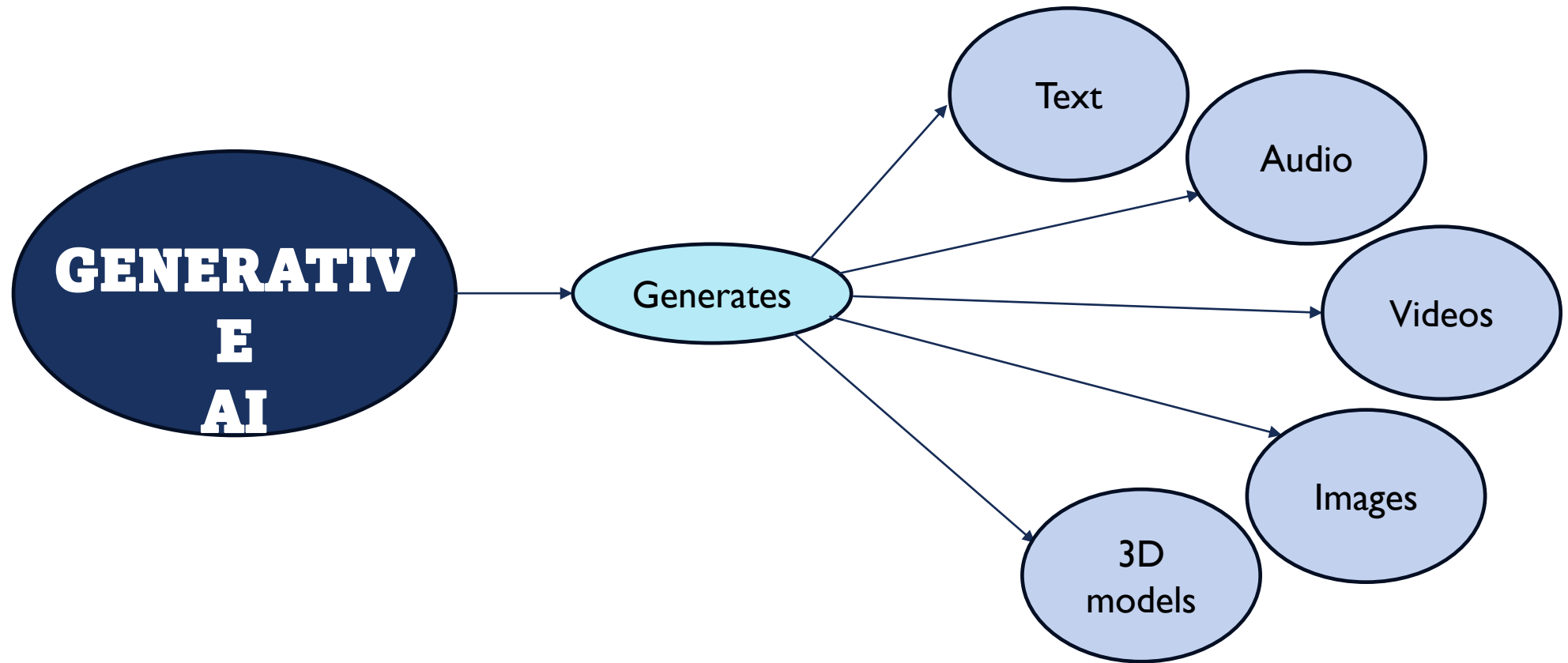


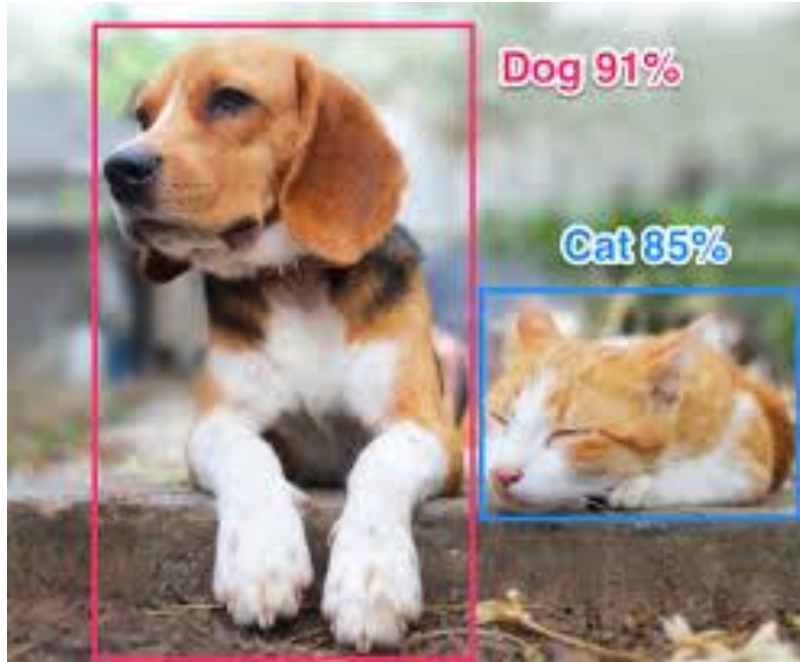


Presenter –
MUSKAN RATHORE
(Ex-DATA AND APPLIED
SCIENTIST, MICROSOFT
IITK'22)


MASTERCLASS - GANS





WHAT IS GENERATIVE AI?





GENERAL MACHINE LEARNING OR AI


 **You**
What should I eat for food in evening?

 **ChatGPT**
How about a cozy bowl of creamy tomato soup paired with a grilled cheese sandwich? It's a classic comfort food combo perfect for a relaxing evening.

 **You**
I want to have Indian today

 **ChatGPT**
How about indulging in some flavorful Indian cuisine? Treat yourself to a savory plate of butter chicken paired with fragrant basmati rice and warm, fluffy naan bread. It's sure to be a delicious and satisfying choice for your evening meal.

 **You**
i am a vegetarian

 **ChatGPT**
No worries! You can enjoy a delicious vegetarian Indian meal. How about trying some paneer tikka masala served with aromatic jeera rice and accompanied by freshly baked garlic naan? It's a mouthwatering combination that celebrates the rich flavors of Indian cuisine without any meat.



GENERATIVE AI



You

Write 2 lines involving a dog and harry becoming friends from a harry potter novel



ChatGPT

In the hallowed halls of Hogwarts, amidst spells and charms, Harry found an unexpected ally in a loyal pup, their bond forged in the magical crucible of friendship.



You

Write 2 lines about Mars from a Shakesperean play



ChatGPT

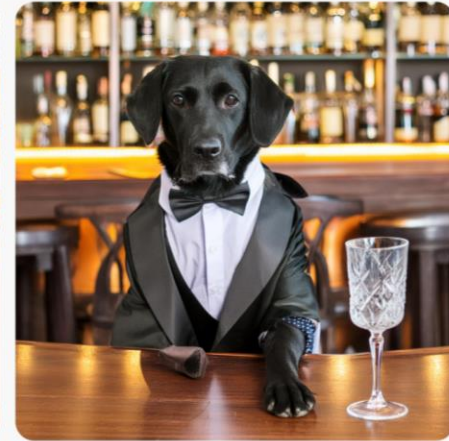
"Upon yon celestial stage, Mars doth twinkle bright,
A fiery orb, casting shadows in the starry night."



GENERATIVE AI

Prompt

A black dog in a tuxedo sitting in a bar



GENERATIVE AI

Lip-Sync Results



Video Input
(Unsynced)



