Supplementary: Bidirectional One-Shot Unsupervised Domain Mapping

Tomer Cohen
Tel Aviv University
tomer104@gmail.com

Lior Wolf
Facebook AI Research and Tel Aviv University
wolf@cs.tau.ac.il

Number of parameters

See Tab. 1 for a comparison of network sizes. Our network requires 56% less trainable parameters than [3] and 3.7% less than [4]. Compared to OST [1], our network has 42% less parameters for Phase I, since we use only four residual blocks instead of six, and we do not use a discriminator. In Phase II, our network is 11% larger due to the separate weights. Note that the majority of training is in Phase I, while Phase II is at most three epochs. Consequentially, our method is at least 23% faster to train. Fig. 1 depicts the convergence of both methods, and Tab. 2 details the training times. Finally, [2] has significantly less trainable parameters than all other methods. However, it is less capable for the given task and it relies on supervision in the form of a pre-trained VGG network with an additional 20M parameters.

Ours	OST[1]	Cycle[4]	MUNIT[3]	Gatys[2]
10.2/20.4M	17.7/18.4M	21.2M	46.5M	0.2M

Table 1. Number of trainable parameters for different methods. Our model is slightly larger than [1], but it is faster to train nevertheless.

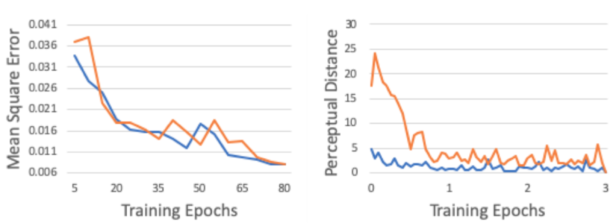

Figure 1. **Left:** Reconstruction error during training Phase I. **Right:** Perceptual distance of A to B mappings to final mapped image, during training Phase II. Blue: our method. Orange: OST.

	Duration	Ours	OST [1]
Stage A	80 epochs	2min/epoch	2.8min/epoch
Stage B	3 epochs	10.2min/epoch	8.9min/epoch
Total	83 epochs	3.2 hours	4.18 hours

Table 2. Comparing training times between OST and our method.

Additional figures

Fig. 2, 3, 4, and 5 present a comparison of sample results obtained by our method, and by multiple baseline methods. These include OST [1], MUNIT [3], CycleGAN [4] and Neural Style Transfer [2]. The results are shown for different domain mapping tasks: Winter2Summer (Fig. 2), Summer2Winter (Fig. 3), Photo2Monet (Fig. 4) and Monet2Photo (Fig. 5).

References

[1] Sagie Benaim and Lior Wolf. One-shot unsupervised cross domain translation. In *NeurIPS*, 2018.

[2] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *CVPR*, 2016.

[3] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. Multimodal unsupervised image-to-image translation. In *ECCV*, 2018.

[4] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *IEEE International Conference on Computer Vision*, 2017.

