# One-Sided Unsupervised Domain Mapping

Sagie Benaim and Lior Wolf

NIPS 2017

# Latest Trends

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

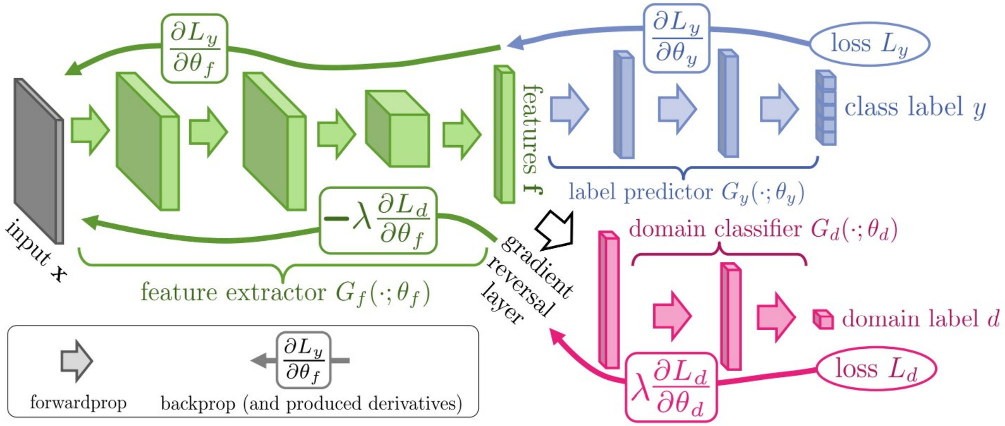
**Not semantic**

# Latest Trends

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

**Not semantic**

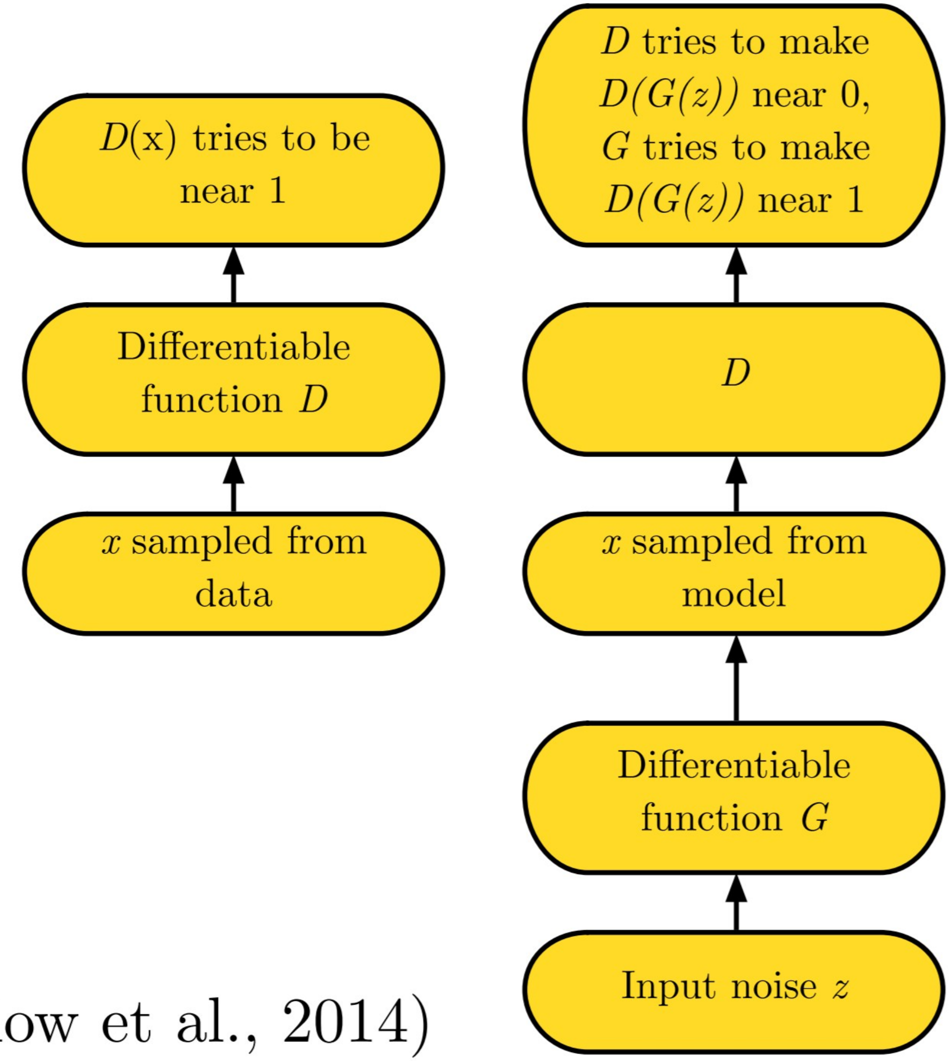
2. Domain adaptation
    - “Domain-adversarial training of neural networks” Ganin et al.




