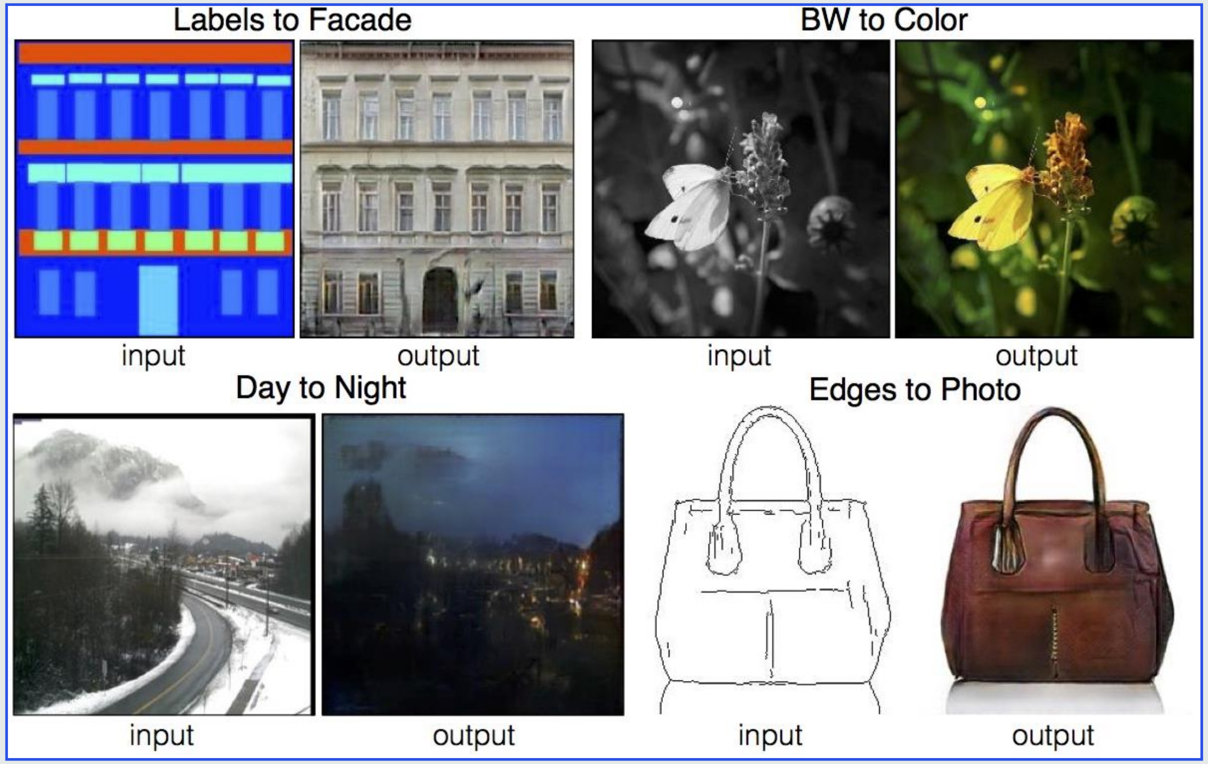
Cycle GAN for Image-to-Image Translation

NICOLA GUGOLE
ID: 619625
INTELLIGENT SYSTEMS FOR PATTERN RECOGNITION
UNIVERSITA' DI PISA

Image to Image translation? PIX2PIX!

