Supplementary Material for RamGAN: Region Attentive Morphing GAN for Region-Level Makeup Transfer

A RamGAN without RAMM

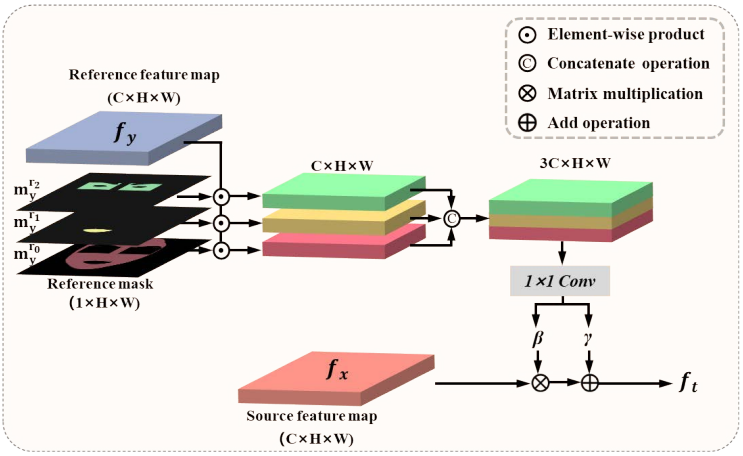

Fig. 1: An overview of RamGAN without RAMM.

For ablation study, when RAMM is removed, the makeup transfer process is shown in Fig. 1. Instead of calculating the Region Attentive Matrices, we directly multiply the reference feature f_y with reference mask element-wisely and combine the result to produce the Makeup Tensor. Then the output Make Tensor is fed into two 1×1 convolution layers to produce two Region Make Tensors, γ and β. The whole process can be expressed as:

$$\begin{aligned} MT &= \mathit{Cat}\big(f_y \odot m_y^{r_k}\big) \\ \gamma &= \mathit{Conv}_\gamma(MT), \beta = \mathit{Conv}_\beta(MT) \\ f_t &= \gamma f_x + \beta, \end{aligned} \tag{1}$$

where *MT* denotes Makeup Tensor, f_t is the transferred feature, *Cat* and *Conv* represent concatenation and 1×1 convolution, respectively.

B Further Study of RMTs

In addition, we also train the RamGAN to produce two RMTs with single channel, i.e., $\hat{\gamma} \in \mathbb{R}^{1\times H\times W}$ and $\hat{\beta} \in \mathbb{R}^{1\times H\times W}$. We duplicate and expand along the channel dimension like PSGAN to perform makeup transfer for comparasion. The first row

of Fig. 2 shows the partial makeup results with single channel RMTs ($\hat{\gamma}$ and $\hat{\beta}$), our RamGAN (2nd row). One can observe that the makeup style of the results transferred by single channel RMTs are not similar to the reference, especially the color of lips and eye shadow. The comparison suggests that our RMTs γ and β contain more spatial-aware information for region-level makeup transfer.

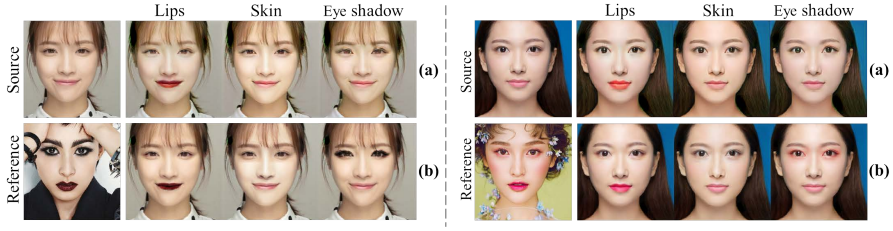

Fig. 2: Comparison of partial makeup transfer between single channel RMTs (a) and the proposed RamGAN (b).

We also calculate the SSIM and FID score to evaluate the quality of images transferred using the two RMTs in Tab. 1.

Table 1: The SSIM/FID of two RMTs.

