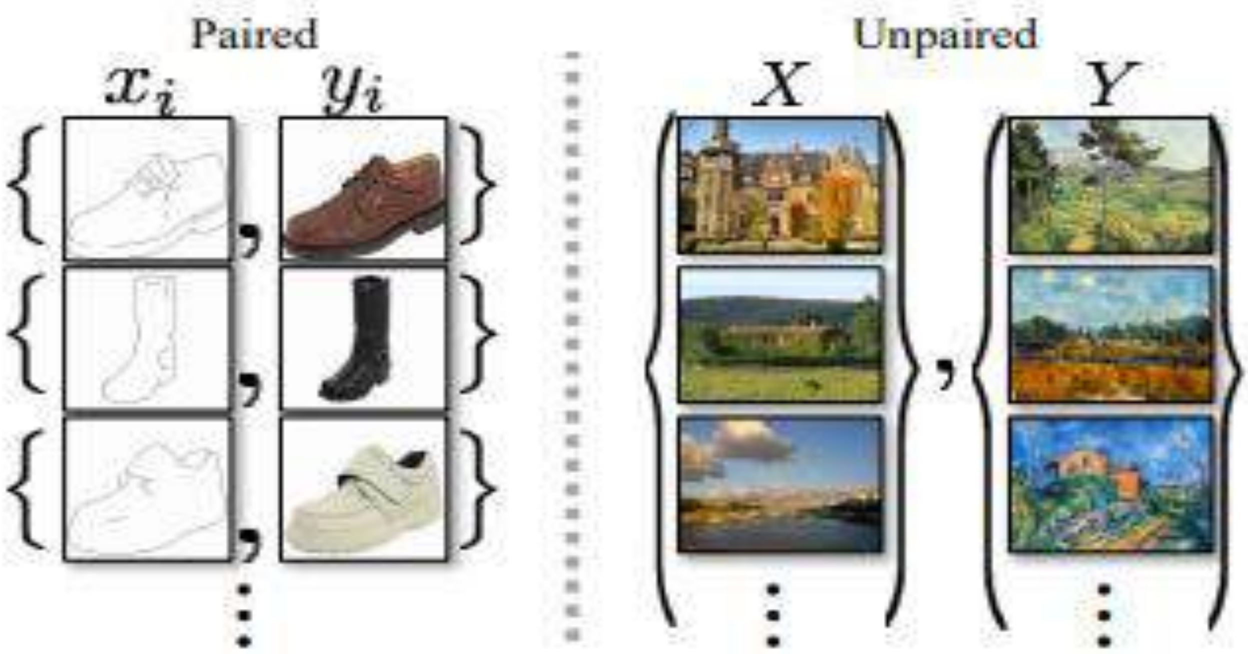
Semester: VIII Subject: Advanced AI Academic Year:2024-2025

Module 1

CycleGAN:

- Image-to-image translation involves generating a new synthetic version of a given image with a specific modification, such as translating a summer landscape to winter.
- Image-to-Image translation involves the controlled modification of an image and requires large datasets of paired images that are complex to prepare or sometimes don't exist.
- These datasets can be difficult and expensive to prepare, and in some cases impossible, such as photographs of paintings by long dead artists.
- CycleGAN is a technique for training unsupervised image translation models via the GAN architecture using unpaired collections of images from two different domains.
- CycleGAN has been demonstrated on a range of applications including season translation, object transfiguration, style transfer, and generating photos from paintings.

The most important feature of this cycle_GAN is that it can do this image translation on an unpaired image where there is no relation exists between the input image and output image.



Architecture:

CycleGAN also has two parts Generator and Discriminator.