**Supervised and not generative**

# Image to Image Translation

# Adversarial Nets Framework

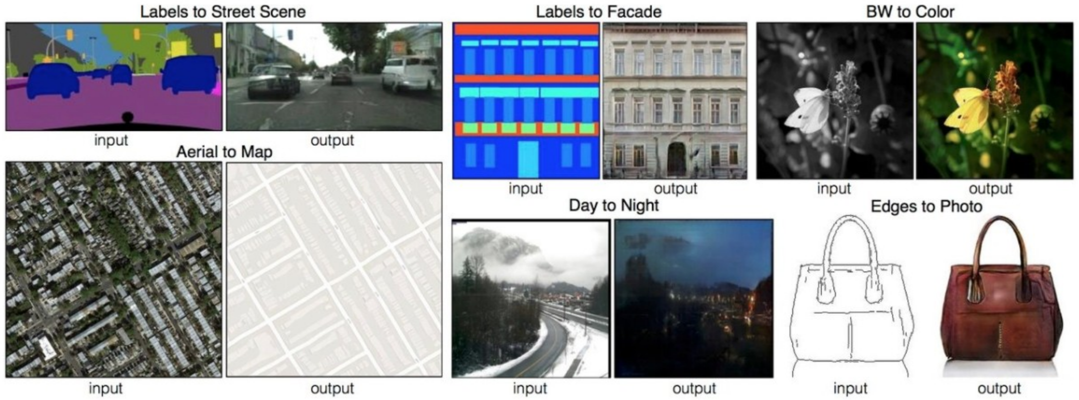



(Goodfellow et al., 2014)

(Karras et. al, 2017)

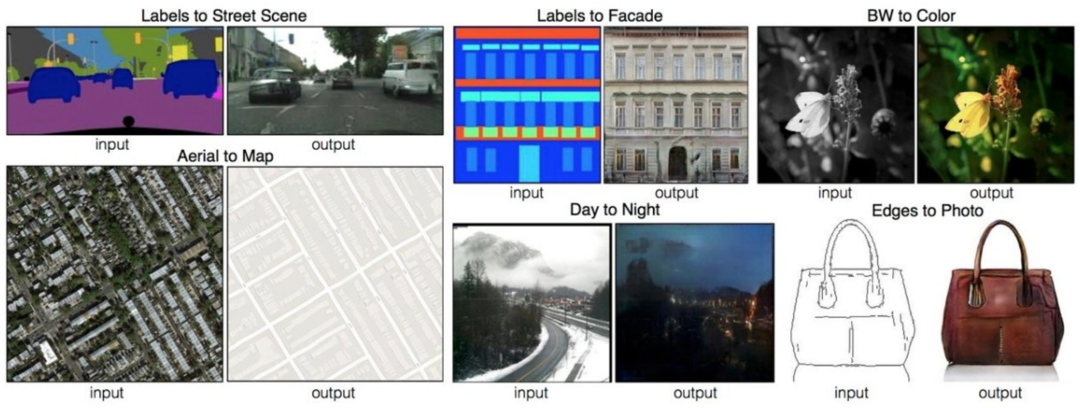
# Fully Supervised Alignment

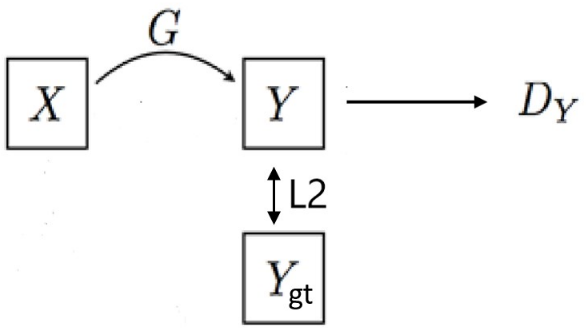
- "Image-to-image translation with conditional adversarial nets" Isola et al (pix2pix)