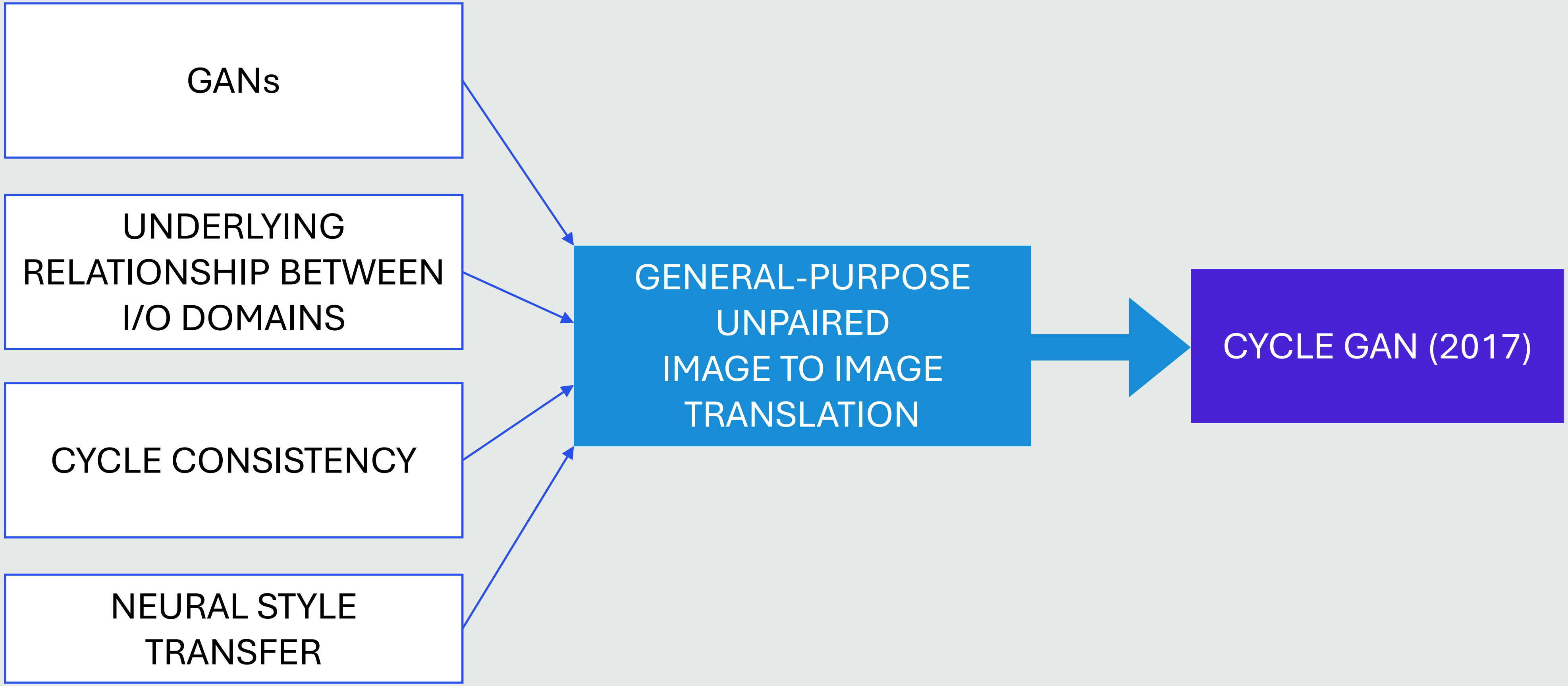
- Task where input image is mapped to a differently styled output image
- PIX2PIX as Conditional GAN
- Needs paired data, which we always have, right? RIGHT??



PIX2PIX (2016): https://arxiv.org/abs/1611.07004

The problem to solve

(& ideas to solve it!)



CycleGAN: double generator/discriminator

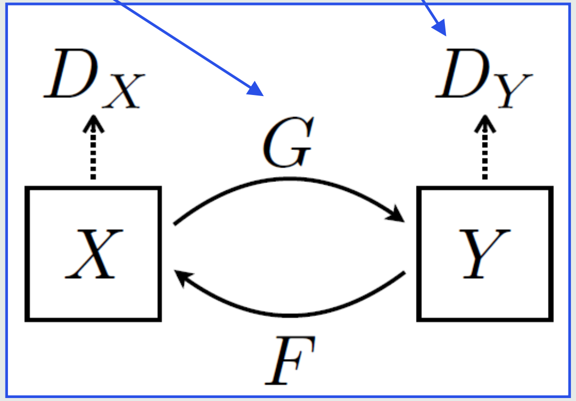
(or how to constrain an under-constrained model optimization)

