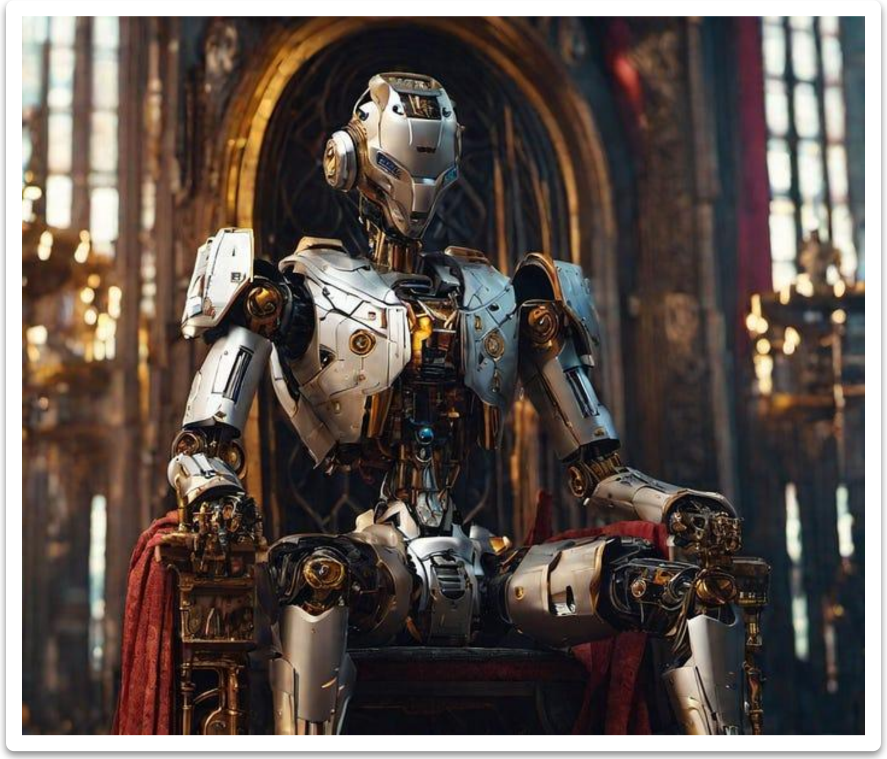
Generative AI for Image Generation

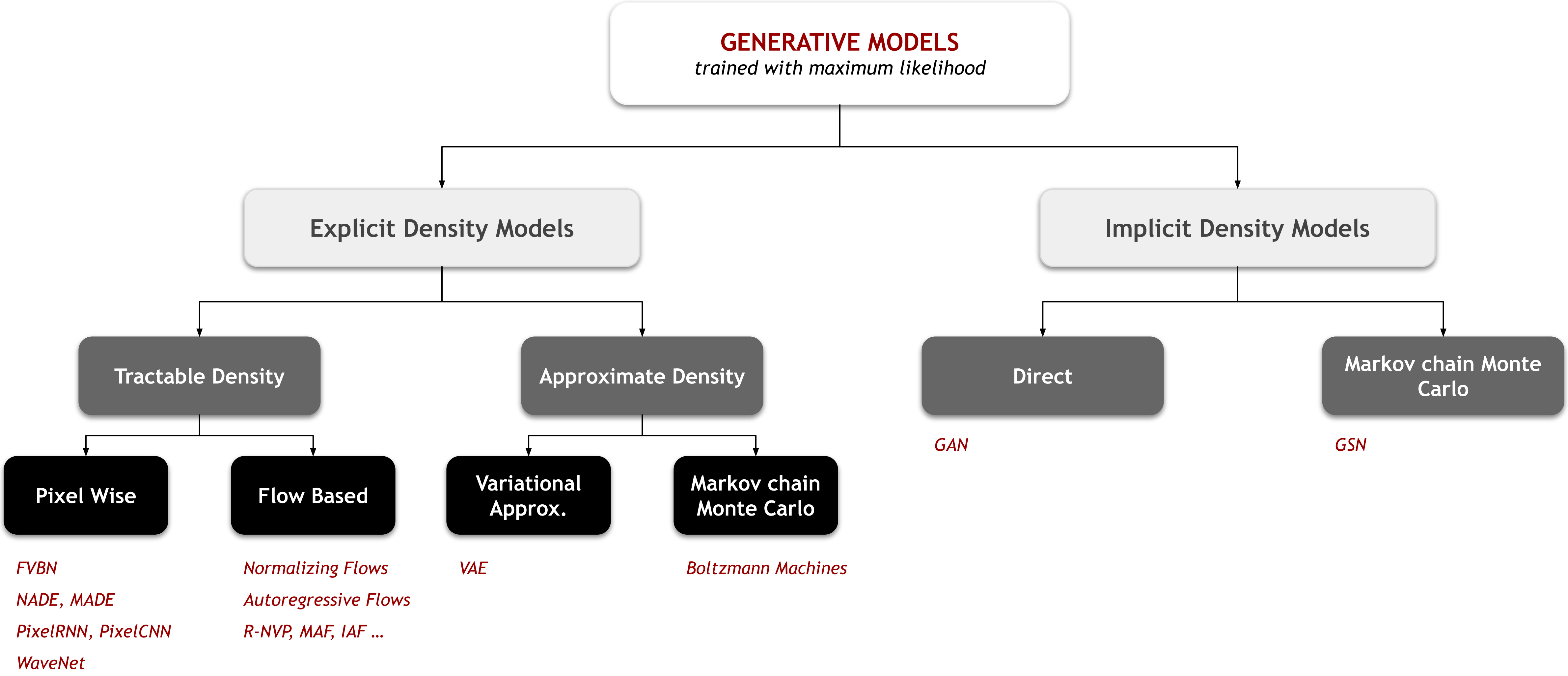
Revolutionizing Visual Creation: Exploring the Potential of Generative AI in Image Generation