Dataset	1-Channel RMTs	Ours
MT	0.72/46.66	**0.94/13.20**
M-Wild	0.62/46.67	**0.95/16.70**
Makeup	0.90/44.34	**0.95/10.67**

We now visualize more example channels of the RMTs (γ and β). As shown in Fig. 3, each example channel of γ or β focuses on different facial regions, which suggests that our RMTs contain more spatial-aware information for region-level makeup transfer.

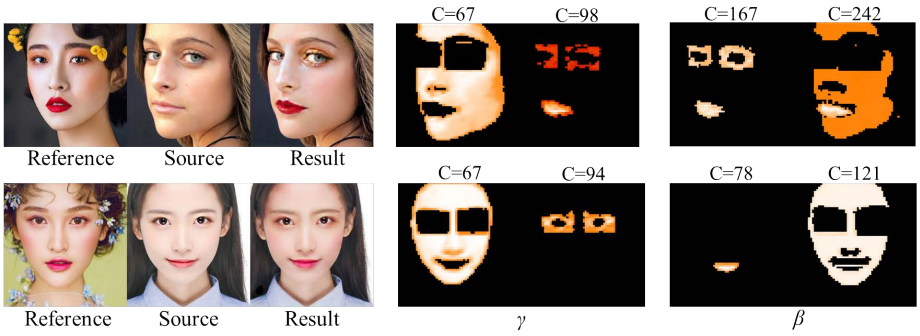

Fig. 3: More visualization of γ and β

C More Visual Results

We now perform more makeup experiments and show more visual results in this section. Firstly, we try to makeup more challenging regions like left faces and right faces,

including more challenging variations like expressions, and show the makeup results in Fig. 4. As shown in the figure, our RamGAN presents good performance in controlling makeup of the challenging regions, e.g. in the left part, only the left/right regions are transferred and the other half faces are well preserved. What's more, the transition boundary between makeup and non-makeup region of results are smooth and natural as well. The eye regions and mouth regions in the right part are well transferred as well, without producing any unnatural boundary between makeup and non-makeup regions, even when there are large pose and expression variations between source and reference faces.

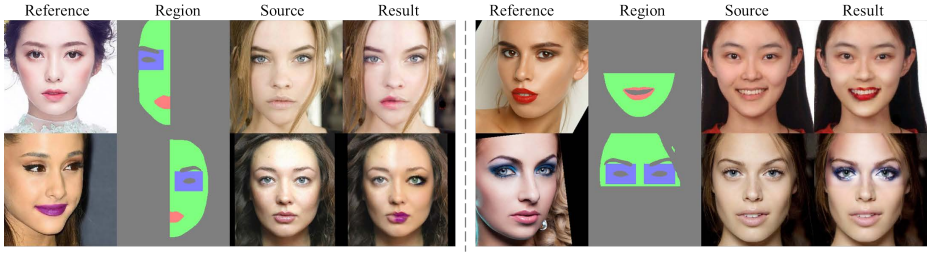

Fig. 4: Partial face makeup transfer.

The spatial region attention mechanism can easily enable us to do makeup interpolations between different reference makeups by a coefficient $\alpha \in [0, 1]$. Given two makeup images $y_1, y_2 \in \mathcal{Y}$, we can separately extract the corresponding makeup-related feature f_{y_1} and f_{y_2} by Feature Extractor. And then we compute $\alpha f_{y_1} + (1-\alpha) f_{y_2}$, and feed the weighted feature into the generator to yield smooth transition between two reference makeup images, by changing the value of α.

Fig. 5: Face level interpolation. The makeup style of middle image is interpolated from two reference styles.

Fig. 5 shows the result interpolated between reference styles shown in the left and right of the figure. A smooth transition can be achieved: e.g. the lip color gradually changes from purple to pink and the color of skin and eye shadow also changes smoothly from the left style to the right style.