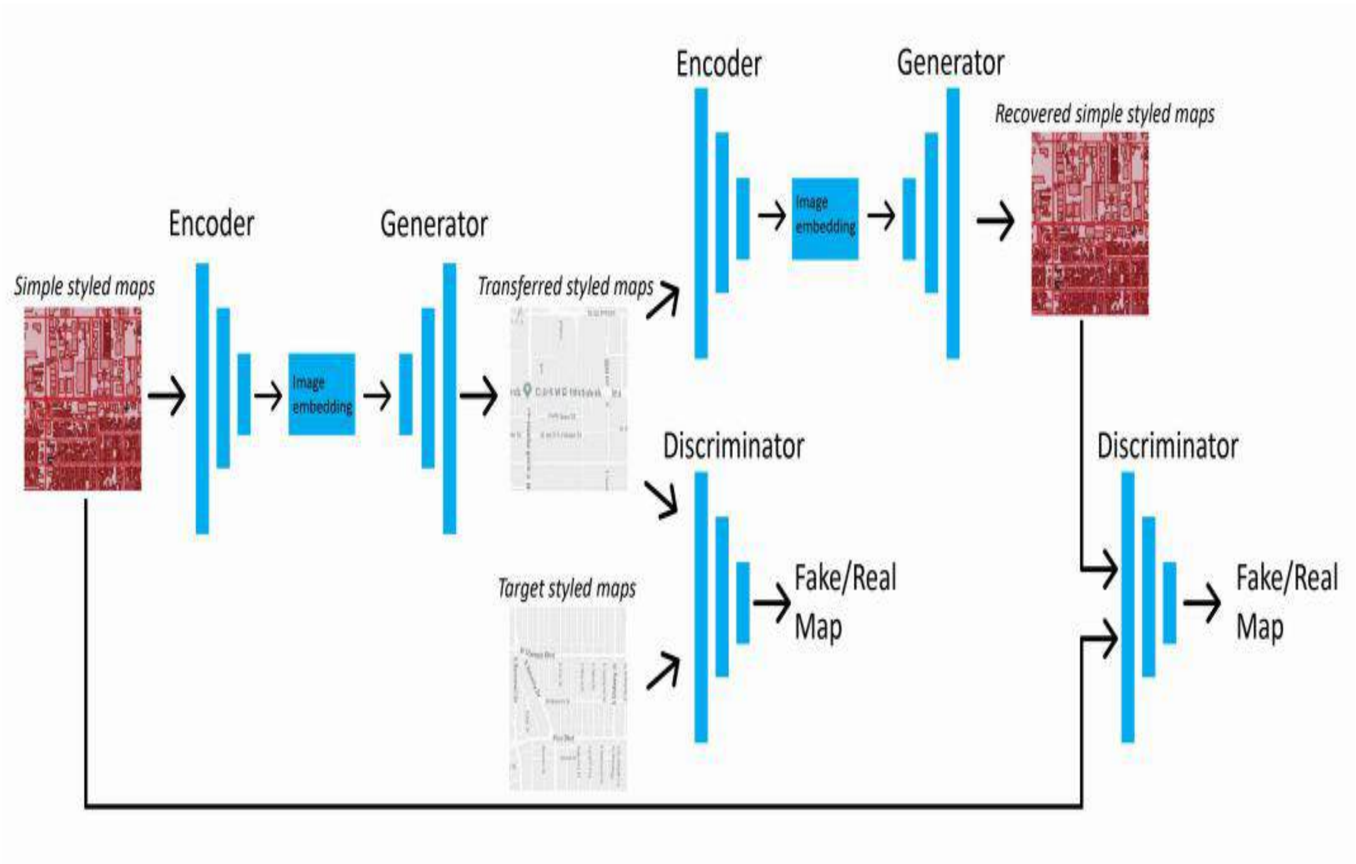
The job of generator to produce the samples from the desired distribution and the job of discriminator is to figure out the sample is from actual distribution (real) or from the one that are generated by generator (fake).

