# The Contextual Loss for Image Transformation with Non-Aligned Data Supplementary Material

Roey Mechrez* , Itamar Talmi*, Lihi Zelnik-Manor

Technion - Israel Institute of Technology
{titamar@campus,roey@campus,lihi@ee}.technion.ac.il

### Limitation and failure cases in style transfer

Out method is not without limitations. In style transfer, our objective aims at generating in the output image a good match for each patch of both the original content image and the style image. In this paper we set this balance fixed to equal contribution, which resulted in some cases, in undesirable results. For example, in Figure 6 (fourth row) the girl's freckles are weakly transfered and in Figure 7 (third row) the global color of the husky is not transfered well. (see in the supp. figures bellow)

To resolve this one could provide the user interactive control over the balance between the content loss and the style loss. Another option is to add a Gram or histogram loss to the objective, which is something we have not tried.

In Figure 1 we show the influence of the bandpass parameter $h$ on the final result. In all the style transfer experiment shown in the paper and in the supplementary we used single value of $h = 0.1$ for the content term and $h = 0.2$ for the style term. Here we show what happen if we move from this working point and by that changing the balance between the content and the style terms.

* indicate authors contributed equally


**Fig. 1. The balance between the style and content terms:** the balance between the content and style terms is controlled by the bandpass parameters, denoted as $h_c$ and $h_s$ for the content and style terms, respectively. High $h_c$ enforces strong content similarity even if $h_s$ is high as well. We found that the most robust and effective working point is $h_c = 0.1$ and $h_s = 0.2$. Top: (a) content image (b) style image and (c) result at working point. Bottom: result at different bandpass parameters values.

### Limitation in Domain Transfer

In the paper we present the use of the contextual loss for domain transfer. Specifically, we show two tasks of male-to-female and female-to-male. Our network achieves good results without the use of GAN just by comparing two random faces, one from each domain, at each iteration. The underlying assumption behind this is that the semantic information of all the images is similar, that is, all images contain a face over some background. This assumption is necessary in order to achieve one-to-one feature matching in the contextual loss.

While this assumption is reasonable when both domains were of faces, it does not generalize nicely to more complex domains. For example, in Figure 2 we show our results on a more complicated dataset: zebra-to-horse. Here the assumption does not hold since each image contains additional objects. Furthermore, the scenes are highly diverse, ranging from a zoom-in on a zebra face to a wide landscape with multiple far-away zebras.

For the zebra-to-horse training without GAN, as we did for gender transfer, failed. Therefore, we adopted the same architecture as in CycleGAN [1], but trained only a single direction, i.e., breaking the cycle, but using GAN. Our loss is a mix of an adversarial loss and the contextual loss between a real horse image and a fake generated horse image. We show in Figure 2 some of our results. More can be found on our web-page.

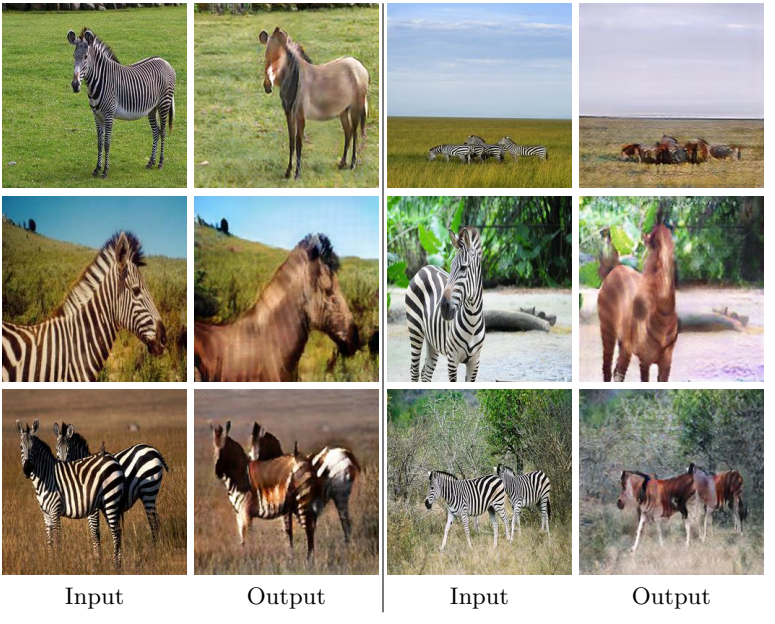


**Fig. 2.** Domain transfer results on zebra-to-horse.

### Symmetry Analysis of the Contextual Loss

The contextual loss definition is not symmetric and $\mathcal{L}_{\text{CX}}(G(s),t) \neq \mathcal{L}_{\text{CX}}(t,G(s))$. nonetheless, in our experiments, we have found that in most cases $\mathcal{L}_{\text{CX}}(G(s),t) \approx \mathcal{L}_{\text{CX}}(t,G(s))$. In Figure 3 we compare between five variations of our loss:

