Makeup based on segmentation and local transfer

Yuanhan Zhang, Haidong Cui, Yidong Li and Zheng Feng

September 11, 2019

Makeup based on segmentation and local transfer

1st Yuanhan Zhang
School of Computer and Information Technology
Beijing Jiaotong University
Beijing, China
18120454@bjtu.edu.cn

2nd Haidong Cui
Taishan Information Technology Co.ltd
Taishan, China
fengzheng@ts-it.cn

3rd Yidong Li
School of Computer and Information Technology
Beijing Jiaotong University
Beijing, China
ydli@bjtu.edu.cn

4th Zheng Feng
Taishan Information Technology Co.ltd
Taishan, China
573470610@qq.com

***Abstract*—In this paper, we introduce an automatic beauty system based on semantic segmentation and local transfer. The user provides a source image and a target image as a reference. Our system create a new makeup upon the face of source image with target image as reference. More precisely, we transfer each makeup areas(lip gloss, eye shadow, eyebrow)to corresponding parts of the face. One major advantages of our approach is we create this new makeup for source image locally, this not only caters to the requirement of users but also renders makeup upon face of source image more naturally. The effectiveness of the proposed are demonstrated in the experiments.**

I. Introduction

At present, idol have gradually become the trend of the fashion, people learn the most popular wear and makeup by learning from their idol. Imagine such a scene, one of your favorite idol wears a makeup that you really like today, you really want to imitate her makeup, but if you want to make up by yourself, this not only requires a lot of time, but also requires some experiences and makeup skills. Before making up by yourself, it would be extremely helpful if you can preview the makeup effects on your own face. However.

In this paper, we proposes an automatic beauty system based on semantic segmentation and local transfer1. Our approach is in view of the experience that makeup can change the color and shape of corresponding areas of the human face. For the imitation makeup, even if the corresponding area of the imitator is exactly the same as the reference picture, it can be done by makeup based on our experiments. So in our approach, we introduce approach of local transfer to create makeup for face in source image. We will introduce our approach as fallow: Firstly we extracts the feature points of the face from the source image and the target image, and then uses the alignment approach to align the target image to the source image. Since the size of the makeup areas in the two images after alignment is not much different, if the makeup areas of the target image can be accurately extracted, after those areas are respectively transferred to the source image, we can achieve the effect of imitation makeup of source image, and such makeup is more natural. We define this kind of transfer as local transfer. For the extraction of makeup areas - eyebrows, eyes, mouth, we use the latest approach of semantic segmentation, because the current semantic segmentation has reached the level of the pixel level, so the segmentation results are very fine, almost no loss of detail. In addition, this local-based transfer approach can better meet the different needs of users. Not every user wants to imitate all the makeup of target image. The mainly flow of our approach are exhibit on this picture2.

Fig. 1. The simple demo of our approach,source image refers to the image you want to makeup,target image refers to the reference image. Makeup areas refers to eyes, eyebrows, mouth

Contribution:For our approach, we introduced a pixel-level image segmentation approach to extract the makeup areas. Each part of the makeup areas in target image is transferred into the source image. Since the image segmentation is pixel level, we can extract each part of the makeup area more accurate than the approach based on feature points, so applying semantic segmentation to extract makeup areas makes the

