Supplemental Materials for MSEmbGAN: Multi-Stitch Embroidery Synthesis via Region-Aware Texture Generation

Xinrong Hu, Chen Yang, Fei Fang[✉], Jin Huang, Ping Li, *Member, IEEE*, Bin Sheng, *Member, IEEE*, and Tong-Yee Lee, *Senior Member, IEEE*

1 Discussion of Color Consistency

For a satisfactory embroidery image, it is necessary to retain the color feature of the original image to the greatest extent. In other words, the colors and shapes of the results should be as similar as possible to those of the input images. In our MSEmbGAN, we provide a colorization network to add the color information of the input image to the resulting image. In the network optimization, the resulting images are encouraged by the color consistency loss to approximate the color of the input images. In this section, we show comparisons of the color feature between the input and output images. In Fig. 1, we enlarge several corresponding regions of the input and resulting images, and readers can also zoom in to compare the consistency of the color between the input and resulting images.It can be seen that the color of our generated results is very close to the color of the original input image.

2 Failure Cases

We find that our network suffers from failed texture generation when the content of the input image is complex. In Fig. 2, when the input image contains much high-frequency information, the texture in the resulting image becomes blurred.

3 More Comparative Experimental Results

Apart from the comparative experimental results presented in Section 4.1 of our main paper, here we show more comparative experimental results. In Fig. 3, we still compare our results with the same state-of-the-art methods as in Section 4.1 of the main paper (i.e., CycleGAN [1], Pix2Pix [2], MUNIT [3], and DRIT++ [4]). As shown in Fig. 3, our method can synthesize better embroidery images than the comparison methods, especially in terms of texture, color, and shape. The results of CycleGAN [1] demonstrate great losses of color information. Besides, this method cannot effectively generate embroidery textures in some regions. Pix2Pix [2] can only retain a part of the color feature of the input image, but the results do not contain embroidery textures. The colors of the MUNIT results [3] are generated from random noise, which randomizes the colors of the embroidery results. DRIT++ [4] can only generate a part of the embroidery textures, and the results are different from the original images in terms of color and shape. The textures of our results are closer to the real embroidery features, and the color of our results are more similar to the input images. Although the results for CycleGAN and DRIT++ are reasonably good, our results are better in terms of maintaining the original input color and embroidery-like texture. Meanwhile, we display the details of our network architecture in Table 1. Some of the network layers, like LeakyReLUConv2d and ReLUINSConv2d, are inspired by the baseline method [4].

4 Intermediate Results

Our MSEmbGAN consists of two sub-networks: a region-aware texture generation network and a colorization network. Given an input image, the region-aware texture generation network generates grayscale embroidery texture images $\hat{e}^L_{fake}$. The colorization network optimizes the color of the full grayscale embroidery texture image $\hat{e}^L_{fake}$ to generate the final embroidery result $\hat{e}_{fake}$. In Fig. 4, we demonstrate examples of intermedeiate results (i.e., the grayscale embroidery texture images generated from the region-aware texture generation network) and the corresponding final resulting images. It can be seen from Fig. 4 that the intermediate grayscale embroidery texture images already have characteristics of the real embroidery textures.

5 Single-Stitch Embroidery Synthesis

The resulting images synthesized by the previous experiments are multi-stitch embroidery images. We also conducted an experiment to synthesize single-stitch embroidery images. The objective of this experiment is to generate embroidery patterns with three independent basic stitches. To achieve this goal, we used MSEmbGAN to generate three corresponding single-stitch embroidery images by adjusting the stitch type parameter st. In Fig. 5, we show the performance of our MSEmbGAN on satin, tatami, and flat-stitch embroidery image synthesis, respectively. As shown

TABLE 1

Some details of our network architecture implementation.