1. $S = \mathcal{L}_{\text{CX}}(G(s),t)$
2. $S + C = \mathcal{L}_{\text{CX}}(G(s),t) + \mathcal{L}_{\text{CX}}(G(s),s)$
3. $S' + C = \mathcal{L}_{\text{CX}}(t,G(s)) + \mathcal{L}_{\text{CX}}(G(s),s)$
4. $S' + C' = \mathcal{L}_{\text{CX}}(t,G(s)) + \mathcal{L}_{\text{CX}}(s,G(s))$
5. $S + S' + C' = \mathcal{L}_{\text{CX}}(G(s),t) + \mathcal{L}_{\text{CX}}(t,G(s)) + \mathcal{L}_{\text{CX}}(s,G(s))$

Our experiments showed that the differences between the five options are marginal suggesting that an additional term, which would make our loss symmetric, is redundant. A similar observation was reported in [2] for template matching.

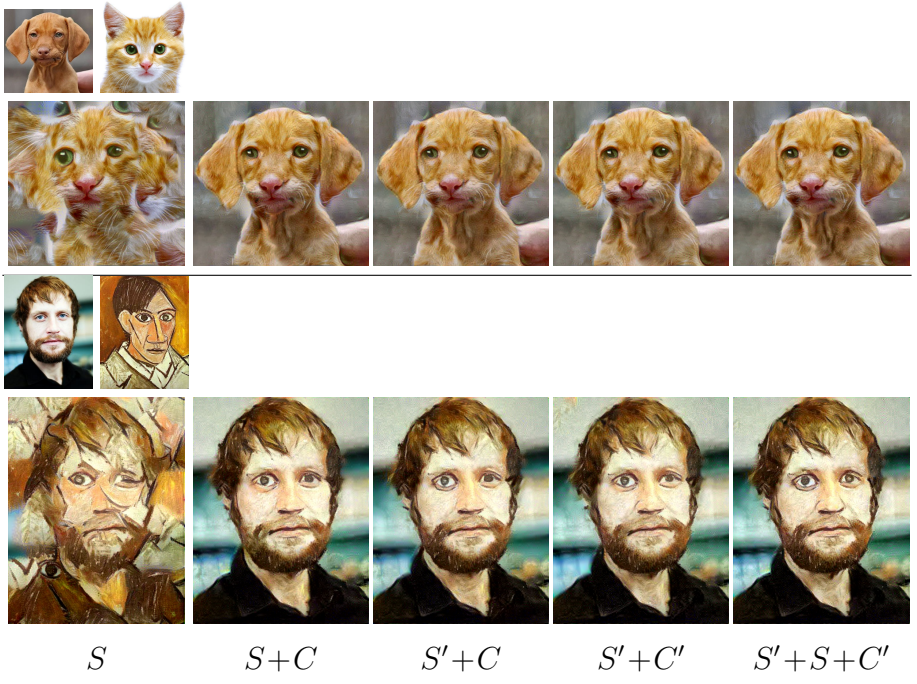


**Fig. 3. Loss symmetry**: Style transfer examples, with five loss variations: $S = \mathcal{L}_{\text{CX}}(G(s),t)$, $S' = \mathcal{L}_{\text{CX}}(t,G(s))$, $C = \mathcal{L}_{\text{CX}}(G(s),s)$,$C' = \mathcal{L}_{\text{CX}}(s,G(s))$. See text for details.

### Ablation Study in the Puppet Control Application

We present an ablation study over the loss function used in Puppet Control. Our objective consists of two loss terms: (i) the contextual loss, and (ii) the perceptual loss. In Figure 4 we show the influence of the perceptual loss on the final result. Generally the observed differences are small and consist mostly of small details. We found that using $\lambda_P = 0.1$ preserves the fine details better, for example, the fingers shape and the eyeglasses. We note, that since the CRN architecture emphasizes the input structure strongly, we do not need an additional loss in order to preserve the input spatial structure.


