Implementation Details

To produce DST results with Gatys, we initialized the output image as the content image resized to have a long side of 256 pixels. During the optimization, we updated the parameters 100 times with a learning rate of 1 with L-BFGS, following the authors’ implementation settings. Since the content and style loss terms and the deformation loss term have vastly different magnitudes, we scaled down the content loss and the style loss by $1/50000$ and $1/100000$, respectively. We also scaled down the deformation parameter gradients by $1/1000000$ to update them at a much smaller rate.

To produce DST results with STROTSS, we followed the authors’ implementation settings and optimized the output image at multiple scales, starting from a low-resolution image scaled to have a long side of 64 pixels. The output image was initialized as the bottom level of a Laplacian pyramid of the content image added to the mean color of the style image. During the optimization, instead of optimizing the pixels directly, we optimized the Laplacian pyramid of the image for faster convergence, in addition to the deformation parameters $\theta$. At each scale, we used the output of the previous scale as the initial image, used bilinear upsampling to increase the resolution, and halved the content weight $\alpha$ that controls the relative importance of content preservation to stylization. We produced the stylized images in the main paper at three scales, starting from content weight $\alpha = 32$. At each scale, we made 350 updates of the parameters using stochastic gradient descent with a learning rate of 0.2. More details can be found in our published code: https://github.com/sunniesuhyoung/DST.

Human Evaluation

As described in Section 5.2 of the main paper, we conducted a human evaluation study using Amazon Mechanical Turk on a set of 75 diverse style/content pairs. We show the evaluation interfaces in Supplementary Figures 1 and 2.

Image Credits

Figure 1

How would you respond to the following statement?

"Image A represents the same scene as Image B"

Supplementary Fig. 1: Evaluation interface for measuring content preservation.

Which of Image A or Image B better matches the style of the reference?

Supplementary Fig. 2: Evaluation interface for measuring stylization.

Figure 2
3. Original artist unknown: ‘Photo of Man in Suit’, Unknown Source

Figure 3
Figure 4

Figure 5
3. Original artist unknown: ‘Husky Puppy Photo’, Unknown source

Figure 6
7. Original artist unknown: ‘Horse Photo’. https://twitter.com/Bbelisle7/status/1215411108921724929/photo/1
Figure 7
1. Original artist unknown: ‘Photo of Woman with Wavy Hair’.  
http://www.faculty.idc.ac.il/arik/site/foa/face-of-art.asp
http://www.faculty.idc.ac.il/arik/site/foa/face-of-art.asp
5. Original artist unknown: ‘Photo of Man with Short Hair’.  
http://www.faculty.idc.ac.il/arik/site/foa/face-of-art.asp

Figure 9
1. Original artist unknown: ‘Photo of Barred Owls’.  
https://1.bp.blogspot.com/-1weEwEzz2wk/Xb64zZC-2RI/AAAAAAADI-M/pTeSYN7v_KAo2YK3FsPUt6SVe76A5pYaQCLcBGAsYHQ/s640/88l.jpg
http://weknowyourdreams.com/single/surprise/surprise-01
5. Original artist unknown: ‘Husky Photo’. Unknown Source

Supplementary Figure 1
1. Original artist unknown: ‘Black Horse Photo’. Unknown Source

Supplementary Figure 2
1. Original artist unknown: ‘Photo of Woman with Wavy Hair’.  
http://www.faculty.idc.ac.il/arik/site/foa/face-of-art.asp