One-Sided Unsupervised Domain Mapping

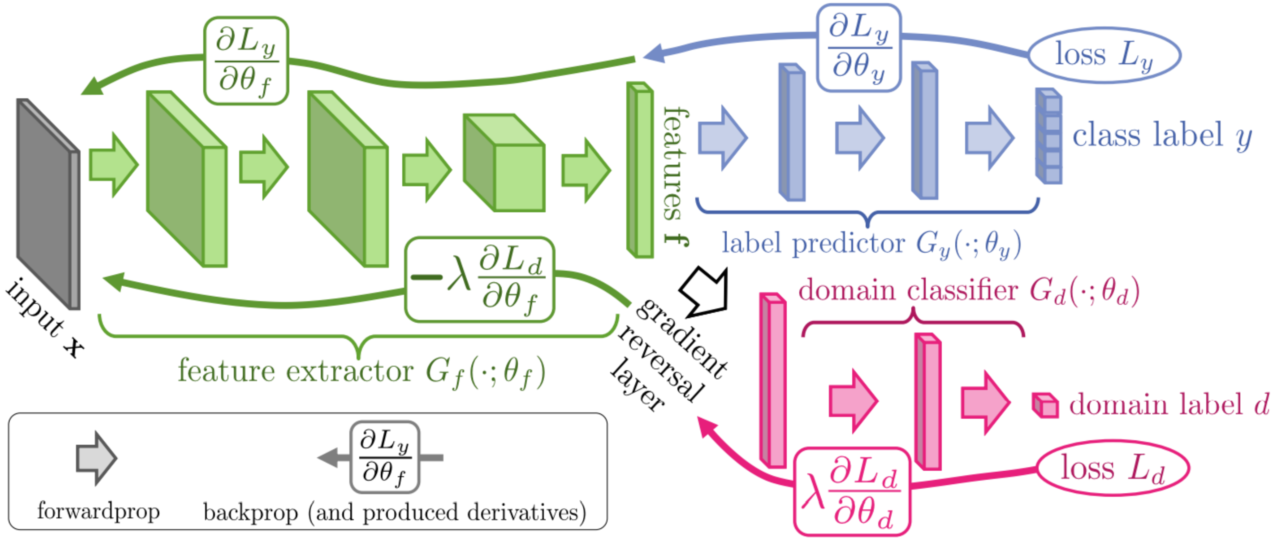
Sagie Benaim and Lior Wolf

Latest Trends

1. Style Transfer (Gatys et al.)
    - Replaces statistics/texture given an exemplar

Not semantic

Latest Trends



2. Domain adaptation
    - "Domain-adversarial training of neural networks" Ganin et al.

Supervised and not generative

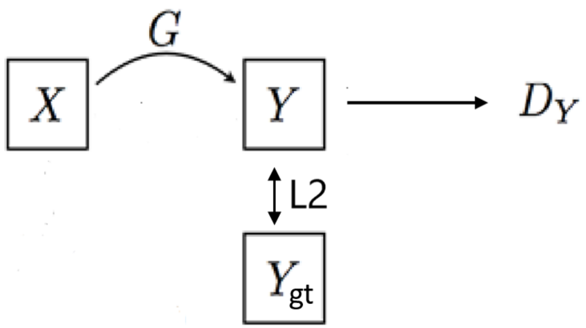
---

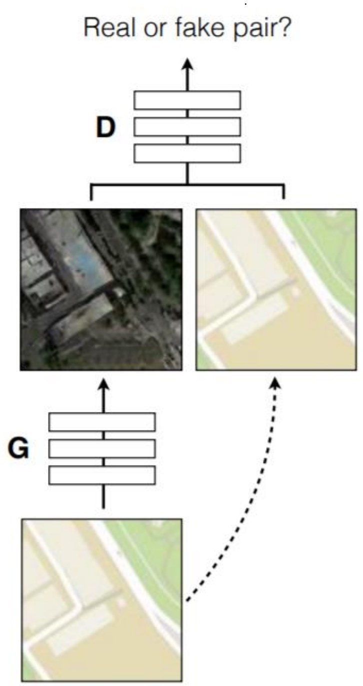
1. Style Transfer (Gatys et al.)
    - Replaces statistics/texture given an exemplar