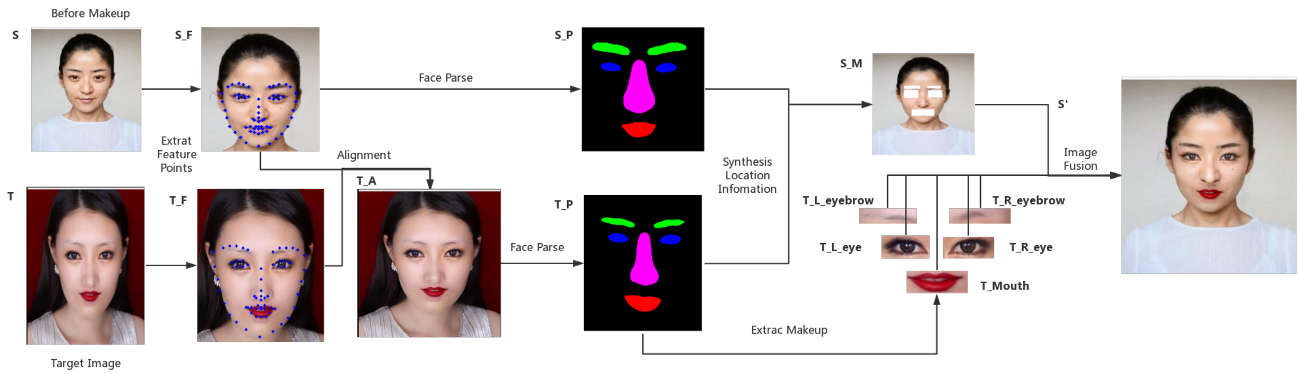

Fig. 2. Our approach is divided into 5 steps. (1) Prepare a source image **S** as a photo for preparing makeup and a target image **T** as a reference photo for makeup. (2) Feature point extraction is performed on the source image and the target image, respectively **S_F,T_F** (3)Align the target image to the source image using feature points.**T_A** (4) Perform a face parse on **S_F,T_A**to get **S_P,T_P** (5) To extract the makeup areas of the target image **T_L_eye,T_R_eye,T_Mouth,T_L_eyebrow,T_R_eyebrow**; combine the makeup/to makeup areas information of **S** and **T** to obtain **S_M**(6) Transfer **T_L_eye,T_R_eye,T_Mouth,T_L_eyebrow,T_R_eyebrow**to **S_M** by image fusion.

later image fusion more precise, and can effectively solve the occlusion problems in other approach. At the same time, the local-based image fusion approach can also make the makeup look more natural and meet different needs of users.

II. Related Work

At present, there are few researches on digital makeup. In the current study which can be divided into deep learning based approach and traditional image processing approach. Next, we will briefly introduce some approach that are similar to our work.

A. Facial Makeup Studies

For the traditional approach of makeup based on image processing, Tong [1], proposed the "cosmetic-transfer" approach, which transfer the makeup to the source image by providing the example pairs as a reference(before makeup image(**b_f**) and after makeup image(**a_f**)). This approach first aligns the **a_f** to the source image using TPS(face warp), and then calculates the pixels relationship between **b_f** and **a_f** in the example pair. By applying this relationship to **a_f** and source image, achieving the effect of makeup. This approach requires two preconditionFirstly, it requires the example pairs, but usually such example pairs are hard to find in real-time application. Secondly the color difference between the source image and the example pair need to be very small. Guo [2] proposed the makeup transfer task, which extracts the source image and the target image by three levels - Face Structure, Skin Detail and Color. By analyzing the source image and target image relationships layer by layer, then utilize these relationships and alignment approach comprehensively to achieve the makeup for the source image. For the Guo's approach, only one target image is needed, instead of a set of example pairs. However, both of these approach actually using the whole face area of after-alignment target image to carry out follow-up work. In fact, such approach will change many details of the no need makeup area upon face and cause the overall makeup effect to be unreal.

For deep learning based approach, liu [3] proposes a approach based on Deep Image Synthesis to achieve the effect of makeup of source image by directly exchanging the corresponding deep features of related area. For deep learning-based algorithms, a large amount of data set are inevitable, and such work also takes a lot of time.

B. Semantic Segmentation

Semantic segmentation is a classification method. In our approach, we classify human face into five classed: eyebrows, eyes, nose, mouth, and background. There are many models for semantic segmentation, Deeplab v3 [4] model from Chen has achieved the best results in the semantic segmentation data set. We use Deeplab v3 as our backbone of semantic segmentation.

III. Methodology

In this Section, we sequentially introduce our makeup system in detail.

