FSGAN: Subject Agnostic Face Swapping and Reenactment
— Supplemental material —

Yuval Nirkin
Bar-Ilan University, Israel
yuval.nirkin@gmail.com

Yosi Keller
Bar-Ilan University, Israel
yosi.keller@gmail.com

Tal Hassner
The Open University of Israel, Israel
talhassner@gmail.com

1. Additional qualitative results

We offer additional quantitative face swapping results in Fig. 1. We have specifically chosen examples of challenging pairs, with partial occlusions, different ethnicities and skin colors, demonstrating the competence of our method on a large variety of subjects. In Fig. 2, we show additional quantitative comparison to Nirkin et al. [3] and DeepFakes [2], and in Fig. 3 we show another comparison to Face2Face [6]. Please also see the attached video for more results.

2. The architecture of the generator CNNs

The architecture of the generators, G_r, G_c, and G_b, is based on the pix2pixHD approach [7], and the layout of the global generator and enhancer is depicted in Fig. 4. The global generator is defined by the number of bottleneck blocks (shown in purple) used in each resolution scale. In our experiments we used only three resolutions. The enhancer is defined by its submodule, that is, either the global generator or another enhancer, and its number of bottleneck layers. The generators are thus given by

$$G_r = G_c = \text{Enhancer}(\text{Global}(2, 2, 3), 2),$$

and

$$G_b = \text{Enhancer}(\text{Global}(1, 1, 1), 1).$$

The face segmentation network G_s is based on the U-Net approach [4], for which we replaced the deconvolution layers with bilinear interpolation upsampling layers.

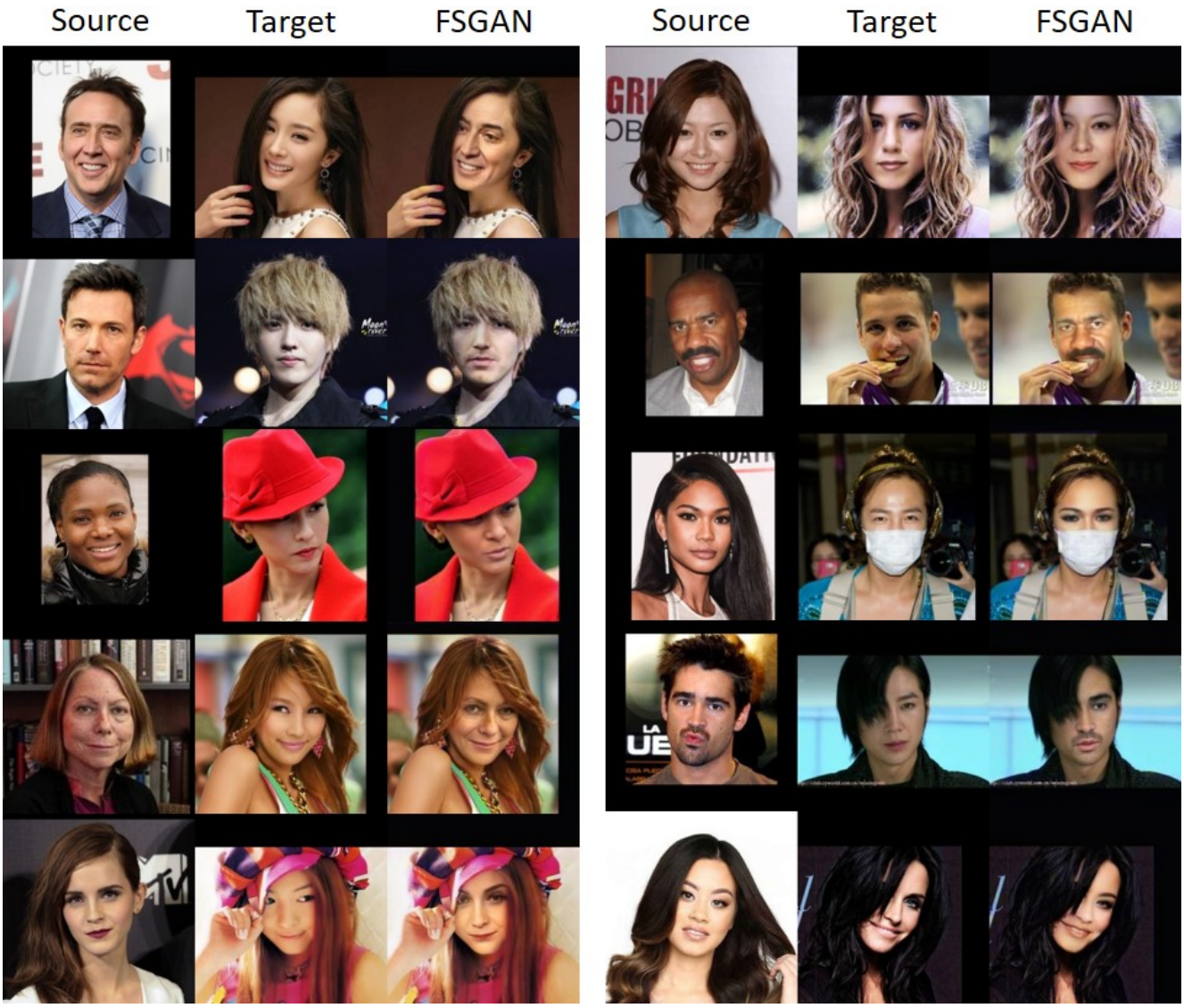

Figure 1: *Additional qualitative face swapping results on on the Caltech Occluded Faces in the Wild (COFW) dataset [1].*