Audio Input



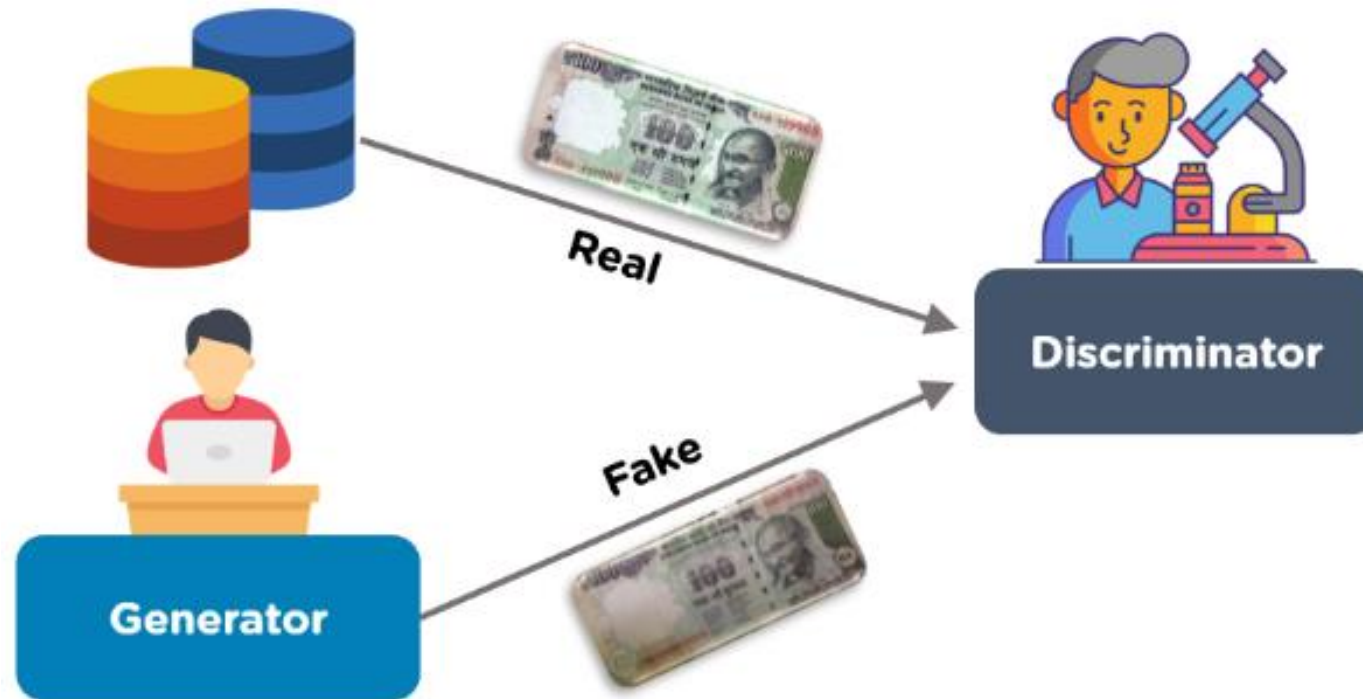
Video Output
(Synced)

Source: Prajwal et al. 2020

GENERATIVE AI

- Learning to mimic people's way of speaking by learning their patterns of speaking and voice modulation by mimic artists.
- Train an artificial author who can write new novels in a similar manner to some famous author, say Dan Brown, Shakespeare, etc.
- You are not able to buy a painting from a famous painter which might be too expensive. Can you create an artificial painter which can paint like any famous artist by learning from his / her past collections?
- "To get better at something, compete with an opponent better than you"
- "Forger vs Investigator": Painting forgers try creating replica paintings of famous artists to scam and earn money, Investigators on the other hand try to detect if a painting is by the true artist or forged. Forger keeps on getting better by learning from the mistakes that got him caught by the investigator, till investigator cannot tell his and artists' paintings apart

INSPIRATION BEHIND GANS



GAN (GENERATIVE ADVERSARIAL NETWORKS)

- Traditional GANs often used fully connected layers, which struggled with large, high-resolution images. DCGANs improve on this by:
 - Replacing fully connected layers with **convolutional layers**, which are better at capturing spatial hierarchies in images.
 - This helps in generating better
 - The generator uses ConvolutionTranspose2D Layers (reverse Convolution layers) to upsample noise vector z to a high dimensional image
 - For instance, the process might go from a small feature map of size 4×4 to 8×8 to 16×16 , and eventually to the desired output size, such as 64×64 or 128×128 .
 - Discriminator uses normal Convolution layers, which takes image as an input, downsamples it to then do binary classification – Real or Fake

```
1 [[1 2]
2  [3 4]]
3
4
5 Layer (type)          Output Shape          Param #
6 =====
7 conv2d_transpose_1 (Conv2DTr (None, 4, 4, 1)      2
8 =====
9 Total params: 2
10 Trainable params: 2
11 Non-trainable params: 0
12 =====
13
14
15 [[1. 0. 2. 0.]
16  [0. 0. 0. 0.]
17  [3. 0. 4. 0.]
18  [0. 0. 0. 0.]]
```

DCGAN (DEEP CONVOLUTIONAL GAN)

- Training is same as basic GAN, only the structure of Generator and discriminator has different layers

Use Cases of DCGAN:

- **Image Synthesis:**

- Generate realistic human faces (e.g., CelebA dataset).

