015] and super-resolution [Dong *et al.*, 2014]. These methods are operated on a single image. NeruralStyle [Gatys *et al.*, 2015a] is the most similar with us. They use CNN to synthesis a new image by combining the structure layer from one image and the style layers of another. The key difference between their work and ours is that our network is applied locally, which can produce more natural results.

3 Approach

In this Section, we sequentially introduce our makeup recommendation and makeup transfer methods in detail.

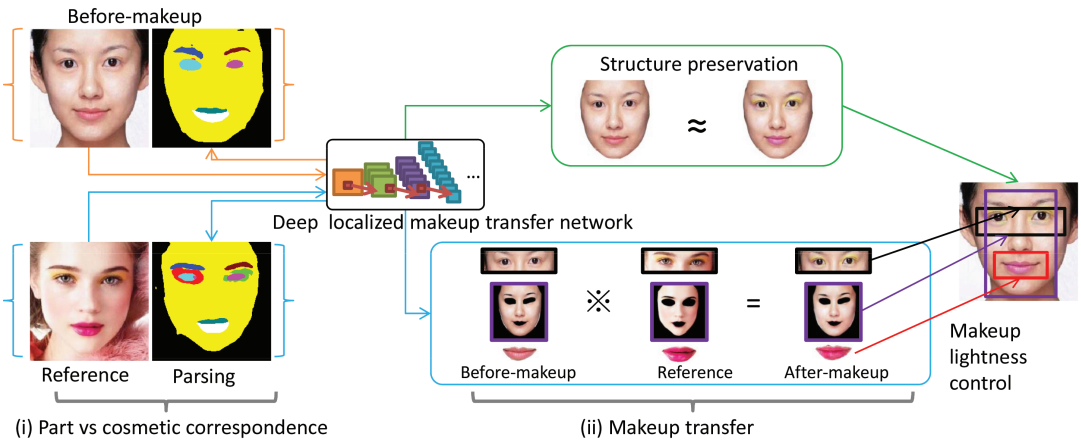

Figure 2: The proposed Deep Localized Makeup Transfer Network contains two sequential steps. (i) the correspondences between the facial part (in the before-makeup face) and the cosmetic (in the reference face) are built based on the face parsing network. (ii) Eye shadow, foundation and lip gloss are locally transferred with a global smoothness regularization.

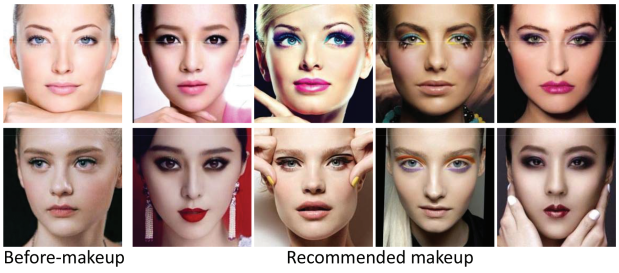

Figure 3: Two examples of makeup recommendation. The first columns are the before-makeup faces, other columns are the recommended reference faces.

3.1 Makeup Recommendation

The most important criterion of makeup recommendation is *personalization* [Liu *et al.*, 2014].

Females that look like are suitable for similar makeups. Given a before-makeup face, we find several similar faces in the reference dataset. The similarities are defined as the Euclidean distances between the deep features extracted by feeding the face into an off-the-shelf face recognition model [Parkhi *et al.*, 2015] named VGG-Face. The deep feature is the concatenation of the two ℓ_2 normalized FC-4096 layers. The VGG-Face model is trained based on VGG-Very-Deep-16 CNN architecture [Simonyan and Zisserman, 2014] and aims to accurately identify different people regardless of whether she puts on makeup, which meets our requirement. Therefore, the extracted features can capture the most discriminative structure of faces. Finally, the retrieved results serve as the reference faces to transfer their cosmetics to the before-makeup face. Figure 3 shows the makeup recommendation results. It illustrates that the recommended reference faces have similar facial shapes with the before-makeup faces, and therefore the recommendation is personalized.