A. Face alignment

In order to ensure that the each part of makeup areas of the target image and corresponding area of the source image are as same as possible in shape and size, we use the face alignment algorithm to align the target image to the source image. After alignment, the makeup areas of the target image are similar to the corresponding areas of the source image. In our work, face alignment consists of two steps, the first step is to extract facial feature points, and second step is feature point alignment. For extracting feature points, Our model are based on FAN-2D Bulat [5], which is the state of art model for face alignment, the model gets 68 facial feature points. Since what we want to do is just the alignment of the makeup area, the makeup area refers to the eyes, nose, mouth, eyebrows, so we only need to use 51 of the 68 feature points, excluding the face contour. And the details will be discussed in the experimental module. For aligning feature points, we use the RANSAC. RANSAC (Random sample consensus) is the basic method of feature point alignment. For any feature point x on the target image, if it corresponds to a feature point y on the source image, then y can be obtained by affine transformation of x.

$$\begin{bmatrix} \vec{y} \\ 1 \end{bmatrix} = H \begin{bmatrix} \vec{x} \\ 1 \end{bmatrix}$$
$$H = \begin{bmatrix} A & \vec{b} \\ 0, ..., 0 & 1 \end{bmatrix} \tag{1}$$

H is called affine matrix, and the purpose of RANSAC is to make the most feature point pairs can be affine transformed with H.

B. Face parse

In order to transfer various makeup areas of the target image to the source image, we need to accurately get the fine segmentation of the makeup areas of the target image and corresponding areas of the source image.

Our face parse model is based on Deeplab v3, Chen [4]. We use the Deeplab v3 model trained on the HELEN dataset [7] [8]. The entire face is classify into 5 classes. For the $output\_stride = 16$ in Deeplab v3, if the input image resolution is too small then the information loss can be too large, so we need to preprocess the input image to 513×513. In the final layer, we used cross entropy to calculate the loss3.

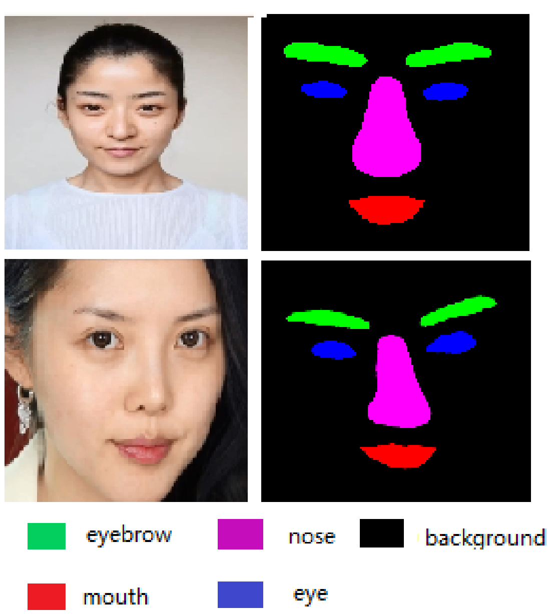

Fig. 3. Two face parse result

C. Local Transfer

We use the image fusion approach locally to transfer various makeup areas of the target image to source image.

Comparing to previous work which using the whole face area of after-alignment target image to carry out follow-up work. We extract each part of the makeup area, and then using each part of the makeup area to carry out follow-up work, making the processing result more natural than previous work. At the same time, this approach also caters to the needs of users. After we use face parse to get the precise areas of each makeup areas of the face, we customize different warp approach. The most tricky part is the eye shadow, because there are certain pixel differences between the eye shadow and the eye. In order to extract the eye shadow of the target image, we also need extract the areas between the eye and the eyebrow. Further, in order to make the fusion result more natural, we take the average of the lower bound of the eyebrow of the target image and the lower bound of the eyebrow of the source image as upper bound of the eye shadow-**down_eyebrow**.

$$down\_eyebrow = \frac{1}{2}(down\_eyebrow\_target + down\_eyebrow\_source) \tag{2}$$

Experiments have shown that such effects are more natural. For eyebrows and the mouth, since the makeup for the eyebrows and the mouth is upon the eyebrows and the mouth, so these two parts of makeup can be directly transfer for fusion. After extracting the makeup of the target image, we can fuse the makeup of the target image to the source image as needed. For image fusion, we use the Poison Blending approach, Patrick [6].

IV. Experiments

In this section, we will discuss in detail the reasons why different parameters or methods lead to different experimental results. And we will exhibit comparisons with state of arts approach of makeup(Excluding approach of deep learning).