- **Data Augmentation:**

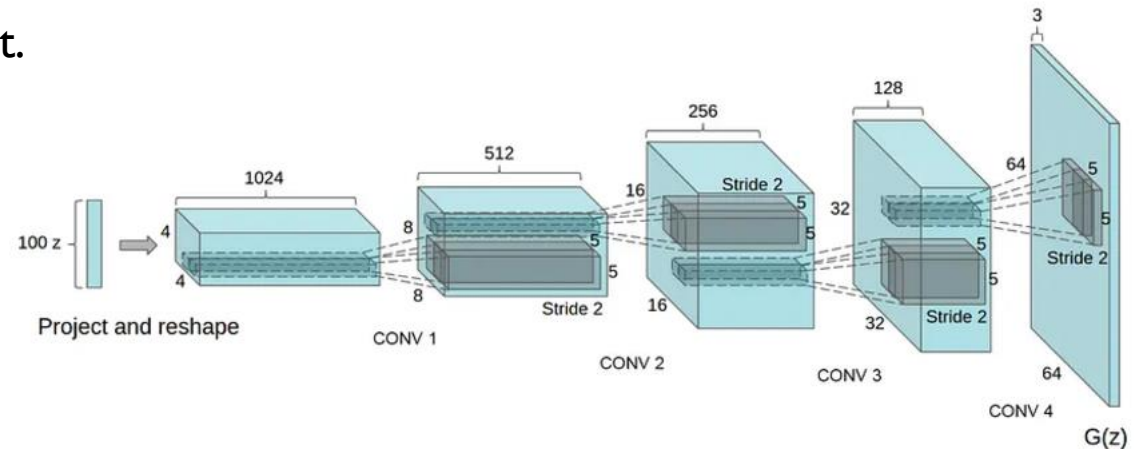
- Generate synthetic data to augment training datasets.

- **Art and Design:**

- Create textures, patterns, and abstract art.

- **Image Inpainting:**

- Reconstruct missing parts of images.



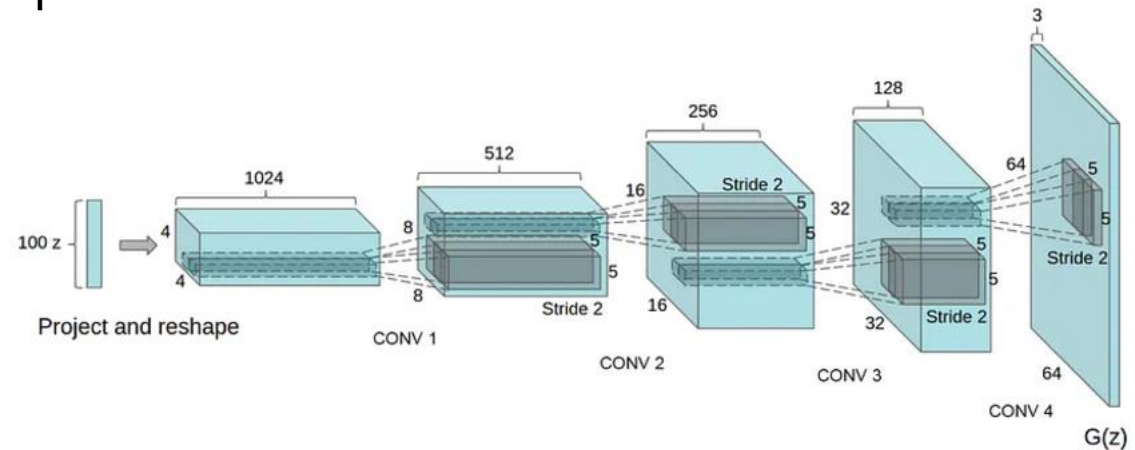
DCGAN (DEEP CONVOLUTIONAL GAN)

Strengths:

- Leverages convolutional layers for better spatial feature learning.
- Produces higher-quality images compared to traditional GANs.
- Relatively simple to implement and train.

Challenges:

- Still prone to mode collapse (generator outputs limited variations).
- Sensitive to hyperparameters (e.g., learning rate, architecture).
- Struggles with very high-resolution images or complex datasets.



DCGAN (DEEP CONVOLUTIONAL GAN)

- A **Conditional GAN (cGAN)** modifies this basic GAN structure by introducing a condition c that guides both the generator and discriminator. This condition can represent specific information, such as:
 - **Class labels** (e.g., digits in MNIST).
 - **Text descriptions** (e.g., "a bird with a red belly").
 - **Images** (e.g., translating images from one domain to another).
- **Generator $G(z, c)$**
 - Combines the random noise vector z with the condition c and generates data based on this combination.
- **Discriminator $D(x, c)$**
 - Takes the data x and the condition c as input and determines whether x is real or fake, ensuring that the generated data matches the given condition.

CGAN (CONDITIONAL GAN)

- Text-to-Image Synthesis:**

- Goal:** Generate images based on text descriptions.

- Example:** Input: "A yellow bird with black wings"; Output: A realistic image of such a bird.