**Fig. 4.** Loss ablation test for puppet control

# Additional Results
# 1. Style Transfer


**Fig. 5.**

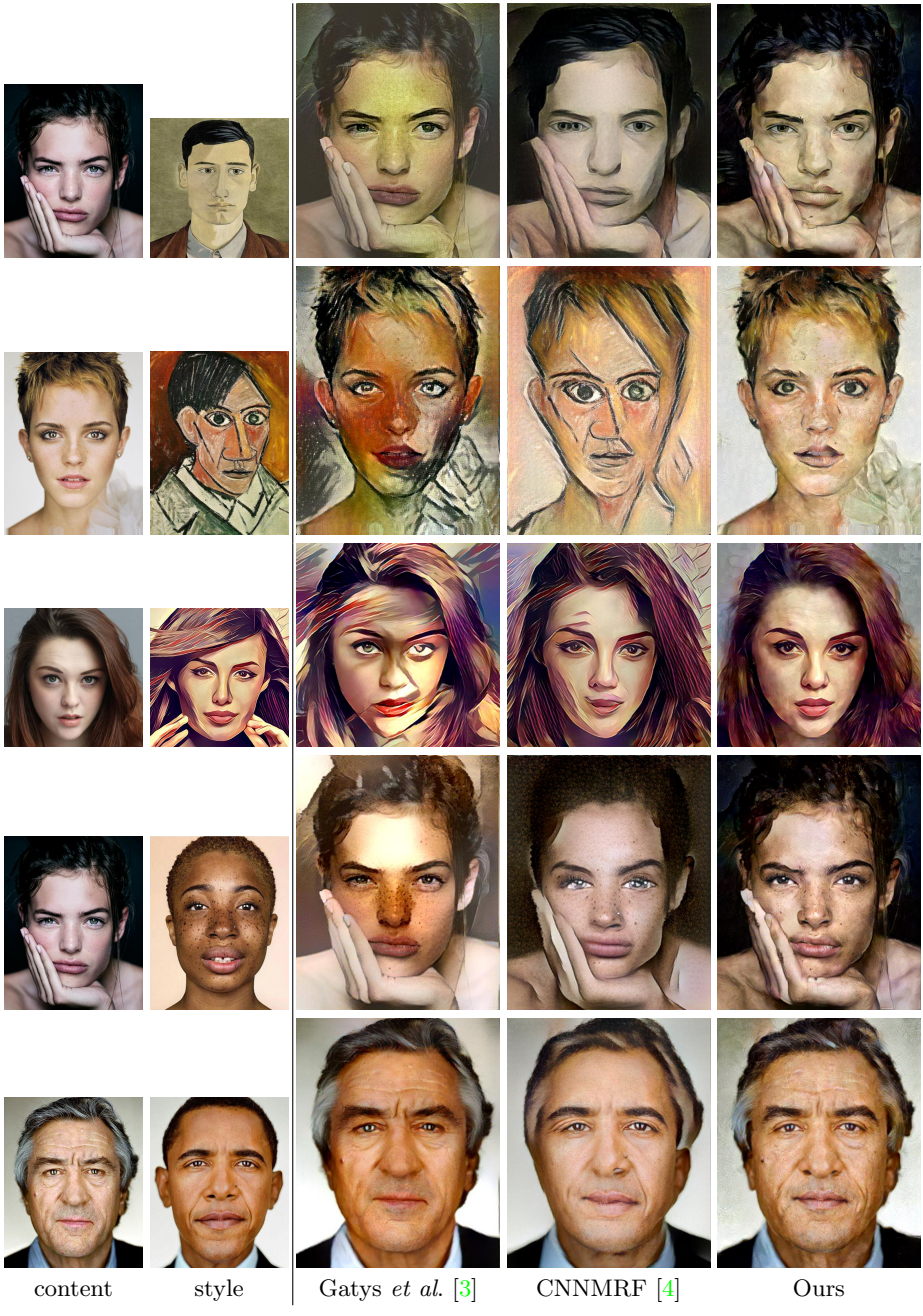


**Fig. 6.**


**Fig. 7.**

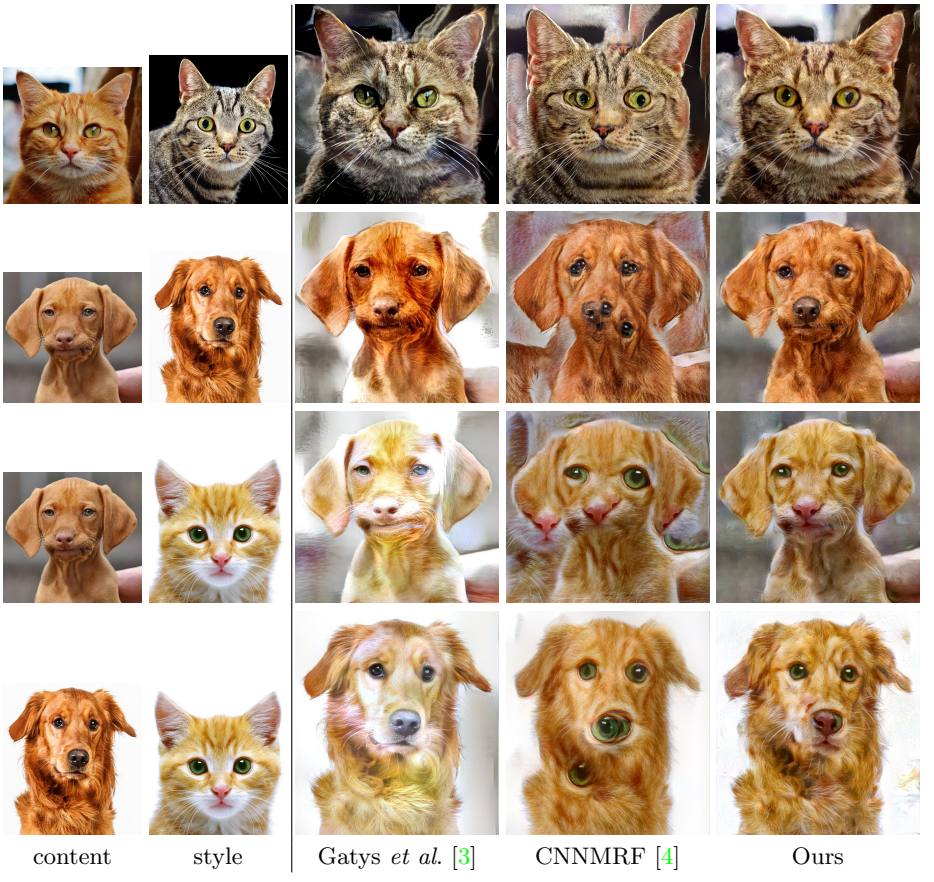