Fig. 6 shows the interpolated results for three regions, i.e., skin, lips and eye shadow. In the figure, the first two images in the first row are source and reference image, respectively. Each row/column depicts a smooth and natural transition for different facial regions.

Fig. 7 **(a)** shows the transfer results compared with PSGAN++ implemented by us and PSGAN, both qualitatively and quantitatively. The FID and SSIM are only calculated on the MT test set. We also provide a qualitative comparison of step-by-step makeup transfer with PSGAN, PSGAN++ and CPM, as shown in Fig. 7 **(b)**.

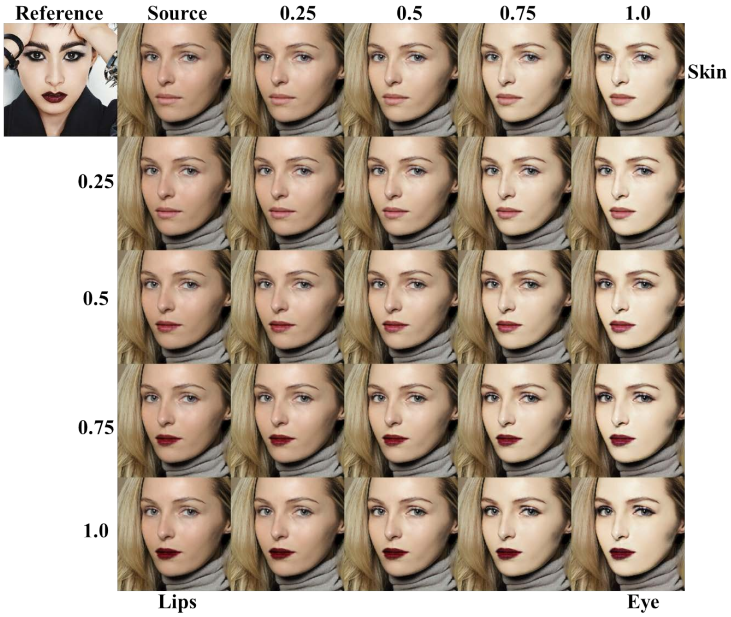

Fig. 6: Region-level interpolation results.

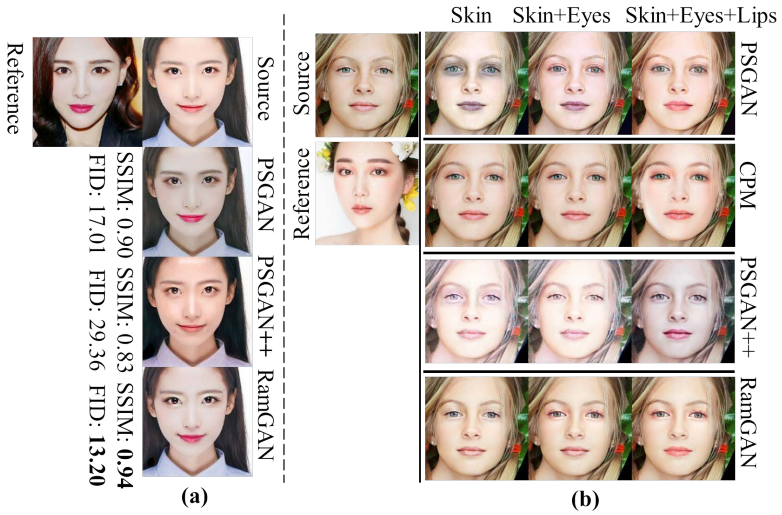

Fig. 7: **(a)** Qualitative and quantitative comparisons of PSGAN, PSGAN++ and RamGAN. **(b)** Step-by-step makeup transfer.

It is a meaningful but challenging task to transfer makeup for a person in the video, since the pose and expression of a face in the video are continuously changing. To examine the generalization of our RamGAN, we randomly select several frames in the video and perform makeup transfer, as shown in Fig. 8 **(a)**. We also perform makeup transfer with different makeup styles for an example man, as shown in Fig. 8 **(b)**

Fig. 8: **(a)** Video makeup transfer results. **(b)** A man makeup with different styles.

