1 Supplementary Results

The anonymous homepage for our project can be found here.

1.1 Supplementary results of diffusion-based data amplifier

As illustrated in Fig. 2, the diffusion-based data amplifier can learn five distinct makeup styles, each with its characteristic LoRA. It can apply makeup styles with stable and high quality to non-makeup images while retaining the facial features of the original portrait.

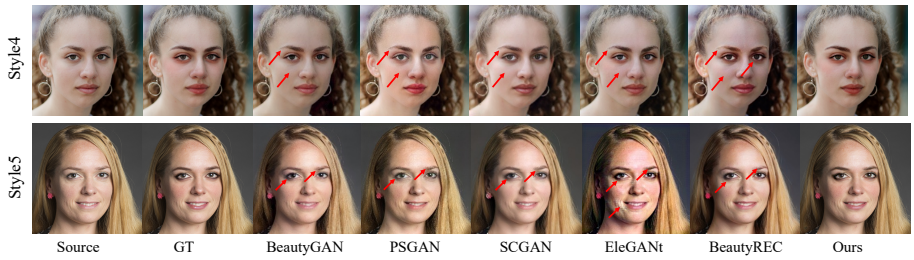

Fig. 1: Visual comparison of TineBeauty and competing methods on the FFHQ [4] image, in Style4 and Style5.

1.2 Supplementary facial makeup results on the FFHQ dataset

Results from our lightweight makeup model and competing methods for Style4 and Style5 on the FFHQ [4] dataset are displayed in Fig. 1. Our method exhibits notable advantages over prior approaches along three primary dimensions. Firstly, makeup around the eye region is applied with heightened precision by our TinyBeauty. Secondly, TinyBeauty possesses the capability to color and define eyebrow regions realistically. Thirdly, delicate cosmetic effects such as blush and contouring are delineated with accurate localization.

1.3 Supplementary facial makeup results on the MT dataset

While our model training utilized solely the FFHQ dataset due to the constrained size of the MT dataset, we performed supplementary testing of our FFHQ-trained model on the MT dataset, as depicted in Fig. 7, to conduct a fairer comparative assessment of TinyBeauty against preceding techniques. Observation of the results indicates that regardless of the exclusion of the MT dataset from training, the generated makeup outputs produced by our approach still surpass those of previous methods in visual quality. This suggests that the representation of facial attributes learned from FFHQ enabled our solution to more precisely depict makeup allocation, demonstrating its ability to generalize to other face image domains beyond the dataset it was exclusively trained on.

Fig. 2: Additional results generated by diffusion-based data amplifier.

1.4 Facial makeup results on high-resolution images.

Prior generative networks adopted an approach with direct mapping from input non-makeup portraits to output makeup depictions. This paradigm necessitates substantial computational costs for high-resolution cosmetic transfer due to processing the entire face image. Additionally, such transformations may induce unintended alterations to facial content properties during image synthesis. To circumvent these limitations, our lightweight model is formulated to yield residual representations specific to sole makeup attributes rather than fully reconstructed portraits.

Specifically, prior techniques exclusively operated on facial portraits resized to 256×256 pixels. To conduct makeup transfer at larger 1024×1024 resolutions, their networks would either necessitate retraining with enlarged 1024×1024 inputs, drastically increasing computational overhead, or rely on upsampling 256×256 outputs to 1024x1024 pixels, incurring information loss. In contrast, our TinyBeauty model retains a 256×256 input dimension, requiring only resizing its makeup residual layers to 1024×1024 before reconstructing the final output by summation with the original 1024×1024 non-makeup portrait. This avoids retraining demands and precludes the degradation in granularity associated with

Fig. 3: Facial makeup results on high-resolution (1024×1024) images.

resizing low-resolution generation results. Critically, it maintains computational efficiency equal to 256×256 processing since the primary network architecture undergoes no modification to accommodate higher resolutions.

