Makeup Style Transfer on Low-quality Images with Weighted Multi-scale Attention

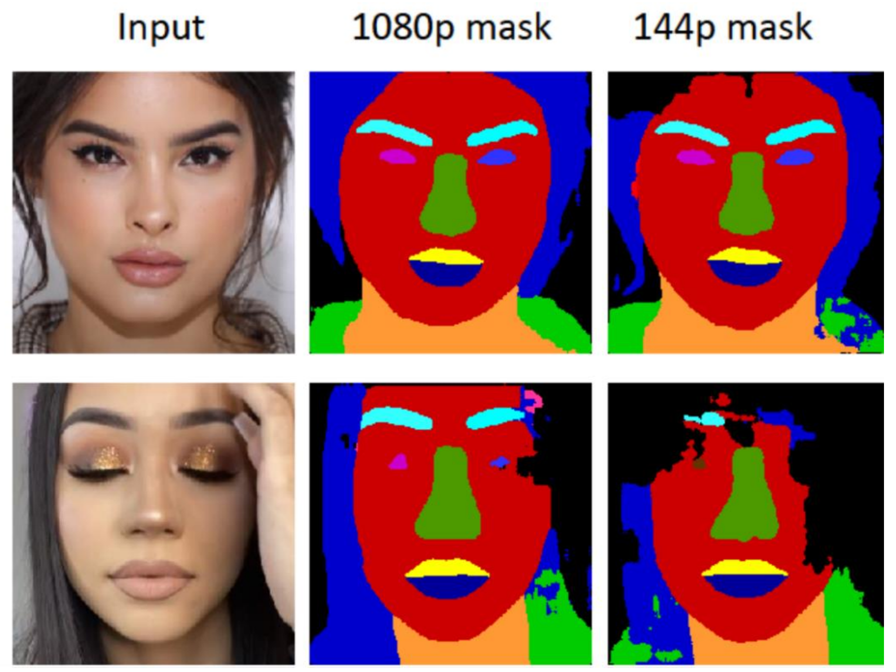
Daniel Organisciak*, Edmond S. L. Ho*, Hubert P. H. Shum†

*Northumbria University, †Durham University

Abstract

Makeup style transfer state-of-the-art models often depend on the Face Parsing Algorithm, which segments a face into parts to extract makeup features. However, this algorithm can only work well on high-definition. We propose an end-to-end holistic approach to effectively transfer makeup styles between two low-resolution images. The idea is built upon a novel weighted multi-scale spatial attention module, which identifies salient pixel regions on low-resolution images in multiple scales and uses channel attention to determine the most effective attention map. We develop an Augmented CycleGAN network that embeds our attention modules at selected layers to most effectively transfer makeup. Our system is tested with the FBD data set, which consists of many low-resolution facial images, and demonstrate that it outperforms state-of-the-art methods.

Problems with State-of-the-Art



- Most state-of-the-art methods depend on systems like Face Parsing algorithm (above) to learn a) the makeup style, and b) where to transfer it.
- FPA performs poorly on low quality images, due to issues such as resolution, lighting, occlusion and pose angle (second row).
- To overcome these issues, we replace the FPA with a multi-scale soft attention module to transfer makeup style in a holistic, end-to-end manner.