Networks	layers	input dim	output dim	kernel	strike	padding
Embroidery Encoder E^{L}_{emb}	ReflectionPad2d	-	-	3	-	-
	ReLUConv2d	1	64	7	1	0
	ReflectionPad2d	-	-	1	-	-
	ReLUConv2d	64	128	4	2	0
	ReflectionPad2d	-	-	1	-	-
	ReLUConv2d	128	256	4	2	0
	ReflectionPad2d	-	-	1	-	-
	ReLUConv2d	256	256	4	2	0
	AdaptiveAvgPool2d	-	-	1	-	-
	Conv2d	256	8	1	1	0
Content Encoder E^{L}_{con}	LeakyReLUConv2d	1	64	7	1	3
	ReLUINSConv2d	64	128	3	2	1
	ReLUINSConv2d	128	256	3	2	1
	INSResBlock	256	256	-	-	-
	INSResBlock	256	256	-	-	-
	INSResBlock	256	256	-	-	-
Luminance Generator G^{L}	LinearReLU	8	256	-	-	-
	LinearReLU	256	256	-	-	-
	MisINSResBlock	256	256	-	-	-
	MisINSResBlock	256	256	-	-	-
	MisINSResBlock	256	256	-	-	-
	ReLUINSConvTranspose2d	256	128	3	2	1
	ReLUINSConvTranspose2d	128	64	3	2	1
	ConvTranspose2d	64	1	1	1	0
Texture Discriminator D^{L}	LeakyReLUConv2d	1	64	3	2	1
	LeakyReLUConv2d	64	128	3	2	1
	LeakyReLUConv2d	64	128	3	2	1
	LeakyReLUConv2d	128	256	3	2	1
	LeakyReLUConv2d	256	512	3	2	1
	Conv2d	512	1	1	1	0
Luminance Encoder E^{L}_{lum}	ReflectionPad2d	-	-	3	-	-
	ReLUConv2d	1	64	7	1	0
	ReLUConv2d	64	64	4	2	0
	Conv2d	64	8	1	1	0
Chrominance Encoder E^{ab}_{chr}	LeakyReLUConv2d	1	64	7	1	3
	ReLUINSConv2d	64	128	3	2	1
	ReLUINSConv2d	128	256	3	2	1
	INSResBlock	256	256	-	-	-
Chrominance Generator G^{Lab}	LinearReLU	8	256	-	-	-
	LinearReLU	256	256	-	-	-
	MisINSResBlock	256	256	-	-	-
	MisINSResBlock	256	256	-	-	-
	ReLUINSConvTranspose2d	256	128	3	2	1
	ReLUINSConvTranspose2d	128	64	3	2	1
	ConvTranspose2d	64	1	1	1	0

Fig. 1: Color consistency between the input and resulting images. For each input image, we zoom in two regions and place the enlarged regions on the right. Meanwhile, the corresponding regions in the resulting images are zoomed in and placed on the left.

TABLE 2

Composition of the Multi-Stitch Embroidery Dataset.

Stitch type	Multiple	Satin	Tatami	Flat
Aligned	6006	3936	1916	1338
Unaligned	8500	4922	2288	1480
Total	14506	8858	4204	2818

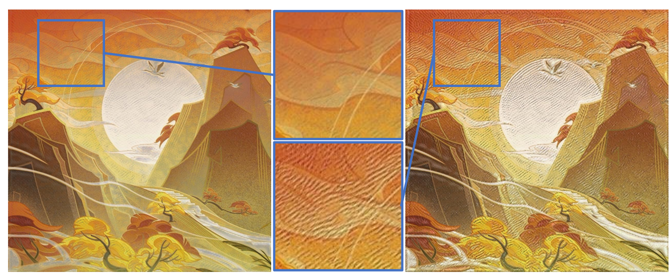

Fig. 2: Example of the failure case.

in Fig. 5, our results are similar to the ground truth images in terms of texture and color. Furthermore, the single-stitch embroidery images generated by our network can well reflect the styles of the corresponding stitches.