Not semantic

Fully Supervised Alignment

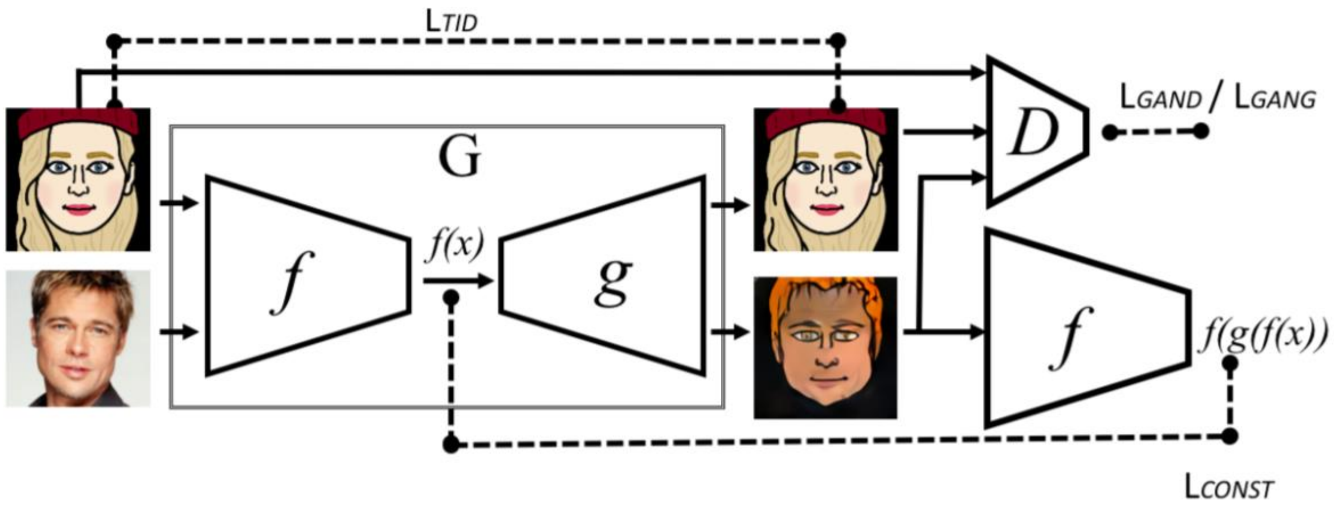
- "Image-to-image translation with conditional adversarial nets" Isola et al (pix2pix)





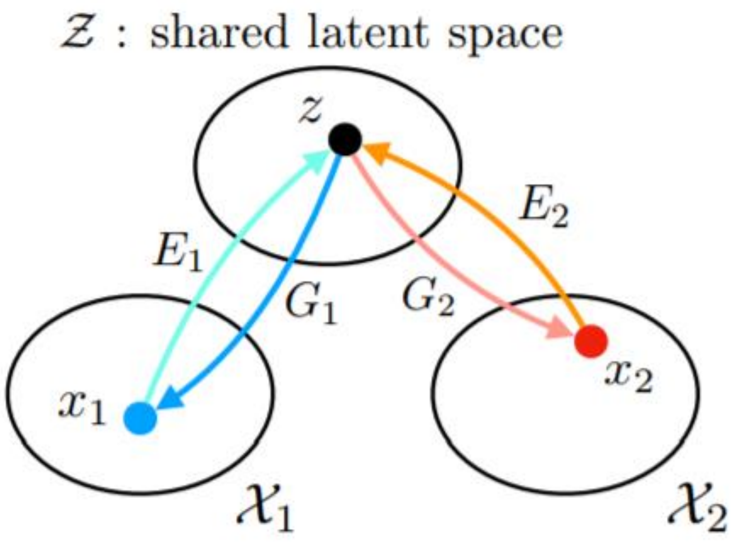
Partially Supervised Alignment

- "Unsupervised Cross-Domain Image Generation" Taigman et al.



Unsupervised Alignment

- Highly related domains
  - “Unsupervised Image-to-Image Translation Networks” Liu et al.



Circular GANs

DiscoGAN: "Learning to Discover Cross-Domain Relations with Generative Adversarial Networks". Kim et al. ICML'17.

CycleGAN: "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks". Zhu et al. arXiv:1703.10593, 2017.

