A. Pseudo Code

We provide the pseudo-code of DreamMakeup for RGB and textual guidance in Algorithm 1. Makeup transfer based on a reference image can be easily implemented using the same method, substituting the transformation $\mathcal{T}_{RGB}$ with $\mathcal{T}_{ref}$ and employing warping and histogram matching algorithms instead of RGB matching.

Algorithm 1 DreamMakeup with RGB and text guidance

Input: Source image $\boldsymbol{x}_0$, early-stop timestep $t^* \leq T$, RGB scaling coefficient $0 \leq \alpha \leq 1$, target RGB color μ_{tgt}, reference makeup image $\boldsymbol{x}_{ref}$, textual prompts for (inversion, editing) C_{inv}, $\{C_{edit,s}\}_{s=1}^{N}$, degree of composition $\{\alpha_s\}_{s=1}^{N}$.
Output: Image with makeup transformation $\tilde{x}_0$.

$\boldsymbol{z}_0 = \mathcal{E}(\boldsymbol{x}_0)$

1. Early-stopped DDIM inversion
for $t = 1$ **to** t^* **do**
  $\boldsymbol{z}_t = \text{DDIM-Inv}(\boldsymbol{z}_{t-1}, t, c)$ (Sec. 4.1)
end for
$\hat{\boldsymbol{z}}_0(t^*) = \frac{1}{\sqrt{\bar{\alpha}_{t^*}}}(\boldsymbol{z}_{t^*} - \sqrt{1-\bar{\alpha}_{t^*}}\boldsymbol{\epsilon}_\theta(\boldsymbol{z}_{t^*}, t^*))$.
$\hat{\boldsymbol{x}}_0(t^*) = \mathcal{D}(\hat{\boldsymbol{z}}_0(t^*))$

2. Pixel-domain Diffusion Guidance
$\hat{\boldsymbol{x}}_{new} = \mathcal{T}_{RGB}\big(\mu_{src}(\hat{\boldsymbol{x}}_0(t^*)), \mu_{tgt}; \alpha\big)$
$\tilde{\boldsymbol{z}}_{t^*} = \sqrt{\bar{\alpha}_t}\mathcal{E}(\hat{\boldsymbol{x}}_{new}) + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}(\boldsymbol{z}_{t^*}, t^*, c)$

3. Reverse sampling with cross attention composition
for $t = t^*$ **to** 1 **do**
  $\tilde{z}_{t-1} \leftarrow \text{ReverseDDIM}\Big(\tilde{z}_t; t, \text{Composition}\big(\{\alpha_s\}_{s=1}^{N}, \{C_{edit,s}\}_{s=1}^{N}\big)\Big)$
end for
$\tilde{x}_0 = \mathcal{D}(\tilde{z}_0)$

B. Additional analysis

We provide in-depth analysis on the core components of DreamMakeup. Additional qualitative results are also provided in Fig. 19.

B.1. Ablation studies

Effects of gradation smoothing. In the generating process of eye mask, gradation smoothing is essential. Without gradation smoothing, the edges of eye masks are accentuated, resulting in an unnatural outcome. Fig. 11 demonstrates that graduation smoothing makes the edge of the eye shadow natural and realistic.

Figure 11. The effect of the gradation smoothing of eyeshadow mask. Please zoom in for detailed inspection.

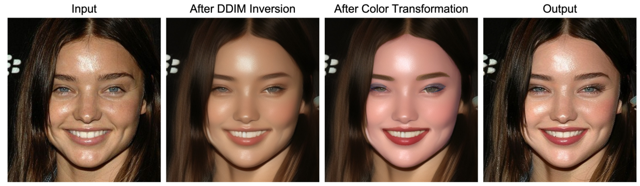

Figure 12. Ablation study on DDIM inversion, coloring, and reverse sampling. Text prompts guide the harmonization of unnatural color transitions into a cohesive aesthetic style during sampling.

B.2. LoRA variation

We mainly utilized Dreamshaper[1], ArienMixXL [2], and BKG1[3] LoRA weights. Fig. 13 shows the experimental results of using other LoRA weights. For comparison, asian beauty v2[4] , Korean Alike[5], Asian Cute Face[6], koreanDoll-Likeness v15[7], PMN 2[8] are used. The results are made with only text guidance, where the prompts are `"deep red lip"` and `"heavy eye makeup"`. The results demonstrate how diverse makeup styles can be achieved by varying LoRA weights. In this paper, we mainly leverage BKG1 LoRA which shows better identity preservation and semantic alignment.

B.3. Additional Results

To demonstrate the robustness of our method, we test its performance on a variety of challenging conditions, including images with low resolution, occlusions, and dark lighting, etc with results shown in Fig. 16. Although our primary focus is on generating natural, daily makeup, the framework's flexibility allows it to be applied to more extreme and artistic makeup styles as well (Fig. 17). Furthermore, the user can customize the target region for color transformation, extending the application of DreamMakeup beyond facial makeup to related tasks such as hair, eyebrow, and pupil coloring (Fig. 18).

[1] https://civitai.com/models/4384/dreamshaper
[2] https://civitai.com/models/118913/sdxl-10-arienmixxl-asian-portrait
[3] https://civitai.com/models/203947/beautiful-korean-girl-bkgv1
[4] https://civitai.com/models/76883/2731-pretty-asian-face-asian-beauty-faces
[5] https://civitai.com/models/193777/korean-alike-by-noerman
[6] https://civitai.com/models/26914?modelVersionId=32215
[7] https://civitai.com/models/26124/koreandolllikeness-v20
[8] https://civitai.com/models/106028/korean-beauty

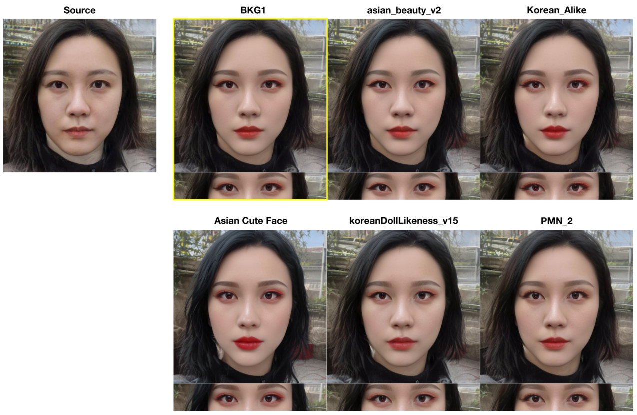

Figure 13. Results of using various LoRA weights.

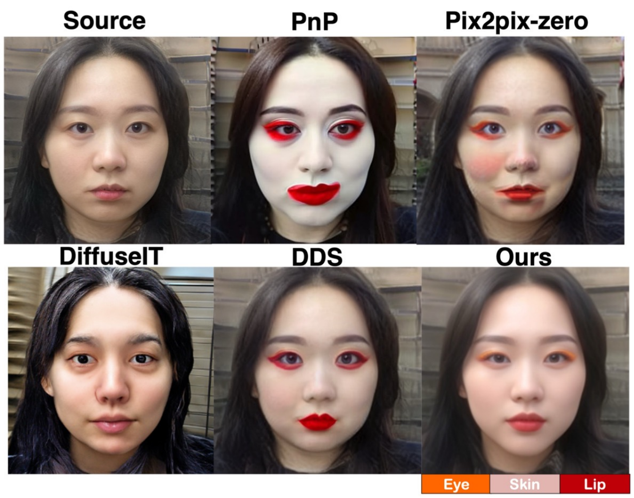

Figure 14. comparison with diffusion editing methods.

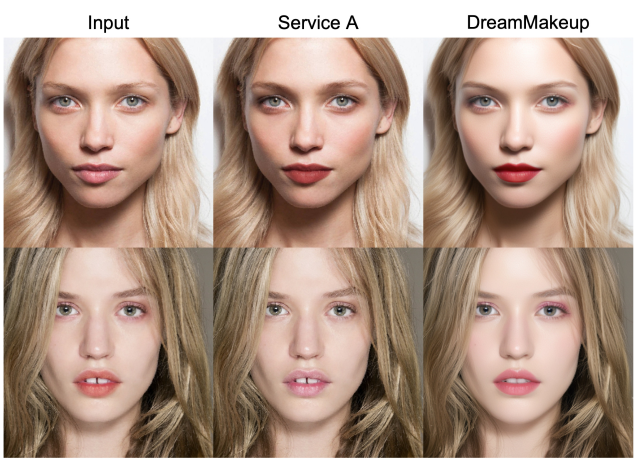

Figure 15. Comparisons of DreamMakeup with other global mobile AI makeup services.

C. Experimental details

We use DDIM scheduler and set the early-stop inversion step ranging from $t^* = 200$ to $t^* = 400$. The number of reverse steps is set to 30. LoRA scale s is set to 0.2. To

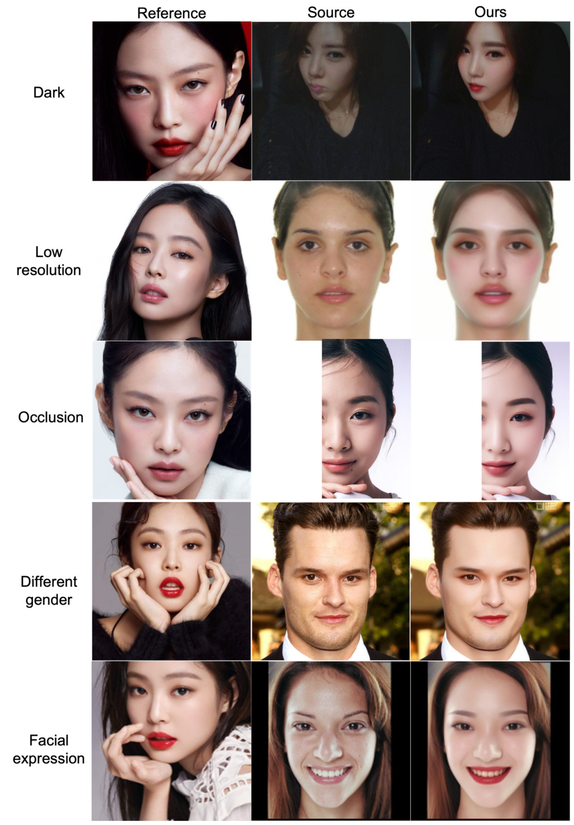

Figure 16. DreamMakeup results on extreme conditions.

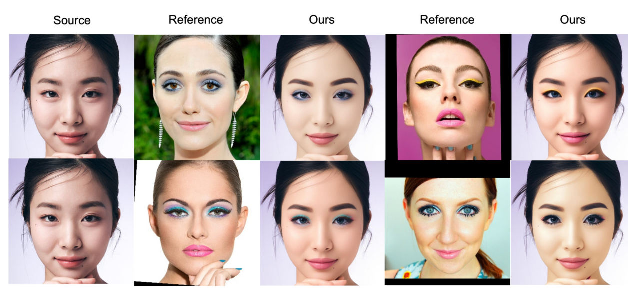

Figure 17. DreamMakeup results on extreme makeup. Reference images are from the LADN[5] dataset.

smooth eye shadow masks, we employed a cross-shaped kernel with the size of $(12, 7)$ and performed 2 iterations of mask dilation. The textual prompts commonly used in cross attention composition are as follows:

- natural lips, natural makeup, fair skin, asian skin
- korean makeup, korean style, korean beauty, (A Classy and Cute Korean girl:1.3), cute, (Korean idol), K-pop, skm_misoo, beautiful
- 32K, high-res, (masterpiece:1.3), best quality, 8K.HDR, smooth face, 1 girl,close up face, (photorealistic:1.6), [:(detailed face:1.2):0.3]
- (Glossy lips:1.6), Gleaming lips, (fair skin:1.4), sharp focus, blusher
- (Goddess smile:1.3)

Figure 18. DreamMakeup results on RGB matching for eyebrows, pupils, and hair. We transform the color of each facial attribute within segmentation mask area.

- (worst_quality:2.0) low quality, blur, deformed ugly, pixelated, cgi, illustration, cartoon, deformed, distorted, disfigured, poorly drawn

The directional degree of s-th composition, $\alpha_s \leq 0$, is assigned $0.1, 0.1, 0.3, 0.7, 0.1, -0.1$ for each prompt.

C.1. LLM

To train the language model, we constructed a QnA dataset containing information matching makeup and facial attributes. The dataset consists of $460,000$ pairs of questions and answers. Below is an example of the makeup dataset.

```
### Instruction: Which lip colors are suitable
for women with the following condition? \nbronze
skin, square face, angular jaw
### Response: deep red or vivid red or dark red.
```

As a base model, we utilized dolly-v2-3b[9] and fine-tuned the model for 3 epochs using the makeup dataset. To prevent the model from forgetting language proficiency during fine-tuning, we also incorporated the natural language dataset used to train this base language model. The training objective is to generate the subsequent tokens based on the tokenized instructions in an autoregressive manner.

D. Limitations

While our proposed method offers an efficient, training-free framework for face makeup application via early-stop DDIM inversion, it requires multiple sampling timesteps to generate the final output. Since our method utilizes BiSeNet [23] for facial segmentation and a pre-trained diffusion model for the generative process, our approach may inherit the intrinsic limitations of these foundational models.

[9]Databricks, Free dolly: Introducing the world's first truly open instruction-tuned llm, https://github.com/databrickslabs/dolly, 2023.

Figure 19. Addition results on makeup transfer.