Introduction to Image Generation

Image generation using AI involves using algorithms and deep learning models to create realistic and novel images from scratch, based on input text (prompt).



Various Image Generation Models



Let’s Focus on…

Variational Encoders (VAE)

Autoregressive Models

Generative Adversarial Networks (GAN)

Various Available Models

Variational Encoders (VAE)

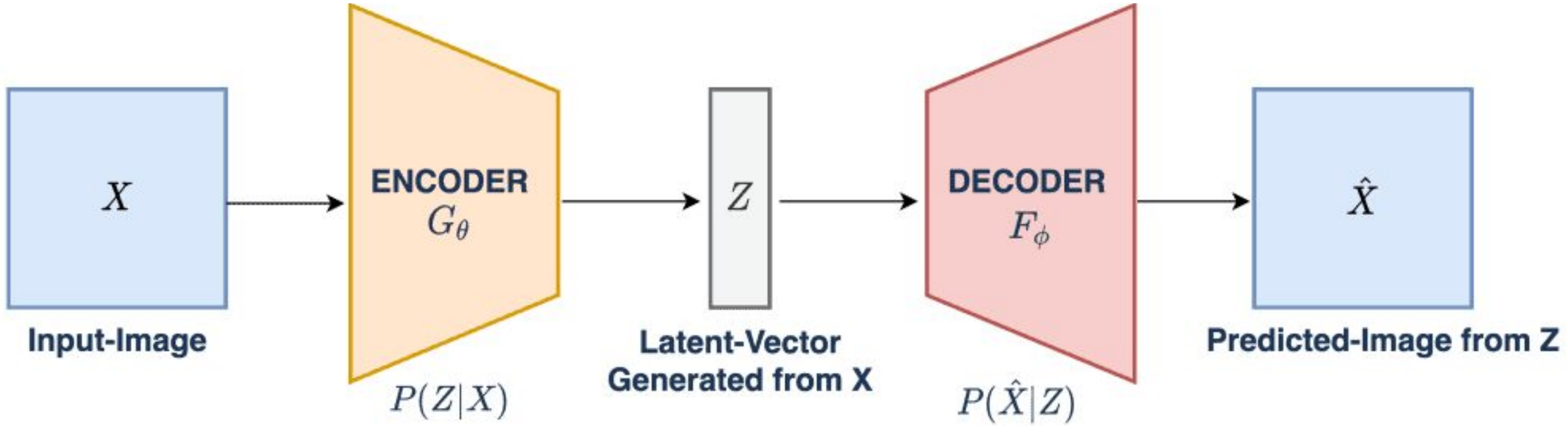
- Variational Autoencoders (VAEs) are generative models that learn a dataset's underlying probability distribution and generate new samples
- They use an encoder-decoder architecture, where the encoder maps the input data to a latent representation, and the decoder tries to reconstruct the original data from this latent representation
- The VAE is trained to minimize the difference between the original data and the reconstructed data, allowing it to learn the underlying distribution of the data and generate new samples that follow that same distribution
- One of the key advantages of VAEs is that they can generate new data samples similar to the training data
- This is because the latent space learned by the VAE is continuous, which allows the decoder to generate new data points that are smoothly interpolated between the training data points