- Image-to-Image Translation:**

- Goal:** Transform an image from one domain to another using conditions.

- Example:** Semantic segmentation maps ↔ realistic images.

- Data Augmentation:**

- Goal:** Generate additional data for specific categories.

- Example:** Creating synthetic images for rare classes in datasets.

- Super-Resolution:**

- Goal:** Generate high-resolution images from low-resolution inputs.

- Condition:** The low-resolution image serves as c.

Advantages:

- Control Over Output:** Unlike traditional GANs, cGANs allow specific outputs based on conditions.

- Versatile Applications:** Can be applied to multiple tasks like text-to-image, image translation, and more.

- Enhanced Learning:** Conditioning helps both the generator and discriminator learn meaningful features.

Cons:

- Data Labeling:** Requires labeled data for the conditioning variable c.

- Training Instability:** Similar to other GANs, cGANs suffer from issues like mode collapse.

- Condition Misalignment:** If the generator fails to incorporate the condition, it may produce irrelevant outputs

CGAN (CONDITIONAL GAN)

- Cyclic Gan learns to generate map images between 2 domains X and Y.
- Has 2 Generators : G and F; and 2 Discriminators: Dx and Dy
- Loss has 3 components: Loss of Dx + Loss of Dy + λ * Cycle Consistency Loss
- Cycle consistency loss makes sure that if an image is constructed from X to Y then it can be reconstructed back to X domain.
- Applications:
 - Style Transfer:**
 - Example: Translating between **photos and paintings** in styles like Monet or Van Gogh.
 - Domain Adaptation:**
 - Example: Converting **synthetic images to real images** for tasks like object detection.
 - Image Enhancement:**
 - Example: Enhancing **low-light images** or converting black-and-white photos to color.
 - Object Transfiguration:**
 - Example: Converting **horses to zebras, cats to dogs**, etc.
 - Medical Imaging:**
 - Example: Translating between different imaging modalities (e.g., MRI to CT scans).

CYCLIC GAN

- **StyleGAN (Style Generative Adversarial Networks)** is a type of **Generative Adversarial Network (GAN)** designed to improve the quality and control of generated images, with a focus on **high-quality image synthesis** and the ability to manipulate various visual aspects of images, such as **style** and **content**. It was introduced by **NVIDIA** in 2018.
- StyleGAN introduced the ability to **separate content and style** during the image generation process, allowing for more flexible control over the output image.
- Enables AdaIN (Adaptive Instance Normalization) - allows for the manipulation of image styles by adjusting the statistics of the feature maps at different layers of the network
- StyleGAN employs **progressive growing**, which gradually increases the resolution of generated images during training. This technique helps stabilize training and ensures that high-quality images are produced from low to high resolutions.

Generator Architecture

1. Mapping Network:

- I. Transforms the input latent vector z (sampled from a standard Gaussian distribution) into an intermediate latent code w . This disentangles entangled features in z making w more meaningful for style control.

2. Synthesis Network:

- I. The generator starts with a constant 4×4 learnable tensor and uses w to modulate the AdaIN layers at different scales, progressively building up the image.

Discriminator Architecture:

- The discriminator is similar to that in standard GANs but uses **progressive growing** to classify real vs. fake images at multiple resolutions.

STYLE GAN

Applications of StyleGAN

1.High-Resolution Image Generation:

- 1. Generating photo-realistic faces, objects, or scenes (e.g., portraits that look indistinguishable from real ones).

2.Image Editing:

- 1. Intuitive manipulation of attributes like hairstyle, age, or expression by altering specific layers or latent vectors.

3.Creative Industries:

- 1. Generating artwork, textures, and assets for games and animations.

4.Medical Imaging:

- 1. Simulating medical data (e.g., MRIs) for training models without exposing real patient data.

5.Research and Education:

- 1. Used as a baseline for understanding disentangled representation learning and GAN architectures.

STYLE GAN

Applications of StyleGAN

1.High-Resolution Image Generation:

- 1. Generating photo-realistic faces, objects, or scenes (e.g., portraits that look indistinguishable from real ones).

2.Image Editing:

- 1. Intuitive manipulation of attributes like hairstyle, age, or expression by altering specific layers or latent vectors.

3.Creative Industries:

- 1. Generating artwork, textures, and assets for games and animations.

4.Medical Imaging:

- 1. Simulating medical data (e.g., MRIs) for training models without exposing real patient data.

5.Research and Education:

- 1. Used as a baseline for understanding disentangled representation learning and GAN architectures.

STYLE GAN



QUESTIONS?



THANK YOU !