1.5 Facial makeup results on hard cases.

To further evaluate TinyBeauty's tolerance to challenging input deviations, supplementary comparisons against BeautyREC [7] were conducted on irregular example portraits. Situations explored included: (a) a non-ideal centered face, (b) a face partially captured outside the bounds of the image, and (c) an occluded face. These archetypes simulate varied real-world acquisition scenarios undermining standard assumptions.

As shown in Fig. 4, TinyBeauty can successfully apply makeup without prerequisite face detection or localization pre-processing, unlike BeautyREC which

Table 1: Results of TinyBauty and competing methods on FFHQ and MT datasets, in Style2.

Method	FFHQ			MT		
	PSNR↑	FID↓	LPIPS↓	PSNR↑	FID↓	LPIPS↓
BeautyGAN [5]	25.95	45.38	0.0595	25.84	31.44	0.0537
PSGAN [3]	25.31	35.88	0.0624	26.99	18.31	0.0419
SCGAN [1]	26.84	38.24	0.0527	25.61	35.96	0.0572
EleGANt [8]	29.34	26.16	0.0441	30.47	13.83	0.0296
BeautyREC [7]	24.09	28.32	0.0591	25.46	23.24	0.0508
TinyBeauty	**33.331**	**10.327**	**0.0258**	**32.445**	**11.313**	**0.0335**

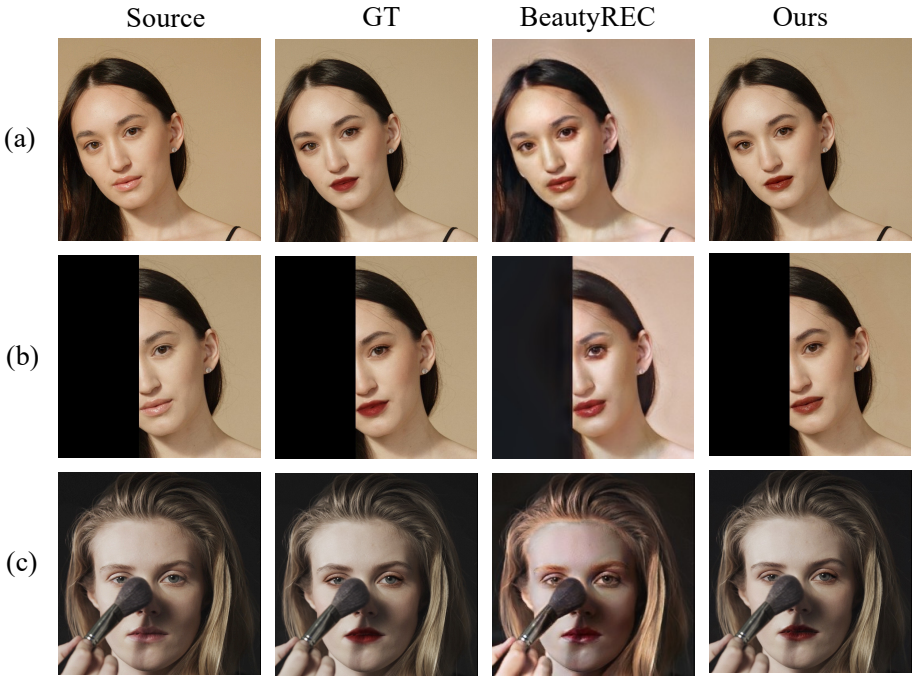

Fig. 4: Visual comparison of TineBeauty and BeautyREC [7] on challenging out-of-distribution examples: (a) a non-ideal centered face, (b) a face partially captured outside the bounds of the image, and (c) an occluded face.

failed to properly segment features. This validates TinyBeauty's end-to-end generation ability from raw inputs alone and TinyBeauty's superior tolerance to input variations.

1.6 Supplementary quantitative results.

In addition to the quantitative analyses on Style 1 reported in the primary manuscript, equivalent objective assessments were performed for the remaining four makeup styles. As depicted in Tab. 1, Tab. 2, Tab. 3, Tab. 4, our approach achieves favorable performance as evaluated by PSNR, FID [2], and LPIPS [9] on both the FFHQ and MT datasets, surpassing prior cosmetic transfer techniques by a substantial margin according to each numerical metric. Our lightweight model consistently outperforms predecessors across all styles assessed and datasets, as evidenced by higher PSNR and lower FID and LPIPS scores. These results substantiate TinyBeauty's demonstrated superior makeup simulation ability and position it as the current leading solution among facial makeup models.

