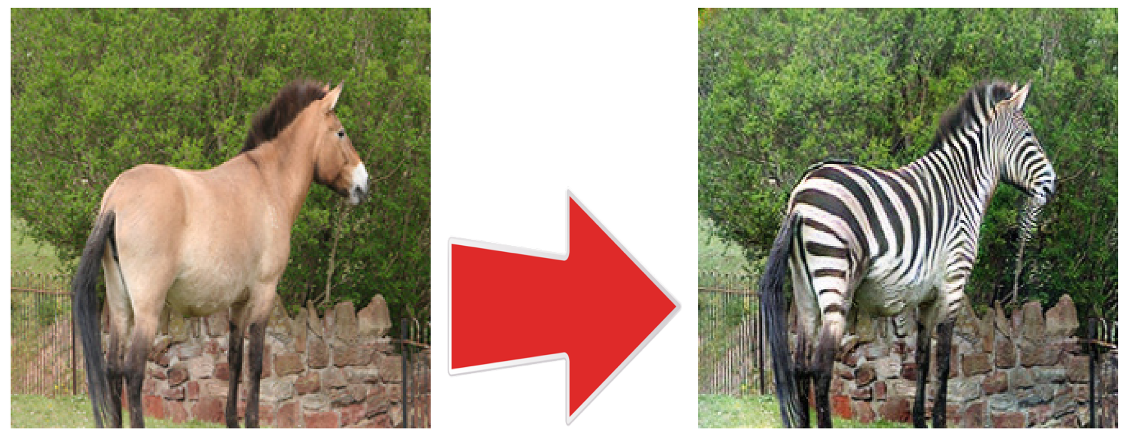
ESTHER: Extremely Simple Image Translation Through Self-Regularization

Chao "Harry" Yang*, Taehwan Kim†, Ruizhe Wang†, Hao Peng† and C.-C. Jay Kuo*
*University of Southern California
†OBEN Inc.