Multi-scale Attention

Augmented CycleGAN is used as the backbone architecture

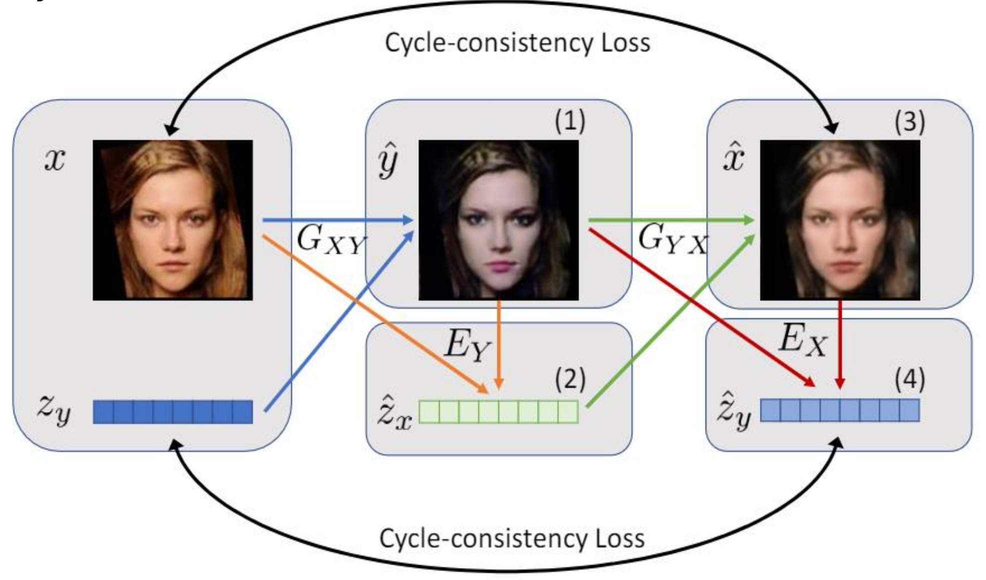
a) Apply makeup style z_y onto image x

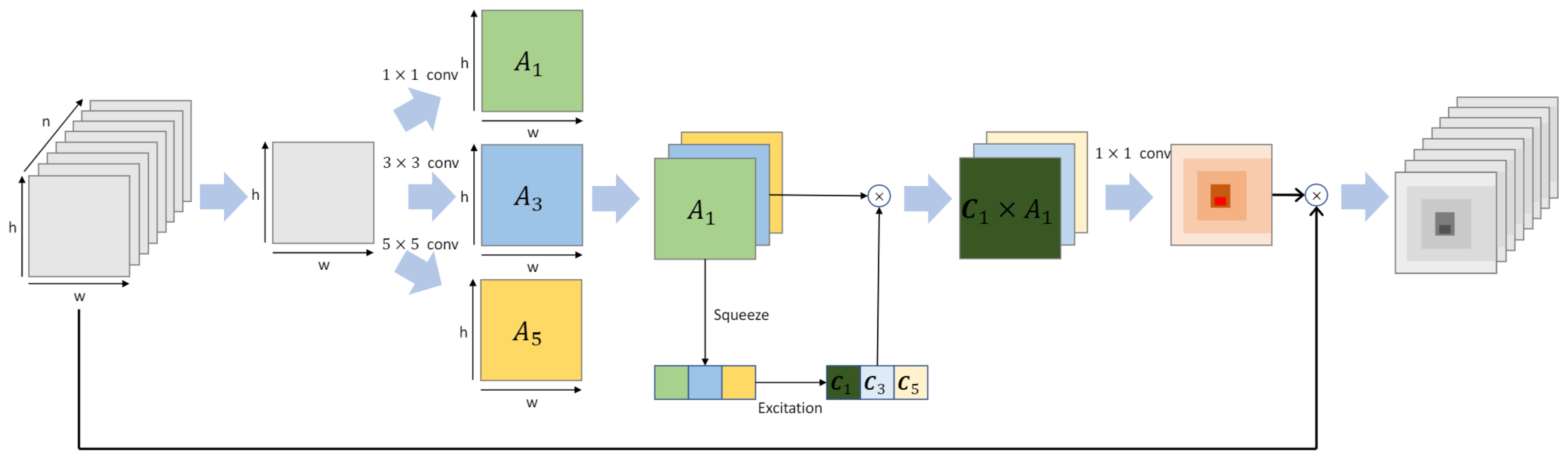
b) From image x and fake image $\hat{y}$ infer de-makeup style $\hat{z}_x$

c) Apply fake de-makeup style $\hat{z}_x$ to image $\hat{y}$

d) From image $\hat{x}$ and fake image $\hat{y}$ infer makeup style $\hat{z}_y$

The architecture maintains cycle-consistency between i) x and $\hat{x}$; ii) z_y and $\hat{z}_y$.