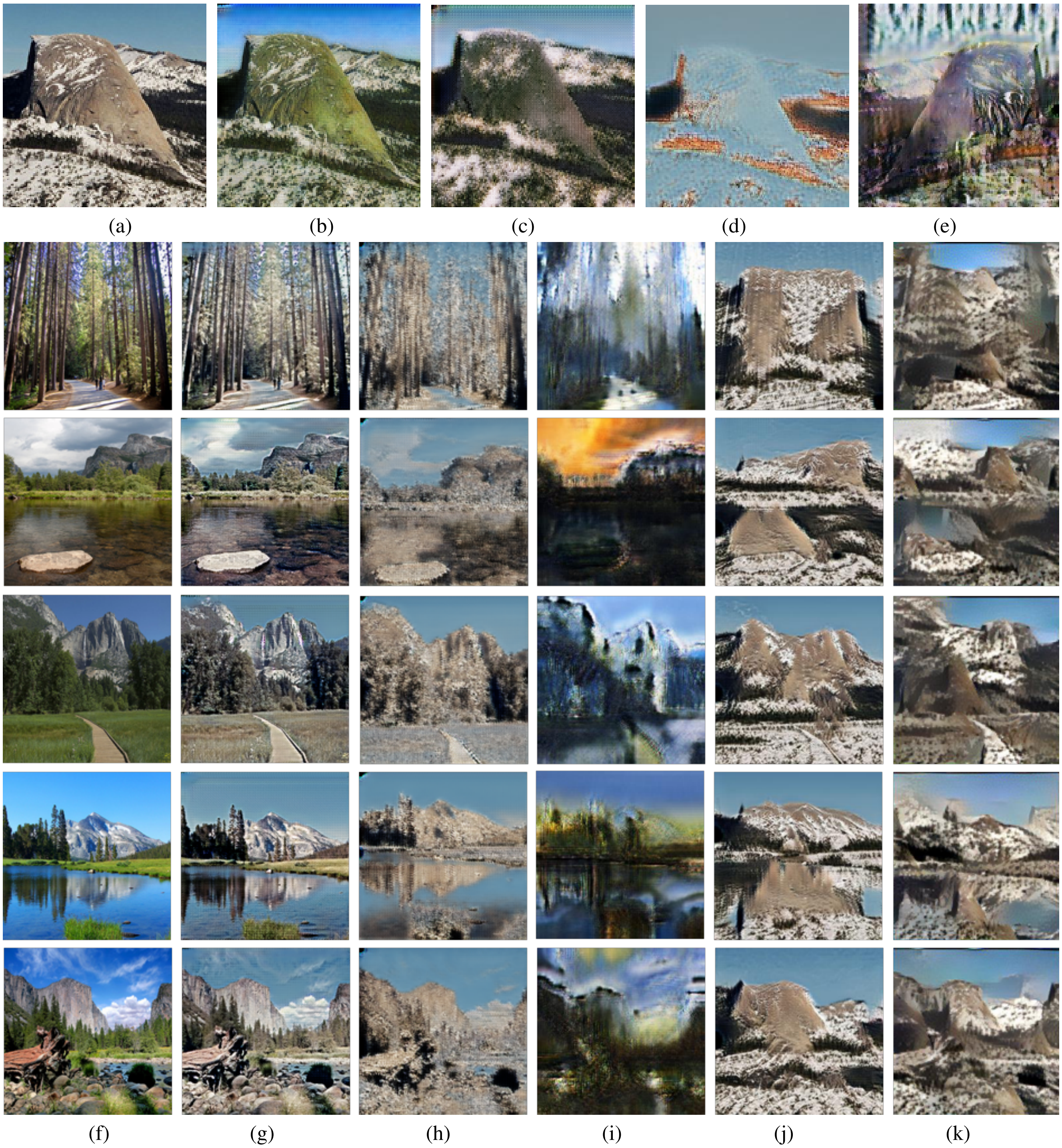

Figure 2. Mapping winter (domain A) to summer (domain B). (a) The sample $x \in A$. (b) Our result for mapping to B. (c) The result of OST. (d) The result of MUNIT. (e) The result of CycleGAN. (f) Two samples $s \in B$. (g) Our result mapping in the other direction, using the same learned model. (h-j) This mapping for the baseline methods in the same order as above. (k) The results of [2], which are only shown in this direction since they cannot benefit from multiple images in B.