To avoid *mode collapse* and have more relevant results

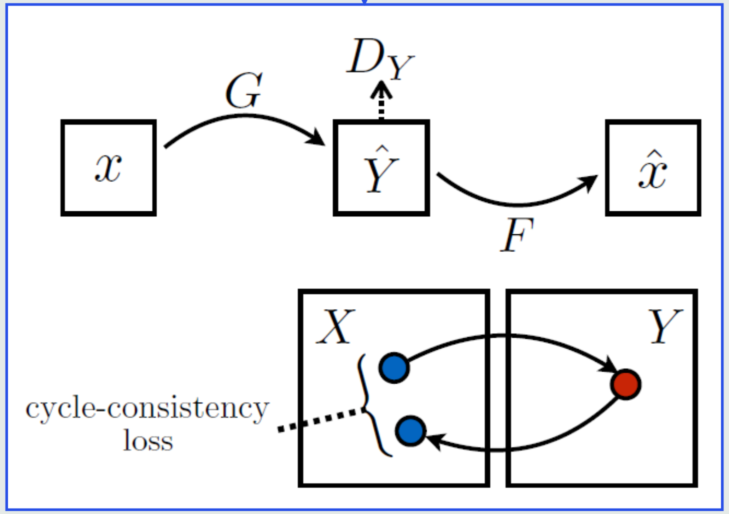
Dy distinguishes G(X) from Y

'Special' kind of autoencoders

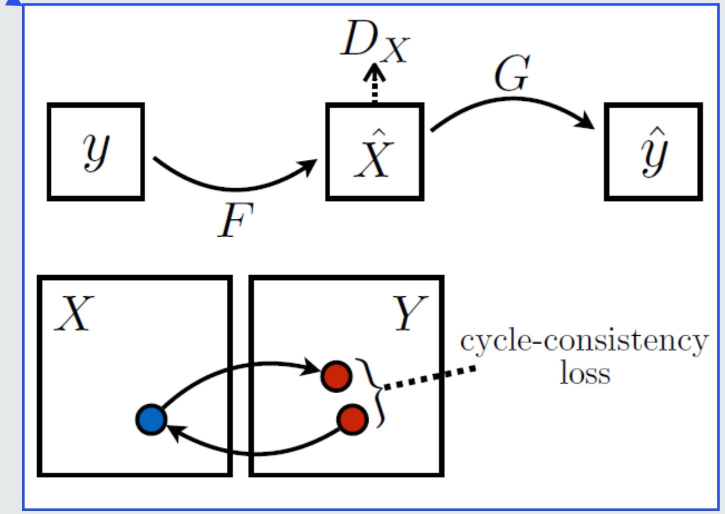
G : X ⟶ Y
F : Y ⟶ X



Adversarial losses



Forward cycle consistency loss