The CycleGAN architecture is different from other GANs in a way that it contains 2 mapping function (*G and F*) that acts as generators and their corresponding Discriminators (Dx and Dy): The generator mapping functions are as follows:

$$G: X \rightarrow Y$$

$$F: Y \rightarrow X$$

where X is the input image distribution and Y is the desired output distribution



CycleGAN Generator:

A CycleGAN generator is an autoencoder that takes an input image, extracts features from it, and generates another image. The generator network consists of three main stages:

- Encoder (convolutional block)
- Transformer (residual block)
- Decoder (transposed convolutional block)

The encoder stage includes three 2D convolutional layers followed by an instance normalization layer and an activation function (ReLU). The encoder extracts features from input images using convolutions, reducing the representation by 25 percent of the input image size.
The encoder output passes through the transformer stage which mainly consists of 6 to 9 residual blocks. Each block is a set of 2D convolutional layers, with every two layers followed by an instance normalization layer with a momentum.

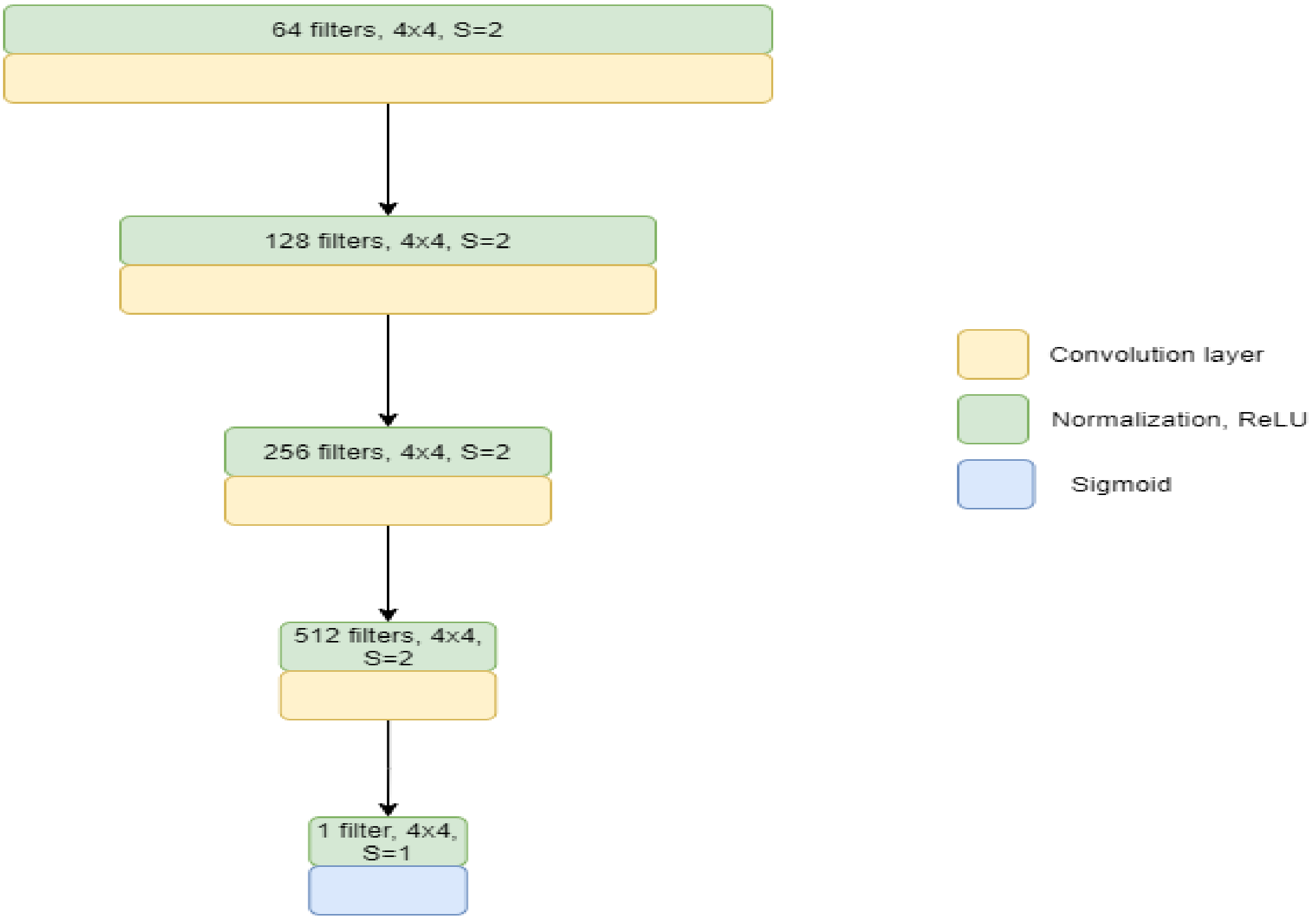
The transformer output will pass through the decoder stage, which consists of two upsampling blocks. Each upsampling block is a transposed convolution layer followed by a ReLU activation function. These two deconvolution blocks increase the size of representation of the processed images coming from the transformer to its original value.
The generator features a final 2D convolutional layer that uses the tanh activation function. This layer allows the generation of images of size equal to the size of the original input images.

CycleGAN Discriminator:

A CycleGAN discriminator is a typical CNN that includes multiple convolutional layers. This network takes an input image and classifies it as real or fake.
The CycleGAN discriminator is different from that used in the regular GAN. The latter maps input from a 256 by 256 image to a single scalar output, which represents "real" or "fake."
The CycleGAN discriminator maps from the 256 by 256 image to an N by N array of outputs X.
In that array, each Xij signifies whether the patch ij in the image is real or fake.

Semester: VIII Subject: Advanced AI Academic Year:2024-2025

The diagram below shows the typical architecture of a CycleGAN discriminator.



CycleGAN Loss Functions:

A CycleGAN has two loss functions:

- Adversarial Loss
- Cycle-Consistency Loss

Adversarial Loss: This loss is similar to the one used in the regular GAN. However, in CycleGAN, adversarial loss is applied to both generators that are trying to generate images of their corresponding domains. A generator aims to minimize the loss against its discriminator in order to finally generate real images. The adversarial loss is calculated as follows:

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

Where G is the "X to Y" domain, and F is the inverse "Y to X" domain. Dx and Dy are the discriminators.

Cycle-Consistency Loss:

To understand this loss, we should first understand the cycle-consistency approach practiced in the CycleGAN.

It represents the inverse mapping, F: Y→X.

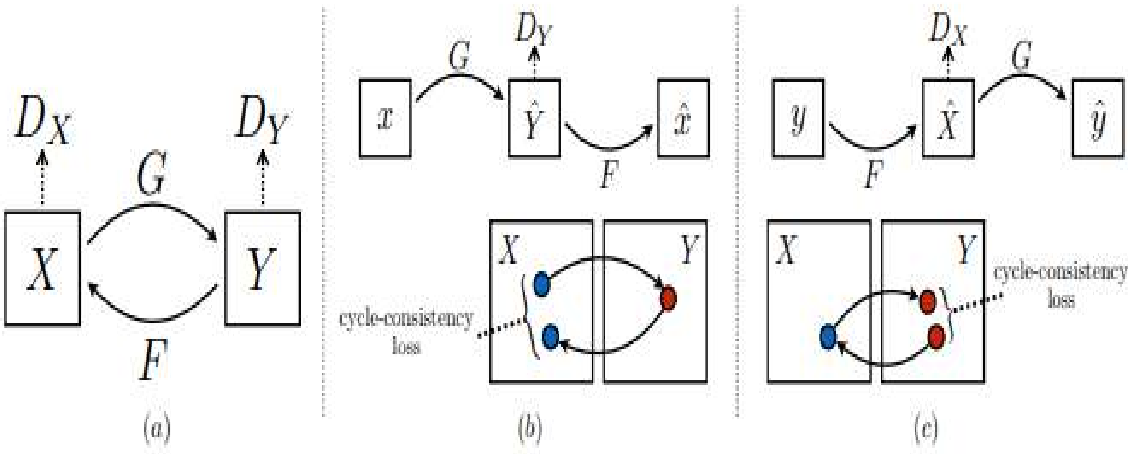
Cycle-consistency was meant to make the inverse mapping of a mapped image produce the same results as the original image.

