AI for image Reading Group

Image-to-Image translation with GANs

Beuve Nicolas
Nicolas.Beuve@insa-rennes.com

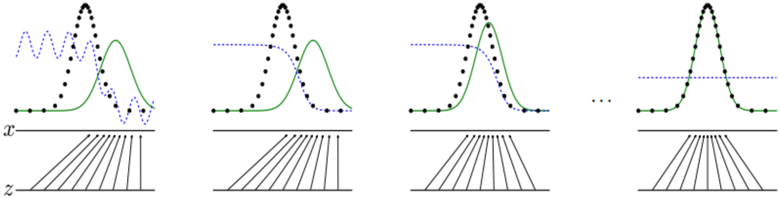
VAADER IETR research team
IETR
INSA

BEUVE Nicolas

- PhD student since October 2020 (Nice timing!)
- Subject: "Automatic detection of deepfake videos"
- Advisors : Wassim Hamidouche and Olivier Deforges
- Funding granted by DGA

1. Generative Adversarial Nets
2. Image-to-Image translation
3. CGAN
4. CycleGAN

Motivation

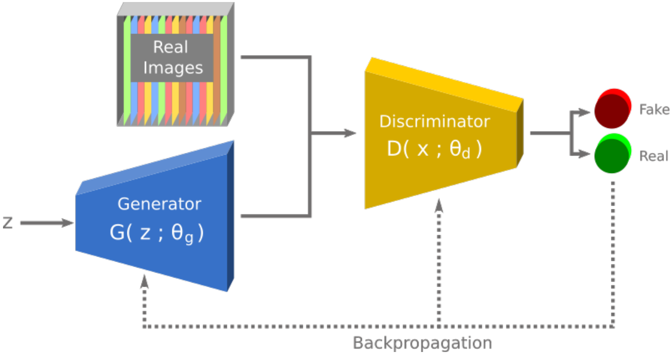


$p_{data}(x)$ Real distribution

$p_g(x)$ Generated distribution

$p_z(z)$ Latent distribution

Generative Adversarial Nets[1]



$$\min_G \max_D \mathcal{L}(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[log\ D(x)] + \mathbb{E}_{z\sim p_z(z)}[log(1 - D(G(z))]$$

[1] Goodfellow et al. 2014

Optimal discriminator

For G fixed, the best discriminator is:

$$D^* = \frac{p_{data}(x)}{p_{data}(x) + p_g(x)} = \frac{1}{2}$$

Optimal generator

For D^*, the best generator G^* is reached when:

$$p_g(x) = p_{data}(x)$$

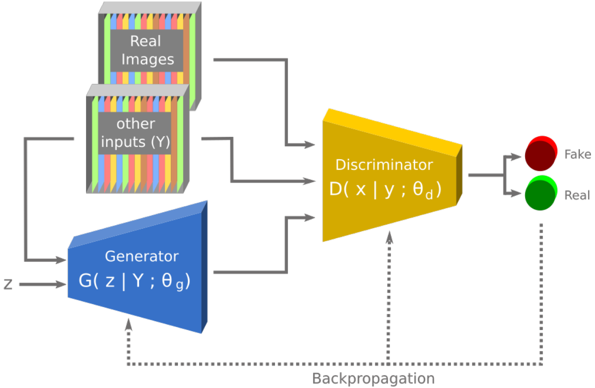
The state (D^*, G^*) is a Nash equilibrium. Meaning that each player has no interest in changing their strategy if the others don't.

Learning the mapping between two image representation.

Source images

Target images

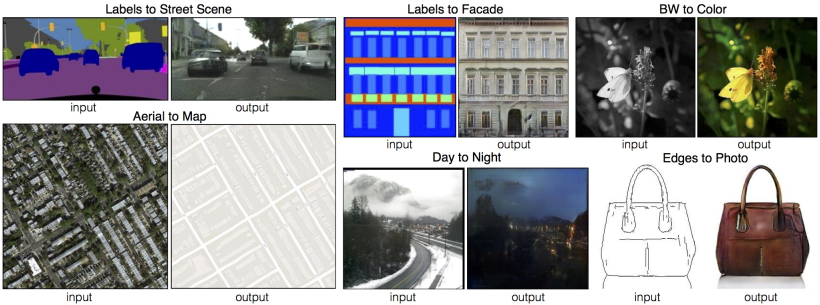
Conditional Generative Adversarial Nets[1]



$$\min_G \max_D \mathcal{L}(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[log D(x|y)] + \mathbb{E}_{z\sim p_z(z)}[log(1 - D(G(z|y)|y))]$$