3.2 Facial Parts vs. Cosmetics Correspondence

In order to transfer the makeup, we need build the correspondence between facial parts of the before-makeup face and the cosmetic regions of the reference face. As a result, the cosmetic can be between the matched pairs. Most of the correspondences can be obtained by the face parsing results, e.g., "lip" vs "lip gloss". The only exception is the eye shadow transfer. Because the before-makeup face does not have eye shadow region and the shape of the eyes are different, we need to warp the eye shadow shape of the reference face.

Face Parsing: Our face parsing model is based on the Fully Convolution Network (FCN) [Long *et al.*, 2014]. It is trained using both the before-makeup and reference faces. The 11 parsing labels are shown in Table 1. The network takes input of arbitrary size and produces correspondingly-sized output with efficient inference and learning.

When training the face parsing model, we pay more attention to the makeup relevant labels. For example, compared with "background", we bias toward the "left eye shadow". Therefore, we propose a *weighted cross-entropy loss* which is a weighted sum over the spatial dimensions of the final layer:

$$\ell(x;\theta) = \sum_{ij} \ell'\left(\mathrm{y}_{ij}, p\left(x_{ij};\theta\right)\right) \cdot w\left(y_{ij}\right), \qquad (1)$$

where ℓ' is the cross entropy loss defined on each pixel. y_{ij} and $p\left(x_{ij};\theta\right)$ are the ground truth and predicted label of the pixel x_{ij}, and $w\left(y_{ij}\right)$ is the label weight. The weight is set empirically by maximizing the F1 score in the validation set.

Because the faces in the collected datasets are in frontal view, and preprocessed by face detection and alignment[3]. In the testing phase, we enforce the symmetric prior and replace the prediction confidence of both point p and its horizontally mirrored counterparts $f\left(p\right)$ by their average $x_{p,c} = \frac{1}{2}\sum\left(x_{p,c} + x_{f(p),c}\right)$, where c denotes the channels. Like [Chen *et al.*, 2015; Papandreou *et al.*, 2015], the symmetric prior is only added in the testing phase currently. In the further we will explore how to enforce the structure prior in

[3] www.faceplusplus.com/

Labels	L/R eye	L/R eyebrow	inner mouth	L/R eye shadow	Up/Low lip (lip gloss)	background	Face (Foundation)
Subsets	both	both	both	ref	before (ref)	both	before (ref)

Table 1: 11 Labels from both before-makeup (referred to as "before") and reference face sets (referred to as "ref"). Certain labels, such as "L eye", belong to both dataset, while certain labels such as "L eye shadow" belong to reference face set only. "L", "R", "Up" and "Low" stands for "Left", "Right", "Upper" and "Lower" respectively.

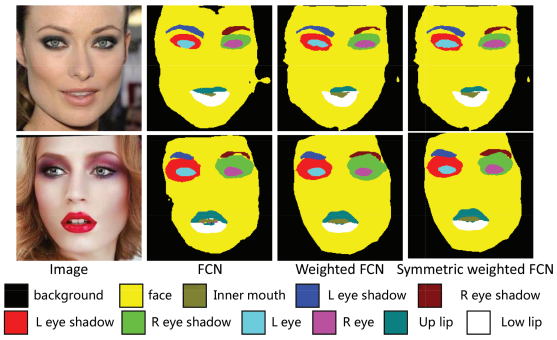

Figure 4: Two face parsing results. The input image, the results parsed by FCN, weighted FCN and symmetric weighted FCN are shown sequentially.

the training phase. Figure 4 shows the parsing results of the original FCN, weighted FCN, and symmetric weighted FCN.

Eye Shadow Warping: Based on the parsing results, most correspondences, e.g., "face" vs "foundation", are built. However, there is no eye shadow in the before-makeup face, therefore we need to generate an eye shadow mask in the before-makeup face. Moreover, because the shapes of eyes and eye brows are different in the face pair, the shape of eye shadow need to be slightly warped. More specifically, we get 8 landmarks on eyes and eye brow regions, including inner, upper-middle, lower-middle and outer corner of eyes and eye brows. Then the eye shadows are warped by the thin plate spline [Bookstein, 1989].

3.3 Makeup Transfer

Makeup transfer is conducted based on the correspondences among image pairs. Next we will elaborate how to transfer eye shadow, lip gloss and foundation. Keeping the facial structure should also be considered and incorporated into the final objective function.

Eye Shadow Transfer needs to consider both the shape and color. Let take the left eye shadow as an example. The right eye shadow is transferred similarly. Suppose s_r is the binary mask of left eye shadow in the reference face while s_b is the warped binary mask in the before-makeup face. Note that after eye shadow warping, s_r and s_b are of the same shape and size. Technically, eye shadow transfer is to replace s_b's deep feature representation in a certain layer (conv1-1 in this paper) with s_r. The mathematical form for the loss of left eye shadow transfer is $R_l(A)$:

$$\begin{aligned} A^* &= \arg\min_{A \in R^{H\times W\times C}} R_l(A) \\ &= \arg\min_{A \in R^{H\times W\times C}} \left\| P\left(\Omega^l\left(A\left(\mathrm{s}'_b\right)\right)\right) - P\left(\Omega^l\left(R\left(\mathrm{s}'_r\right)\right)\right)\right\|_2^2 \end{aligned} \tag{2}$$

Figure 5: Two results of eye shadow transfer. The before-makeup, reference and after-makeup eye shadow are shown.

where H, W and C are the height, width and channel number of the input image. $\Omega^l : R^{H\times W\times C} \rightarrow R^d$ is the d-dim feature representation of the conv1-1 layer of the face parsing model, A and R are the after-makeup face and reference face, respectively. s'_b and s'_r are achieved by mapping s_b to s_r from the data layer to the conv1-1 layer via the convolutional feature masking [Dai *et al.*, 2015; He *et al.*, 2015]. $A\left(\mathrm{s}'_b\right)$ and $R\left(\mathrm{s}'_r\right)$ are the eye shadow regions corresponding to the masks s'_b and s'_r. Similarly, we can define the loss function for right eye shadow $R_r(A)$. The results for both eye shadow transfer are shown in Figure 5, where both the color and shape of the eye shadows are transferred.

Lip Gloss and Foundation Transfer require transferring color and texture. The lip gloss $R_f(A)$ is defined as in (3).

$$\begin{aligned} A^* &= \arg\min_{A \in R^{H\times W\times C}} R_f(A) \\ &= \arg\min_{A \in R^{H\times W\times C}} \sum_{l=1}^{L} \left\| \Omega^l_{ij}\left(A\left(\mathrm{s}'_b\right)\right) - \Omega^l_{ij}\left(R\left(\mathrm{s}'_r\right)\right)\right\|_2^2 \end{aligned} \tag{3}$$

Here, L is the number of layers used. Technically, we use 5 layers, including conv1-1, conv2-1, conv3-1, conv4-1 and conv5-1. The Gram matrix $\Omega^l \in R^{N_l\times N_l}$ is defined in (4), where N_l is the number of feature maps in the l-th layer and Ω^l_{ij} is the inner product between the vectorised feature map i and j in layer l:

$$\Omega^l_{ij} = \sum_k \Omega^l_{ik}\Omega^l_{jk} \tag{4}$$

The results for foundation transfer are shown in Figure (6). The girl's skin is exquisite after the foundation transfer.

The upper lip gloss loss $R_{up}(A)$ and lower lip gloss loss $R_{low}(A)$ are defined in similar way as (3). And the lip gloss transfer results are shown in Figure 7. After lip gloss transfer, the colors of the lip are changed to the reference lip.

Structure Preservation term $R_s(A)$ is defined as in (2). The only difference is that every elements of s_b and s_r are 1.

Overall Makeup Transfer considers eye shadow, lip gloss and foundation, and also preserves the face structure.

$$\begin{aligned} A^* = \arg\min_{A \in R^{H\times W\times C}} & \lambda_e\left(R_l(A) + R_r(A)\right) + \lambda_f R_f(A) \\ & + \lambda_l\left(R_{up}(A) + R_{low}(A)\right) + \lambda_s R_s(A) + R_{V^\beta}(A) \end{aligned} \tag{5}$$

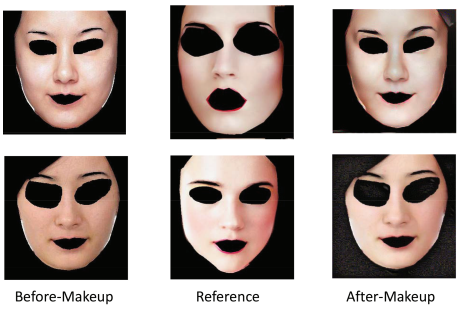

Figure 6: Two exemplars of foundation transfer. In each row, before-makeup, reference and after-makeup are shown.

Figure 7: Two exemplars of lip gloss transfer results. In each row, before-makeup, reference and after-makeup are shown.

To make the results more natural, the total variance term $R_{V^\beta} = \sum_{i,j} \left((A_{i,j+1} - A_{ij})^2 + (A_{i+1,j} - A_{ij})^2 \right)^{\frac{\beta}{2}}$ is added. $R_l(A)$, $R_r(A)$, $R_f(A)$, $R_{up}(A)$, $R_{low}(A)$ and $R_s(A)$ are the left, right eye shadow, foundation, upper, lower lip gloss and face structure loss. And λ_e, λ_f, λ_l and λ_e are the weights to balance different cosmetics. By tuning these weights, the lightness of makeup can be adjusted. For example, by increasing the λ_e, the eye shadow will be darker.

The overall energy function (5) is optimized via Stochastic Gradient Descent (SGD) by using momentum [Mahendran and Vedaldi, 2015]:

$$\mu_{t+1} \leftarrow m\mu_t - \eta_t \nabla E(A) \quad A_{t+1} \leftarrow A_t + \mu_t \qquad (6)$$

where the μ_t is a weighed average of the last several gradients, with decaying factor $m = 0.9$. A_0 is initialized as the before-makeup face.

4 Experiments

4.1 Experimental Setting

Data Collection and Parameters: We collect a new dataset with 1000 before-makeup faces and 1000 reference faces. Some before-makeup faces are nude makeup or very light makeup. Among the 2000 faces, 100 before-makeup faces and 500 reference faces are randomly selected as test set. The remaining 1300 faces and 100 faces are used as training and validation set. Given one before-makeup test face, the most similar ones among the 500 reference test faces are chosen to transfer the makeup. The weights $[\lambda_s \ \lambda_e \ \lambda_l \ \lambda_f]$ are set as $[10 \ 40 \ 500 \ 100]$. The weights of different labels in the weighted FCN are set as $[1.4 \ 1.2 \ 1]$ for {eyebrows, eyes, eye shadows}, {lip, inner mouth} and {face, background}, respectively. These weights are set empirically by the validation set.

Figure 8: The lightness of the makeup can be controlled.

Baseline Methods: To the best of our knowledge, the only makeup transfer work is Guo and Sim [Guo and Sim, 2009]. We also compare with two variants of Gatys *et al.*' method [Gatys *et al.*, 2015a]. They use CNN to synthesis a new image by combining the content layer from one image and the style layers of another image. The first variant is called NerualStyle-CC treating both before-makeup and reference faces as content. Another variant is named as NerualStyle-CS which uses the before-makeup as content and reference face as style. We cannot compare with other related makeup methods, such as, Tong et al. [Tong *et al.*, 2007], Scherbaum et al. [Scherbaum *et al.*, 2011] or Beauty E-expert system [Liu *et al.*, 2014] require before-and-after makeup image pairs, 3D information or extensive labeling of facial attributes. The proposed model can transfer the makeup in 6 seconds for an 224×224 image pair using TITAN X GPU.

4.2 Makeup Lightness

In order to show our method can generate after-makeup face with various makeup lightness, ranging from light makeup to dark makeup, we gradually increase certain makeup weights λ_e, λ_f and λ_l. Four results are shown in Figure 8. The first two rows use the same before-makeup and reference faces. The girl's eye shadows become gradually darker in the first row. In the second row, the lip color becomes redder while other cosmetics keep unchanged. The third and fourth rows show another example. The eye shadow and lip gloss are increasingly darker in the third and last row, respectively.