A. Face alignment

Facial Feature Points In the study of makeup, the previous work directly depends on all the facial feature points for alignment, including facial contours, nose, mouth, eyes, eyebrows-five areas.(in the 68 feature points, face contour:17; each eye:6; nose:9; each eyebrow:5; mouth:20) Since we only need to align the corresponding makeup areas, we exclude the face contour feature points and keep the rest of 51 feature points. Most importantly, since the facial contour feature points (17) account for a large proportion of the 68 feature points, RANSAC based on all feature points has a great influence on the alignment effect of the makeup areas we need. As can be seen from 4, in order to align 68 feature points, the eye of the target image after alignment appears to be significantly smaller than our approach.

Alignment: In the current makeup work, the most commonly used alignment approach is Thin Plate Spline (TPS), but this approach will cause serious distortion5of the target image. because this approach forcibly guarantees the feature point pairs to align. Although RANSAC approach only aligns

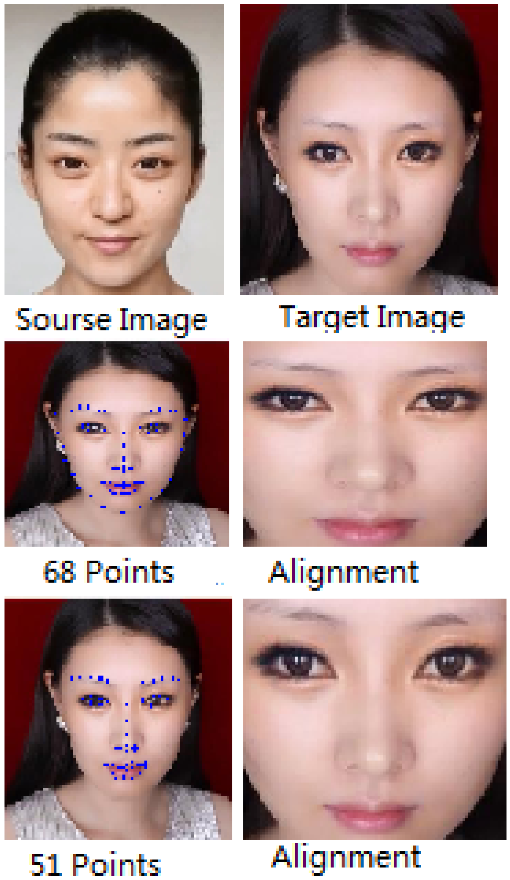

Fig. 4. The effect of the number of different feature points on alignment

most of the feature points, there is no distortion of the target image.

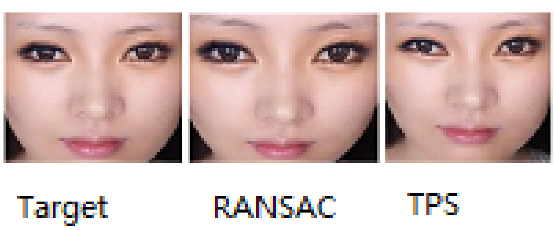

Fig. 5. The effect of different alignment approach

B. Extract Makeup Areas

Our approach for the extraction of the target image Makeup areas based on the location information of source image. More accuratelythe location information of makeup areas of source image. Since the target image and the source image have been aligned, the location of facial makeup areas from two images are basically the same. But in order to make the image fusion more natural on the source image, we choose the location information of makeup areas in the source image.

There are two approach for extracting makeup area of the target image. One approach based on facial feature points, and the other based on face parse. Because face parse uses pixel-level semantic segmentation which can do finer segmentation, this approach can be further used to solve the problem of occlusion of the source image. Suppose a scene where the source image's eyes and parts of the eyebrows are obscured by the hair, and the user does not want the makeup to cover that hair when wearing makeup. As can be seen from 6, the image itself has an occlusion area (we use a black box to simulate blocking by the hair), we use **SO** to represent the occluded source image. We segment the face of **SO** with two

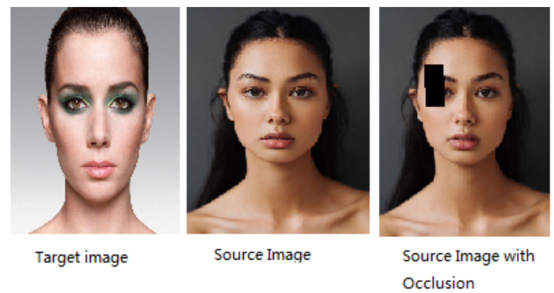

Fig. 6. Source image with occlusion

different segmentation approach. It can be seen from 7 that the segmentation approach based on the facial feature points (above) also causes makeup to cover the occlusion area, and the segmentation approach based on the semantic segmentation (down) does not; The makeup effect based on the face parse

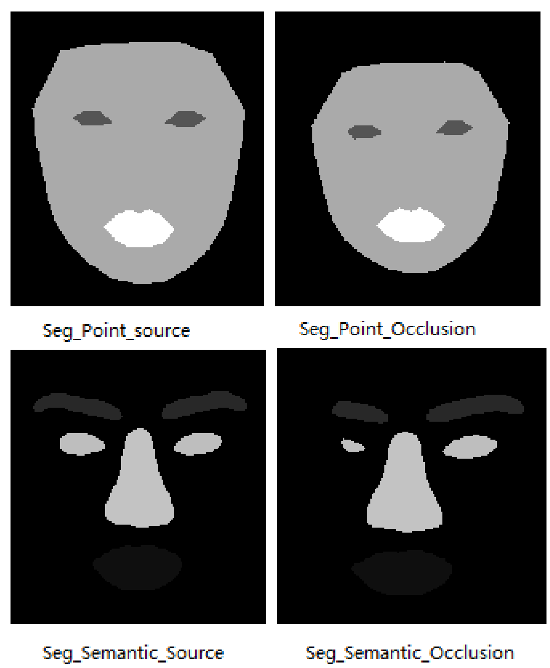

Fig. 7. segmentation result of source image with occlusion

segmentation approach can deal with occlusion8

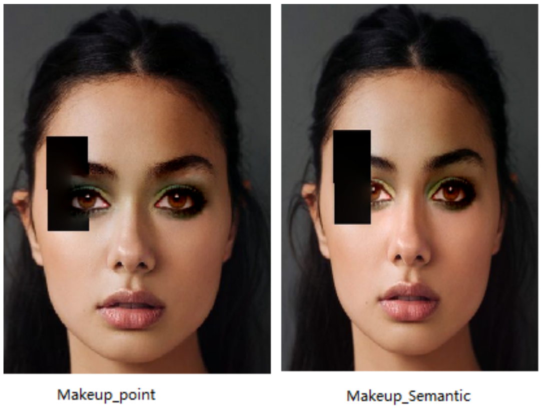

Fig. 8. makeup result with occlusion

C. Comparison with state of arts

As far as we know, the current no-deep learning makeup works are mainly Guo [2]and Tong [1], but Tong's work requires two reference images, obviously this work is very inconvenient and does not match our application scenario, so in the comparison test 9, we only use the work of Guo [2] [1]. From various experimental results, we can have the following conclusions: 1. From the perspective of the overall skin color, since our approach is a local-based approach, the facial skin color of the experimental results does not have too much influence. However, Guo's approach obviously makes the skin color of the experimental results very unnatural. 2. For the eye shadow and lip color, our approach makes the eye shadow and lip color of the experimental result closer to the target image, but the approach of Guo makes the eye shadow of the experimental result shallower than target image, and even let the teeth paint some lip color. 3. Guo's approach does not pay attention to the change of eyebrow shape. The change of eyebrow shape also plays a very important role in makeup. 4. For occlusion scene of the source image, our approach is also significantly more natural than the approach of Guo. 10.

V. Conclusion

This paper proposes a semantic segmentation-based automatic beauty system we introduced a pixel-level image segmentation approach to extract the face. Each organ of the face in target image is integrated into the source image. Since the image segmentation is pixel level, we can makes the later image fusion more precise, and we can also effectively solve the occlusion scene of the source image. At the same time, the local-based image fusion approach can also make the makeup look more natural and meet the different needs of users.

Acknowledgment

This work is supported by Natural Science Foundation of China (NSFC) under Grants 61672088, 61790573 and 61790575,The corresponding author is Yidong Li.

[1] We use the codes https://github.com/Siddharth24Khera/Auto_MakeUp_Transfer.