"**DualGAN**: Unsupervised Dual Learning for Image-to-Image Translation". Zili et al. arXiv:1704.02510, 2017.

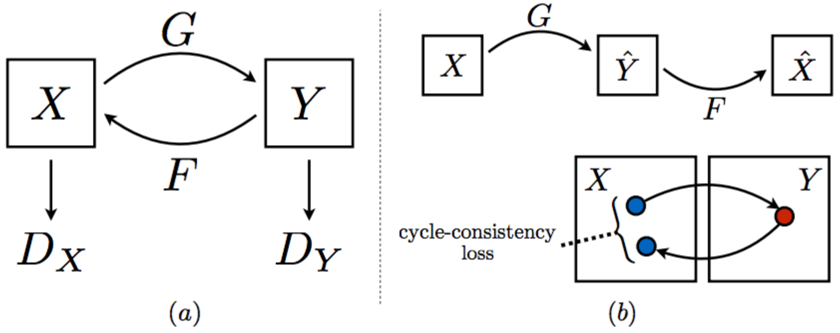
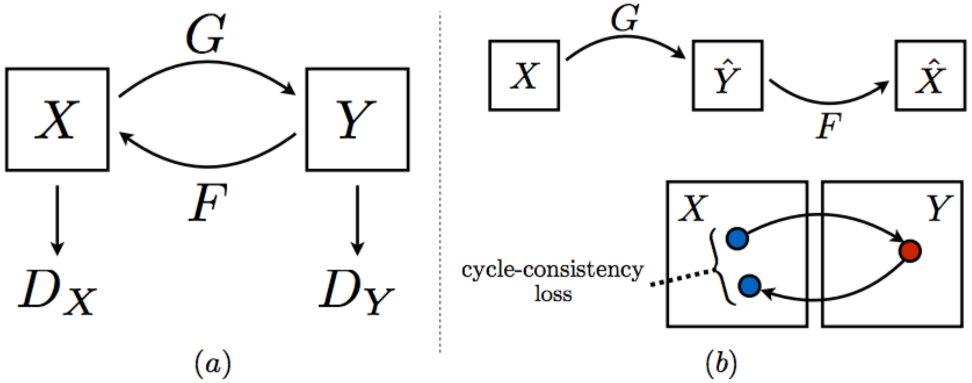


Image credit: Zhu et al.

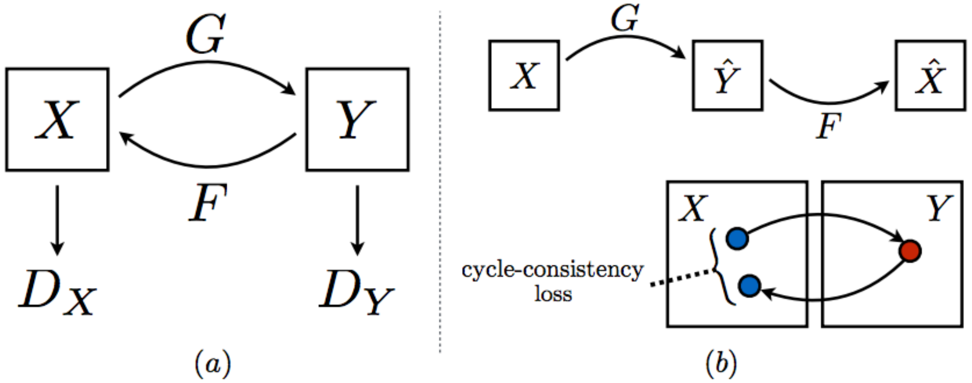
Circular GANs



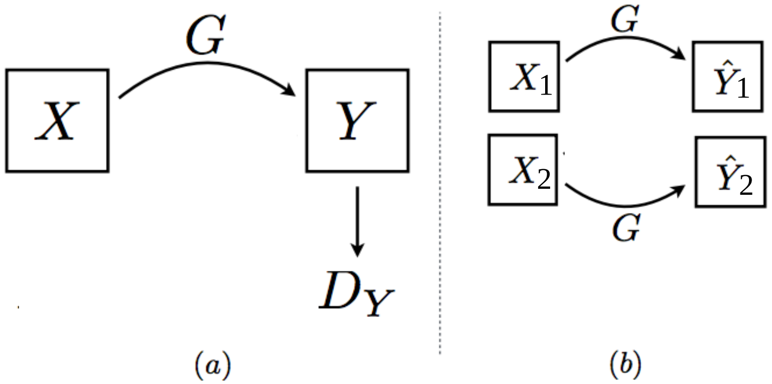
Circular GANs (DiscoGAN, CycleGAN, DualGAN)

$$x \sim F(G(x))$$
$$y \sim G(F(y))$$

Introducing DistGAN



Circular GANs (DiscoGAN, CycleGAN, DualGAN)



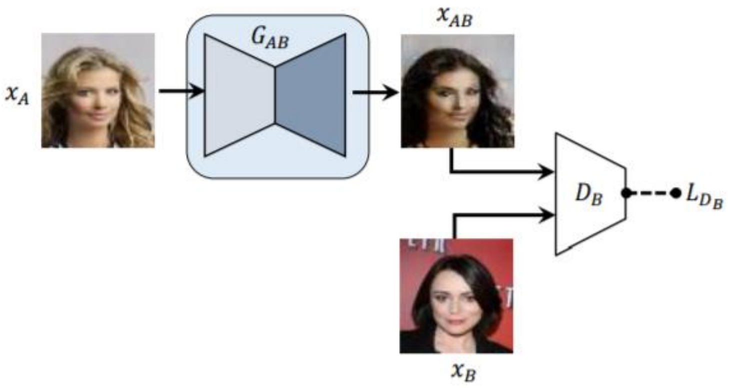
DistGAN

Sagie Benaim (TAU), Lior Wolf.
"One-Sided Unsupervised Domain Mapping"
arXiv 1706.00826, 2017. Public code on Github

$$x \sim F(G(x))$$
$$y \sim G(F(y))$$

$$|x_1 - x_2|_1 \sim |G(x_1) - G(x_2)|_1$$

Basic Building Block: Conditional GAN

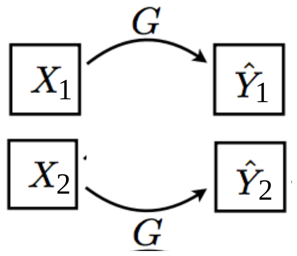


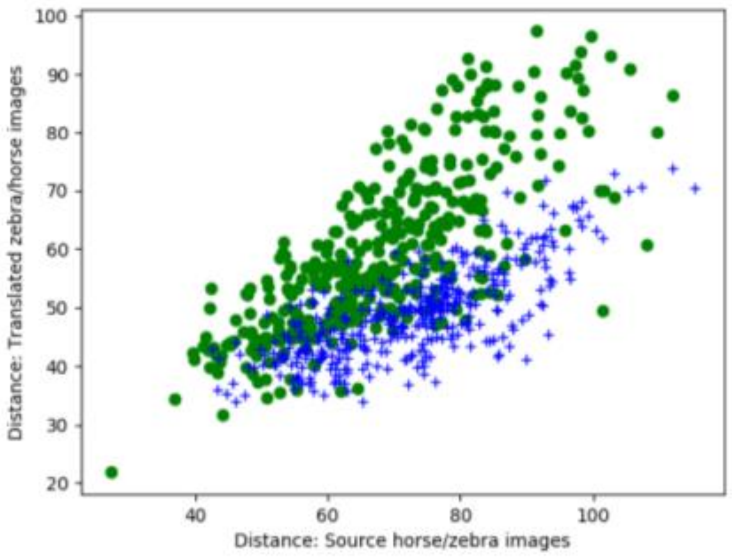
$$\mathcal{L}_{\mathrm{GAN}}(G_{AB}, D_B, \hat{p}_A, \hat{p}_B) = \mathbb{E}_{x_B \sim \hat{p}_B}[\log D_B(x_B)] + \mathbb{E}_{x_A \sim \hat{p}_A}[\log(1 - D_B(G_{AB}(x_A))]$$

- Other GAN variants can be used: w-gan, improved w-gan, BEGAN, etc.

Correlation of distances between X and Y

- A pair of images of a given distance are mapped to a pair of outputs with a similar distance
- $\left|x_i - x_j\right|_1$ and $\left|G(x_i) - G(x_j)\right|_1$ are highly correlated.





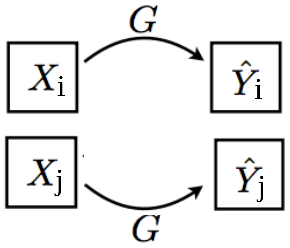
Analysis of CycleGAN's horse to zebra results

x $G(x)$ $\Pi(x)$

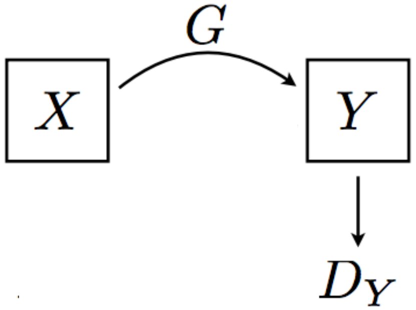
Non-negative matrix approx. of DiscoGAN's bag to shoe

The loss used

- A distance correlation loss (L1 more stable than multiplication):
  - $\sum_{x_i, x_j} |d_1 - d_2|$
  - $d_1 = \frac{1}{\sigma_A}(\left|x_i - x_j\right|_1 - \mu_A)$
  - $d_2 = \frac{1}{\sigma_B}(\left|G(x_i) - G(x_j)\right|_1 - \mu_A)$



- A GAN loss on Y



Losses Tradeoff

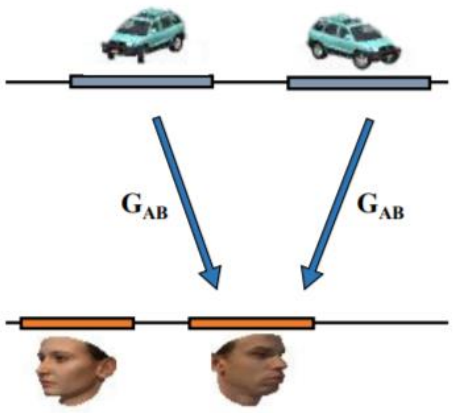
$$\alpha_{1A}\mathcal{L}_{\text{GAN}}(G_{AB}, D_B, \hat{p}_A, \hat{p}_B) + \alpha_{1B}\mathcal{L}_{\text{GAN}}(G_{BA}, D_A, \hat{p}_B, \hat{p}_A) + \alpha_{2A}\mathcal{L}_{\text{cycle}}(G_{AB}, G_{BA}, \hat{p}_A) +$$
$$\alpha_{2B}\mathcal{L}_{\text{cycle}}(G_{BA}, G_{AB}, \hat{p}_B) + \alpha_{3A}\mathcal{L}_{\text{distance}}(G_{AB}, \hat{p}_A) + \alpha_{3B}\mathcal{L}_{\text{distance}}(G_{BA}, \hat{p}_B) +$$
$$\alpha_{4A}\mathcal{L}_{\text{self-distance}}(G_{AB}, \hat{p}_A) + \alpha_{4B}\mathcal{L}_{\text{self-distance}}(G_{BA}, \hat{p}_B)$$

Table 1: Tradeoff weights for each experiment.

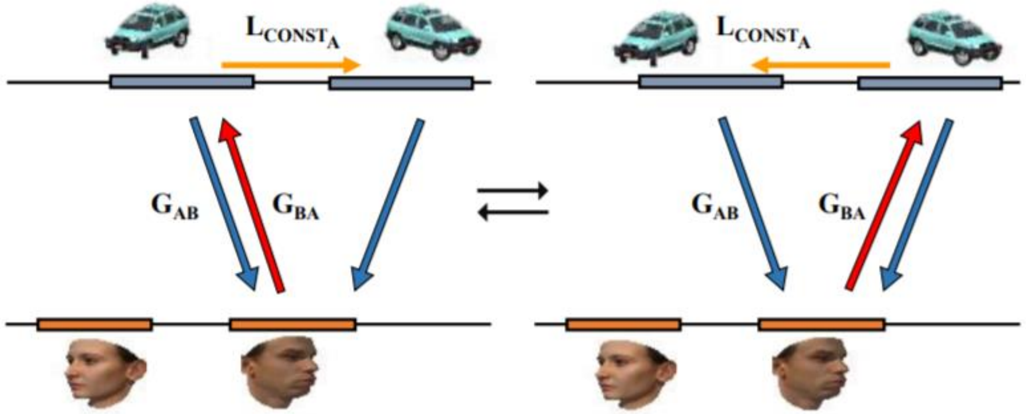
Experiment	α_{1A}	α_{1B}	α_{2A}	α_{2B}	α_{3A}	α_{3B}	α_{4A}	α_{4B}
DiscoGAN	0.5	0.5	0.5	0.5	0	0	0	0
Distance $\rightarrow$	0.5	0	0	0	0.5	0	0	0
Distance $\leftarrow$	0	0.5	0	0	0	0.5	0	0
Dist+Cycle	0.5	0.5	0.5	0.5	0.5	0.5	0	0
Self Dist $\rightarrow$	0.5	0	0	0	0	0	0.5	0
Self Dist $\leftarrow$	0	0.5	0	0	0	0	0	0.5