4.3 Comparison with State-of-the-arts

We compare with Guo and Sim[4] [Guo and Sim, 2009], NerualStyle-CC and NerualStyle-CS[5] [Gatys *et al.*, 2015a].

[4]We sincerely thank the authors for sharing their code.

[5]we use the code https://github.com/jcjohnson/neural-style and fine-tune the parameter for best visual effects.

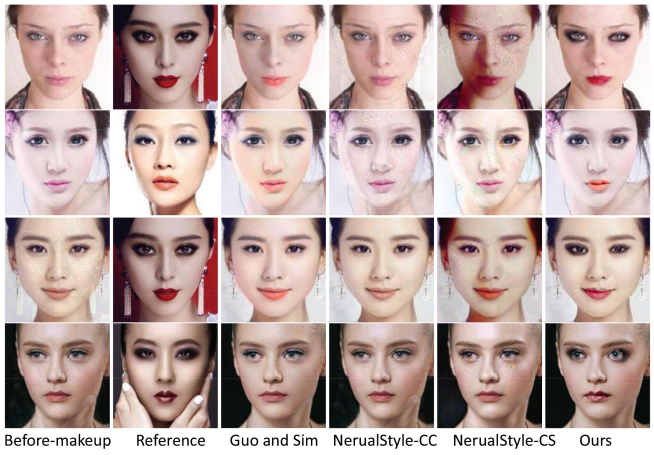

Figure 9: Qualitative comparisons between the state-of-the-arts and ours.

Follwing the evaluation metric in [Liu *et al.*, 2014], the comparison is conducted both qualitatively and quantitatively.

The **qualitative** results are shown in Figure 9. Both Guo and Sim [Guo and Sim, 2009] and ours produce naturally looking results. However, our result transfers the lip gloss and eye shadows with the same lightness as the reference faces, while Guo and Sim [Guo and Sim, 2009] always transfers much lighter makeup than the reference face. For example, in the first row, the lip gloss of both the reference face and our results are deep red. But the lip gloss of Guo and Sim is orange-red. In addition, the eye shadows of both the reference face and our results are very dark and smoky. However, Guo and Sim can only produce very light eye shadow. Compared with NerualStyle-CC and NerualStyle-CS [Gatys *et al.*, 2015a], our after-makeup faces contain much less artifacts. It is because our makeup transfer is conducted between the local regions, such as lip vs lip gloss while the NerualStyle-CC and NerualStyle-CS [Gatys *et al.*, 2015a] transfer the makeup globally. Global makeup transfer suffers from the mismatch problem between the image pair. It shows the advantages of our deep localized makeup transfer network.

The **quantitative** comparison mainly focuses on the quality of makeup transfer and the degree of harmony. For each of the 100 before-makeup test faces, five most similar reference faces are found. Thus we have totally 100×5 after-makeup results for each makeup transfer method. We compare our method with each of the 3 baselines sequentially. Each time, a 4-tuple, i.e., a before-makeup face, a reference face, the after-makeup face by our method and the after-makeup face by one of the baseline, are sent to 20 participants to compare. Note that the two after-makeup faces are shown in random order. The participants rate the results into five degrees: "much better", "better", "same", "worse", and "much worse". The percentages of each degree are shown in Table 2. Our method is much better or better than Guo and Sim in 9.7% and 55.9% cases. We are much better than NerualStyle-CC and NerualStyle-CS in 82.7% and 82.8% cases.

	much better	better	same	worse	much worse
Guo and Sim	9.7%	55.9%	22.4%	11.1%	1.0%
NerualStyle-CC	82.7%	14.0%	3.24%	0.15%	0%
NerualStyle-CS	82.8%	14.9%	2.06%	0.29%	0%

Table 2: Quantitative comparisons between our method and three other makeup transfer methods.