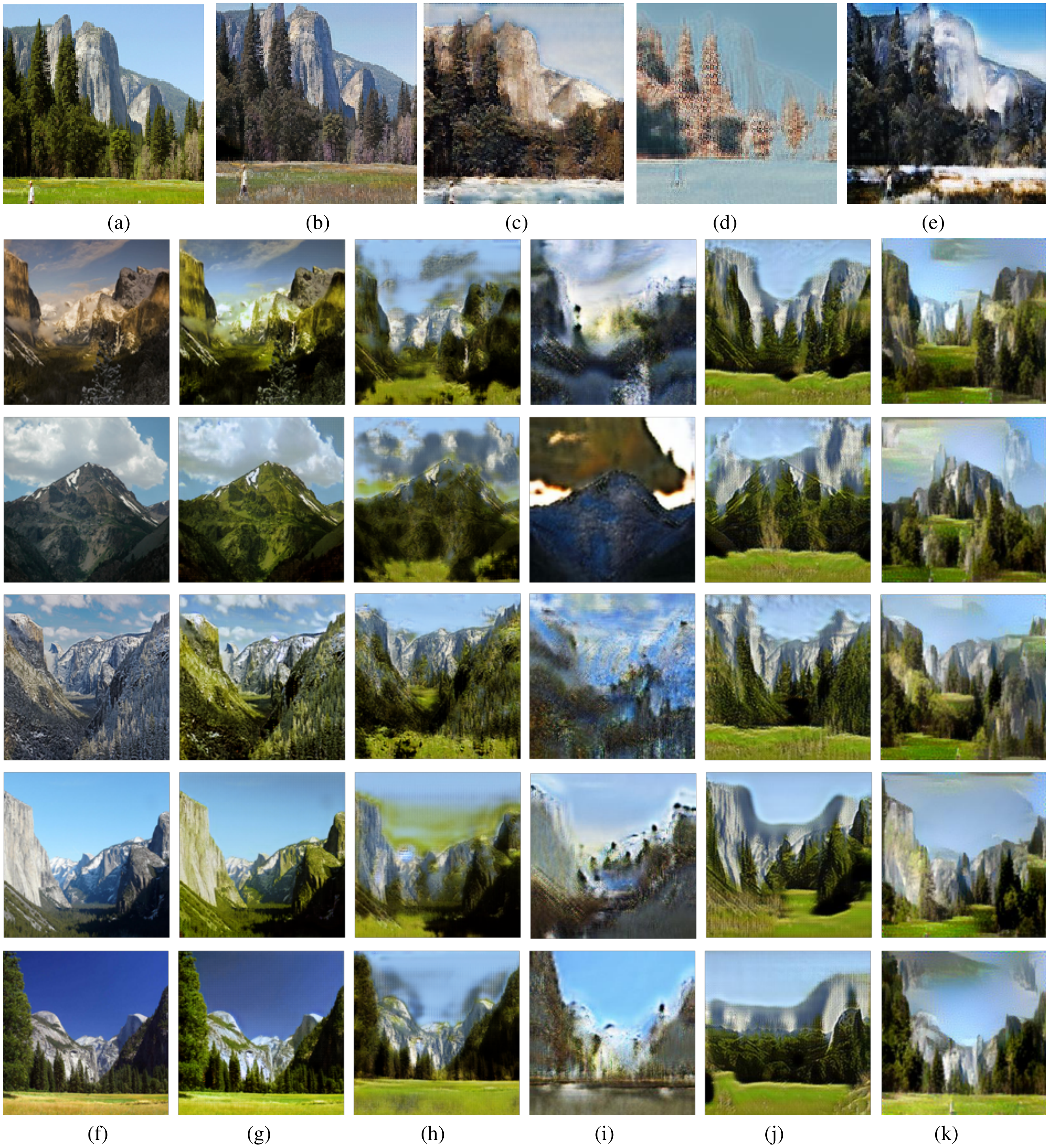

Figure 3. Mapping summer (domain A) to winter (domain B). (a) The sample $x \in A$. (b) Our result for mapping to B. (c) The result of OST. (d) The result of MUNIT. (e) The result of CycleGAN. (f) Two samples $s \in B$. (g) Our result mapping in the other direction, using the same learned model. (h-j) This mapping for the baseline methods in the same order as above. (k) The results of [2], which are only shown in this direction since they cannot benefit from multiple images in B.