In other words, if we want to transform an image from horse to a zebra, and then translate it back from zebra to horse, we should get the initial image.

CycleGAN uses **two cycle-consistency losses** to regularize the mappings:

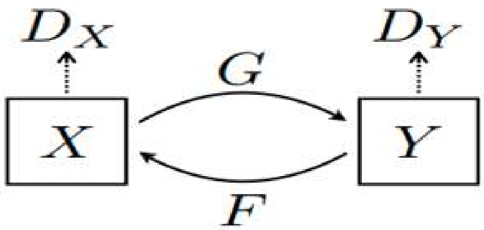
- **Forward Cycle-Consistency Loss**
- **Backward Cycle-Consistency Loss**

These two are calculated in order to make sure that if an image is translated from one domain to another (A to B) and then back (B to A), the same results will be obtained.



There are 2 generators (G and F) and 2 discriminators (X and Y) being trained here.

- Generator `G` learns to transform image `X` to image `Y`. $(G : X -> Y)$
- Generator `F` learns to transform image `Y` to image `X`. $(F : Y -> X)$
- Discriminator `D_X` learns to differentiate between image `X` and generated image `X` (`F(Y)`).
- Discriminator `D_Y` learns to differentiate between image `Y` and generated image `Y` (`G(X)`).



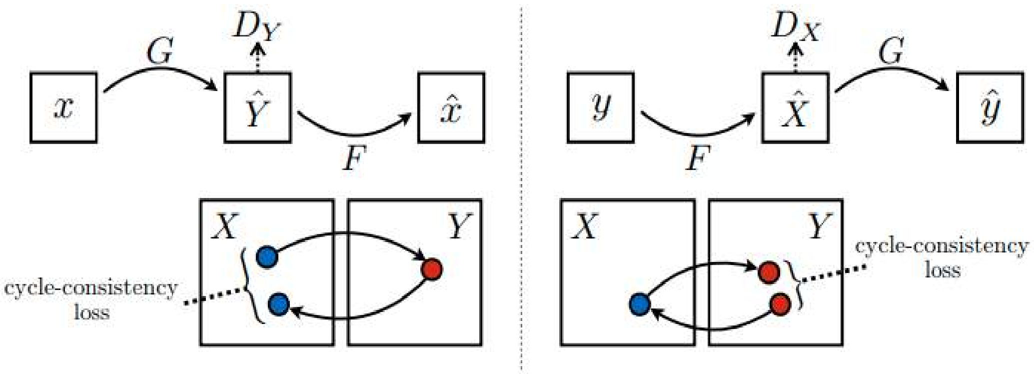
Cycle consistency means the result should be close to the original input. For example, if one translates a sentence from English to French, and then translates it back from French to English, then the resulting sentence should be the same as the original sentence.

In cycle consistency loss,

- Image X is passed via generator G that yields generated image $\hat{Y}$.
- Generated image $\hat{Y}$ is passed via generator F that yields cycled image $\hat{X}$.
- Mean absolute error is calculated between X and $\hat{X}$.

$$forward\ cycle\ consistency\ loss : X- > G(X)- > F(G(X)) \sim \hat{X}$$

$$backward\ cycle\ consistency\ loss : Y- > F(Y)- > G(F(Y)) \sim \hat{Y}$$



CycleGAN Advantages:

1. **Accurate** — CycleGAN is more accurate than its predecessors because it takes advantage of unpaired data, meaning that it can produce better results without the need for a large number of paired training images. This is important because paired training images are often difficult to obtain.
2. **Robust** — CycleGAN is more robust to domain shifts in the data, meaning that it can perform well even if the input images come from different domains. This is important because it allows CycleGAN to be used for a range of image-to-image translation tasks, such as changing the style of an image, translating from one domain to another, or even translating from one language to another.
3. **Small dataset** — CycleGAN can be used to generate high-quality images with relatively little training data. This is important for tasks in which training data is limited, such as medical image-to-image translation.
4. **Easy to train** — CycleGAN is easy to train and use, meaning that it is accessible to practitioners who may not have access to the complex infrastructure required for other GAN models.

CycleGAN Disadvantages:

1. The CycleGAN can be slow to train, particularly when there are a large number of training images.
2. The CycleGAN can be prone to overfitting, which can lead to poor results on unseen data.

Applications:

1. Style Transfer

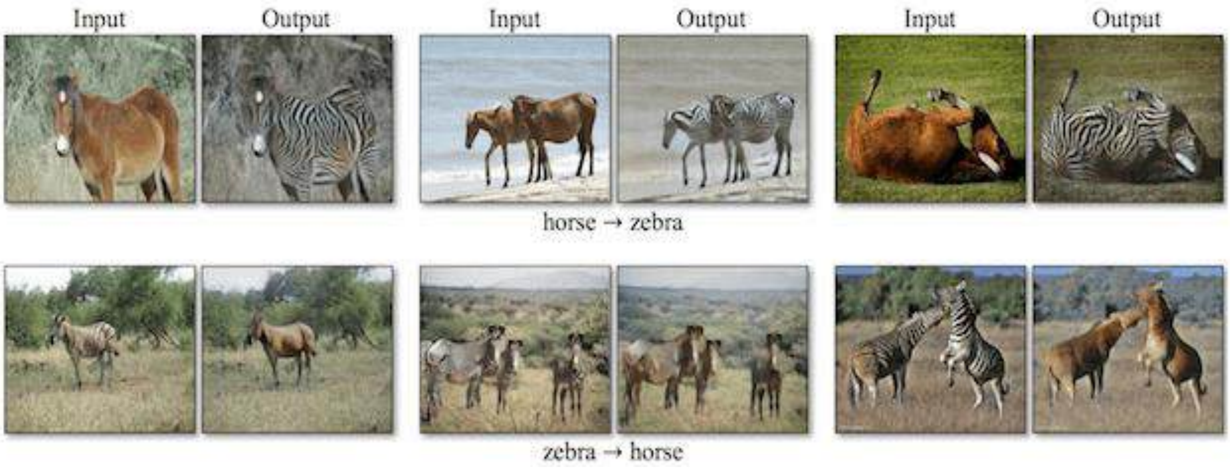
Style transfer refers to the learning of artistic style from one domain, often paintings, and applying the artistic style to another domain, such as photographs.

The CycleGAN is demonstrated by applying the artistic style from Monet, Van Gogh, Cezanne, and Ukiyo-e to photographs of landscapes.

2. Object Transfiguration

Object transfiguration refers to the transformation of objects from one class, such as dogs into another class of objects, such as cats.

The CycleGAN is demonstrated transforming photographs of horses into zebras and the reverse: photographs of zebras into horses. This type of transfiguration makes sense given that both horse and zebras look similar in size and structure, except for their coloring.



3. Season Transfer

Season transfer refers to the translation of photographs taken in one season, such as summer, to another season, such as winter.

The CycleGAN is demonstrated on translating photographs of winter landscapes to summer landscapes, and the reverse of summer landscapes to winter landscapes.

4. Photograph Generation From Paintings:

Photograph generation from paintings, as its name suggests, is the synthesis of photorealistic images given a painting, typically by a famous artist or famous scene. The CycleGAN is demonstrated on translating many paintings by Monet to plausible photographs.