Solves asymmetry problem: Mode Collapse

- GAN:

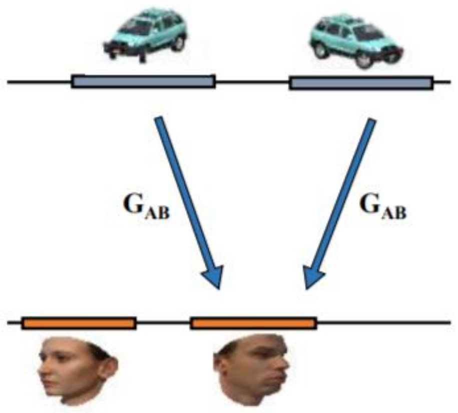


Cycle:

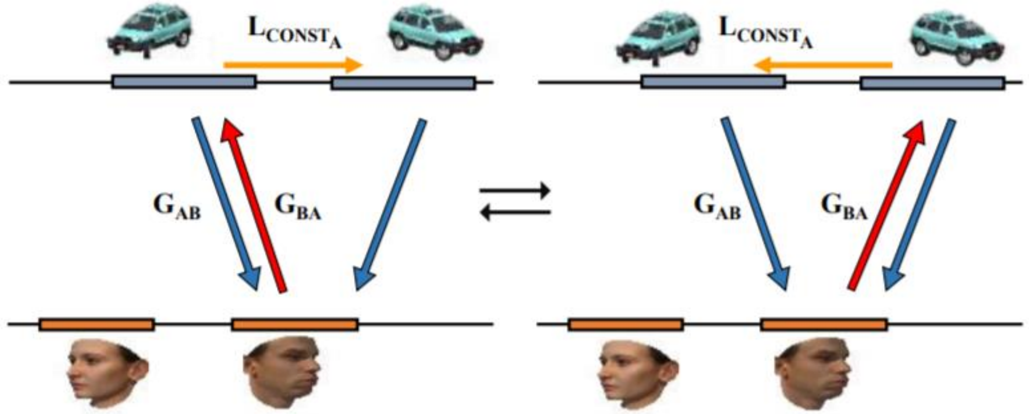


Solves asymmetry problem: Mode Collapse

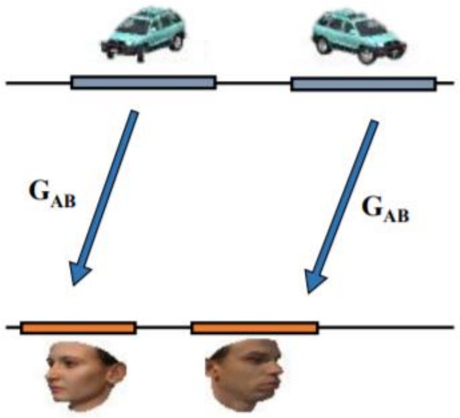
- GAN:



Cycle:



- Distance:



GAN Architecture

- DiscoGAN based (64 bits):
  - Generator: Encoder-Decoder, Based on DCGAN
  - Discriminator: Simple Decoder

- CycleGAN based(128-256 bits):
  - Based on “Perceptual losses for real-time style transfer and super-resolution” Johnson et al.
  - Generator: Use of additional Residual blocks
  - Discriminator: Use of 70*70 Patch-GAN

Experiments

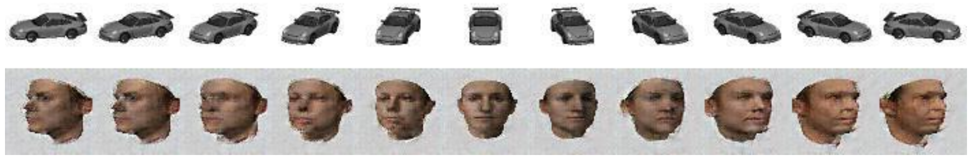


Table 2: Normalized RMSE between the angles of source and translated images.