4.4 More Makeup Transfer Results

In Figure 10, for each reference face, we select five most similar looking before-makeup girls. Then the same makeup is applied on different before-makeup girls. It shows that the eye shadow, lip gloss and foundation are transferred successfully to the eye lid, lip and face areas. Note that our method can handle the makeup transfer between different facial expressions. For example, in Figure 10, the second girl in the left panel is grinning. However, the reference face is not smiling. Thanks to the localization property of our method, the lip gloss does not transfer to the teeth in the after-makeup face.

In Figure 11, for each before-makeup face, we select five most similar looking reference girls. This function is quite useful in real application, because the users can virtually try different makeup and choose the favorite one.

5 Conclusion

In this paper, we propose a novel Deep Localized Makeup Transfer Network to automatically transfer the makeup from a reference face to a before-makeup face. The proposed deep transfer network has five nice properties: with complete function, cosmetic specific, localized, producing naturally looking results and controllable makeup lightness. Extensive experiments show that our network performs better than the state-of-the-arts. In the future, we plan to explore the extensibility of the network. For example, one before-makeup face can be combined with two reference faces. The after-makeup face has the eye shadow of one reference face and lip gloss of another reference face.

Acknowledgments

This work was supported by National Natural Science Foundation of China (No.61572493, Grant U1536203), 100 Talents Programme of The Chinese Academy of Sciences, and "Strategic Priority Research Program" of the Chinese Academy of Sciences (XDA06010701).

References

[Bookstein, 1989] Fred L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (6):567–585, 1989.

[Chen *et al.*, 2015] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. In *International Conference on Learning Representations*, 2015.

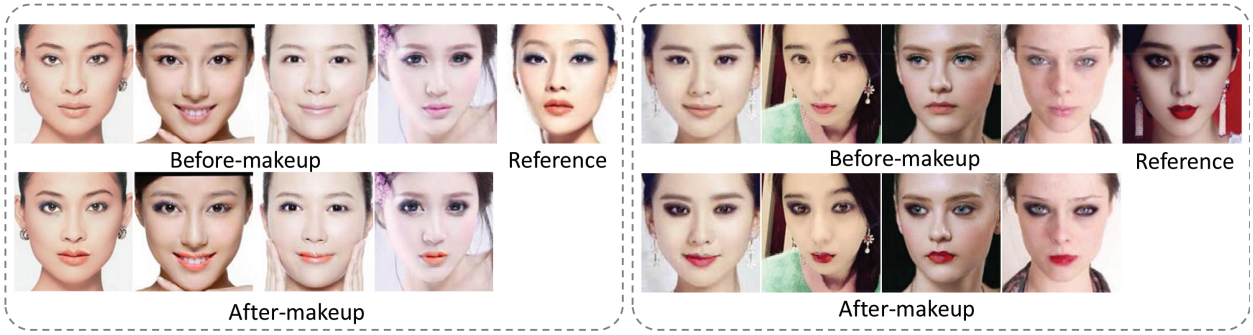

Figure 10: Different girls wear the same makeup.

Figure 11: The same girl wears different makeups.

[Cheng *et al.*, 2015] Zezhou Cheng, Qingxiong Yang, and Bin Sheng. Deep colorization. In *IEEE International Conference on Computer Vision*. 2015.

[Dai *et al.*, 2015] Jifeng Dai, Kaiming He, and Jian Sun. Convolutional feature masking for joint object and stuff segmentation. *Computer Vision and Pattern Recognition*, pages 3992–4000, 2015.

[Dong *et al.*, 2014] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *European Conference on Computer Vision*, pages 184–199. 2014.

[Dosovitskiy and Brox, 2015] Alexey Dosovitskiy and Thomas Brox. Inverting convolutional networks with convolutional networks. *CoRR*, abs/1506.02753, 2015.

[Dosovitskiy *et al.*, 2015] Alexey Dosovitskiy, Jost Tobias Springenberg, and Thomas Brox. Learning to generate chairs with convolutional neural networks. *Computer Vision and Pattern Recognition*, pages 1538–1546, 2015.

[Gatys *et al.*, 2015a] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. A neural algorithm of artistic style. *CoRR*, abs/1508.06576, 2015.