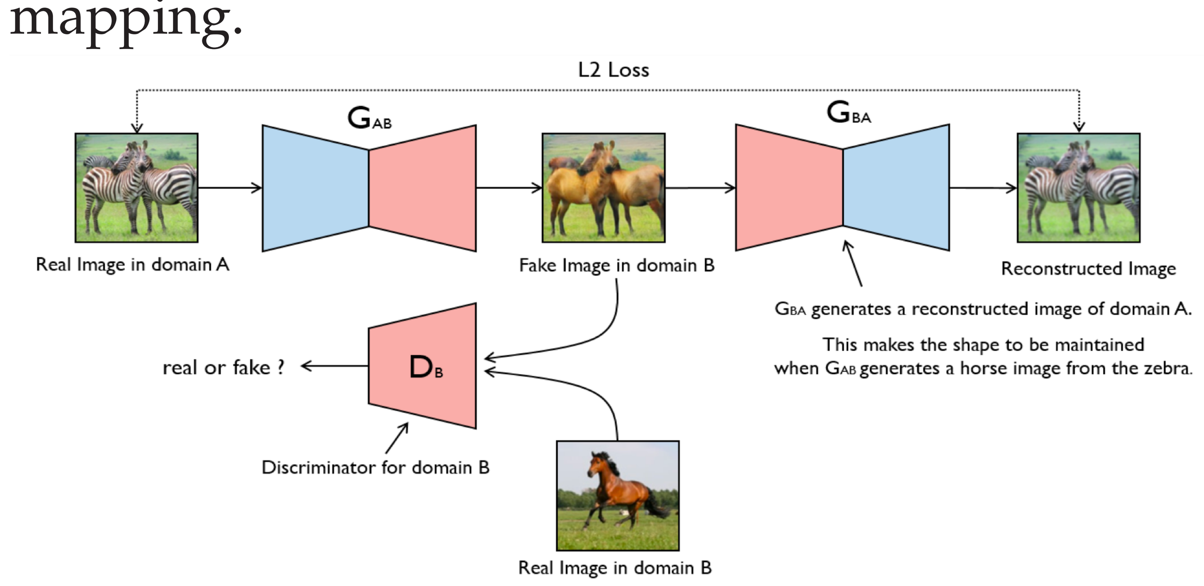
Overview



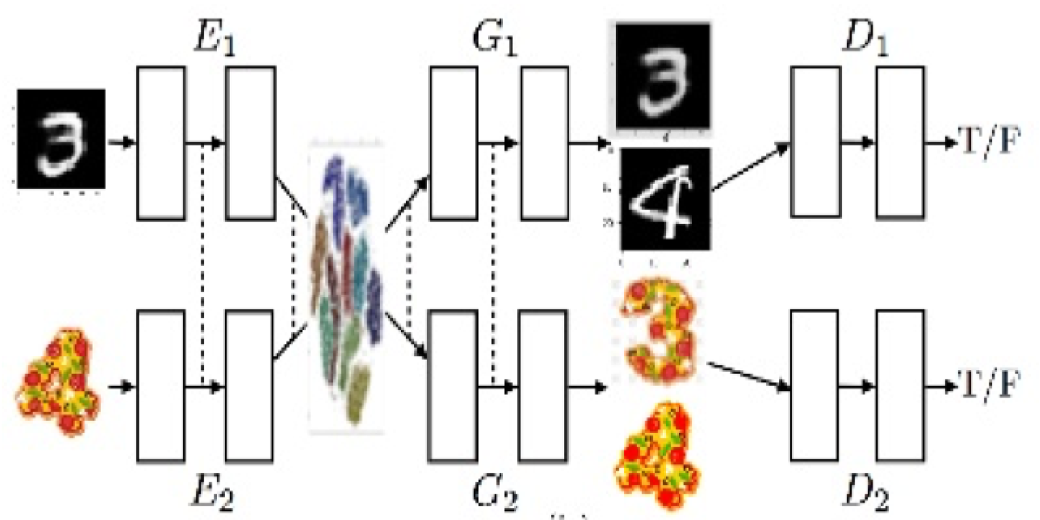
Problem: *Unpaired image translation between two domains*, where the goal is to learn the mapping from an input image in the source domain to an output image in the target domain.
Challenge: An ill-posed task (no paired data)!
Method: An extremely simple yet effective image translation approach, which consists of a *single generator* and is trained with a *self-regularization term* and an *adversarial term*.
Results: Our model achieves better performance than other methods on a broad range of tasks and applications.