**Fig. 8.**

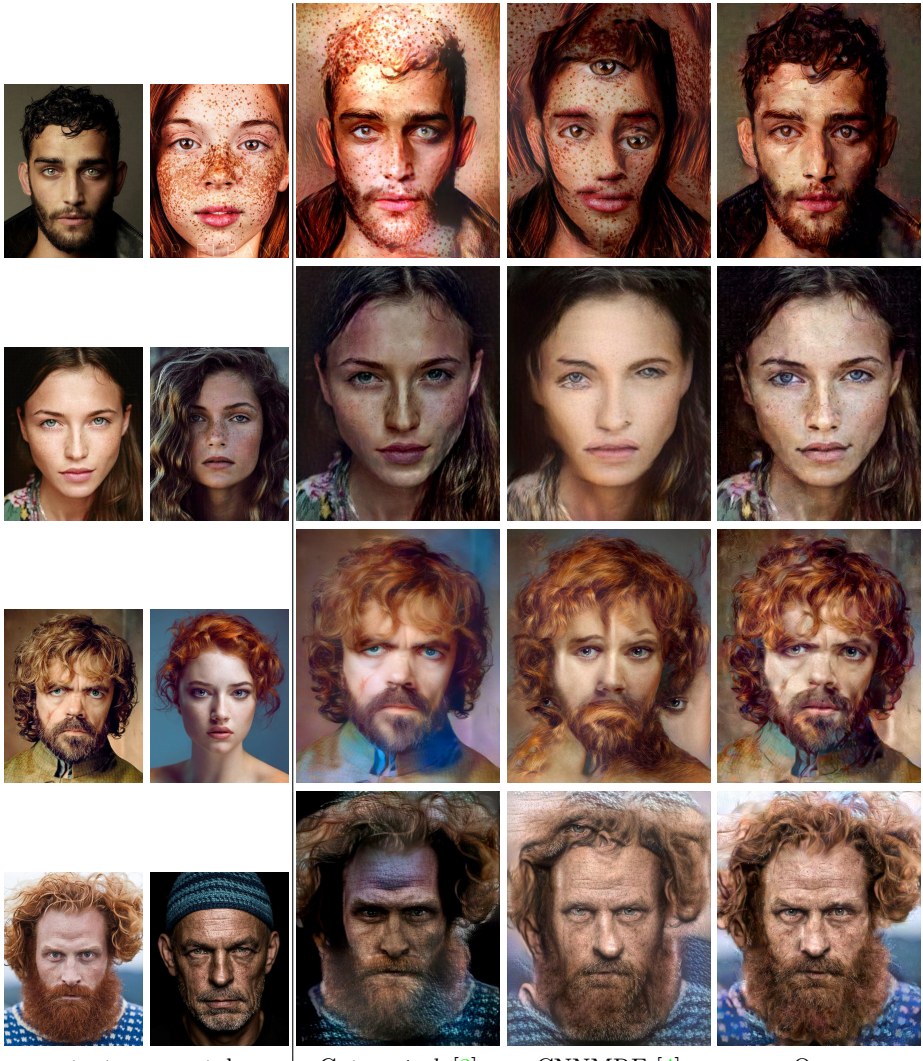


**Fig. 9.**

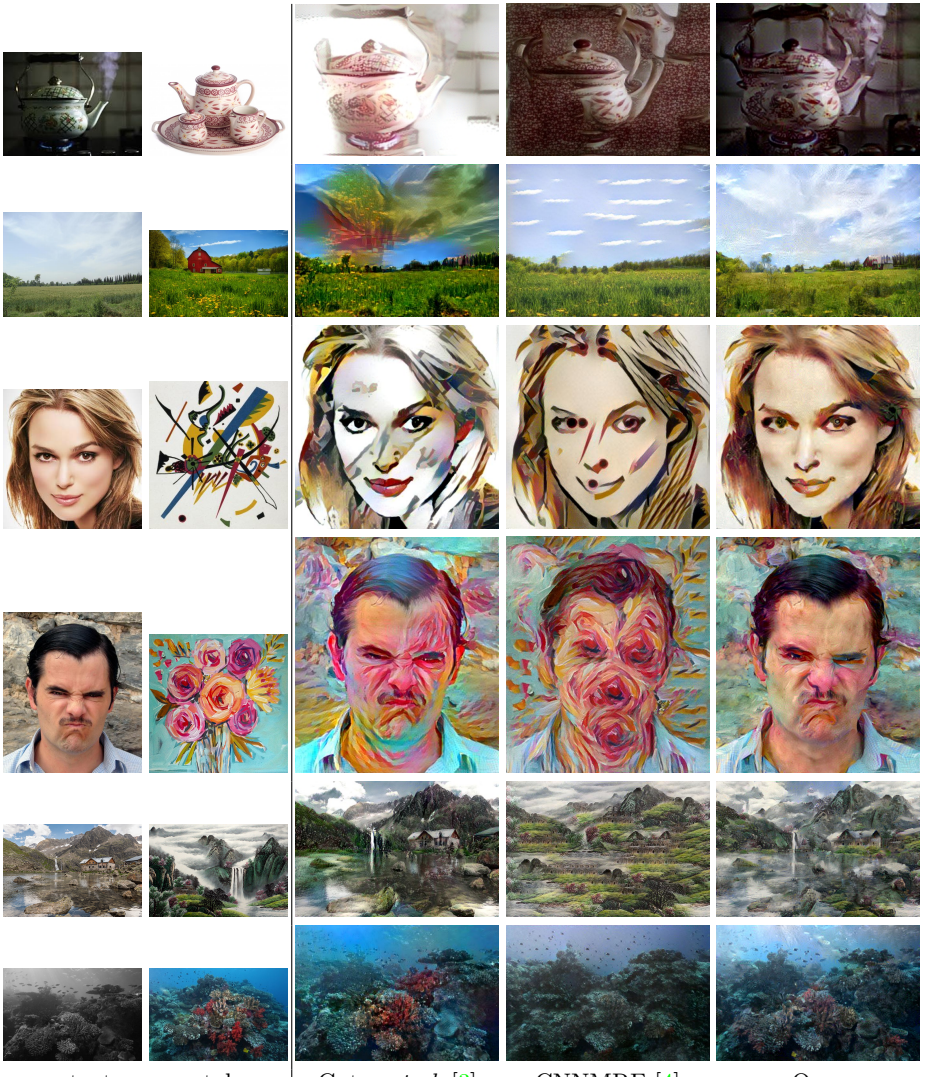


**Fig. 10.**

**Fig. 11.**

# 2. Puppet Control

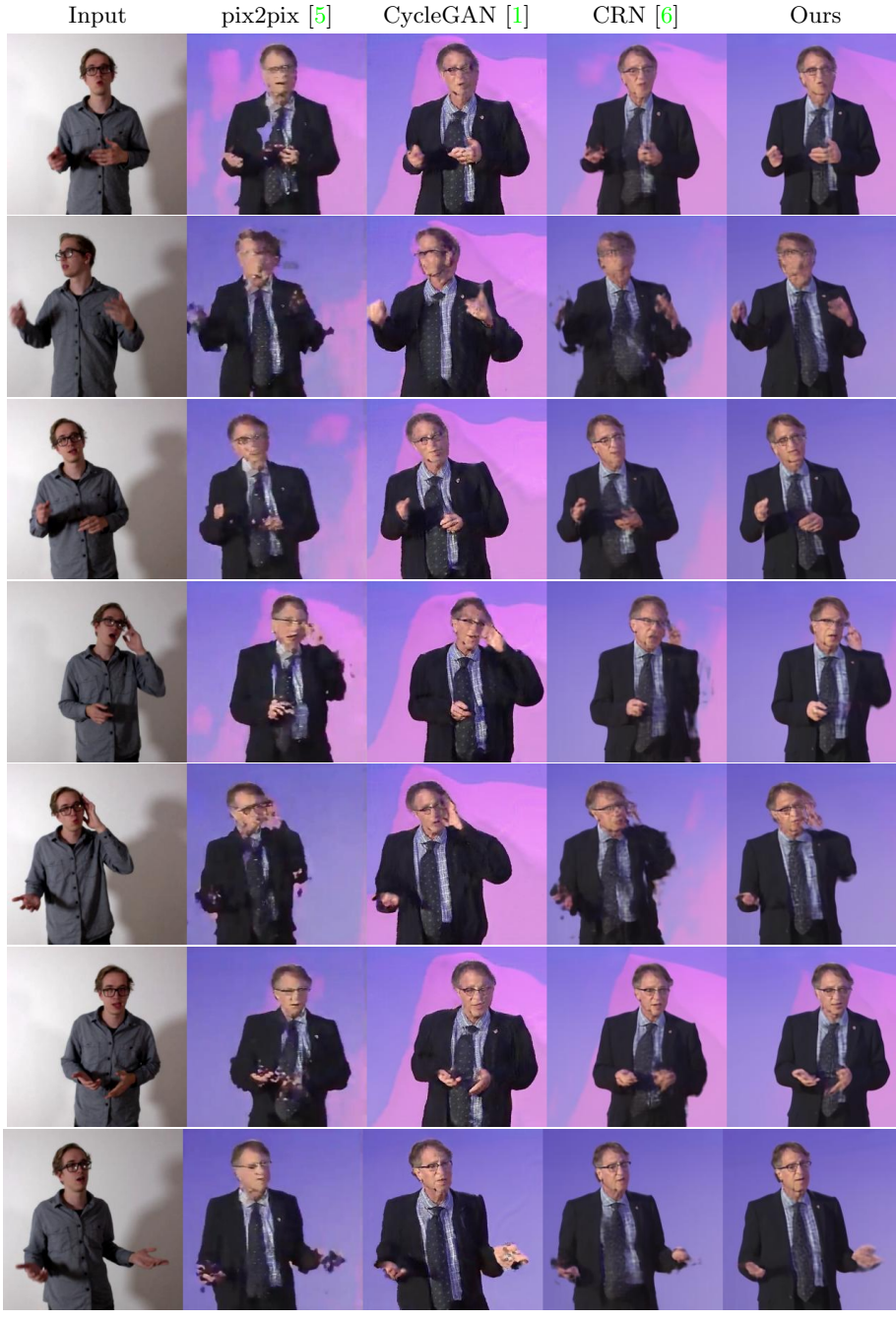


**Fig. 12.**


**Fig. 13.