Various Available Models

Variational Encoders (VAE)

VAEs are a type of Generative Models that can learn to encode and decode data, enabling the generation of new and realistic data based on input

- Encoding
- Sampling
- Decoder



Various Available Models

Variational Encoders (VAE)

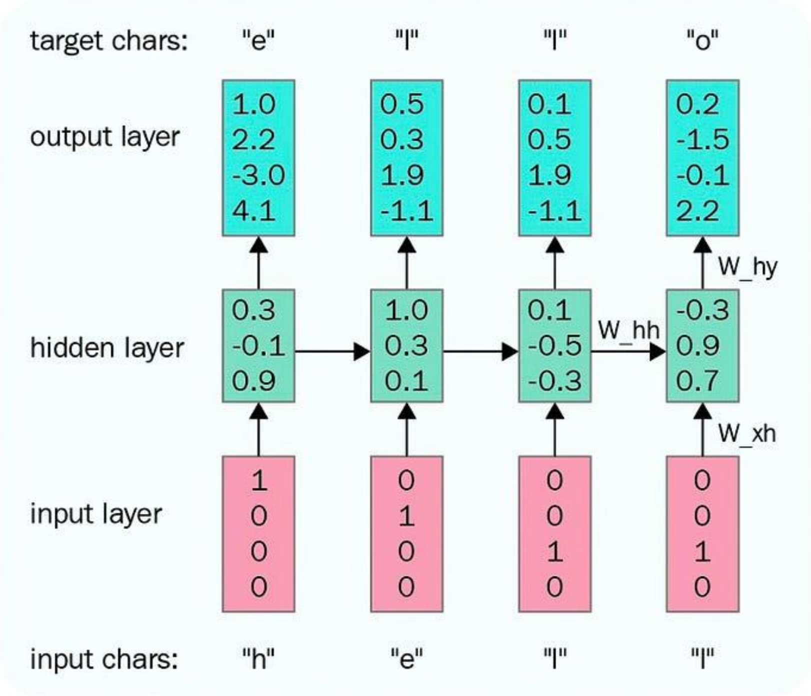
When not to use VAE:

- Avoid VAEs for small datasets or datasets with simple structures, as they may struggle to capture meaningful latent representations
- VAEs may not be ideal for highly noisy or high-dimensional input spaces, where simpler models might be more effective
- If real-time processing is crucial, consider alternatives to VAEs to minimize computational overhead associated with probabilistic sampling
- Choose other models if precise control over the latent space or accurate reconstruction without probabilistic components is a priority
- When domain-specific constraints, interpretability, or prior knowledge are critical, explore models that better accommodate such requirements, as VAEs may lack flexibility in these aspects

Various Available Models

Autoregressive Models

- The AutoRegressive (AR) model generates images from the random noises or latent vectors in the Variational Autoencoders (VAEs)
- These generate images by treating an image as a sequence of pixels



Various Available Models

Autoregressive Models

- Autoregressive models, in the context of image generation, refer to a class of generative models that generate images pixel by pixel, typically from left to right and top to bottom
- These models define a probabilistic distribution over the entire image by modeling the joint distribution of pixel intensities, conditioned on the previously generated pixels
- In other words, the generation process is autoregressive because each pixel's value is generated based on the values of the pixels that have been generated before it
- One prominent example of autoregressive models for image generation is the **PixelRNN** (Pixel Recurrent Neural Network) and its variant, **PixelCNN** (Pixel Convolutional Neural Network)
- These models use neural networks to model the conditional probability distribution of each pixel given the already generated pixels in a sequential manner

Various Available Models

Autoregressive Models - PixelRNN

- An effective approach to model such a network is to use probabilistic density models (like Gaussian or Normal distribution) to quantify the pixels of an image as a product of conditional distributions
- This approach turns the modeling problem into a sequence problem wherein the next pixel value is determined by all the previously generated pixel values