Table 2: Results of TinyBauty and competing methods on FFHQ dataset and MT dataset, in Style3.

2*Method	FFHQ			MT		
	PSNR↑	FID↓	LPIPS↓	PSNR↑	FID↓	LPIPS↓
BeautyGAN [5]	26.41	48.67	0.0645	27.51	47.55	0.0628
PSGAN [3]	25.67	38.97	0.0677	27.71	38.10	0.0518
SCGAN [1]	27.65	44.52	0.0541	27.17	53.16	0.0598
EleGANt [8]	29.91	30.55	0.0483	32.06	37.20	0.0368
BeautyREC [7]	24.55	33.12	0.0626	26.62	42.22	0.0601
TinyBeauty	**34.72**	**11.14**	**0.0288**	**33.788**	**12.029**	**0.0233**

Table 3: Results of TinyBauty and competing methods on FFHQ dataset and MT dataset, in Style4.

Method	FFHQ			MT		
	PSNR↑	FID↓	LPIPS↓	PSNR↑	FID↓	LPIPS↓
BeautyGAN [5]	26.75	44.73	0.0568	27.65	28.96	0.0467
PSGAN [3]	25.68	35.69	0.0607	27.94	14.49	0.0345
SCGAN [1]	27.97	35.35	0.0453	27.29	30.68	0.0615
EleGANt [8]	30.05	24.18	0.0404	32.22	10.86	0.0200
BeautyREC [7]	24.83	27.02	0.0546	27.00	19.78	0.0424
TinyBeauty	**34.737**	**9.29**	**0.0225**	**33.658**	**10.756**	**0.0378**

1.7 Comparison of convergence curves.

To further validate that TinyBeauty's training procedure is simpler than previous methods, we compared the rising PSNR curves of TinyBeauty and BeautyREC [7] during 50 epochs of training on 200 images. Specifically, as shown in Fig. 5, TinyBeauty approached convergence within 15 training iterations, while BeautyREC did not approach convergence until 40 iterations. Additionally, TinyBeauty's final PSNR score far exceeded that of BeautyREC. These optimization results quantitatively demonstrate that TinyBeauty requires substantially fewer parameter updates than BeautyREC but reaches both fast convergence and superior performance, benefiting from direct L1 loss.

2 Additional Losses

Perceptual Similarity Loss. To maintain visual fidelity between the input face image and synthesized makeup result, we introduce a perceptual similarity loss $\mathcal{L}_{per}$ following previous works [3, 5, 7]. This loss measures the Euclidean

Table 4: Results of TinyBauty and competing methods on FFHQ dataset and MT dataset, in Style5.

2*Method	FFHQ			MT		
	PSNR↑	FID↓	LPIPS↓	PSNR↑	FID↓	LPIPS↓
BeautyGAN [5]	26.47	46.20	0.0617	27.21	28.87	0.0519
PSGAN [3]	25.57	36.31	0.0646	27.70	18.47	0.0388
SCGAN [1]	26.97	42.62	0.0572	26.60	31.32	0.0515
EleGANt [8]	29.88	27.27	0.0442	32.50	14.28	0.0266
BeautyREC [7]	24.46	29.72	0.0590	26.38	22.45	0.0503
TinyBeauty	**35.652**	**9.18**	**0.0143**	**34.684**	**9.28**	**0.0216**

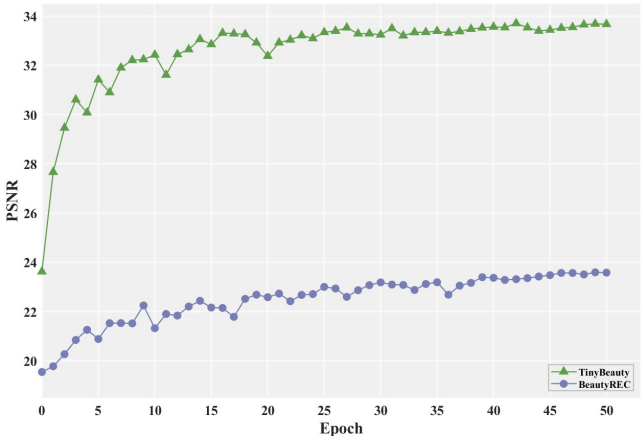

Fig. 5: Comparison of PSNR convergence curves between TinyBeauty and BeautyREC during training.

distance in feature space between the conv4 layer activations of the input and output images when fed through a pre-trained VGG-19 network [6]:

$$\mathcal{L}_{per} = ||\phi_{vgg}(x) - \phi_{vgg}(y')||_2, \tag{1}$$

where ϕ_{vgg} denotes the convolutional features extracted from the conv4 layer before activation. By minimizing this loss, our model generates makeup results that align with the input face image semantically.

Adversarial Loss. To help the makeup model generate realistic makeup outputs, we incorporate global and local adversarial losses to further constrain the makeup outputs. The ultimate formulation of the adversarial loss is:

$$\mathcal{L}_{adv} = \mathcal{L}_{adv}^{global} + \mathcal{L}_{adv}^{eyes} + \mathcal{L}_{adv}^{eyebrows} + \mathcal{L}_{adv}^{skin} + \mathcal{L}_{adv}^{lips}, \tag{2}$$

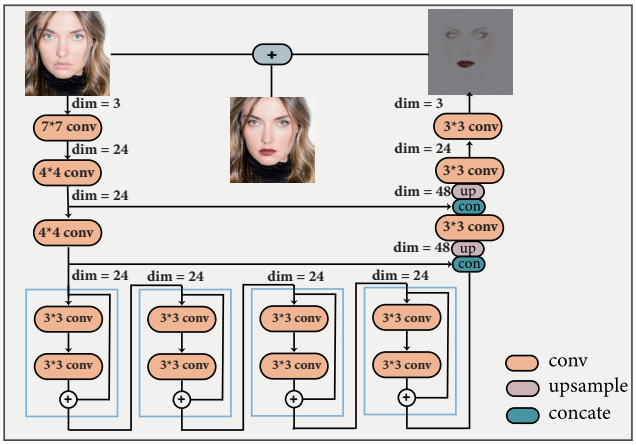

Fig. 6: Architecture of TinyBeauty Model.

where $\mathcal{L}adv^{global}$ penalizes deviations from the ground truth makeup style globally across the entire facial region. The localized losses $\mathcal{L}adv^{eyes}$, $\mathcal{L}adv^{eyebrows}$, $\mathcal{L}adv^{skin}$, and $\mathcal{L}_{adv}^{lips}$ further optimize consistency specifically within semantically meaningful regions. Through this formulation, we aim to refine the makeup transfer at both global and localized levels, enforcing fine-grained fidelity to the target style distribution over key facial components.

Total Loss. The total loss is a combination of the above-mentioned losses and losses in the main paper, which can be expressed as:

$$\mathcal{L}_{total} = \mathcal{L}_{rec} + \mathcal{L}_s + \lambda_{per}\mathcal{L}_{per} + \lambda_{adv}\mathcal{L}_{adv}, \tag{3}$$

where $\lambda_{per} = 0.005$ and $\lambda_{adv} = 0.5$ are the corresponding weights for balancing the magnitudes of losses.

3 Structure of TinyBeauty Model

To show the network structure of our TinyBeauty model more clearly, we drew the structural configurations of the convolutional layers comprising the network, as shown in Fig. 6. Particularly, we employ ReLU as an activation layer after each convolutional stage for nonlinear mapping, alongside Instance Normalization to regulate the distribution of input features. By leveraging such operation-efficient modules, our model achieves markedly expedited inference amenable to deployment on mobile devices with constrained computational budgets.