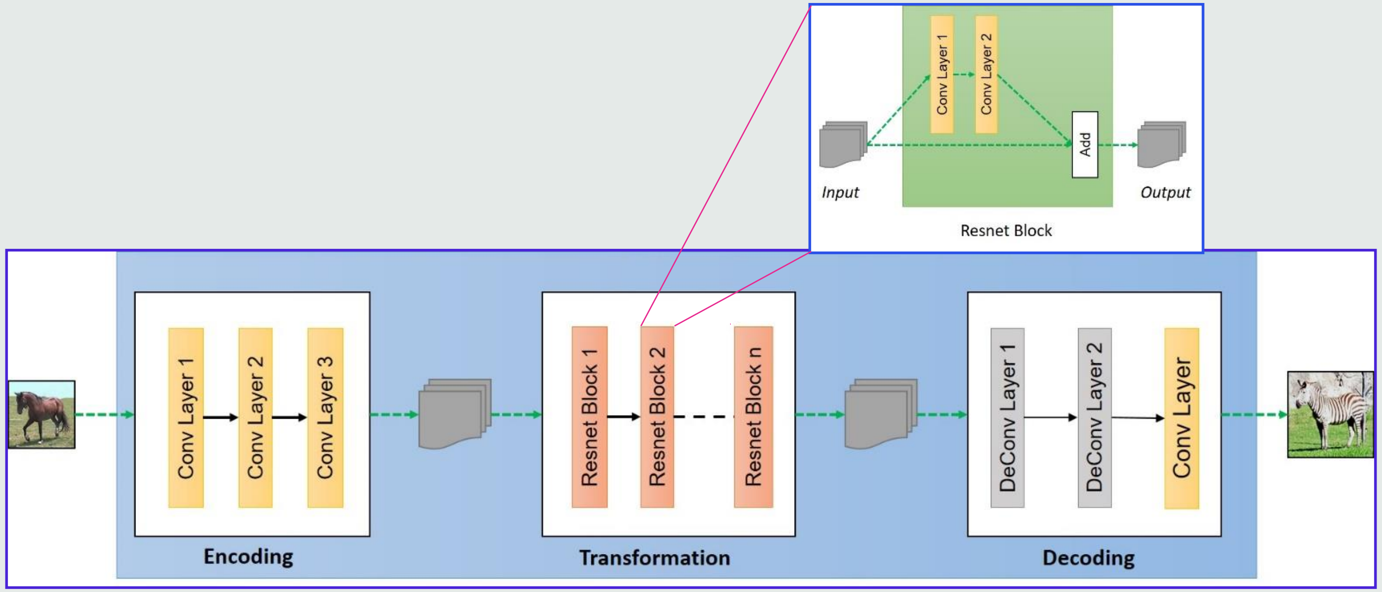


Backward cycle consistency loss

Model architecture

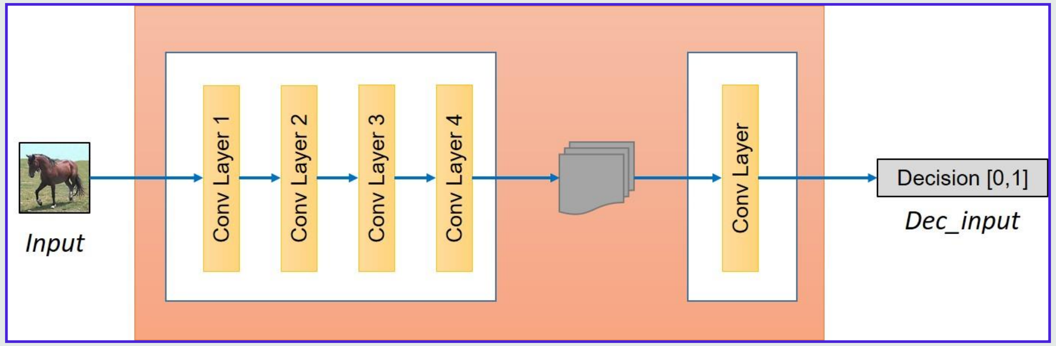
Generator has an encoder-decoder structure inspired by previous works on NST

Residual blocks for a better gradient flow, different number of blocks depending on image size