Related Works

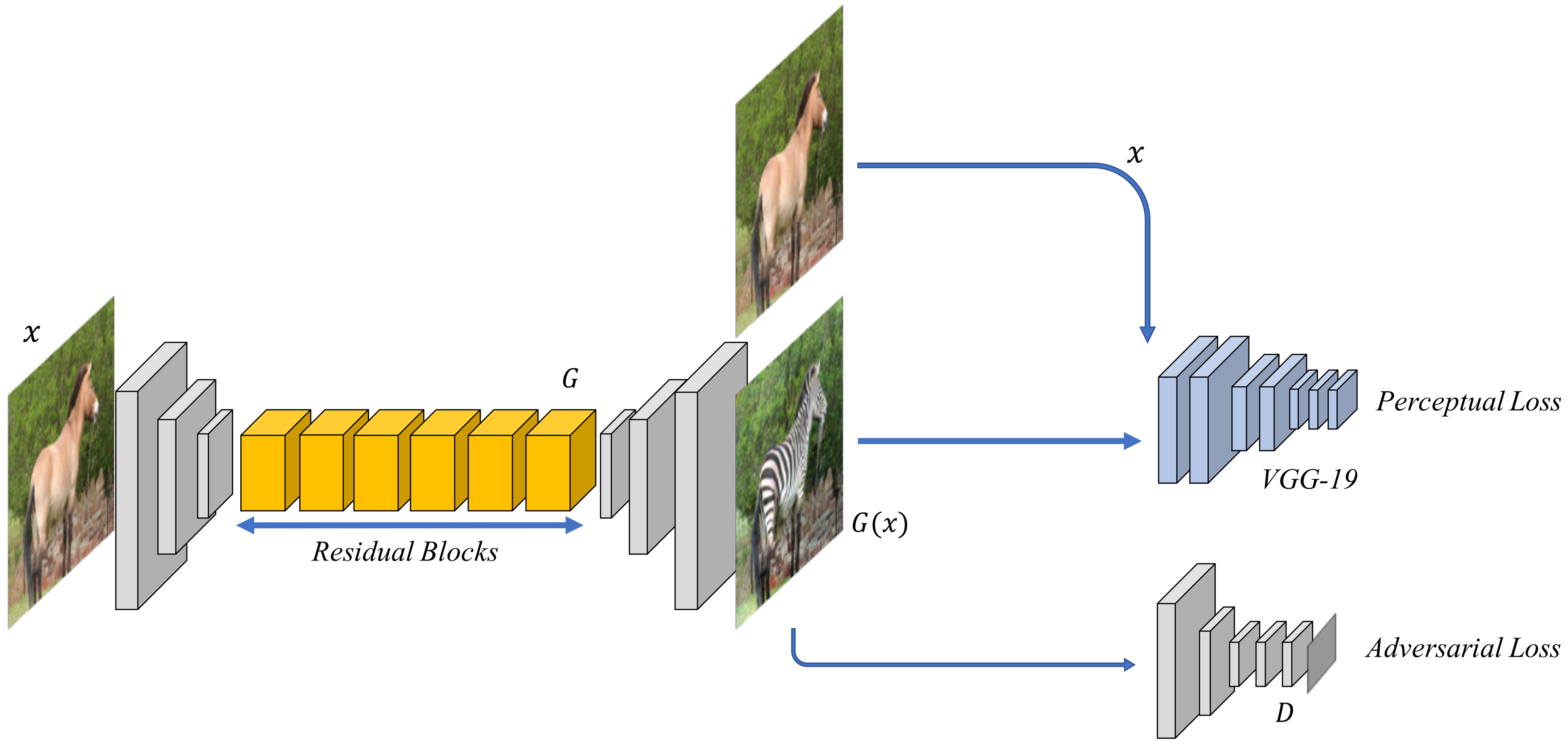
CycleGAN: relies on "laziness" of neural network and cycle-consistency to ensure 1-1 mapping.



UNIT: relies on shared latent space and cycle-consistency.



Our Model



Our model consists of a single generator G that maps an image from one domain to another domain. We train G using the self-regularization perceptual loss and the adversarial loss:

$$\mathcal{L}_G(\theta) = \ell_{adv}(G_\theta(x), Y) + \lambda \ell_{reg}(x, G_\theta(x)).$$

Training Details

Training Losses:

- Image domain adversarial loss (with PatchGAN):

$$\ell_{adv}(G_\theta(x), Y) = E_{y\sim Y}[\log D(y)] + E_{x\sim X}[\log(1 - D(G(x)))].$$

- The self-regularization loss (with Perceptual Loss):

$$\ell_{reg}(y', x) = \sum_{l=1}^{3} \frac{1}{H_l W_l} \sum_{h,w} (\| w_l^T \circ (\hat{F}(x)^l_{hw} - \hat{F}(y')^l_{hw}) \|_2^2).$$

Notes: we use VGG-19 pre-trained on ImageNet as F. We obtained the best results by using the first three layers of VGG-19, which preserves the low-level traits of the input during translation.

Domain Adaptation Results

From photo to maps:

Method	Accuracy
Pix2Pix	43.18%
CycleGAN	45.91%
Ours	**46.72%**

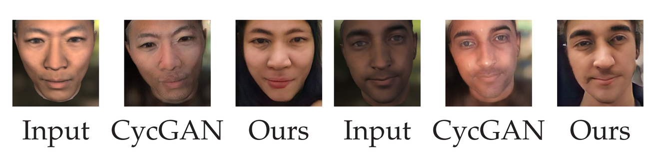
Figure 1: Visualization (left) and quantitative evaluation (right) based on image to map translation results.

From MNIST to USPS and MNIST-M:

Method	USPS	MNIST-M
CoGAN	95.65%	-
PixelDA	95.90%	98.20%
UNIT	95.97%	-
CycleGAN	94.28%	93.15%
Target-only	96.50%	96.40%
Ours	**96.80%**	**98.33%**

Figure 2: Visualization (left) and classification accuracy (right) based on domain adaptation from MNIST to USPS and MNIST-M results.