6 Composition of Multi-Stitch Embroidery Dataset

Our multi-stitch embroidery dataset contains more than 30K high-quality embroidery images, including over 13K aligned content-embroidery images and more than 17K unaligned images. Meanwhile, the images in the dataset contain three types of single stitches and multiple stitches mixed with the three single stitches. See Table 2 for the specific dataset composition. In MSEmbGAN, we pre-trained the region-aware texture generation network with three types of aligned single-stitch images labeled with stitch types. Then the color feature extractor was pre-trained with the aligned multi-stitch images. Although our MSEmbGAN needs single-stitch embroidery images for pre-training, the number of needed images is relatively small. In order for MSEmbGAN to capture the texture feature of each type of stitch, a number of single-stitch embroidery images are necessary. We also trained the colorization network on unaligned multi-stitch datasets. During training, our network only required unaligned multi-stitch embroidery images that are easy to collect.

References

[1] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," in *International Conference on Machine Learning*, 2017, pp. 1857–1865.

[2] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1125–1134.

[3] X. Huang, M.-Y. Liu, S. Belongie, and J. Kautz, "Multimodal unsupervised image-to-image translation," in *European Conference on Computer Vision*, 2018, pp. 172–189.

[4] H.-Y. Lee, H.-Y. Tseng, Q. Mao, J.-B. Huang, Y.-D. Lu, M. Singh, and M.-H. Yang, "DRIT++: Diverse image-to-image translation via disentangled representations," *International Journal of Computer Vision*, vol. 128, no. 10, pp. 2402–2417, 2020.