Figure 2: *Additional qualitative face swapping comparison to Nirkin et al. [3] and DeepFakes [2] on FaceForensics++ [5].*

Figure 3: *Additional qualitative face reenactment comparison to Face2Face [6] on FaceForensics++ [5].*

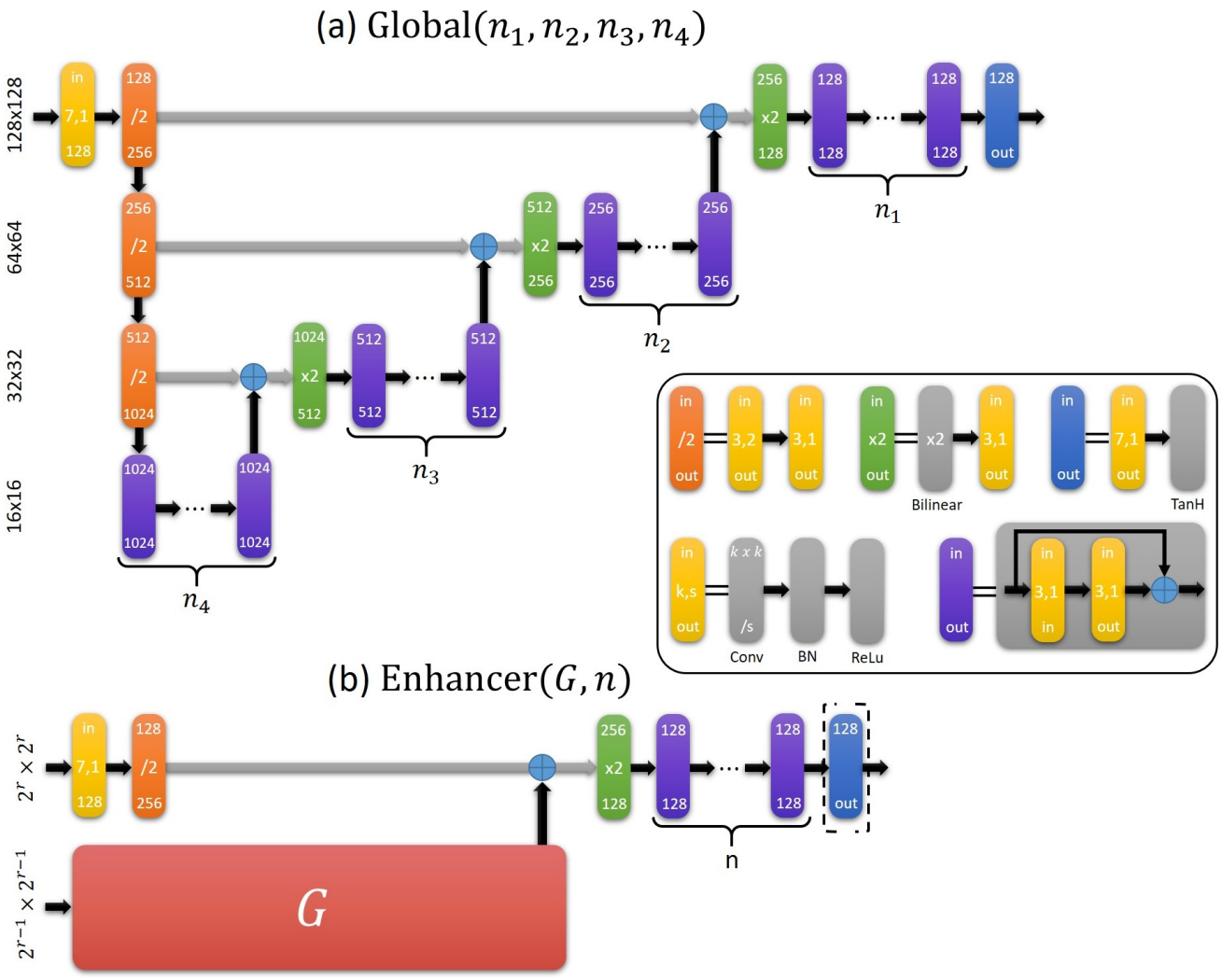

Figure 4: *Generator architectures.* (a) The global generator is based on a residual variant of the U-Net [4] CNN, using a number of bottleneck layers per resolution. We replace the simple convolutions with bottleneck blocks (in purple), the concatenation with summation (plus sign), and the deconvolutions with bilinear upsamplnig following by a convolution. (b) The enhancer utilizes a submodule and a number of bottleneck layers. The last output block (in blue) is only used in the enhancer of the finest resolution.

References

[1] X. P. Burgos-Artizzu, P. Perona, and P. Dollár. Robust face landmark estimation under occlusion. In *Proc. Int. Conf. Comput. Vision*, pages 1513–1520. IEEE, 2013.

[2] DeepFakes. FaceSwap. `https://github.com/deepfakes/faceswap`. Accessed: 2019-02-06.

[3] Y. Nirkin, I. Masi, A. T. Tuan, T. Hassner, and G. Medioni. On face segmentation, face swapping, and face perception. In *Automatic Face & Gesture Recognition (FG 2018), 2018 13th IEEE International Conference on*, pages 98–105. IEEE, 2018.

[4] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[5] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner. Faceforensics++: Learning to detect manipulated facial images. *arXiv*, 2019.

[6] J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt, and M. Nießner. Face2face: Real-time face capture and reenactment of rgb videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2387–2395, 2016.

[7] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.