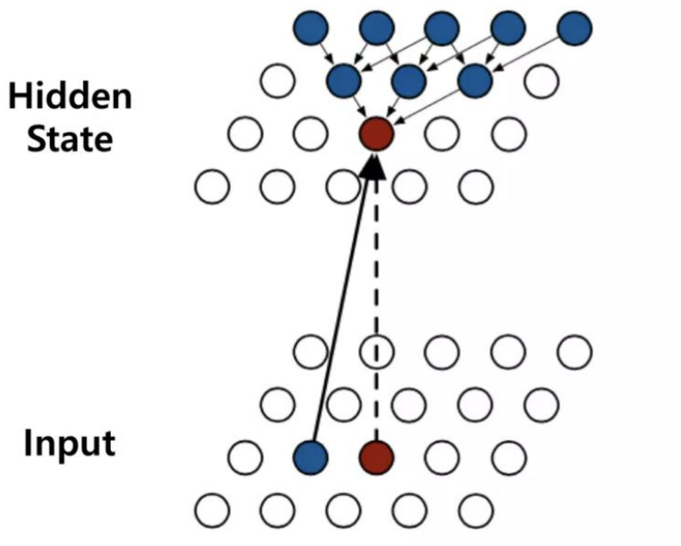
PixelCNN	**Row LSTM**	**Diagonal BiLSTM**
7×7 conv mask A		
Multiple residual blocks: (see fig 5)		
Conv 3×3 mask B	Row LSTM i-s: 3×1 mask B s-s: 3×1 no mask	Diagonal BiLSTM i-s: 1×1 mask B s-s: 1×2 no mask
ReLU followed by 1×1 conv, mask B (2 layers)		
256-way Softmax for each RGB color (Natural images) or Sigmoid (MNIST)		

Table 1. Details of the architectures. In the LSTM architectures i-s and s-s stand for input-state and state-state convolutions.