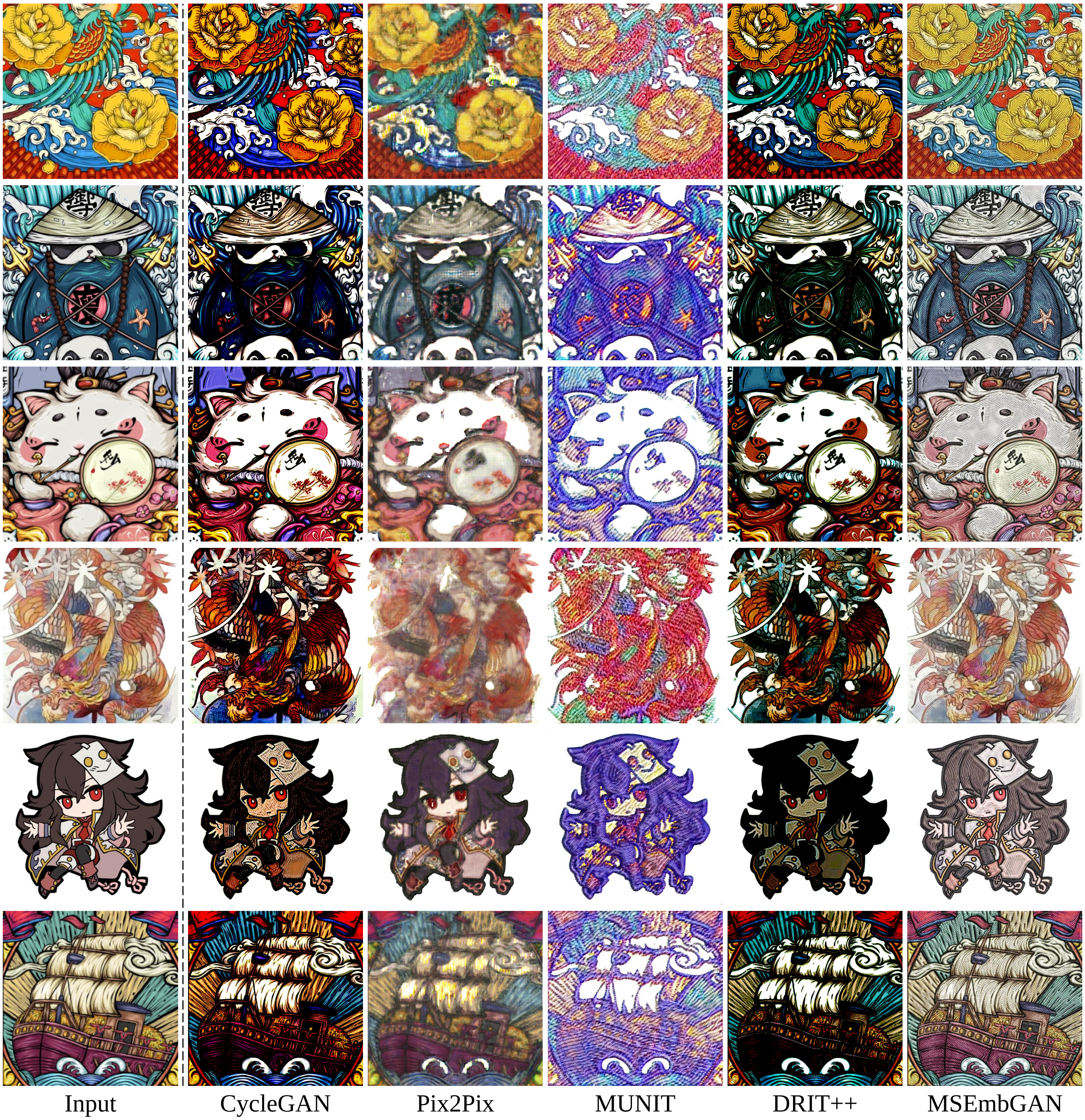

Fig. 3: More comparative experimental results between MSEmbGAN and four SOTA style-transfer networks (i.e., CycleGAN [1], Pix2Pix [2], MUNIT [3], and DRIT++ [4]).

