TextureGAN Supplementary Material

Layer	Type	In	Out	Kernel Size	Stride	Pad
1	conv [skip connection]	5	32	3	1	1
2	residual block	32	32	3	1	1
3	conv	32	64	3	2	1
4	residual block	64	64	3	1	1
5	conv	64	128	3	2	1
6	residual block	128	128	3	1	1
7	conv	128	256	3	2	1
8	residual block	256	256	3	1	1
9	residual block	256	256	3	1	1
10	residual block	256	256	3	1	1
11	residual block	256	256	3	1	1
12	residual block	256	256	3	1	1
13	conv + biup	256	128	3	1	1
14	residual block	128	128	3	1	1
15	residual block	128	128	3	1	1
16	conv + biup	128	64	3	1	1
17	residual block	64	64	3	1	1
18	residual block	64	64	3	1	1
19	conv + biup	64	32	3	1	1
20	residual block	32	32	3	1	1
21	conv [skip connection]	32 + 5	64	3	1	1
22	residual block	64	64	3	1	1
23	residual block	64	64	3	1	1
24	conv	64	3	3	1	1

Table 1: Generator network. Every conv layer is followed by batch normalization and relu activation. Residual blocks consist of 2 conv layers with stride of 1 (table 4), in which the input to the block is added to the output as proposed in [1]. Biup refers to the upsampling with scale of 2x, using bilinear interpolation. We also add a single skip connection from input to layer 21.

Layer	Type	Input Channels	Output Channels	Kernel Size	Stride	Pad
1	conv	1	32	9	2	1
2	conv	32	64	5	2	1
3	conv	64	256	5	2	1
4	residual block	256	256	3	1	1
5	residual block	256	256	3	1	1
6	conv	256	128	4	2	1
7	conv	128	1	4	2	1

Table 2: Global Discriminator network. The input is a single L channel normalized to 0/1 range (equivalent to greyscale image).

Layer	Type	Input Channels	Output Channels	Kernel Size	Stride	Pad
1	conv	2	32	3	2	1
2	conv	32	128	3	2	1
3	residual block	128	128	3	1	1
4	residual block	128	128	3	1	1
5	conv	128	64	3	2	1
6	conv	64	1	3	2	1

Table 3: Texture Discriminator network. The input to the network is a pair of texture patches, of size 50×50.

Layer	Type	Kernel Size	Stride	Pad
1	conv	3	1	1
2	conv	3	1	1
3	+ input	-	-	-

Table 4: Residual Block. Each conv is followed by batch normalization, then relu activation.

References

[1] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

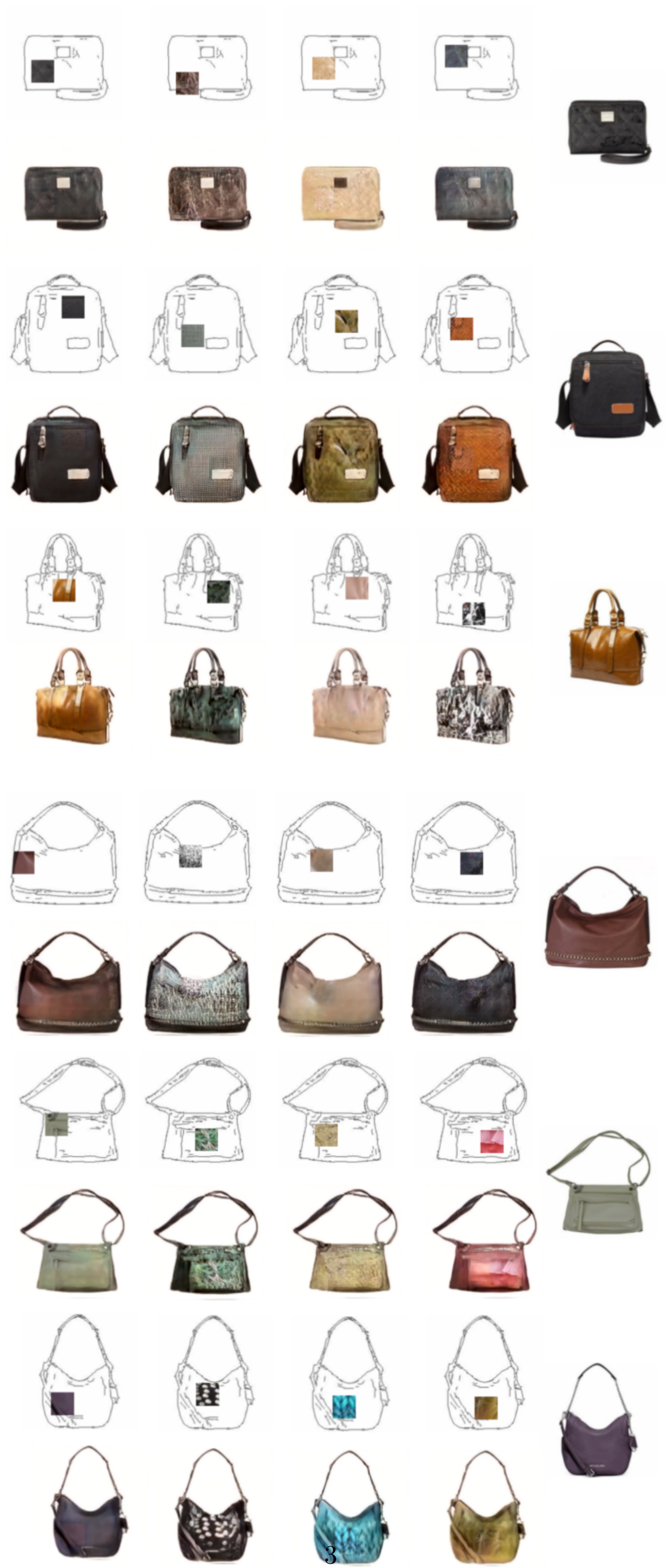

Figure 1: Additional results on held out handbag sketches [152x152]. On the right is the "ground truth" photo from which the sketch was synthesized. On the far left, a texture patch is also sampled from the original handbag. We show three additional results with diverse textures.

Figure 2: Additional results on held out shoes sketches [152x152]. On the right is the "ground truth" photo from which the sketch was synthesized. On the far left, a texture patch is also sampled from the original shoe. We show three additional results with diverse textures.

Figure 3: Additional results on held out clothes sketches [256x256].

Figure 4: Comparison of results before and after fine-tuning on an external texture database. The base model is trained on handbag and shoe sketches with ground truth images. The fine-tuned model with local texture loss improves texture propagation on difficult texture with high contrast and strong regularity, e.g. striped, dotted, checkerboard patterns, which are rarely seen in the ground-truth.

Figure 5: In some cases, especially with complicated or rare texture patterns, the network fails to propagate texture throughout the object. In these cases, the input texture is blended into the object instead. A user can alleviate such effect by placing more texture swatches.

Figure 6: The effect of varying input patch sizes during testing.

Figure 7: Comparison **RGB** versus **Lab** inputs. Enforcing feature, texture and adversarial losses on the L channel and adding a separate color loss on the ab channels significantly improve the color propagation.