- Our proposed weighted, multi-scale attention module:
  - a) squeeze along the channel dimension to obtain the representation matrix;
  - b) Convolve the representation matrix through different sized kernels to extract intermediate attention maps;
  - c) Squeeze and excite intermediate attention maps to determine which attention scale is most important for the image being processed
- Low-resolution images can be blurry to different extents, a multi-scale architecture can select the most effective convolution kernel size to implement spatial attention
- Different attention scales extract different types of makeup (fake tan and eyeliner require different scales)

Quantitative Evaluation

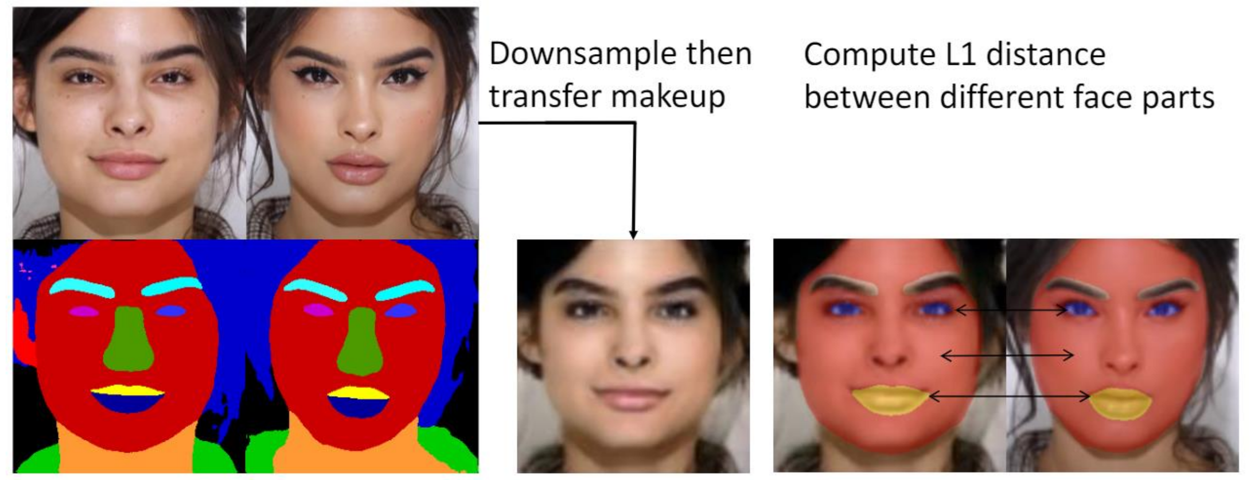


TABLE I
COMPARISON WITH STATE OF THE ART METHODS

Method	Eyes	Skin	Lips	Total
BeautyGAN [2]	0.230	0.086	**0.215**	0.532
DMT [4]	0.238	**0.084**	0.218	0.541
Ours	**0.197**	0.089	0.229	**0.515**