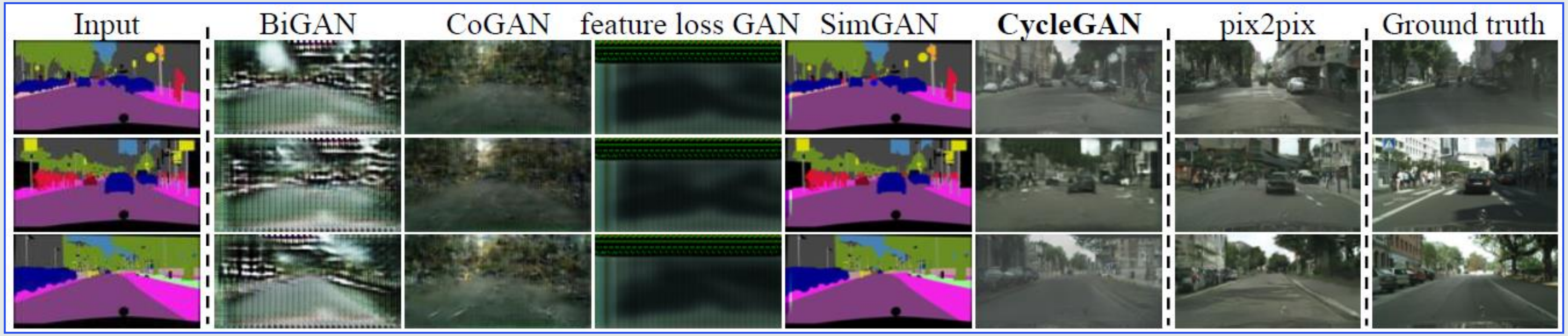


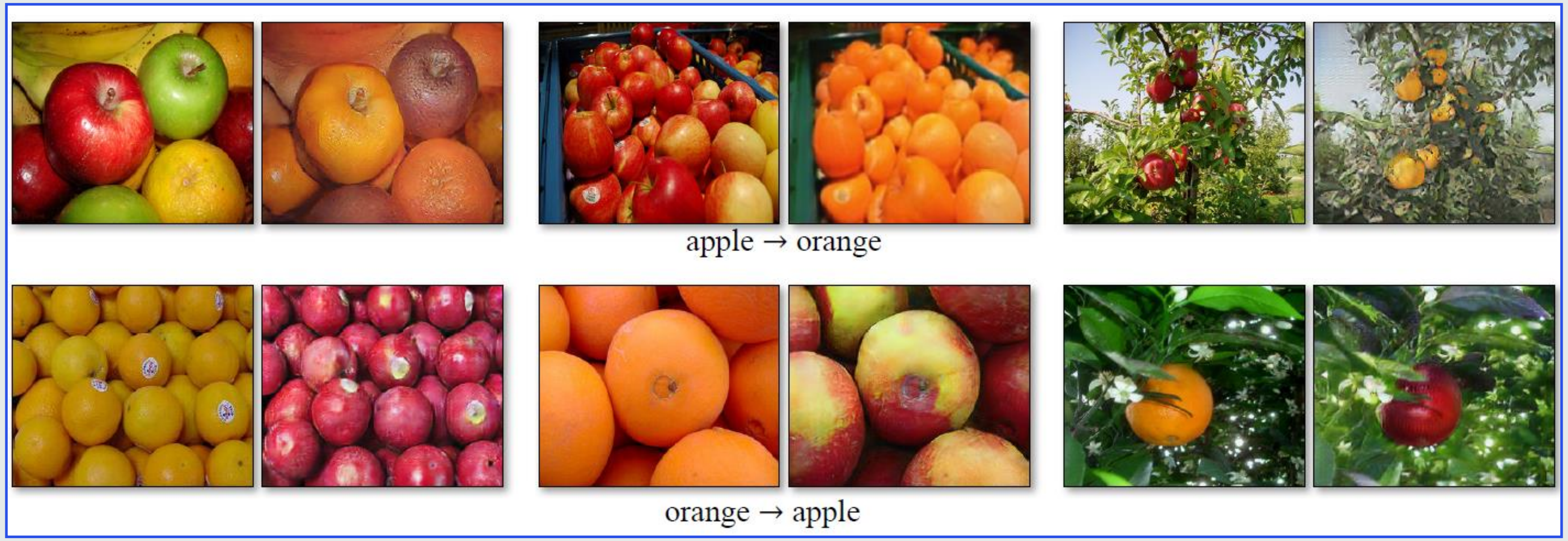
Discriminator is implemented as a 70x70 PatchGAN, classifying as real/fake not the whole image, but overlapping patches

A winning approach, fewer parameters, no dependency on image size!



Qualitative results





Various applications:

- STYLE TRANSFER
- PHOTO ENHANCEMENT
- SEASON TRANSFER
- GAME STYLE TRANSFER
- FACE TO RAMEN (yes, it really exists, and I cannot sleep anymore)

Quantitative results

Perceptual test using Amazon Mechanical Turk, showing supremacy over other concurrents with less general-purpose implementations

Both directions

Loss	Map → Photo % Turkers labeled *real*	Photo → Map % Turkers labeled *real*
CoGAN [32]	0.6% ± 0.5%	0.9% ± 0.5%
BiGAN/ALI [9, 7]	2.1% ± 1.0%	1.9% ± 0.9%
SimGAN [46]	0.7% ± 0.5%	2.6% ± 1.1%
Feature loss + GAN	1.2% ± 0.6%	0.3% ± 0.2%
CycleGAN (ours)	**26.8% ± 2.8%**	**23.2% ± 3.4%**

Table 1: AMT "real vs fake" test on maps↔aerial photos at 256 × 256 resolution.

FCN score evaluating how interpretable results are according to a semantic segmentation algorithm. Comparisons with ablations to show how important it is to have such a structured loss

Loss	Per-pixel acc.	Per-class acc.	Class IOU
Cycle alone	0.22	0.07	0.02
GAN alone	0.51	0.11	0.08
GAN + forward cycle	**0.55**	**0.18**	**0.12**
GAN + backward cycle	0.39	0.14	0.06
CycleGAN (ours)	0.52	0.17	0.11

Table 4: Ablation study: FCN-scores for different variants of our method, evaluated on Cityscapes labels→photo.