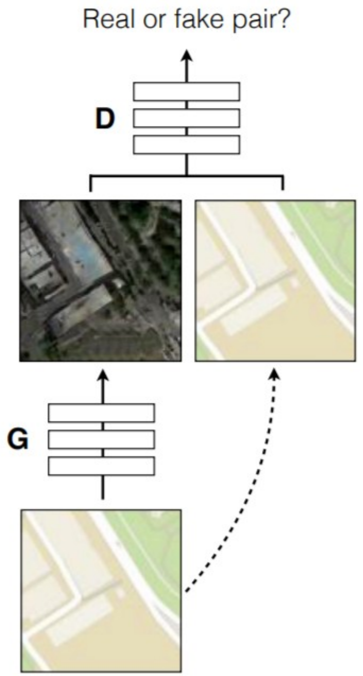


# Fully Supervised Alignment

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

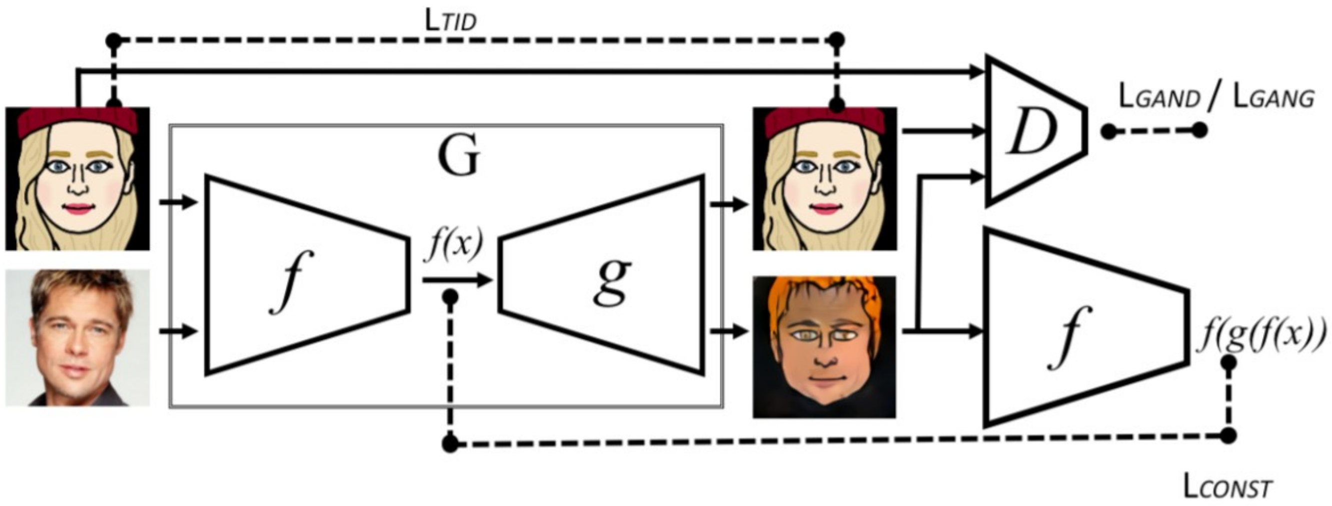


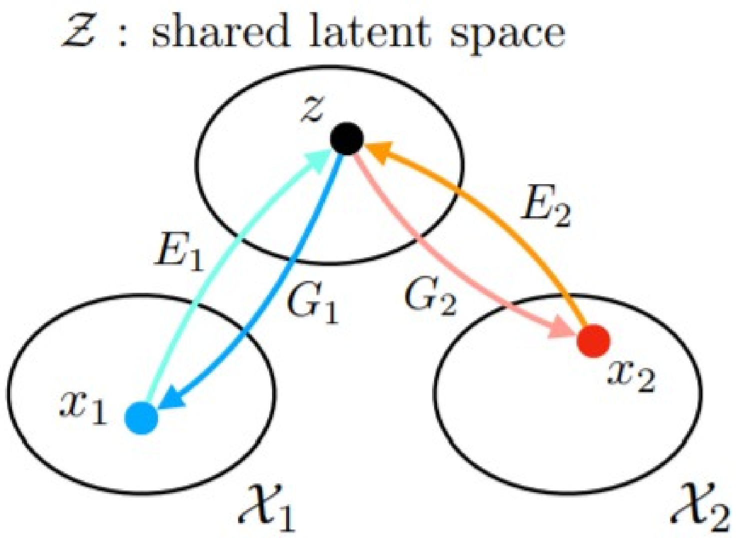
# Partially Supervised Alignment

- “Unsupervised Cross-Domain Image Generation” Taigman et al.