Lower numbers are better

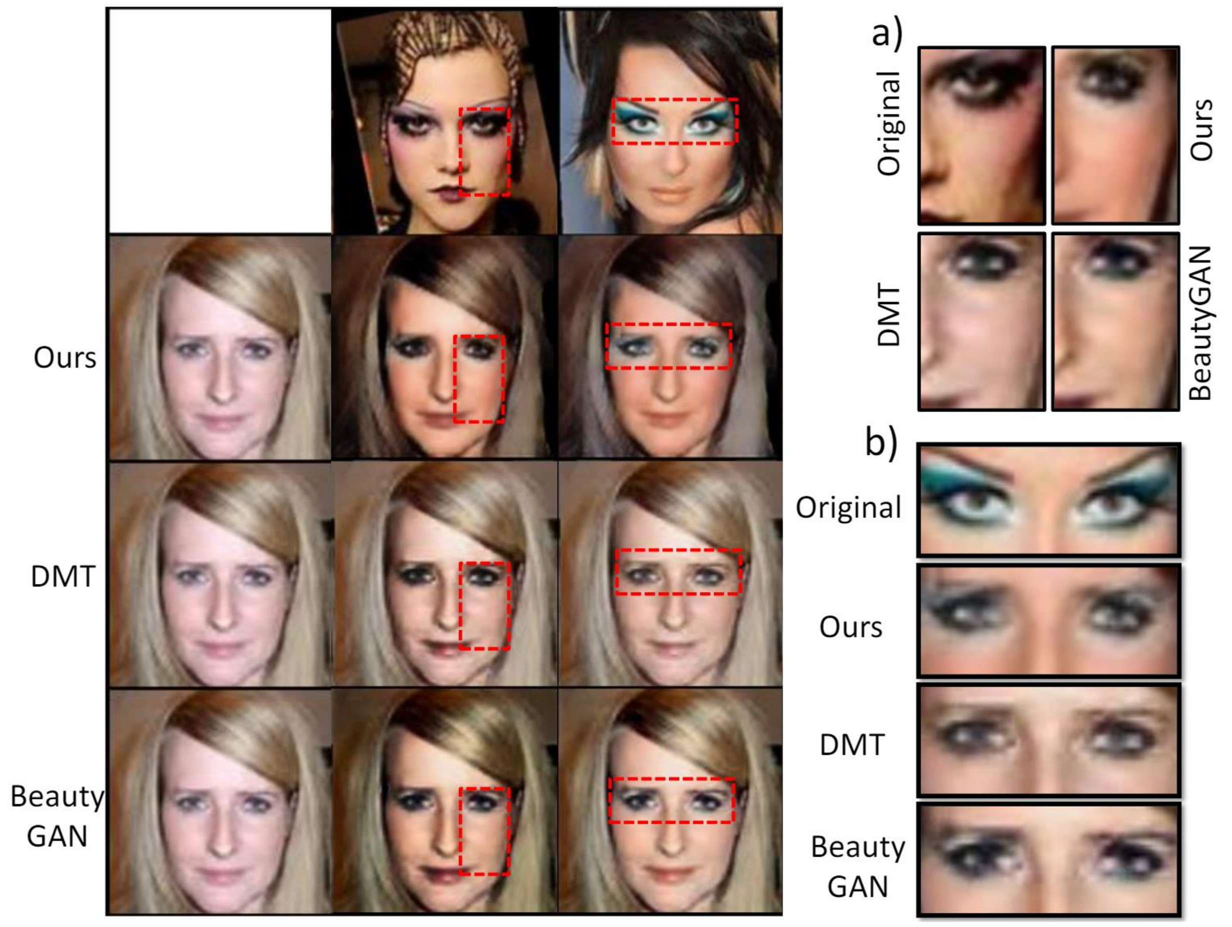
TABLE II
ABLATION STUDY ON ATTENTION

Method	Eyes	Skin	Lips	Total
Ours	0.197	**0.089**	**0.229**	**0.515**
w/o Multi-scale Attention	**0.188**	0.105	0.236	0.529
w/o Any Attention	0.274	0.126	0.247	0.647

Lower numbers are better

- We create a new makeup/non-makeup data set with 10 subjects from YouTube makeup tutorials.
- We design an evaluation metric to quantitatively assess low-resolution makeup style transfer models:
  - a) Apply FPA to 1080p images
  - b) Downsample images and transfer makeup style
  - c) Using FPA masks of real and fake images, compute colour histograms for each face part then calculate the L1 distance

Qualitative Evaluation



Comparison on challenging makeup styles:

a) Our method is best approximates the skin tone colour distribution

b) Our method best transfers fake eyelashes and comes closest to transferring the butterfly wings.

For extensive results, see the full paper and supplementary materials.

Key References

A. Almahairi, S. Rajeswar, A. Sordoni, P. Bachman, and A. Courville, “Augmented CycleGAN: Learning many-to-many mappings from un-paired data,” *arXiv preprint arXiv:1802.10151*, 2018.

T. Li, R. Qian, C. Dong, S. Liu, Q. Yan, W. Zhu, and L. Lin,“BeautyGAN: Instance-level facial makeup transfer with deep generative adversarial network,” in *Proceedings of the 26th ACM International Conference on Multimedia*, p. 645–653.

H. Zhang, W. Chen, H. He, and Y. Jin, “Disentangled makeup transfer with generative adversarial network,” *arXiv preprint arXiv:1907.01144*, 2019.

B. M. Smith, L. Zhang, J. Brandt, Z. Lin, and J. Yang, “Exemplar-based face parsing,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3484–3491.