Loss	Per-pixel acc.	Per-class acc.	Class IOU
Cycle alone	0.10	0.05	0.02
GAN alone	0.53	0.11	0.07
GAN + forward cycle	0.49	0.11	0.07
GAN + backward cycle	0.01	0.06	0.01
CycleGAN (ours)	**0.58**	**0.22**	**0.16**

Table 5: Ablation study: classification performance of photo→labels for different losses, evaluated on Cityscapes.

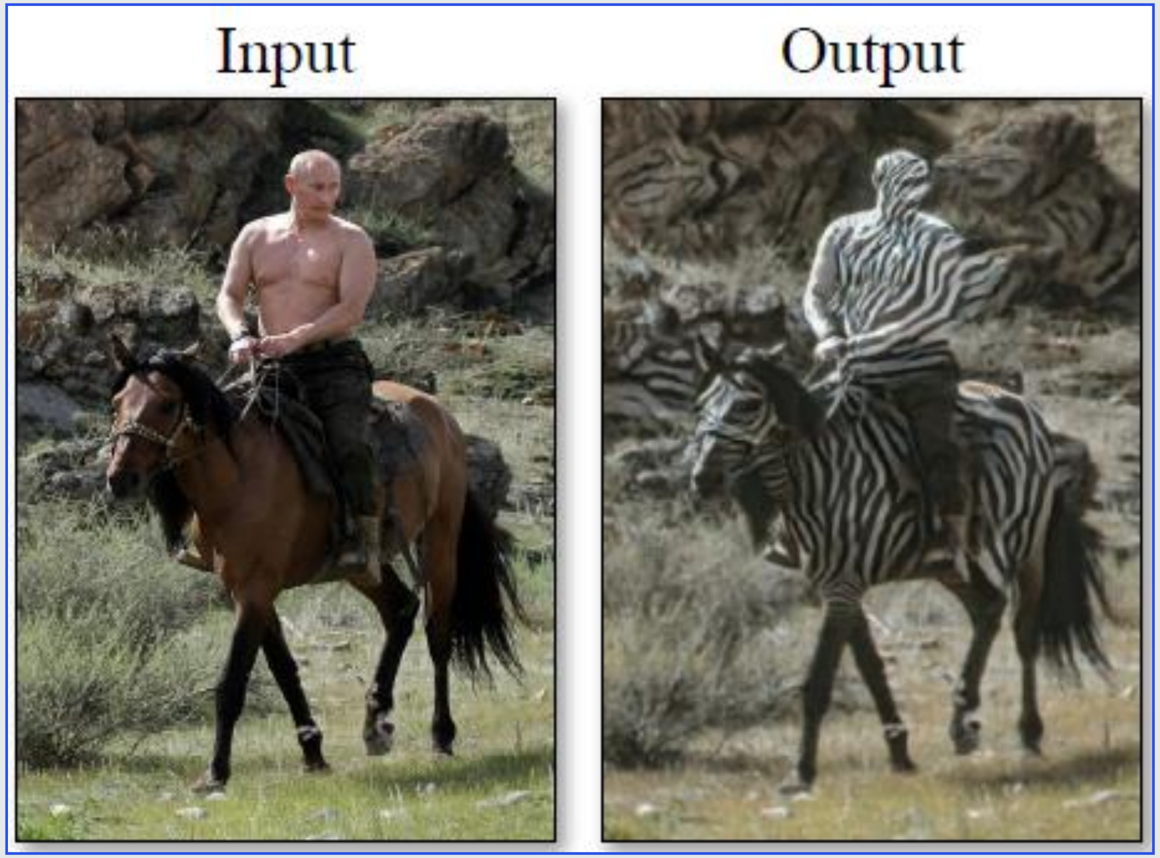
Not everything always works out..

(and Putin would not be happy)

Results depend on the needed mapping:

- Color/texture change: all good!
- Geometric changes: not as fortunate..

All the model has seen is a bunch of zebras and wild horses, not a single photo of horse-riding, no wonder Putin is now a zebra!



Conclusions

- Better use of unpaired data is key to Artificial Intelligence future.
- CycleGAN shows how to close the gap with supervised approaches using more structured learning
- Integration of weak/semi-supervised data may lead to extremely better results, still keeping the annotation cost at a minimum

Thank you for your attention!