References

1. Deng, H., Han, C., Cai, H., Han, G., He, S.: Spatially-invariant style-codes controlled makeup transfer. In: IEEE Conference on Computer Vision and Pattern

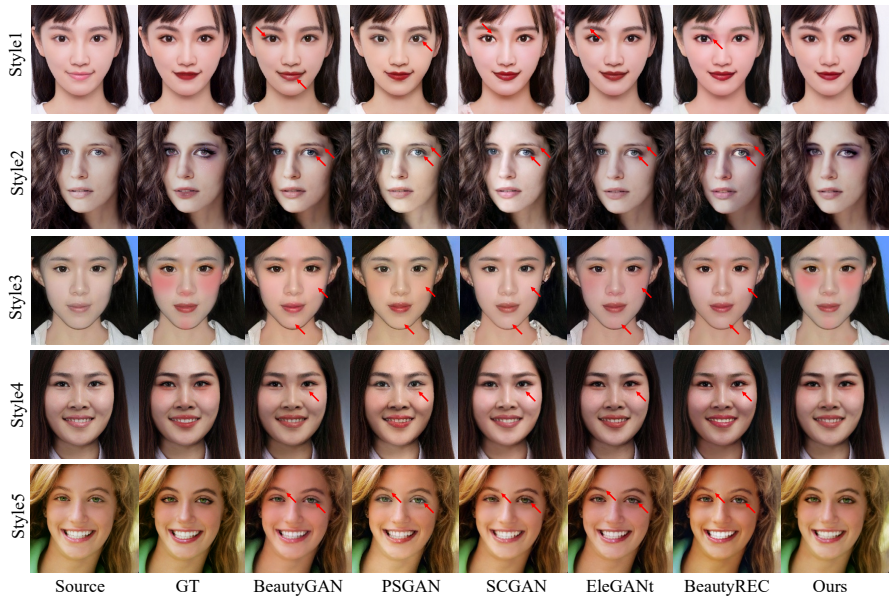

Fig. 7: Visual comparison of TineBeauty and competing methods on the MT [5] images.

Recognition, CVPR 2021, virtual, June 19-25, 2021. pp. 6549–6557. Computer Vision Foundation / IEEE (2021)
2. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. In: Guyon, I., von Luxburg, U., Bengio, S., Wallach, H.M., Fergus, R., Vishwanathan, S.V.N., Garnett, R. (eds.) Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA. pp. 6626–6637 (2017)
3. Jiang, W., Liu, S., Gao, C., Cao, J., He, R., Feng, J., Yan, S.: PSGAN: pose and expression robust spatial-aware GAN for customizable makeup transfer. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020. pp. 5193–5201. Computer Vision Foundation / IEEE (2020)
4. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019. pp. 4401–4410. Computer Vision Foundation / IEEE (2019)
5. 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: 2018 ACM Multimedia Conference on Multimedia Conference, MM 2018, Seoul, Republic of Korea, October 22-26, 2018. pp. 645–653. ACM (2018)
6. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: Bengio, Y., LeCun, Y. (eds.) 3rd International Conference on Learn-

ing Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings (2015), http://arxiv.org/abs/1409.1556
7. Yan, Q., Guo, C., Zhao, J., Dai, Y., Loy, C.C., Li, C.: Beautyrec: Robust, efficient, and component-specific makeup transfer. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023 - Workshops, Vancouver, BC, Canada, June 17-24, 2023. pp. 1102–1110. IEEE (2023)
8. Yang, C., He, W., Xu, Y., Gao, Y.: Elegant: Exquisite and locally editable GAN for makeup transfer. In: Avidan, S., Brostow, G.J., Cissé, M., Farinella, G.M., Hassner, T. (eds.) Computer Vision - ECCV 2022 - 17th European Conference, Tel Aviv, Israel, October 23-27, 2022, Proceedings, Part XVI. Lecture Notes in Computer Science, vol. 13676, pp. 737–754. Springer (2022)
9. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018. pp. 586–595. Computer Vision Foundation / IEEE Computer Society (2018)