# Unsupervised Alignment

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



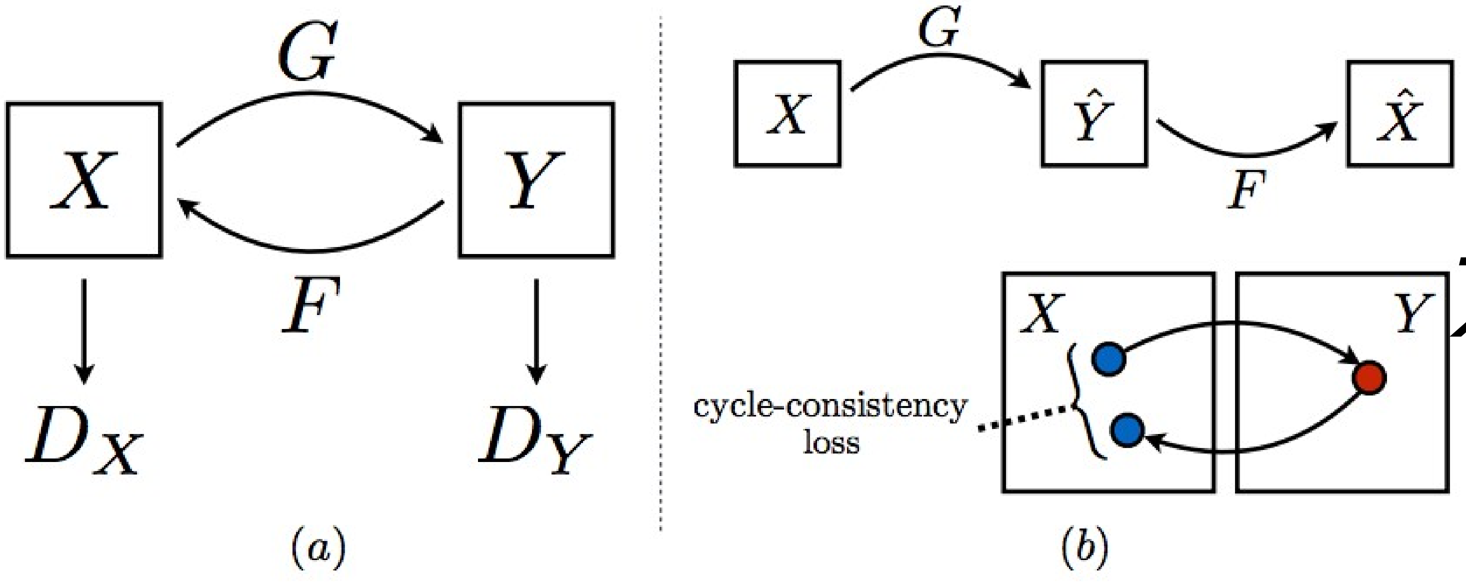
# Circular GANs

**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.

# Circular GANs




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

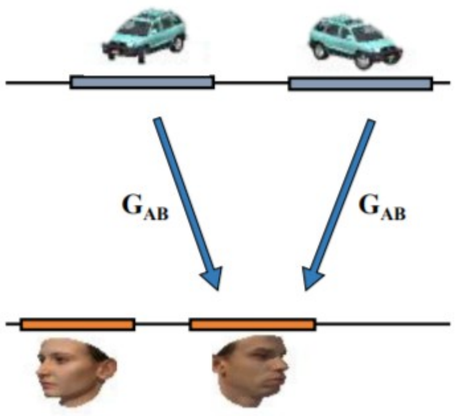
$y \sim G(F(y))$

Circular GANs (DiscoGAN, CycleGAN, DualGAN)

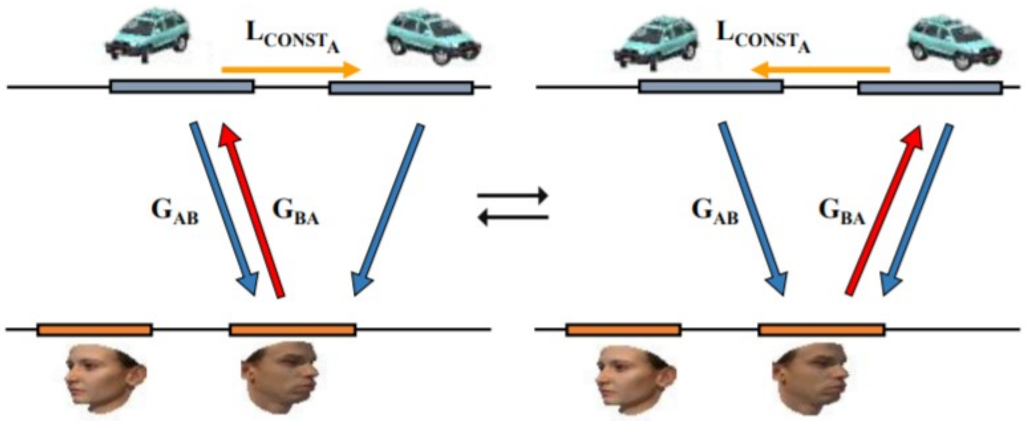
# Approximate Mapping


# Mode Collapse

- GAN:




Cycle:

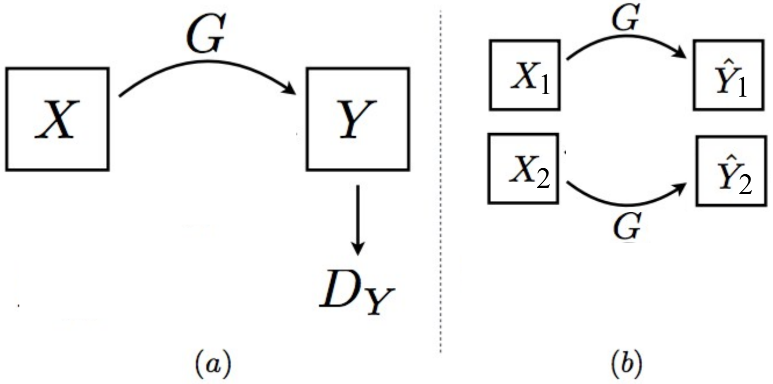


# Introducing DistanceGAN

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

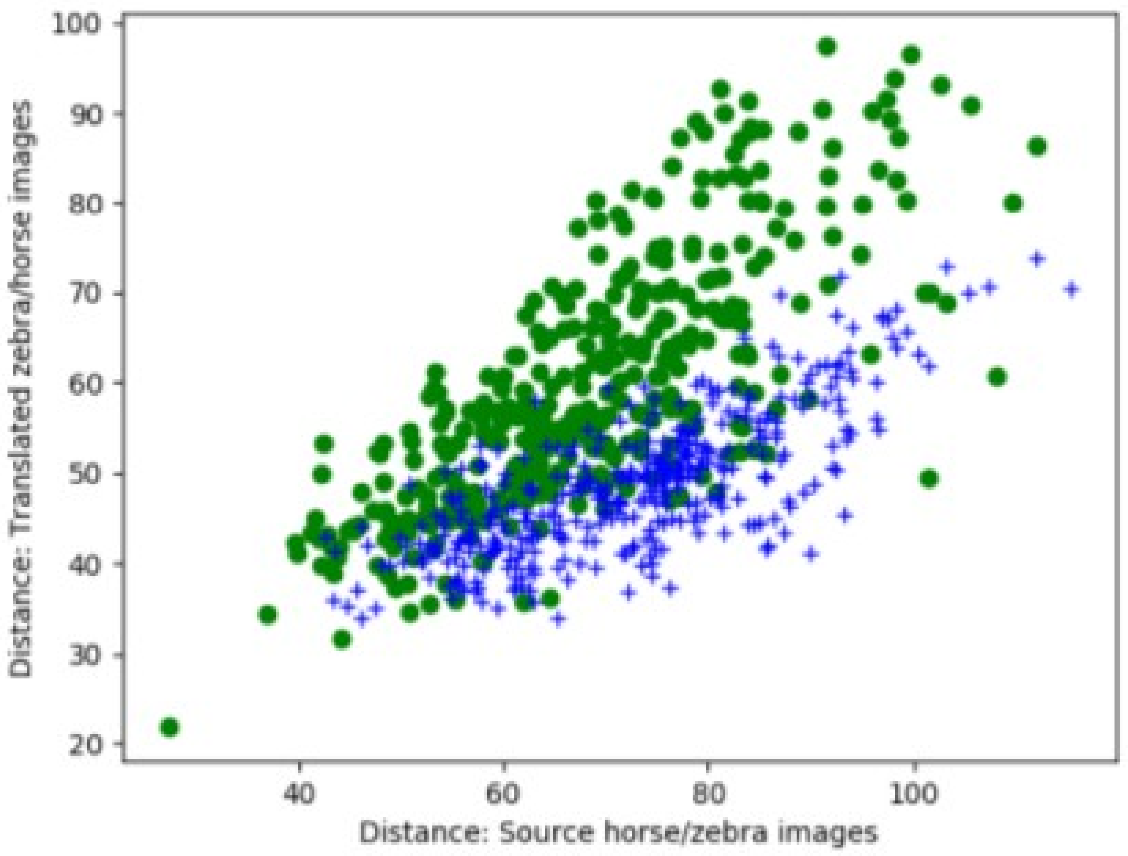
# Introducing DistanceGAN

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




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

# Motivating distance correlations I



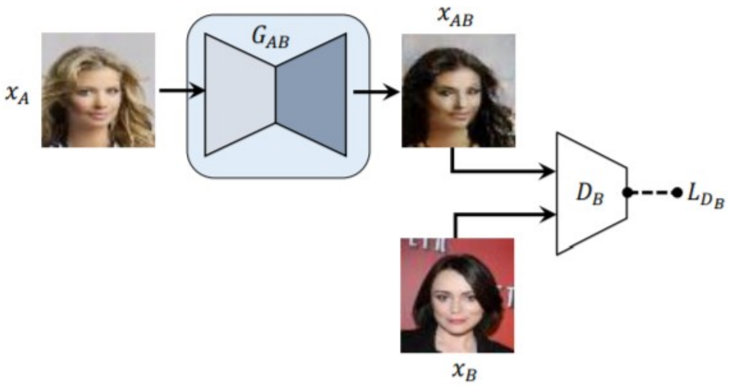

Analysis of CycleGAN's horse to zebra results

# Motivating distance correlations II

$\boldsymbol{x}$ $\quad G(\boldsymbol{x})$ $\quad \Pi(\boldsymbol{x})$

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

# Building Block: Conditional GAN




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

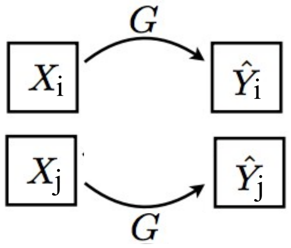
- Other GAN variants can be used: w-gan, improved w-gan, BEGAN, etc.

