Pix 2 Pix Overview

Image to Image translation

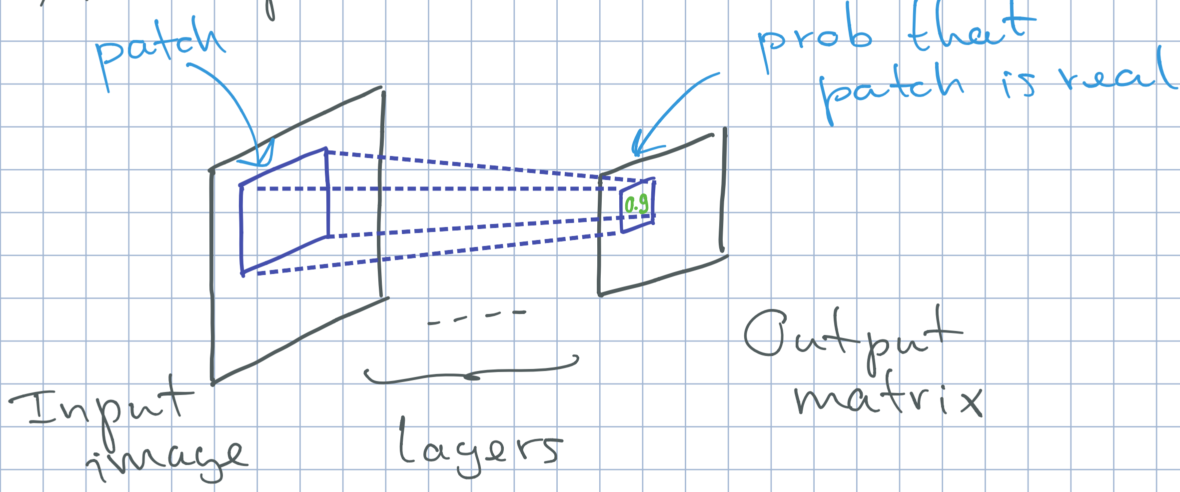
$D = \{(x_1^i, x_2^i)\}_{i=1}^n$, where x_1, x_2 – pair of images

for example x_1 – black & white and x_2 – colored version

or, x_1 – scetch, x_2 – realistic looking object

Discriminator (aka PatchGAN)

Matrix output instead of single valued fake/real prediction



Model can be trained using BCE – for fake image output matrix consists of 0's and for real – from 1's.

Input to Discriminator:

x_1 – scetch
x_2 – real/fake image

} → concat across channels dimension

Generator (U-Net)

Remark : U-Net originaly developed to solve image segmentation problem

(Encoder-Decoder Architecture) with skip-connections

Pixel Distance Loss

Addition to BCE loss we can use

$$\lambda \, | \text{real img} - \text{fake img} |$$

L_1 dist. between pixels

as a loss function term (for Generator)

This will enforce generator to build output which even more close to real image.

Generator "sees" real images.

Unpaired Image-to-Image Translation

Idea:

Instead of having dataset of paired images $\{(x_1^i, x_2^i)\}_{i=1}^n$, we have two piles of images

$X = \{x_i\}$, $Y = \{y_j\}$.

Goal of model is to find content presented in both piles (common part) and styles (unique parts of each pile)

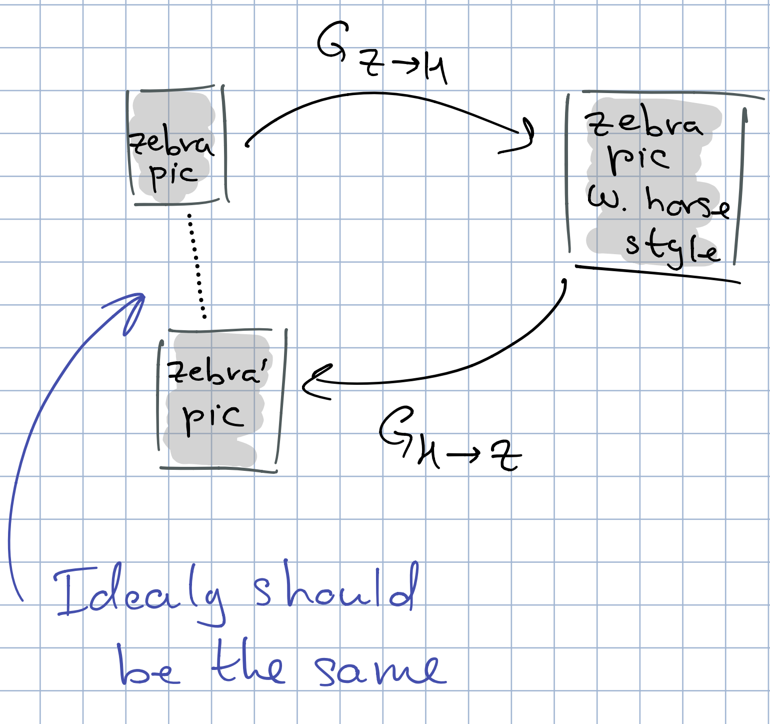
Which will allow us to take content and change its style only.

Example X - pictures of horses, Y - zebras

Model allows us to turn horses on the pic. to zebras and vise versa.

CycleGAN Overview

Intuition:



1) Take $G_{Z \to H}$ and generate horse img from zebra
2) Use generated img and $G_{H \to Z}$ to gen. zebra from fake horse img
3) Idealy, zebra' img should closly resemble original real zebra

Remark Repeat the cycle starting from horse picture.

* Generators : U-Net
* Discriminator : from PatchGAN

Four components in total:

$$G_{Z \to H}, \; D_H, \; G_{H \to Z}, \; D_Z$$

Remark

Unlike Pix2Pix Discriminator takes only an image as input (w/o condition) since dataset is not paired.

Cycle Consistency Loss

Training Loss = Adversarial + $\lambda \cdot \boxed{\text{Cycle Consist. L}}$

$$\| X_{zebra} - X_{zebra'} \|_1 + \| X_{horse} - X_{horse'} \|_1 ,$$

where
$$X_{zebra'} = G_{H \to Z}(G_{Z \to H}(X_{zebra})),$$
$$X_{horse'} = G_{Z \to H}(G_{H \to Z}(X_{horse}))$$

Cycles in both directions

Remark

Without Cycle Consist. term CycleGAN produces poor results.

Remark

For Adversarial Loss instead of using BCE like in Pix2Pix, authors use MSE, to prevent vanishing gradients problem.

Identity Loss (Optional Loss term)

Intuition:

$$G_{Z \to H}(X_{horse}) \approx X_{horse},$$

since input already a horse, $G_{Z \to H}$ should idealy output the same image

We can compute pixel distance betwen X_{horse} and $G_{Z \to H}(X_{horse})$.

Similar to $G_{H \to Z}$.

! Helps to preserve colors