**

# 3. Single Image Animation


**Fig. 14.**


**Fig. 15.**

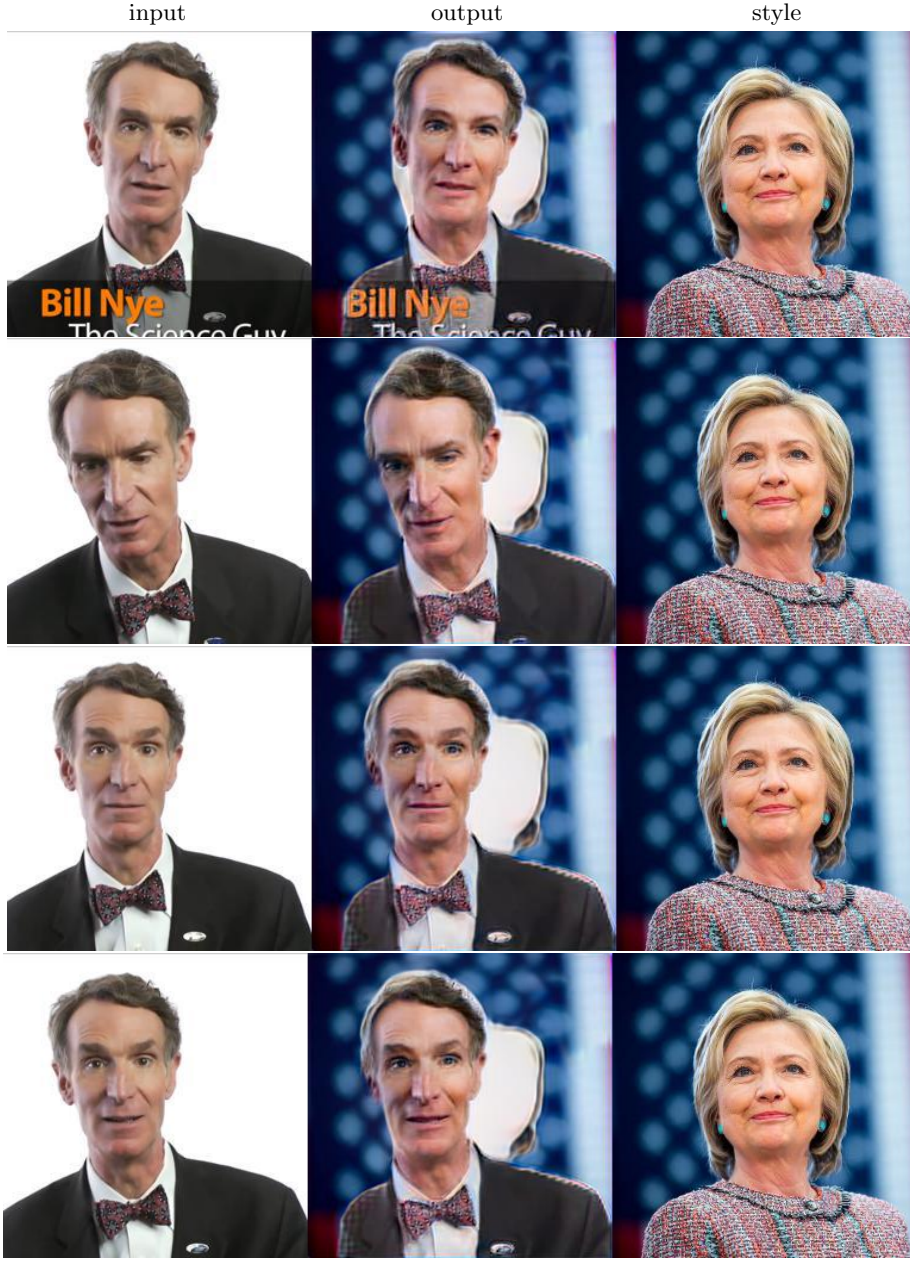


**Fig. 16.**


**Fig. 17.**

# 4. Domain Translation

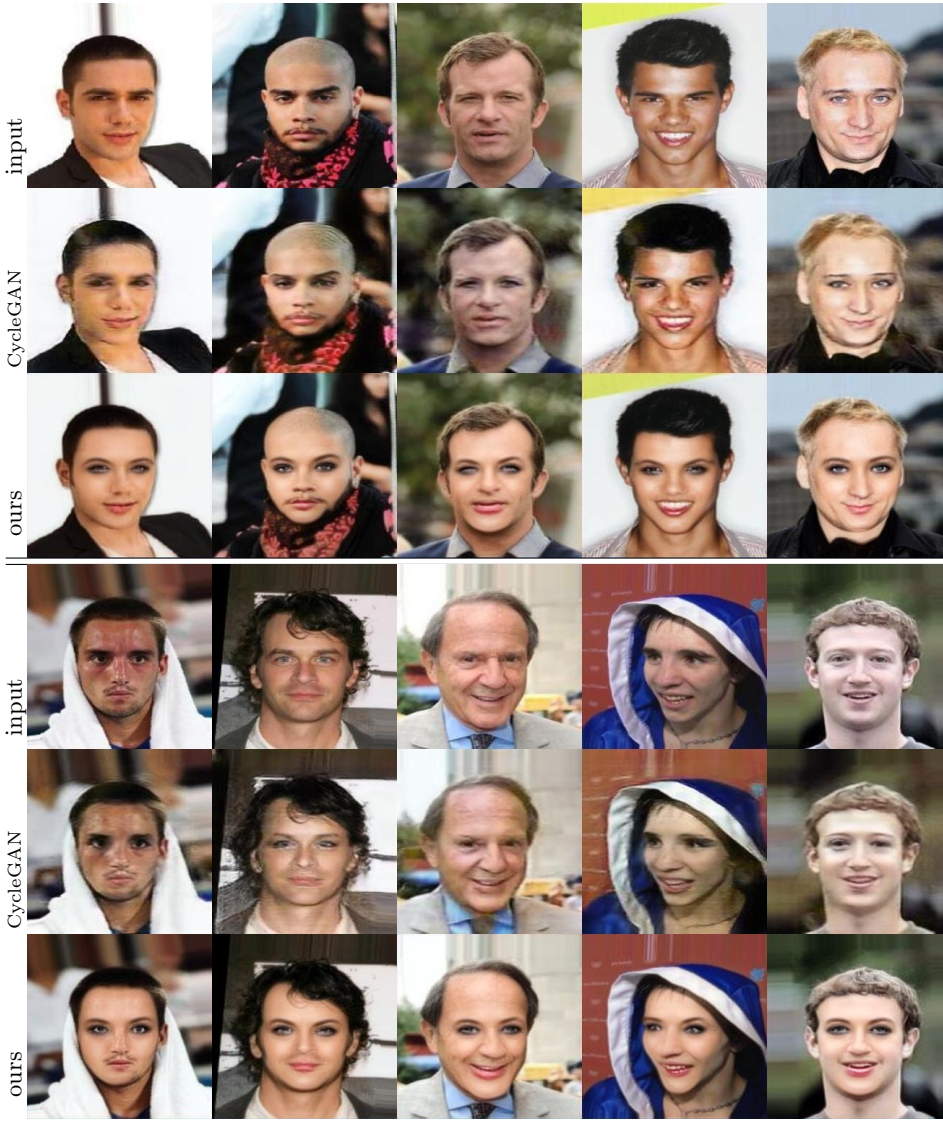


**Fig. 18.** male-to-female


**Fig. 19.** female-to-male

## References

1. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: ICCV. (2017) 3, 15, 16
2. Talmi, I., Mechrez, R., Zelnik-Manor, L.: Template matching with deformable diversity similarity. In: CVPR. (2017) 4
3. Gatys, L.A., Ecker, A.S., Bethge, M.: Image style transfer using convolutional neural networks. In: CVPR. (2016) 8, 9, 10, 11, 12
4. Li, C., Wand, M.: Combining markov random fields and convolutional neural networks for image synthesis. In: CVPR. (2016) 8, 9, 10, 11, 12
5. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: CVPR. (2017) 15, 16
6. Chen, Q., Koltun, V.: Photographic image synthesis with cascaded refinement networks. In: ICCV. (2017) 15, 16