# The loss used

- A distance correlation loss:

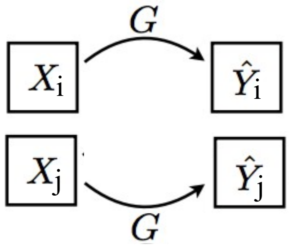
- A distance correlation loss:

  - $\sum_{x_i,x_j} |d_1 - d_2|$
  - $d_1 = \frac{1}{\sigma_A}(\left|x_i - x_j\right|_1 - \mu_A)$
  - $d_2 = \frac{1}{\sigma_B}(\left|G(x_i) - G(x_j)\right|_1 - \mu_B)$

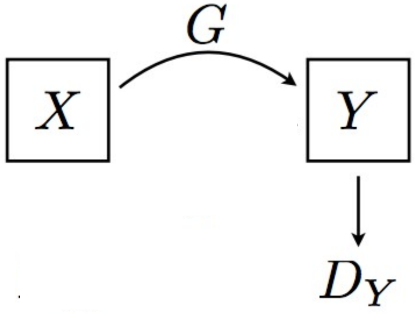


# The loss used

- A distance correlation loss:
- A distance correlation loss:
    - $\sum_{x_i,x_j}|d_1 - d_2|$
    - $d_1 = \frac{1}{\sigma_A}(|x_i - x_j|_1 - \mu_A)$
    - $d_2 = \frac{1}{\sigma_B}(|G(x_i) - G(x_j)|_1 - \mu_B)$




- A GAN loss on $Y$
- A GAN loss on



# Additive is more stable than multiplicative

- Additive Loss: $\sum_{x_i,x_j} |d_1 - d_2|$
- Multiplicative Loss: $-\sum_{x_i,x_j} d_1 d_2$
- Highly correlated. Additive Loss doesn’t dominate optimization.

- Highly correlated. Additive Loss doesn’t dominate optimization.

# GAN Architecture

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

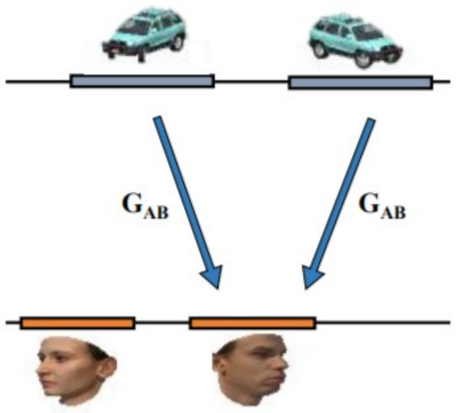
- CycleGAN based (128-256 pixels):
  - Based on “Perceptual losses for real-time style transfer and super-resolution” Johnson et al.
  - Generator: Use of additional Residual blocks
  - Discriminator: Use of 70*70 Patch-GAN

# Variants

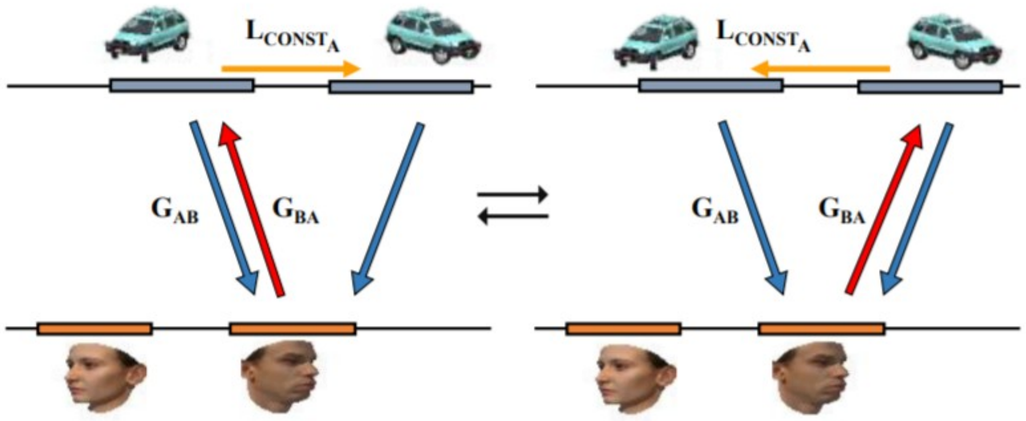
- Distance Loss Only (Either on DiscoGAN arch of CycleGAN arch)
- Distance + Cycle Loss
- Self Distance