References

[1] W.-S. Tong, C.-K. Tang, M. S. Brown, and Y.-Q. Xu. Example-based cosmetic transfer. In Proc. Pacific Confer-ence on Computer Graphics and Applications, 2007.

[2] Dong Guo and Terence Sim. Digital face makeup by example. In Computer Vision and Pattern Recog-nition, pages 7379, 2009.

[3] Liu S , Ou X , Qian R , et al. Makeup like a superstar: Deep Localized Makeup Transfer Network[J]. 2016.

[4] Chen L C , Papandreou G , Schroff F , et al. Rethinking Atrous Convolution for Semantic Image Segmentation[J]. 2017.

[5] Bulat A , Tzimiropoulos G . How far are we from solving the 2D & 3D Face Alignment problem? (and a dataset of 230,000 3D facial landmarks)[J]. 2017.

[6] P'erez, Patrick, Gangnet M , Blake A . Poisson image editing[J]. ACM Transactions on Graphics, 2003, 22(3):313.

[7] Le, V.; Brandt, J.; Lin, Z.; Bourdev, L.; Huang, T. S.Interactive facial feature localization. In: Computer VisionECCV 2012. Lecture Notes in Computer Science, Vol. 7574. Fitzgibbon, A.; Lazebnik, S.; Perona, P.; Sato, Y.; Schmid, C. Eds. Springer, Berlin, Heidelberg, 679692, 2012.

[8] Smith, B. M.; Zhang, L.; Brandt, J.; Lin, Z.; Yang, J. Exemplar-based face parsing. In: Proceedings of the IEEE Conference on Computer Vision

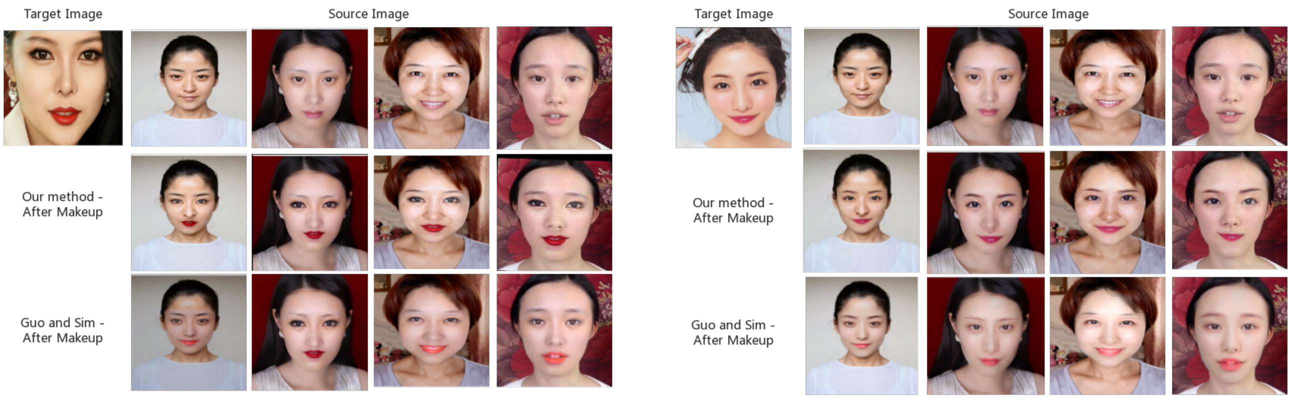

Fig. 9. Qualitative comparisons between the approach of Guo and ours

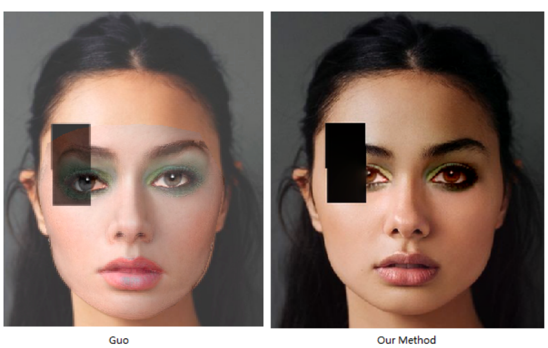

Fig. 10. makeup result with occlusion