Various Available Models

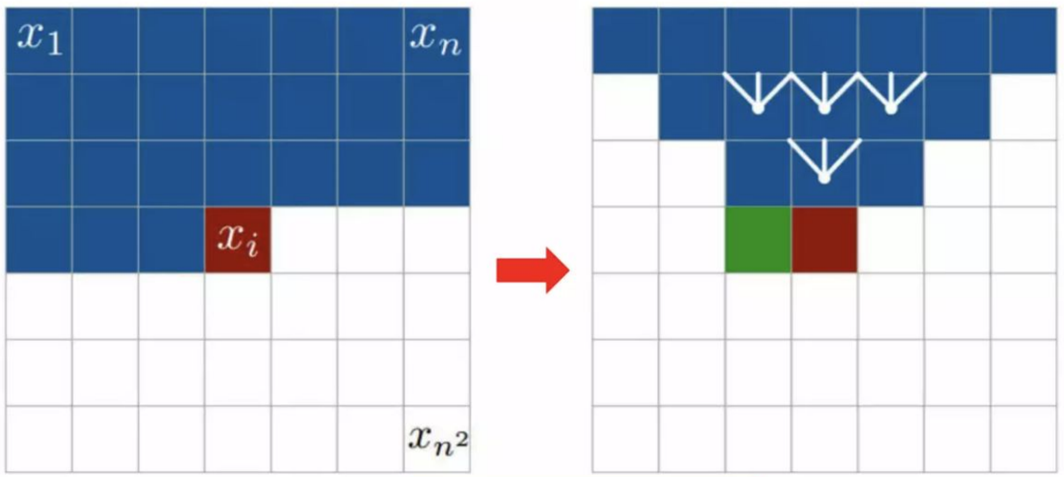
Autoregressive Models - PixelRNN

Row LSTM



Row LSTM

input-to-state & state-to-state

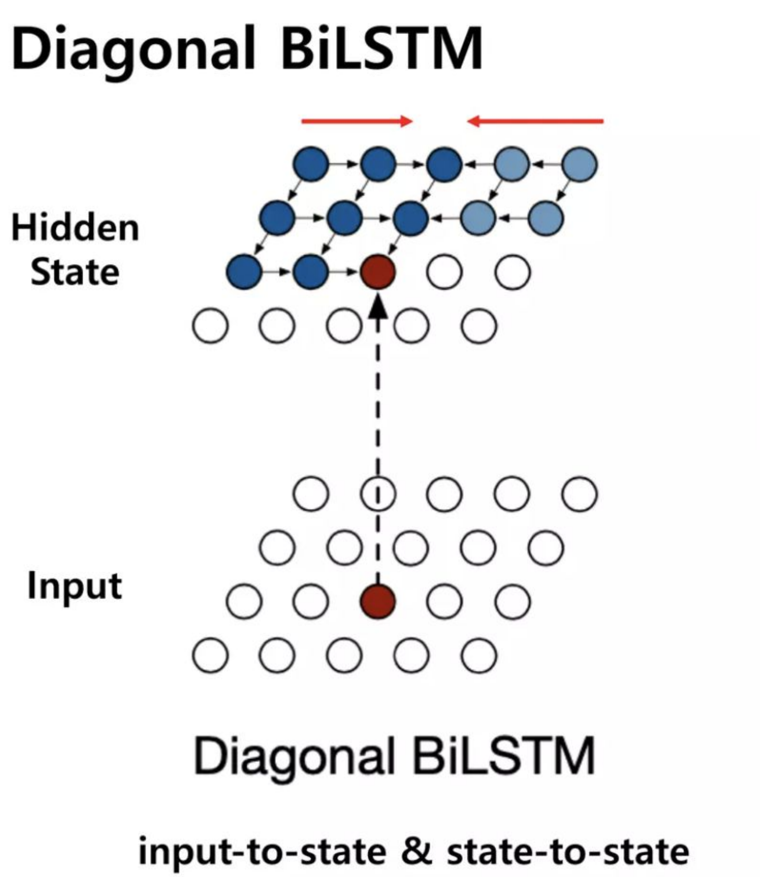


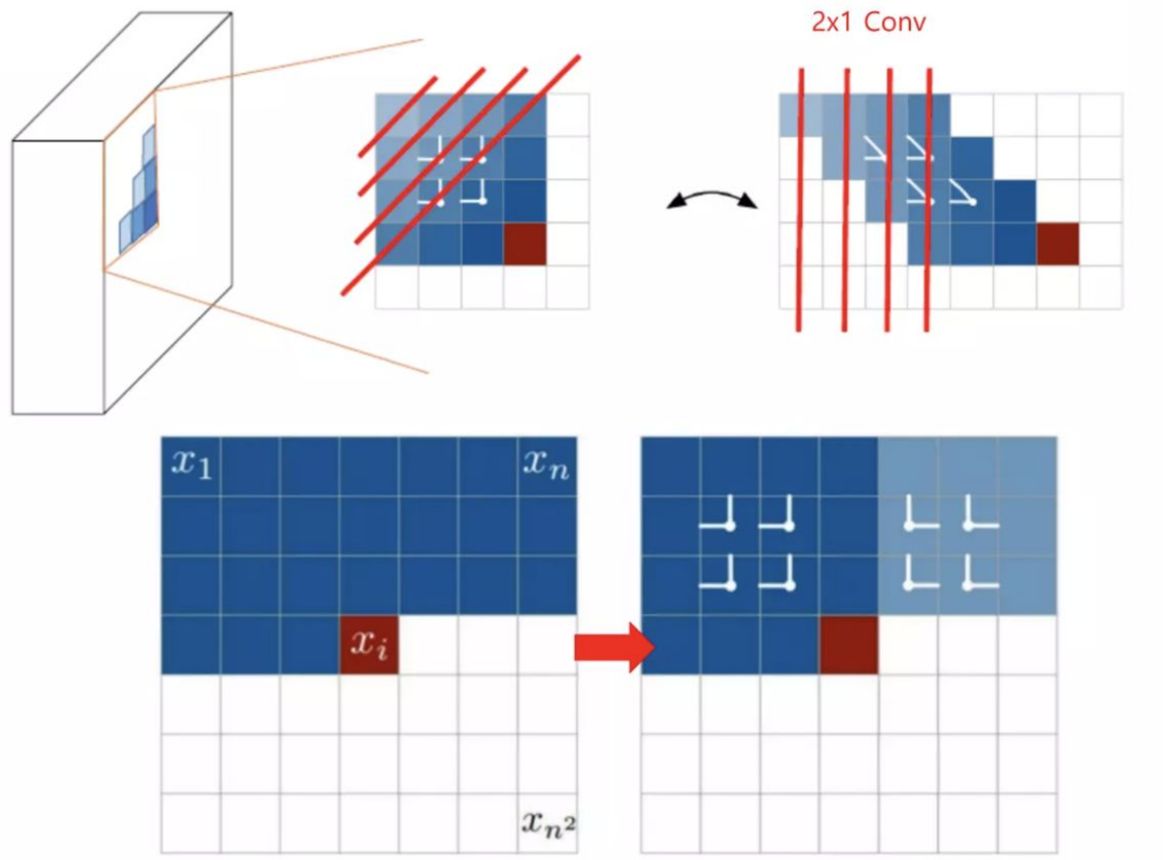
Multiplication → Convolution

$$[\mathbf{o}_i, \mathbf{f}_i, \mathbf{i}_i, \mathbf{g}_i] = \sigma(\mathbf{K}^{ss} \circledast \mathbf{h}_{i-1} + \mathbf{K}^{is} \circledast \mathbf{x}_i)$$
$$\mathbf{c}_i = \mathbf{f}_i \odot \mathbf{c}_{i-1} + \mathbf{i}_i \odot \mathbf{g}_i$$
$$\mathbf{h}_i = \mathbf{o}_i \odot \tanh(\mathbf{c}_i)$$