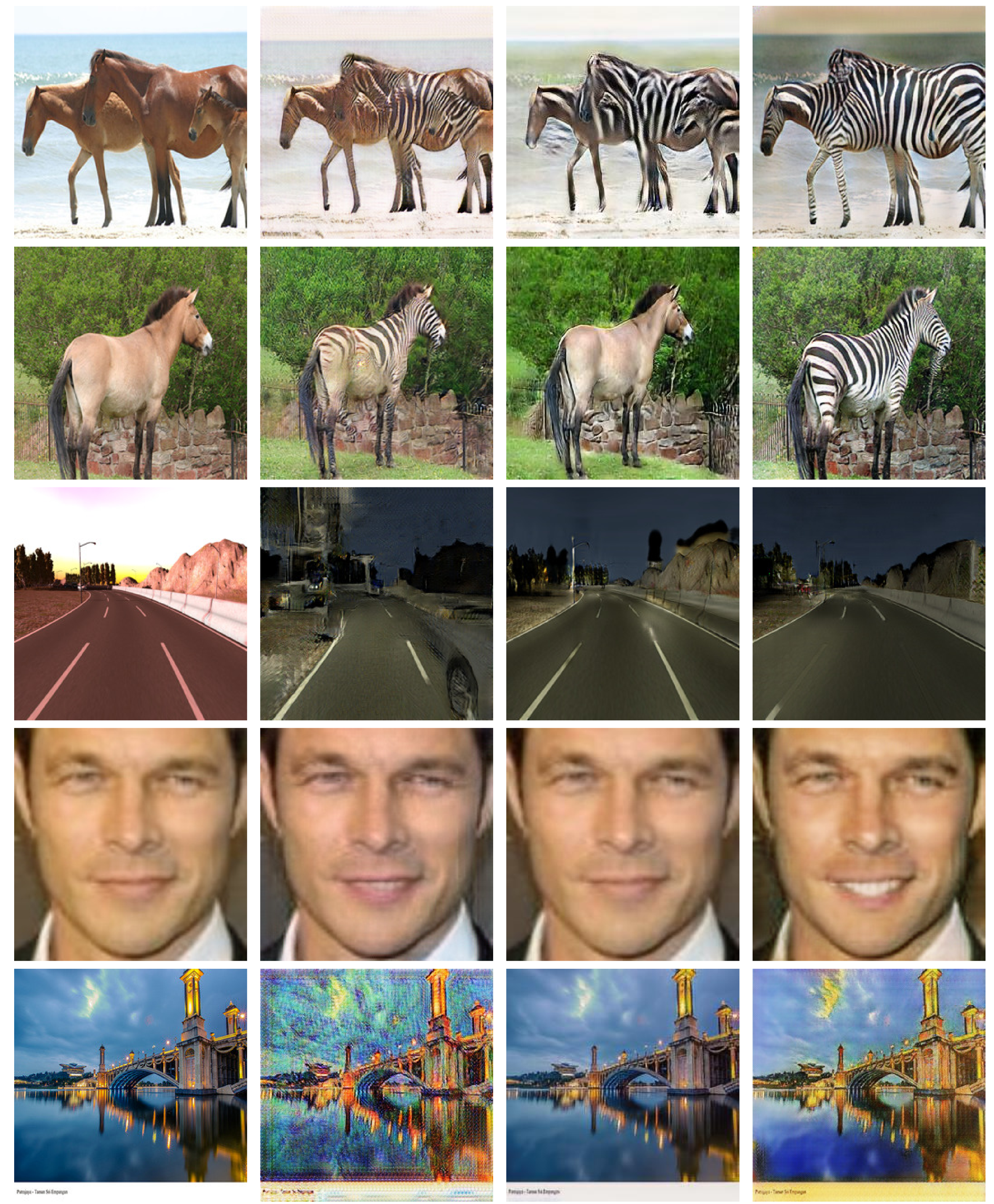
From rendered face to real face:



Method	MSE
Baseline	2.26
CycleGAN	2.04
Ours	**1.97**

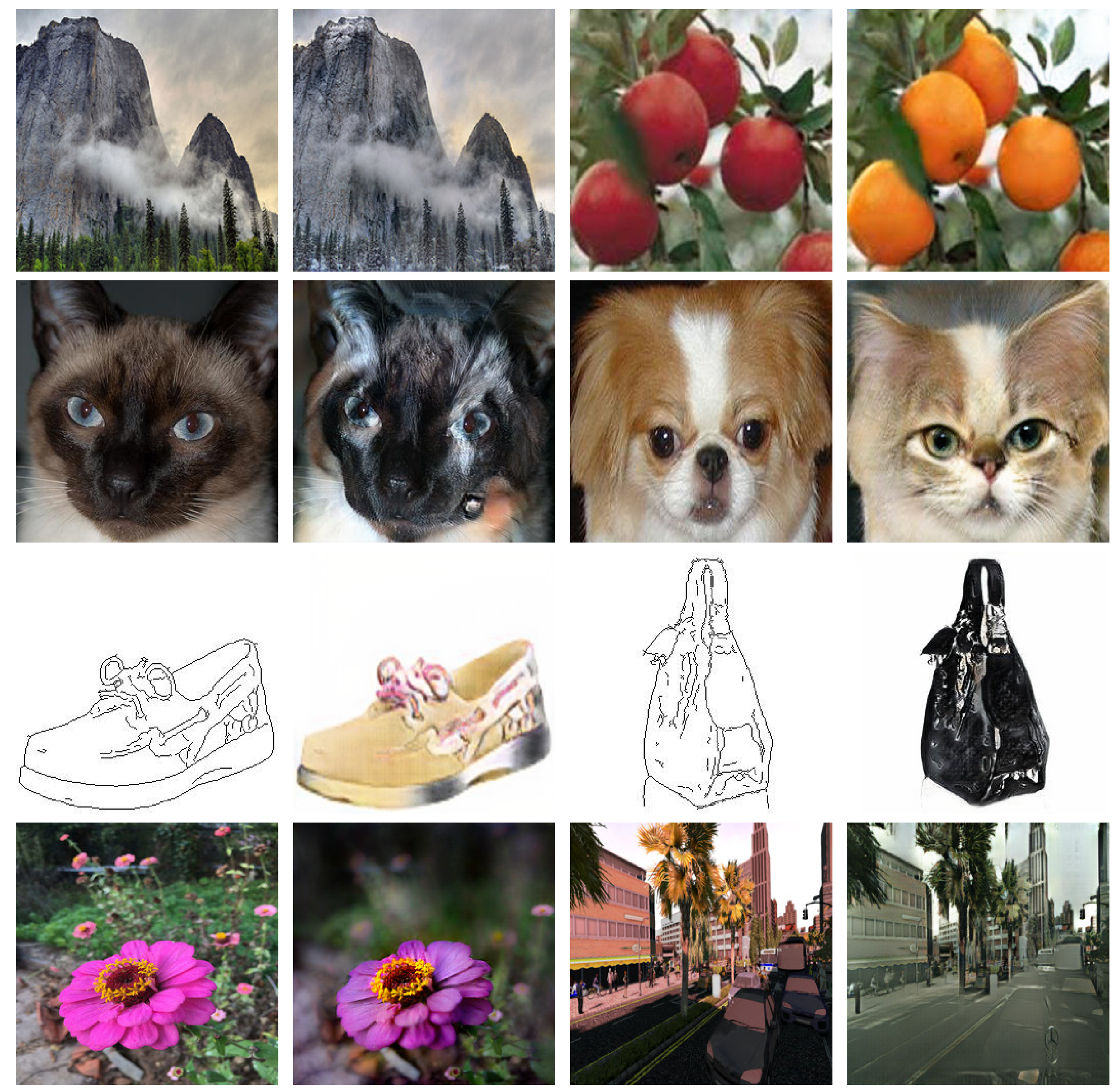
Figure 3: Visualization (left) and 3DMM estimation accuracy (right) based on domain adaptation from rendered faces to real faces results.

Visual Comparisons



(a) Input (b) CycleGAN (c) UNIT (d) Ours

More Results



Please **attend** to `https://arxiv.org/abs/1806.06195` for extended version!