D Hyper Parameters Tunning

We now provide a series of experiments on hyper parameters tunning, both qualitatively and quantitatively. The hyper parameters $\lambda_{adv} = 1$, $\lambda_{cyc} = 10$, $\lambda_{per} = 0.005$ are the same as those in PSGAN. The ablation studies on the other hyper parameters tuning are presented in Fig. 9. The FID and SSIM are only calculated on the MT test set. It can be seen that the improvements are mainly due to the proposed modules (without each module *i.e.*, column $\lambda = 0$ or with each module, *i.e.*, column $\lambda > 0$).

E Analysis of the Attention Matrix in PSGAN

We first review the mathematical formulation of the attention matrix and give another two mathematical formulation to calculate the attention matrix. As proposed in PSGAN, the attention matrix can be calculated as

$$A_{i,j}^{r_k} = \mathit{Softmax}\left(\mathit{Cat}(w f_{x_i}^{r_k}, \mathbf{p}_{x_i}^{r_k})^T \cdot \mathit{Cat}(w f_{y_j}^{r_k}, \mathbf{p}_{y_j}^{r_k})\right), \tag{2}$$

where *Softmax* and *Cat* denote softmax activation layer and concatenation operation, respectively. $\mathbf{p}$ indicates the relative positions to 68 facial landmarks. w is the weight for visual features and set as 0.01 as suggested by the authors of PSGAN. A^{r_k} represents the attention matrix for different regions, and the subscript i and j indicate the i^{th} and j^{th} pixel in the feature map $f_x^{r_k}$ and $f_y^{r_k}$, respectively.

In order to further illustrate that it is unreasonable to take the relative position as primary concern, we only use either the visual feature or relative position to calculate the attention matrix. The two attention matrices can be expressed as

$$A_{i,j}^{r_k} = \mathit{Softmax}\left((f_{x_i}^{r_k})^T \cdot f_{y_j}^{r_k}\right), \tag{3}$$

$$A_{i,j}^{r_k} = \mathit{Softmax}\left((\mathbf{p}_{x_i}^{r_k})^T \cdot \mathbf{p}_{y_j}^{r_k})\right). \tag{4}$$

Fig. 9: Qualitative and quantitative experiments on hyper parameters tuning. The text in red denotes our default hyper parameters.

Fig. 10 shows attention map on reference image (first row) and the corresponding transferred image (second row). In column **(b)** and **(c)**, we use Eq. 4 and Eq. 3 to calculate the attention matrix, respectively. One can observe that the transferred result, generated by using the visual feature (Eq 3) only is better than the result generated by only using relative position (Eq. 4). The comparison further illustrates that the visual features play an important role in makeup transfer. In column **(d)** and **(e)** of the figure, we also show the attention map and the transferred result of PSGAN and our proposed RamGAN.

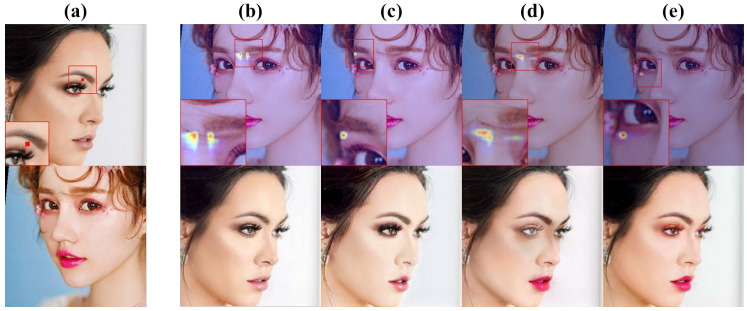

Fig. 10: Attention map on the reference images with different methods. **(a)** Source (first row) and reference (second row). **(b)**-**(e)** Attention map on the reference image (first row) and the transferred images (second row).