Various Available Models

Autoregressive Models - PixelRNN





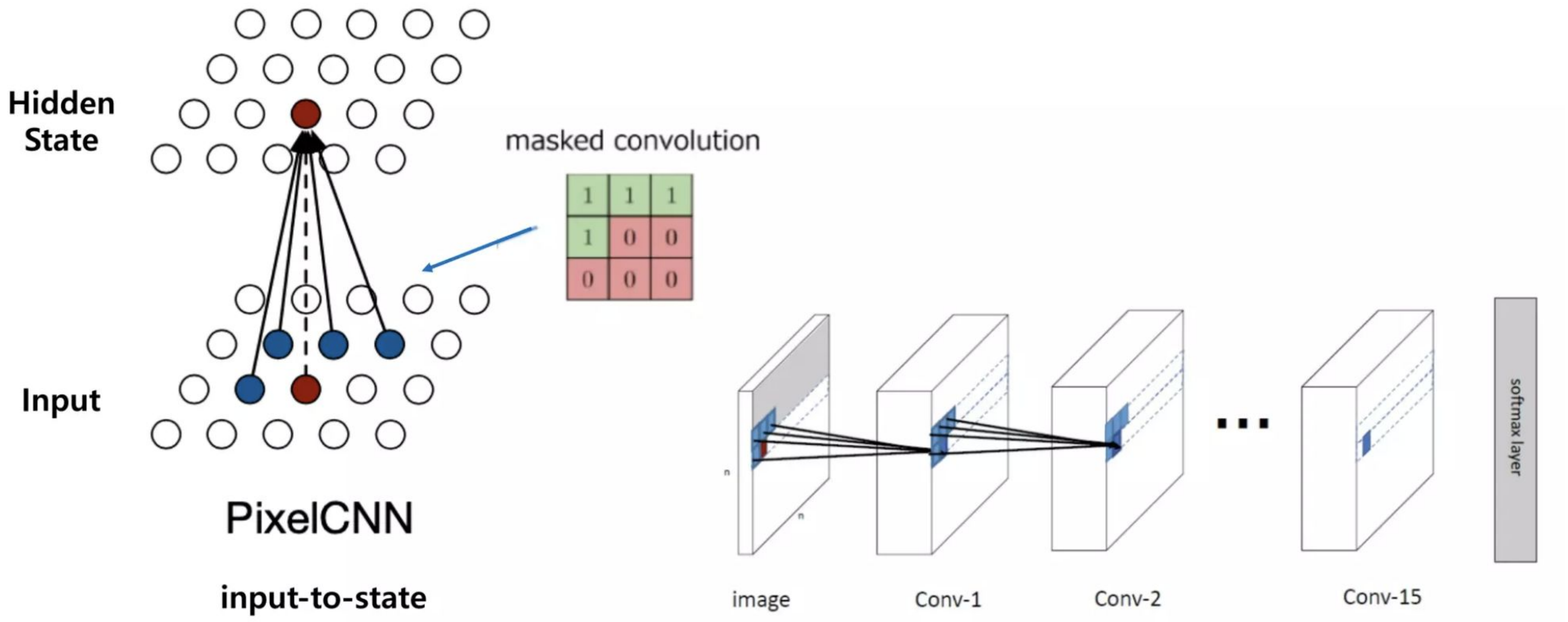
Various Available Models

Autoregressive Models - PixelCNN

- The main drawback of PixelRNN is that training is very slow as each state needs to be computed sequentially
- This can be overcome by using convolutional layers and increasing the receptive field
- PixelCNN uses standard convolutional layers to capture a bounded receptive field and compute features for all pixel positions at once
- It uses multiple convolutional layers that preserve the spatial resolution
- However, pooling layers are not used. Masks are adopted in the convolutions to restrict the model from violating the conditional dependence

Various Available Models

Autoregressive Models - PixelCNN



Various Available Models

Autoregressive Models

Disadvantages of Autoregressive Models:

- One of the main disadvantages is that it is computationally expensive and slow, since it requires generating each pixel sequentially and conditioning on all the previous pixels
- This makes it hard to scale up to large or high-resolution images, or to generate multiple images in parallel