[Gatys *et al.*, 2015b] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. Texture synthesis and the controlled generation of natural stimuli using convolutional neural networks. In *Advances in Neural Information Processing Systems 28*, 2015.

[Goodfellow *et al.*, 2014] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, pages 2672–2680, 2014.

[Goodfellow *et al.*, 2015] Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. In *International Conference on Learning Representations*, 2015.

[Guo and Sim, 2009] Dong Guo and Terence Sim. Digital face makeup by example. In *Computer Vision and Pattern Recognition*, pages 73–79, 2009.

[He *et al.*, 2015] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 37(9):1904–1916, 2015.

[Kulkarni *et al.*, 2015] Tejas D Kulkarni, Will Whitney, Pushmeet Kohli, and Joshua B Tenenbaum. Deep convolutional inverse graphics network. *arXiv preprint arXiv:1503.03167*, 2015.

[Liang *et al.*, 2015] Xiaodan Liang, Si Liu, Xiaohui Shen, Jianchao Yang, Luoqi Liu, Jian Dong, Liang Lin, and Shuicheng Yan. Deep human parsing with active template regression. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2015.

[Liu *et al.*, 2014] Luoqi Liu, Junliang Xing, Si Liu, Hui Xu, Xi Zhou, and Shuicheng Yan. Wow! you are so beautiful today! *ACM Transactions on Multimedia Computing, Communications, and Applications*, 11(1s):20, 2014.

[Liu *et al.*, 2015] Si Liu, Xiaodan Liang, Luoqi Liu, Xiaohui Shen, Jianchao Yang, Changsheng Xu, Liang Lin, Xiaochun Cao, and Shuicheng Yan. Matching-cnn meets knn: quasi-parametric human parsing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[Long *et al.*, 2014] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *arXiv preprint arXiv:1411.4038*, 2014.

[Mahendran and Vedaldi, 2015] Aravindh Mahendran and Andrea Vedaldi. Understanding deep image representations by inverting them. *Computer Vision and Pattern Recognition*, pages 5188–5196, 2015.

[Papandreou *et al.*, 2015] George Papandreou, Liang-Chieh Chen, Kevin Murphy, and Alan L Yuille. Weakly- and semi-supervised learning of a dcnn for semantic image segmentation. *arxiv:1502.02734*, 2015.

[Parkhi *et al.*, 2015] Omkar M Parkhi, Andrea Vedaldi, Andrew Zisserman, A Vedaldi, K Lenc, M Jaderberg, K Simonyan, A Vedaldi, A Zisserman, K Lenc, et al. Deep face recognition. In *British Machine Vision Conference*, 2015.

[Ren *et al.*, 2015] Jimmy SJ Ren, Li Xu, Qiong Yan, and Wenxiu Sun. Shepard convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 901–909, 2015.

[Scherbaum *et al.*, 2011] Kristina Scherbaum, Tobias Ritschel, Matthias Hullin, Thorsten Thormählen, Volker Blanz, and Hans-Peter Seidel. Computer-suggested facial makeup. In *Computer Graphics Forum*, volume 30, pages 485–492, 2011.

[Shen *et al.*, 2015] Li Shen, Teck Wee Chua, and Karianto Leman. Shadow optimization from structured deep edge detection. *arXiv preprint arXiv:1505.01589*, 2015.

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

[Simonyan *et al.*, 2013] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. *CoRR*, 2013.

[Tong *et al.*, 2007] Wai-Shun Tong, Chi-Keung Tang, Michael S Brown, and Ying-Qing Xu. Example-based cosmetic transfer. In *Computer Graphics and Applications*, pages 211–218, 2007.

[Xu *et al.*, 2015] Li Xu, Jimmy Ren, Qiong Yan, Renjie Liao, and Jiaya Jia. Deep edge-aware filters. In *International Conference on Machine Learning*, pages 1669–1678, 2015.

[Yan *et al.*, 2014] Zhicheng Yan, Hao Zhang, Baoyuan Wang, Sylvain Paris, and Yizhou Yu. Automatic photo adjustment using deep neural networks. *CoRR*, abs/1412.7725, 2014.