# Solves asymmetry problem: Mode Collapse

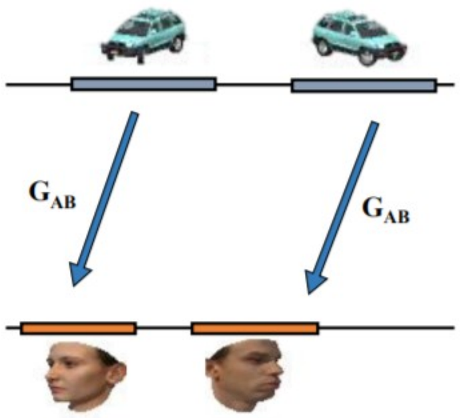
- GAN:




Cycle:




- Distance:



# Experiments

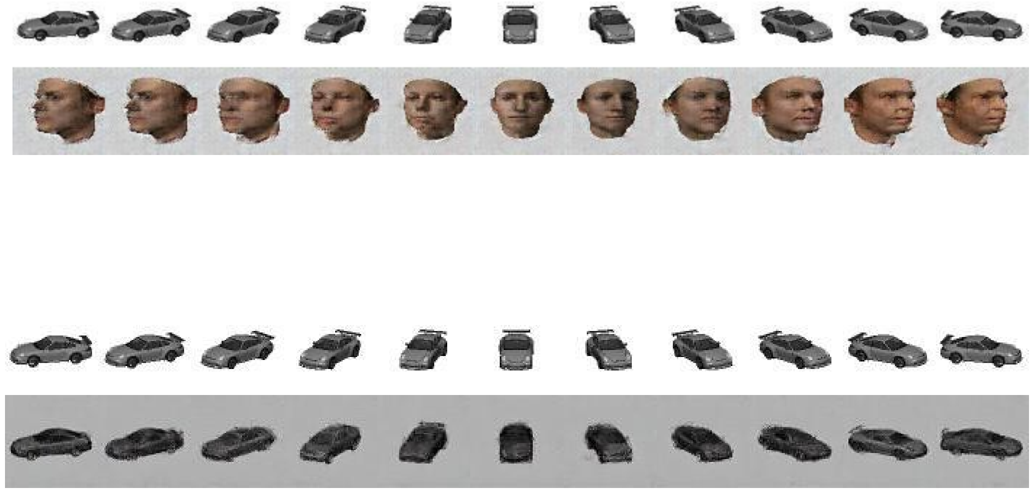


Table 2: Normalized RMSE between the angles of source and translated images.

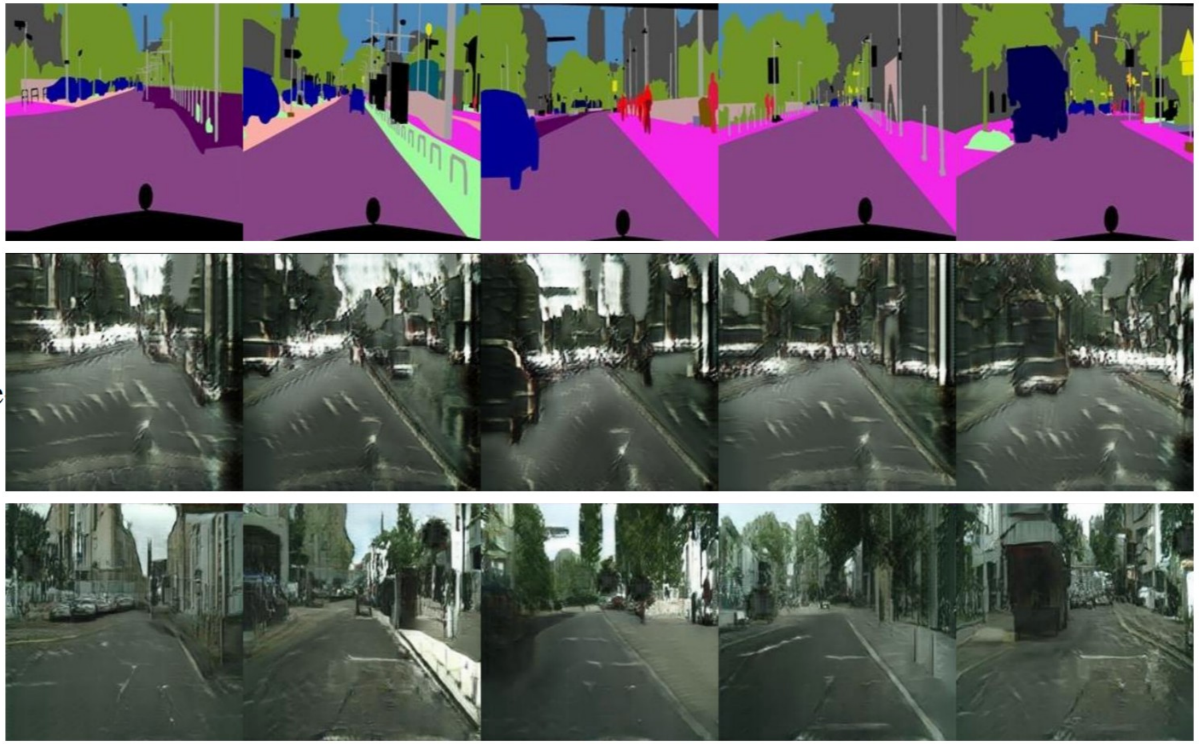
| Method | car2car | car2head |
|---|---|---|
| DiscoGAN | 0.306 | 0.137 |
| Distance | 0.135 | **0.097** |
| Dist.+Cycle | **0.098** | 0.273 |
| Self Dist. | 0.117 | 0.197 |