- Another disadvantage is that autoregressive models can suffer from exposure bias and mode collapse, which means that they tend to generate images that are similar to the training data, and ignore some regions or modes of the data distribution

Various Available Models

Generative Adversarial Networks (GAN)

- Generative Adversarial Networks (GANs) are a powerful class of neural networks that are used for unsupervised learning
- GANs are made up of two neural networks, a discriminator and a generator
- They use adversarial training to produce artificial data that is identical to actual data
- The Generator attempts to fool the Discriminator, which is tasked with accurately distinguishing between produced and genuine data, by producing random noise samples
- Realistic, high-quality samples are produced as a result of this competitive interaction, which drives both networks toward advancement
- GANs are proving to be highly versatile artificial intelligence tools, as evidenced by their extensive use in image synthesis, style transfer, and text-to-image synthesis

Various Available Models

Generative Adversarial Networks (GAN)

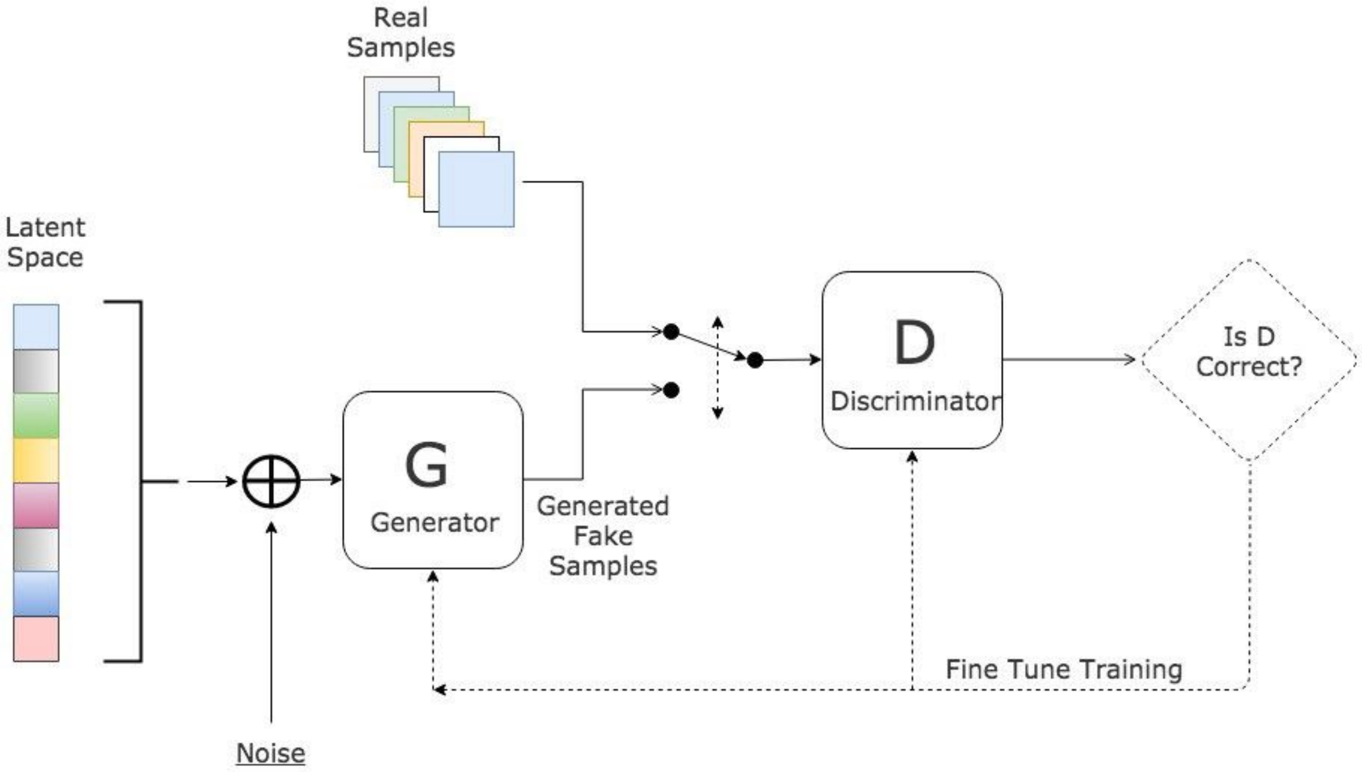
Generative Adversarial Networks (GANs) can be broken down into three parts:

- **Generative:** To learn a generative model, which describes how data is generated in terms of a probabilistic model
- **Adversarial:** The word adversarial refers to setting one thing up against another. This means that, in the context of GANs, the generative result is compared with the actual images in the data set. A mechanism known as a discriminator is used to apply a model that attempts to distinguish between real and fake images
- **Networks:** Use deep neural networks as artificial intelligence (AI) algorithms for training purposes

Various Available Models

Generative Adversarial Networks (GAN)

A Generative Adversarial Network (GAN) is composed of two primary parts, which are the Generator and the Discriminator



Various Available Models

Generative Adversarial Networks (GAN)

Types of GAN

DC GAN

It is a Deep convolutional GAN. It is one of the most used, powerful, and successful types of GAN architecture. It is implemented with help of ConvNets in place of a Multi-layered perceptron. The ConvNets use a convolutional stride and are built without max pooling and layers in this network are not completely connected

Conditional GAN

Conditional GAN is deep learning neural network in which some additional parameters are used. Labels are also put in inputs of Discriminator in order to help the discriminator to classify the input correctly and not easily full by the generator

Least Square GAN

It is a type of GAN that adopts the least-square loss function for the discriminator. Minimizing the objective function of LSGAN results in minimizing the Pearson divergence

Auxiliary Classifier GAN

It is the same as CGAN and an advanced version of it. It says that the Discriminator should not only classify the image as real or fake but should also provide the source or class label of the input image

Various Available Models

Generative Adversarial Networks (GAN)

Types of GAN

DVD GAN

DVD-GAN is a generative adversarial network for video generation built upon the BigGAN architecture. DVD-GAN uses two discriminators: a Spatial Discriminator and a Temporal Discriminator

SR GAN

Its main function is to transform low resolution to high resolution known as Domain Transformation

Cycle GAN

It is released in 2017 which performs the task of Image Translation. Suppose we have trained it on a horse image dataset and we can translate it into zebra images

Info GAN

Advance version of GAN which is capable to learn to disentangle representation in an unsupervised learning approach

Various Available Models

Generative Adversarial Networks (GAN)

Training of GAN

1. Define a Problem
2. Select Architecture of GAN
3. Train Discriminator on Real Data
4. Generate Fake inputs for Generator
5. Train Discriminator on Fake Data
6. Train Discriminator with the output of discriminator

Various Available Models

Generative Adversarial Networks (GAN)

Challenges with GAN:

- The problem of stability between generator and discriminator. We do not want that discriminator should be too strict, we want to be lenient
- Problem to determine the positioning of objects. suppose in a picture we have 3 horse and generator have created 6 eyes and 1 horse
- The problem in understanding the global objects – GANs do not understand the global structure or holistic structure which is similar to the problem of perspective. It means sometimes GAN generates an image that is unrealistic and cannot be possible
- A problem in understanding the perspective – It cannot understand the 3-d images and if we train it on such types of images then it will fail to create 3-d images because today GANs are capable to work on 1-d images


THANK YOU