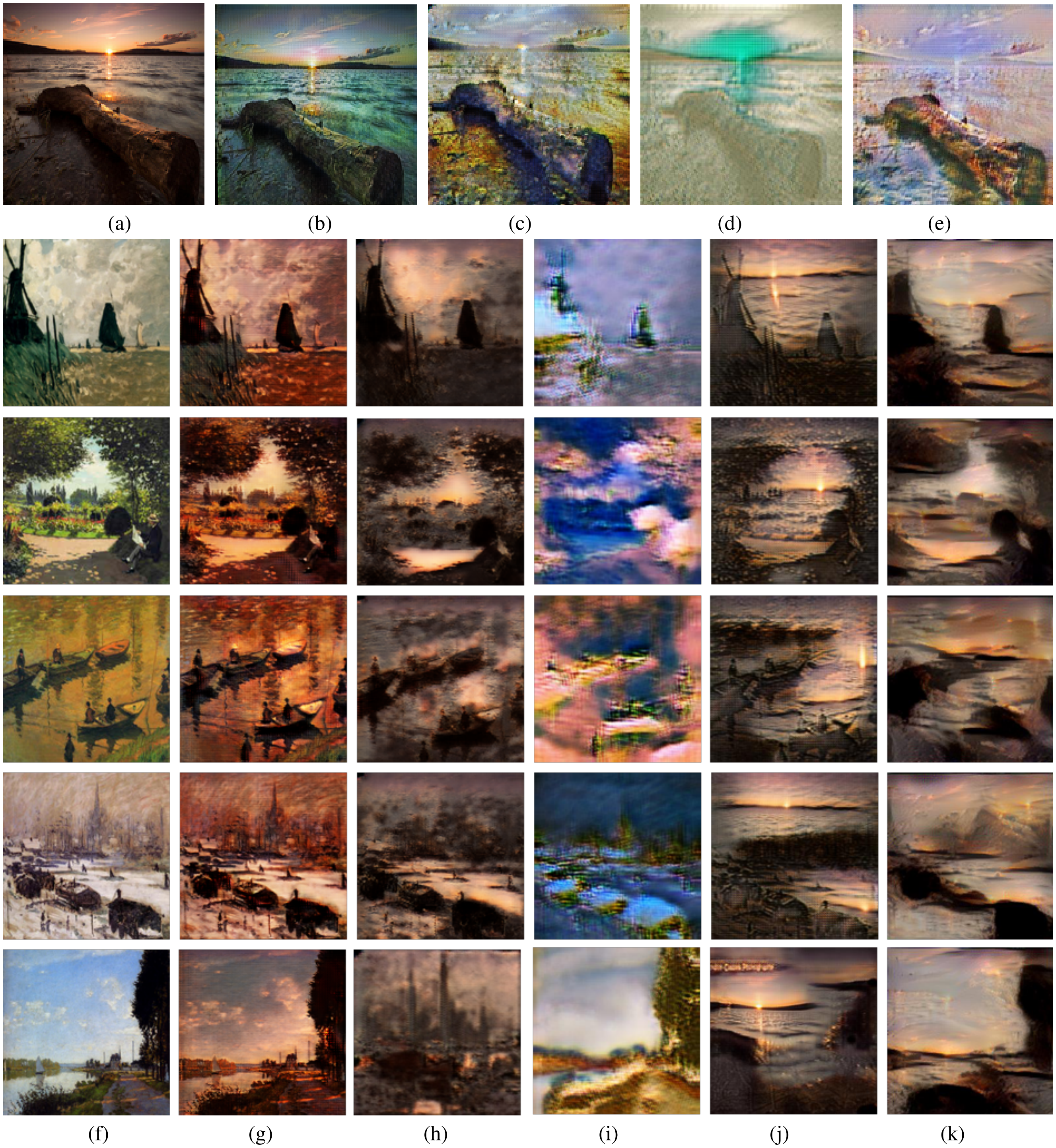

Figure 4. Mapping photo (domain A) to Monet (domain B). (a) The sample $x \in A$. (b) Our result for mapping to B. (c) The result of OST. (d) The result of MUNIT. (e) The result of CycleGAN. (f) Two samples $s \in B$. (g) Our result mapping in the other direction, using the same learned model. (h-j) This mapping for the baseline methods in the same order as above. (k) The results of [2], which are only shown in this direction since they cannot benefit from multiple images in B.