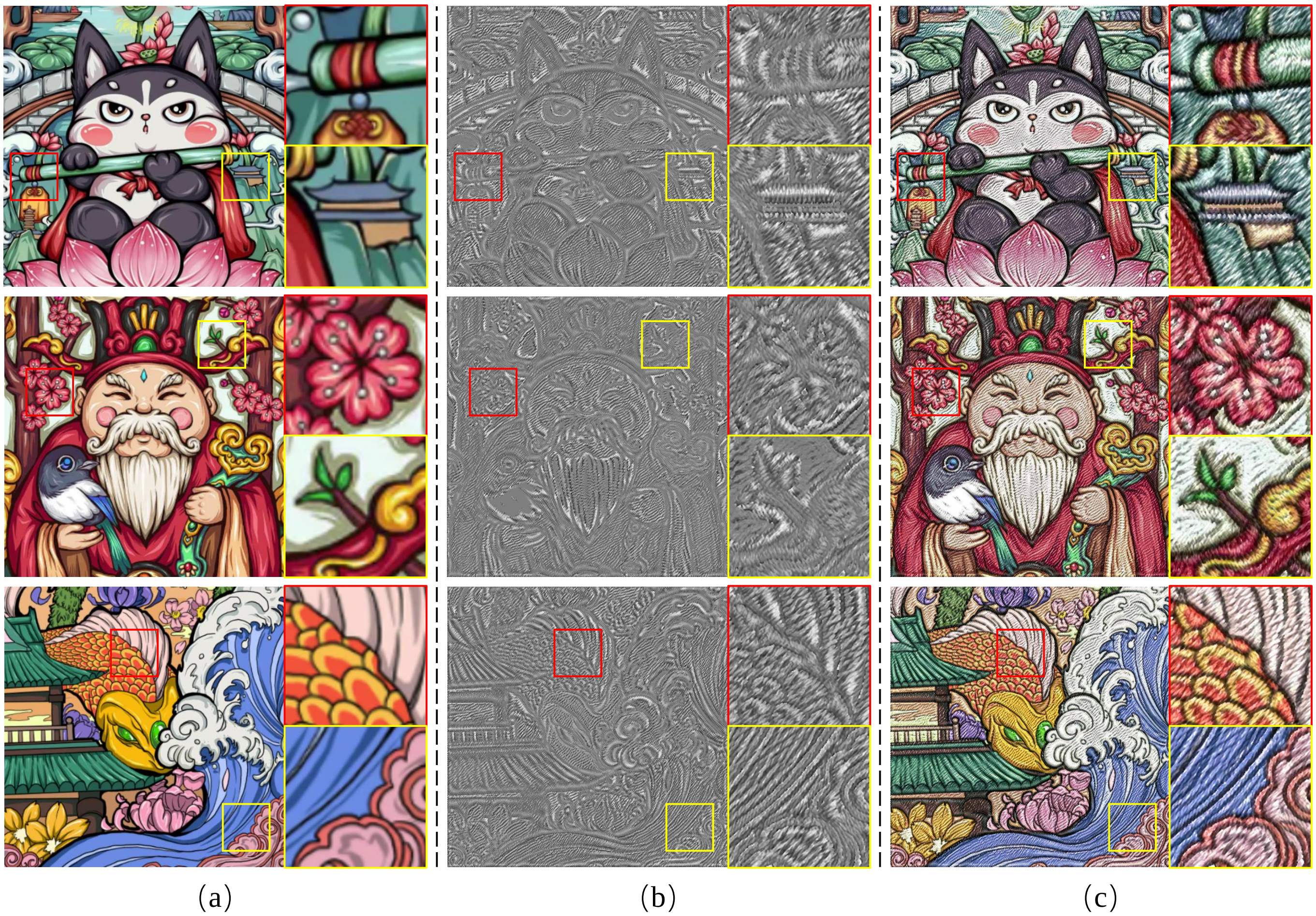

Fig. 4: Examples of intermediate results: (a) original input images $\boldsymbol{c}$; (b) intermediate grayscale embroidery texture images $\hat{\boldsymbol{e}}^{L}_{fake}$; (c) final resulting embroidery images $\hat{\boldsymbol{e}}_{fake}$.

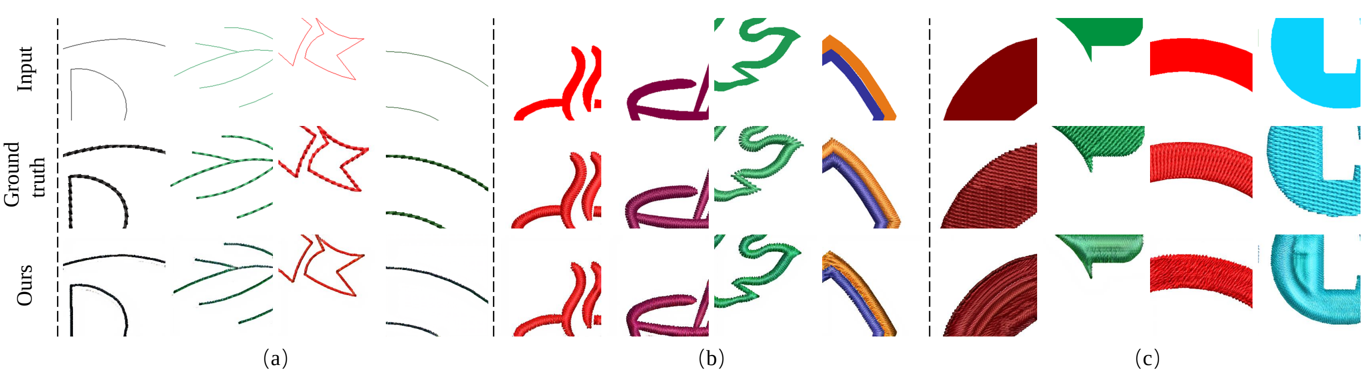

Fig. 5: Single-stitch embroidery synthesis results: (a) flat stitch; (b) satin stitch; and (c) tatami stitch.