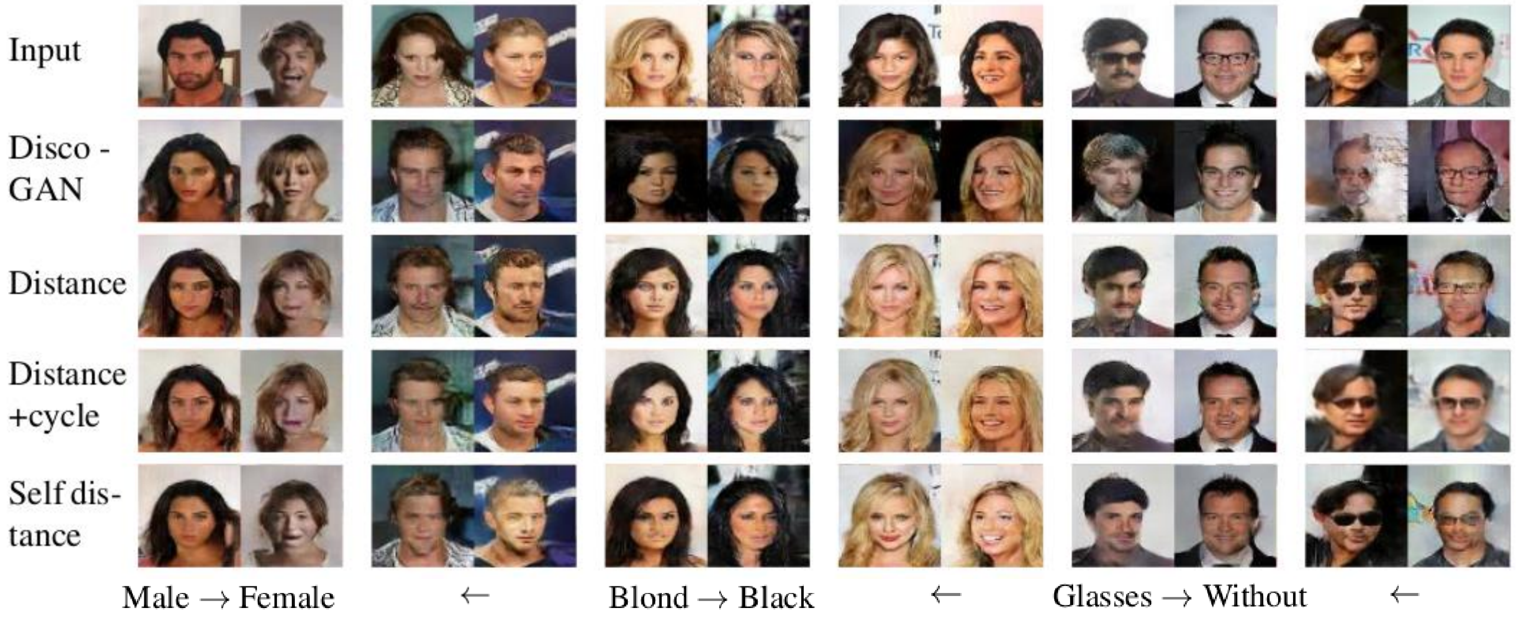
Method	car2car	car2head
DiscoGAN	0.306	0.137
Distance	0.135	**0.097**
Dist.+Cycle	**0.098**	0.273
Self Dist.	0.117	0.197

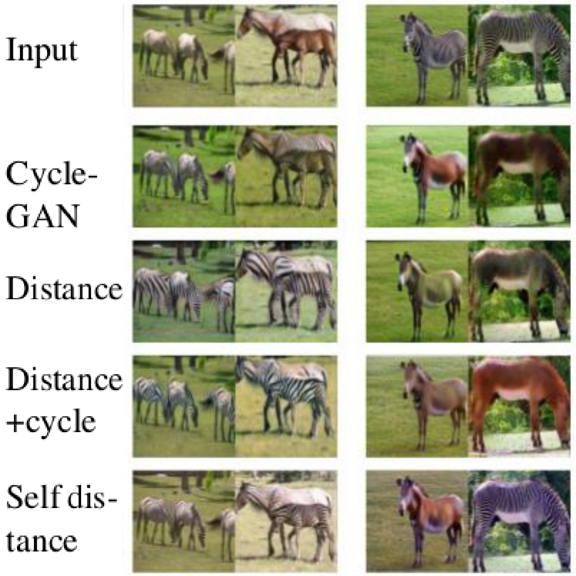
Table 3: MNIST classification on mapped SHVN images.

Method	Accuracy
CycleGAN	26.1%
Distance	**26.8%**
Dist.+Cycle	18.0%
Self Dist.	25.2%

Table 4: CelebA mapping results using the VGG face descriptor.

Method	Male → Female		Blond → Black		Glasses → Without	
	Cosine Similarity	Separation Accuracy	Cosine Similarity	Separation Accuracy	Cosine Similarity	Separation Accuracy
DiscoGAN	0.23	0.87	0.15	0.89	0.13	**0.84**
Distance	0.32	**0.88**	**0.24**	**0.92**	**0.42**	0.79
Distance+Cycle	**0.35**	0.87	**0.24**	0.91	0.41	0.82
Self Distance	0.24	0.86	**0.24**	0.91	0.34	0.80
Other direction						
DiscoGAN	0.22	0.86	0.14	0.91	0.10	**0.90**
Distance	0.26	0.87	**0.22**	**0.96**	**0.30**	0.89
Distance+Cycle	**0.31**	0.89	**0.22**	0.95	**0.30**	0.85
Self Distance	0.24	**0.91**	0.19	0.94	**0.30**	0.81





Comparison of Losses

Table 5: Losses measured for each method on the CelebA dataset.

Method	$\mathcal{L}_{\text{GAN}}(A)$	$\mathcal{L}_{\text{GAN}}(B)$	$\mathcal{L}_{\text{cycle}}(B)$	$\mathcal{L}_{\text{cycle}}(B)$	$\mathcal{L}_{\text{dist}}(A)$	$\mathcal{L}_{\text{dist}}(B)$	$\mathcal{L}_{\text{selfd}}(A)$	$\mathcal{L}_{\text{selfd}}(B)$
(A) Male to (B) Female:								
DiscoGAN	4.300	2.996	0.036	0.024	0.466	0.457	0.441	0.422
Distance	3.702	2.132	0.026	0.026	0.047	0.047	0.038	0.044
Distance+Cycle	4.280	1.651	0.017	0.016	0.046	0.043	0.042	0.040
Self Distance	3.322	3.131	0.092	0.091	0.048	0.050	0.045	0.044
(A) Blond to (B) Black hair:								
DiscoGAN	2.511	3.297	0.019	0.018	0.396	0.399	0.396	0.399
Distance	0.932	2.243	0.021	0.017	0.046	0.042	0.046	0.042
Distance+Cycle	1.045	2.484	0.013	0.012	0.043	0.043	0.043	0.042
Self Distance	0.965	2.867	0.022	0.018	0.049	0.048	0.049	0.048
(A) With or (B) Without eyeglasses:								
DiscoGAN	5.734	3.621	0.110	0.040	0.535	0.337	0.535	0.074
Distance	7.697	0.804	0.046	0.036	0.023	0.065	0.023	0.065
Distance+Cycle	5.730	0.924	0.024	0.017	0.027	0.048	0.028	0.048
Self Distance	8.242	0.795	0.040	0.018	0.029	0.051	0.029	0.050

Maps and Cityscapes

- FCN Score: Better per-class accuracy (Significantly), per-pixel accuracy, Class IOU.
- User Study:
  - Cityscapes Labels to Photos realness (71% of cases better than CycleGAN)
  - Similarity to Ground Truth (68% of cases better than CycleGAN)
  - Similar experiments in DiscoGAN’s Male to Female and Handbags to Shoes.

Extensions and Notes

- Cycle loss: Only approximate. A zebra translated to a horse must loose its stripes.
- Minimal information is required – potentially infinitely many mappings.
- Other domains? Text translation from one embedding to another.

Thank You! Questions?