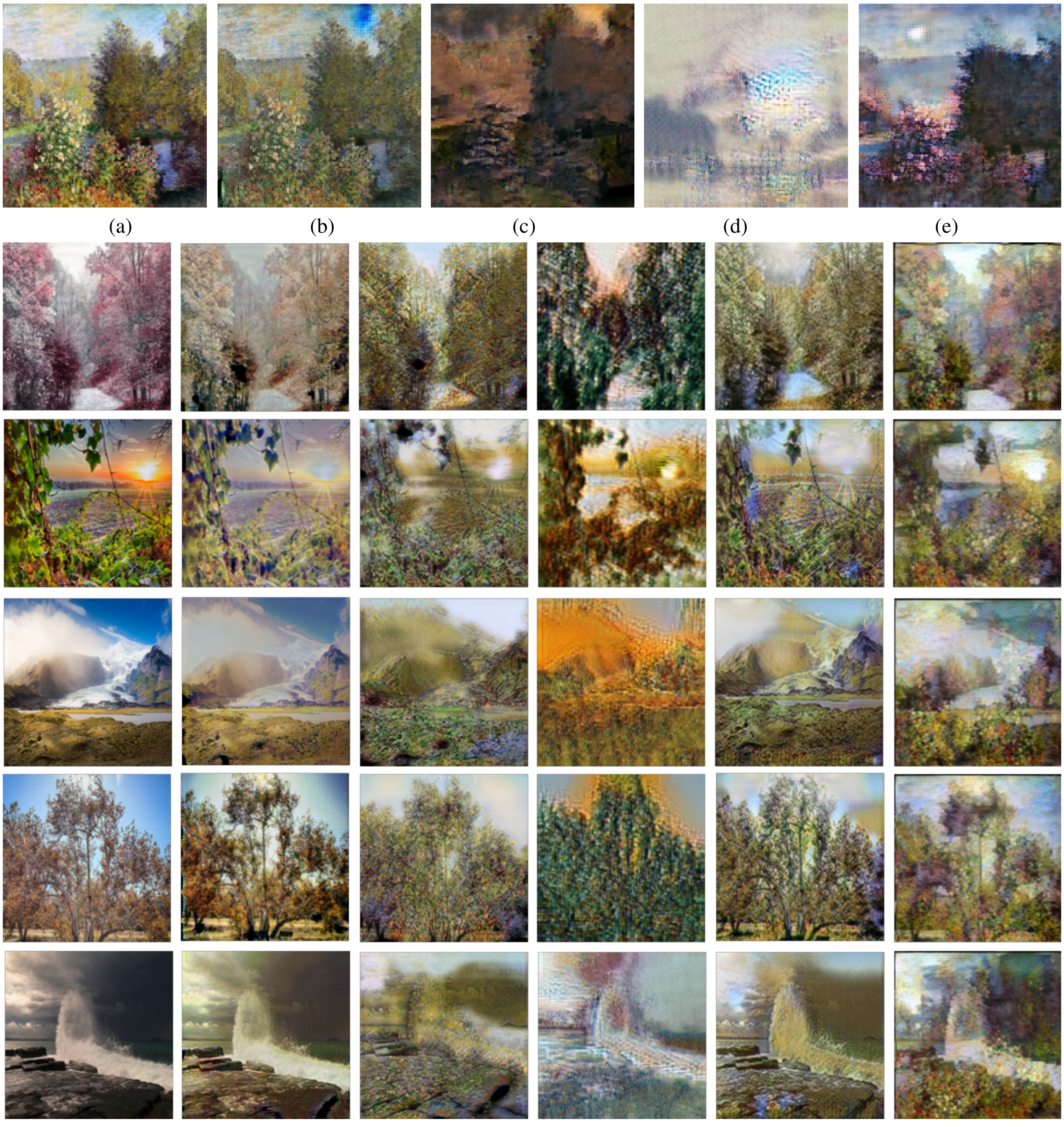

Figure 5. Mapping Monet (domain A) to photo (domain B). (a) The sample $x \in A$. (b) Our result for mapping to B. (c) The result of OST. (d) The result of MUNIT. (e) The result of CycleGAN. (f) Two samples $s \in B$. (g) Our result mapping in the other direction, using the same learned model. (h-j) This mapping for the baseline methods in the same order as above. (k) The results of [2], which are only shown in this direction since they cannot benefit from multiple images in B.