[1] Mirza and Osindero 2014

Image-to-Image Translation with Conditional Adversarial Nets[1]

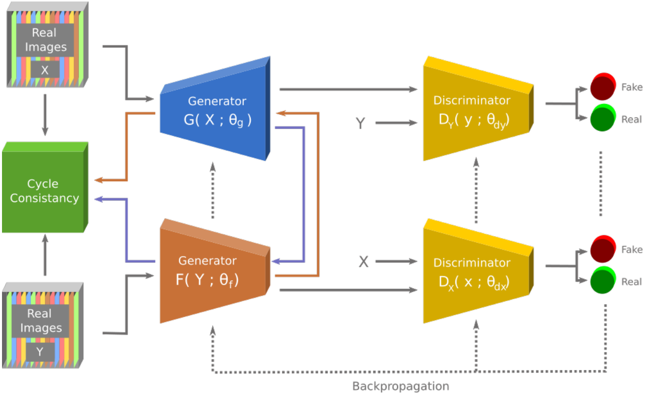


[1] Isola et al. 2016

Unpaired datasets

Source images

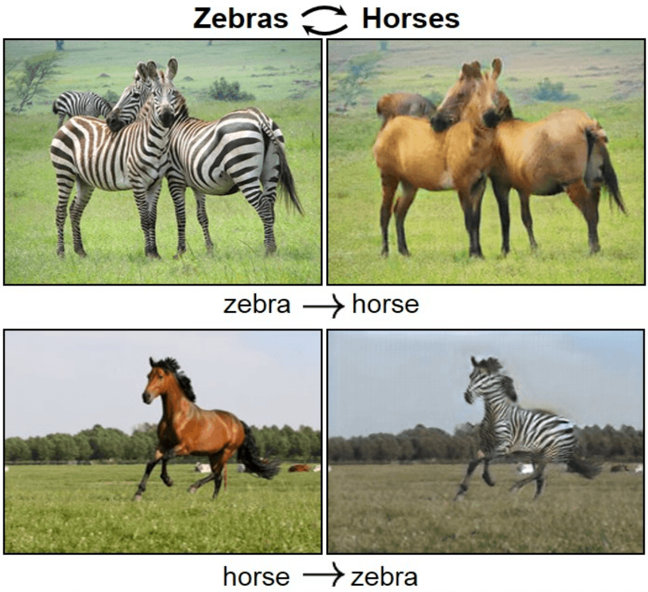
Target images



$$\mathcal{L}_{cyc}(F,G) = \mathbb{E}_{y \sim p_{data}(y)}\left[||G(F(y)) - y||_1\right] + \mathbb{E}_{x \sim p_{data}(x)}\left[||F(G(x)) - x||_1\right]$$

$$\min_{G,F} \max_{D_X,D_Y} V(G,F,D_X,D_Y) = \mathcal{L}_{GAN}(G,D_Y) + \mathcal{L}_{GAN}(F,D_X) + \mathcal{L}_{cyc}(F,G)$$

[1] Zhu et al. 2017


Zebras
Horses
zebra → horse
horse → zebra

Advantages

- Strong mathematical theory
- Very flexible architecture
  - Can use state of the art models as generator or discriminator
  - D2GAN, MGAN, CycleGAN

Drawbacks

- Two models trained at once
- Very sensible training
  - Gradiant vanishing
  - Mode collapsing

Go further

- Improve training with Wasserstein GAN [1]
- Generate HD content with PG-GAN [2]

[1] Arjovsky, Chintala, and Bottou 2017

[2] Karras et al. 2018

Arjovsky, Martin, Soumith Chintala, and Léon Bottou (2017). *Wasserstein GAN*. arXiv: 1701.07875 [stat.ML].

Goodfellow, Ian et al. (2014). "Generative Adversarial Nets". In: *Advances in Neural Information Processing Systems 27*. Ed. by Z. Ghahramani et al. Curran Associates, Inc., pp. 2672–2680. URL: http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf.

Isola, Phillip et al. (2016). "Image-to-Image Translation with Conditional Adversarial Networks". In: *CoRR* abs/1611.07004. arXiv: 1611.07004. URL: http://arxiv.org/abs/1611.07004.

Karras, Tero et al. (2018). *Progressive Growing of GANs for Improved Quality, Stability, and Variation*. arXiv: 1710.10196 [cs.NE].

Mirza, Mehdi and Simon Osindero (2014). "Conditional Generative Adversarial Nets". In: *CoRR* abs/1411.1784. arXiv: 1411.1784. URL: http://arxiv.org/abs/1411.1784.

Zhu, Jun-Yan et al. (2017). "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks". In: *CoRR* abs/1703.10593. URL: