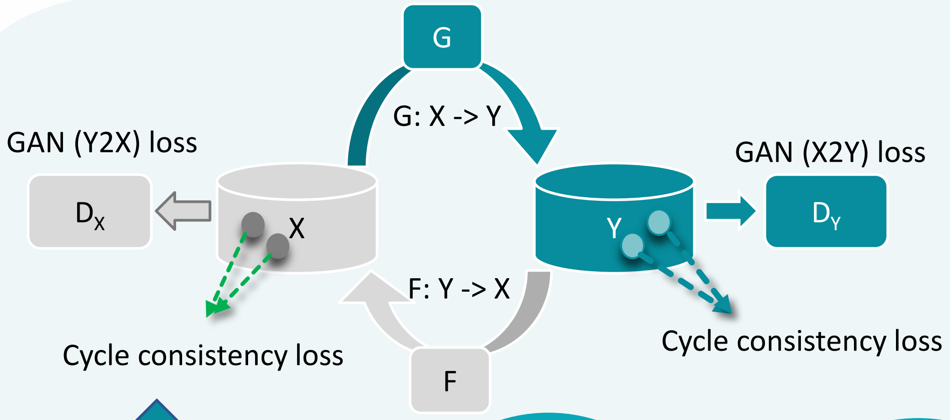
Unpaired Image to Image Translation: CycleGAN

Course Instructor: VLADIMIR PAVLOVIC
Course : **Machine Learning** (16:198:536)

Contents

CycleGAN



Architecture

Loss

Adversarial Loss

We apply adversarial loss to both our mappings of generators and discriminators.

Cycle Consistency Loss

This loss function used in Cycle GAN to measure the error rate of inverse mapping G(x) -> F(G(x)). This causes close mapping of real input (x) and F(G(x))

Cost Functions

Adversarial Loss

$$Loss_{advers}(G, D_y, X, Y) = \frac{1}{m}\sum(1 - D_y(G(x)))^2$$

$$Loss_{advers}(F, D_x, Y, X) = \frac{1}{m}\sum(1 - D_x(F(y)))^2$$

Cycle Consistency Loss

$$Loss_{cyc}(G, F, X, Y) = \frac{1}{m}[(F(G(x_i)) - x_i) + (G(F(y_i)) - y_i)]$$

Overall Loss

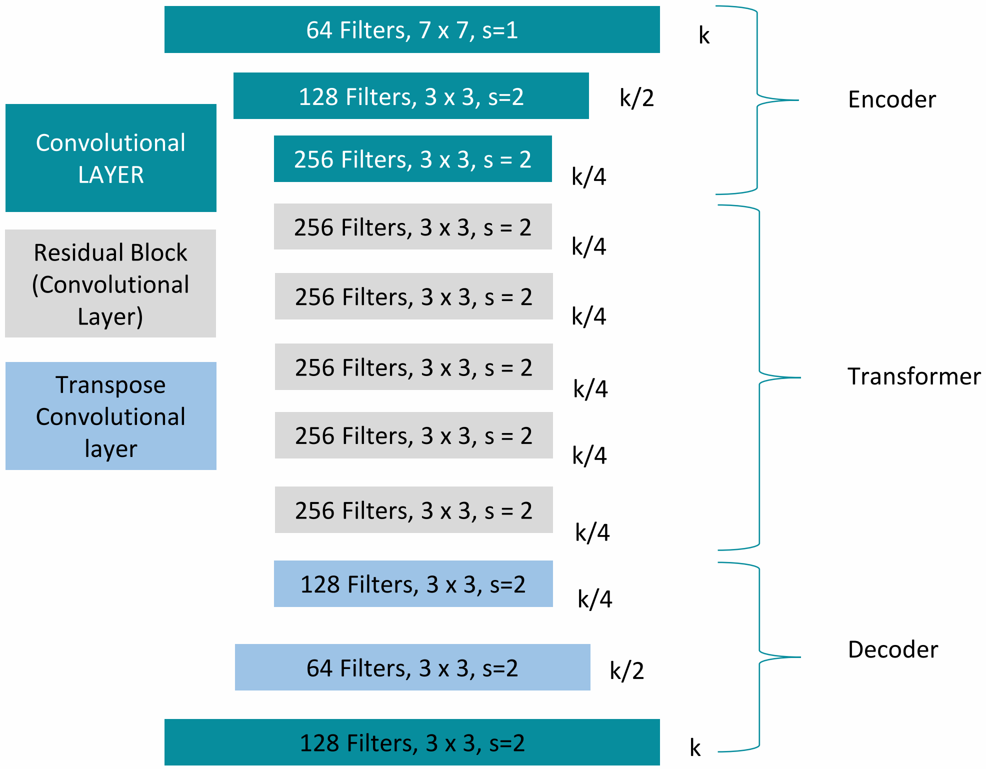
$$L(G, F, D_x, D_y) = L_{advers}(G, D_y, X, Y) + L_{advers}(F, D_x, Y, X) + \lambda L_{cycl}(G, F, X, Y)$$

Goal

$$\underset{G,F}{argmin}\ \underset{D_x,D_y}{max}\ L(G, F, D_x, D_y)$$

RUTGERS

Generator - Architecture



The architecture of generator is:

c7s1-64, d128, d256, R256, R256, R256, R256, R256, R256, u128, u64, c7s1-3

PROCEDURE

The input image is passed into the encoder. The encoder extracts features from the input image by using Convolutions and compressed the representation of image but increase the number of channels.

The encoder consists of 3 convolution that reduces the representation by 1/4 th of actual image size. Consider an image of size (256, 256, 3) which we input into the encoder, the output of encoder will be (64, 64, 256).

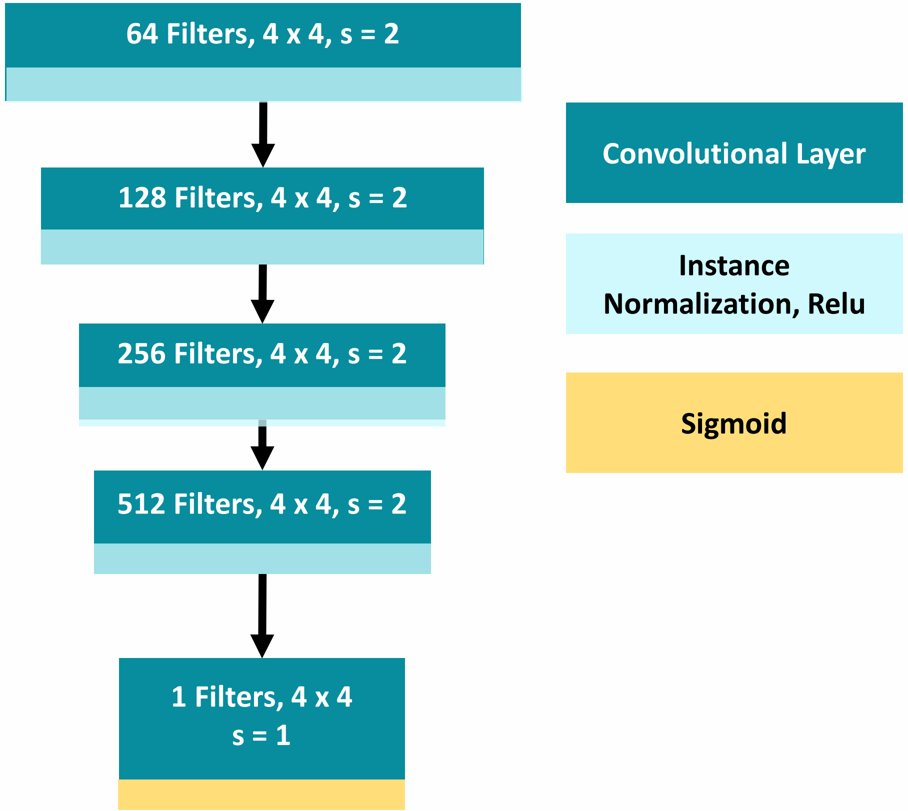
Then the output of encoder after activation function is applied is passed into the transformer. The transformer contains 6 or 9 residual blocks based on the size of input.

The output of transformer is then passed into the decoder which uses 2 -deconvolution block of fraction strides to increase the size of representation to original size

- *c7s1-k* denote a 7×7 Convolution-InstanceNorm-ReLU layer with k filters and stride 1.
- *dk* denotes a 3 × 3 Convolution-InstanceNorm-ReLU layer with *k* filters and stride 2.
- *Rk* denotes a residual block that contains two *3 × 3* convolution layers with the same number of filters on both layer.
- uk denotes a 3 × 3 fractional-strides-Convolution-InstanceNorm-ReLU layer with k filters and stride 1/2 (i.e deconvolution operation)

RUTGERS

Discriminator - Architecture



We use PatchGAN discriminator which maps from 256×256 to an NxN (here 70×70) array of outputs X, where each X_{ij} signifies whether the patch *ij* in the image is real or fake.

The architecture of discriminator is :

C64-C128-C256-C512

- Ck is 4×4 convolution-InstanceNorm-LeakyReLU layer with k filters and stride 2
- We don't apply InstanceNorm on the first layer (C64)
- After the last layer, we apply convolution operation to produce a 1×1 output

Enhancements to Cycle GAN

Global Discriminator

The global discriminator (256x256) along with Patch GAN (70x70)

Adding FID to loss

We use weighted loss with FID being 20% of the contribution to the overall loss

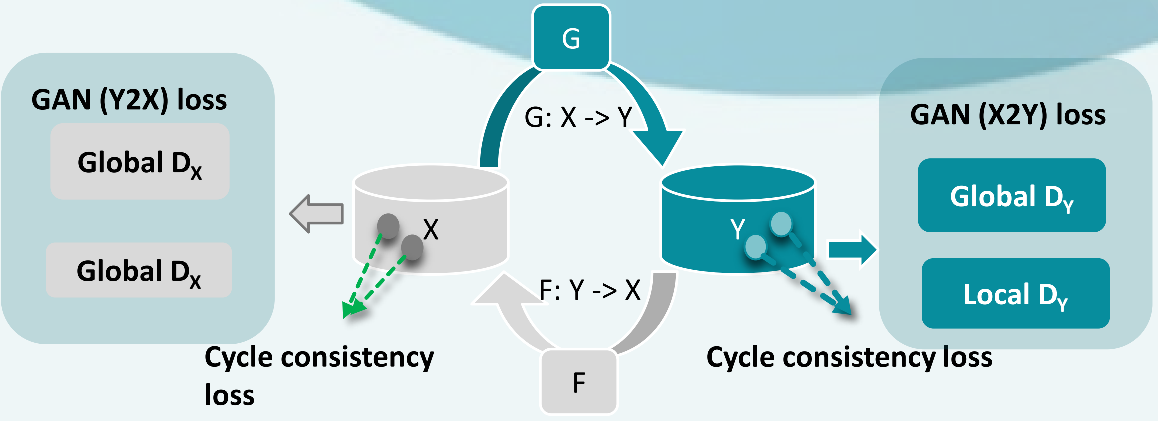


Figure: Improvised Cycle GAN Architecture

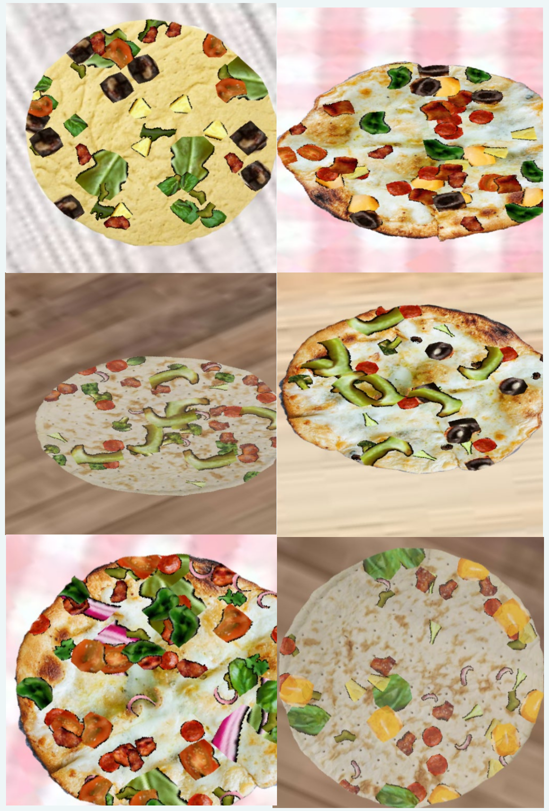
Inferences

- The use of **two discriminators** (one **local** and one **global**). Multi-scale discriminators seem can provide better texture details of the completed images at various scales.
- The key idea of using FID scores to overall loss would serve as a better penalty metric for generators compared to the base Cycle GAN

RUTGERS

Qualitative and Quantitative Evaluation of Original Cycle GAN

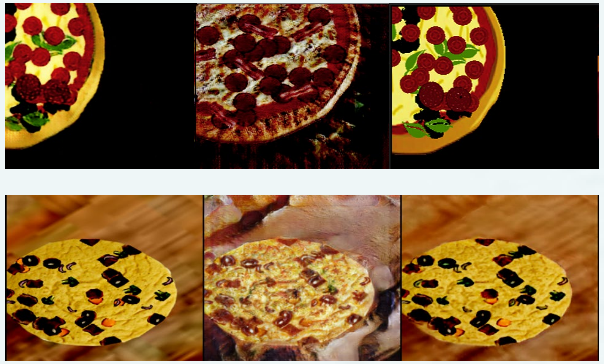
Samples of Live Synthetic Pizza dataset that we generated



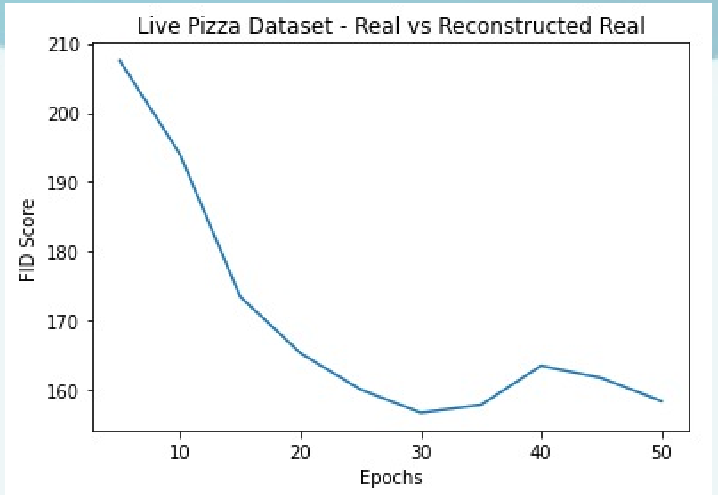
Live Pizza Dataset

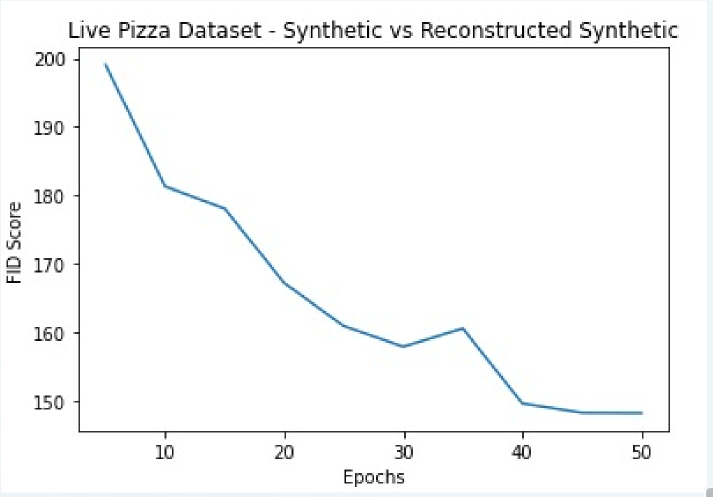
Real -> Synthetic -> Regenerated

Live Pizza Dataset



Synthetic -> Real -> Regenerated





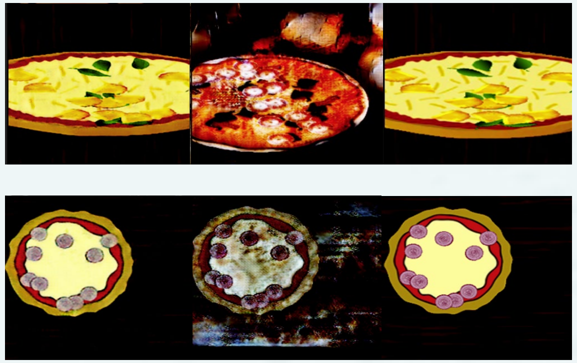
Qualitative and Quantitative Evaluation of Original Cycle GAN

Samples of synthetic pre-recorded Pizza dataset

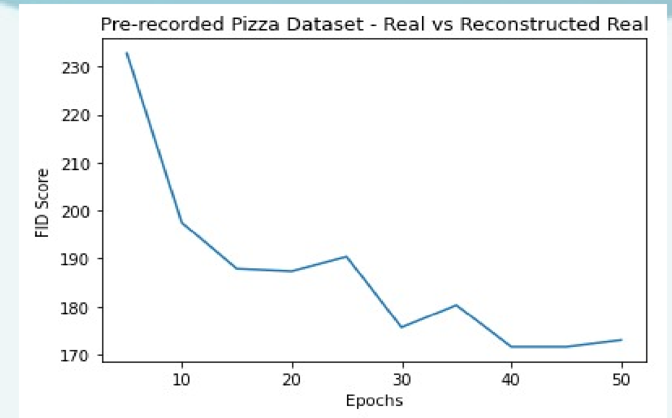
Recorded Pizza Dataset

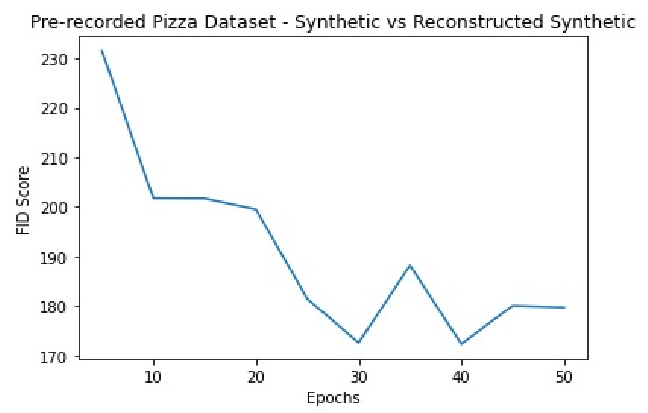
Real -> Synthetic -> Regenerated

Recorded Pizza Dataset



Synthetic -> Real -> Regenerated





Team

Karan

Animesh

Jaini

Harish