(Handbags to shoes) (Shoes to handbags)


(a)
Input
(c)Cycle
GAN
(d) Dis-
tance
(e) Cy-
cle+dist
(f) Self-
distance

# Cityscapes

- FCN Score: Better per-class accuracy (Significantly), per-pixel accuracy, Class IOU



(a) Input

(c)Cycle GAN

(d) Dis- tance


Input
Disco -
GAN
Distance
Distance
+cycle
Self dis-
tance

# CelebA mapping results using the VGG face descriptor

| | Male → Female | |
|---|---|---|
| Method | Cosine Similarity | Separation Accuracy |
| DiscoGAN | 0.23 | 0.87 |
| Distance | 0.32 | **0.88** |
| Distance+Cycle | **0.35** | 0.87 |
| Self Distance | 0.24 | 0.86 |

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 |

Table 3: MNIST classification on mapped SHVN images.

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

# 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

- Minimal information is required – potentially infinitely many mappings.
- Better understanding of semantics: SVHN to MNIST
- Other domains? Text translation from one embedding to another.
- Other Trends: Many to Many Translations (e.g StarGAN)

Thank You! Questions?

# Minimality

- Potentially Infinitely many solutions preserving distance correlations

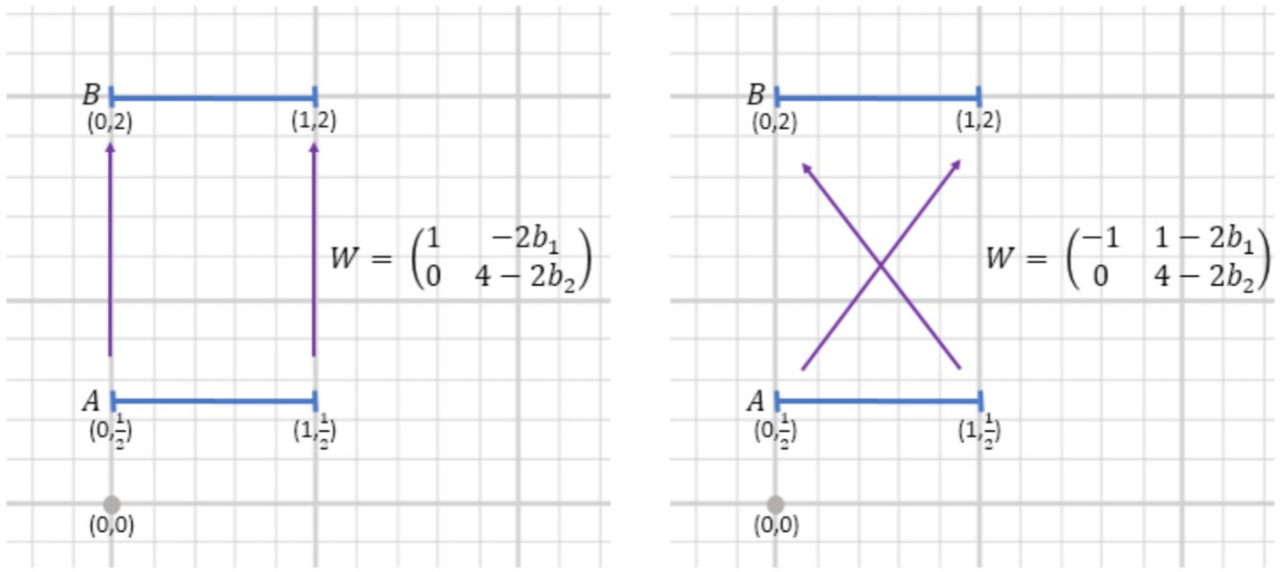



Figure 1: An illustrative example where the two domains are line segments in $\mathbb{R}^2$. There are infinitely many mappings that preserve the uniform distribution on the two segments. However, only two stand out as "semantic". These are exactly the two mappings that can be captured by a neural network with only two hidden neurons and Leaky ReLU activations, i.e., by a function $h(x) = \sigma_a(Wx + b)$, for a weight matrix $W$ and the bias vector $b$.