Homework 2

In your own words, define the basic concept of the following neural networks and specify the main advantage of every network over the network before it. Moreover, define the types of problems (or data) that can be solved with each network:

1. MLP

The Multilayer Perceptron (MLP) is an artificial neural network that is used to learn complex relationships between inputs and outputs. It consists of an input layer, one or more hidden layers, and an output layer. The connections between the neurons in different layers are represented by learnable parameters called weights. MLP is particularly good at learning non-linear relationships between inputs and outputs. It is used for classification (e.g., image classification) and regression (e.g., housing price prediction) problems and can handle structured and unstructured data.

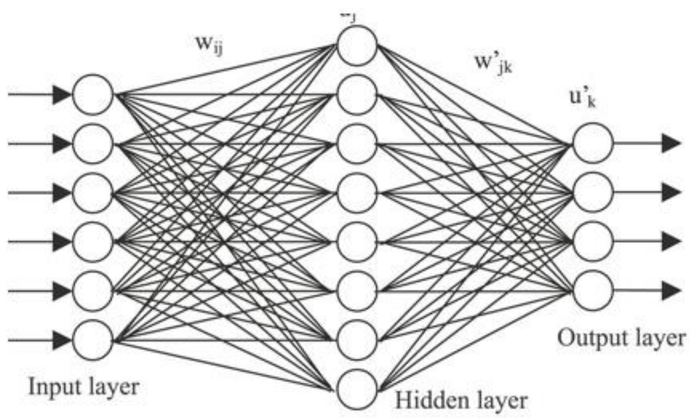

Figure 1 – MLP

2. CNN

A Convolutional Neural Network (CNN) is a type of artificial neural network that is widely used for image recognition tasks. CNNs use convolutional layers to apply a set of filters to the input data, which helps to extract important features. The learnable parameters in a CNN are the filter weights, also known as kernel weights.

One of the key advantages of CNNs is their ability to learn hierarchical representations of data. This means that the network can learn to identify complex features from simpler features, which allows it to recognize more complex patterns in the data. Another advantage of CNNs is that they are location invariant, which means that they can recognize the same pattern regardless of where it occurs in the input image. This is particularly useful in to adding some flexibility to the model and reduce the size of train dataset.

In Contract to NLP, CNNs have shared parameters which reduce the number of network parameters, making them more efficient to train and less prone to overfitting. Neurons in CNNs are only connected to a subset of neurons in the previous layer, unlike traditional neural networks where every neuron is fully connected.

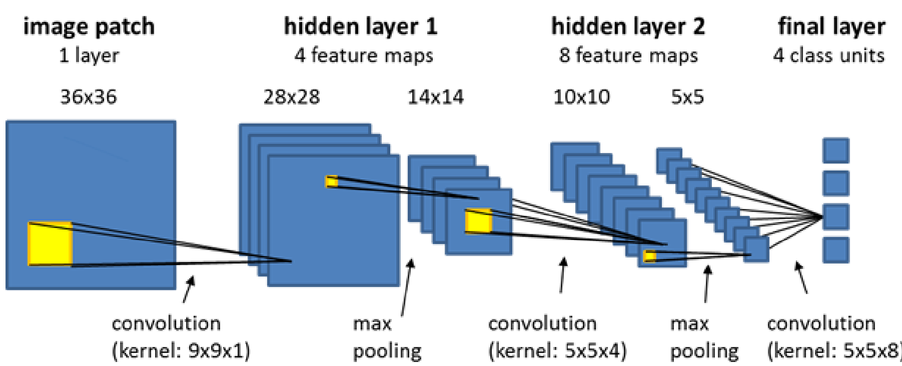

Figure 2 – CNN

3. RNN

RNN, short for Recurrent Neural Network, is a type of neural network that used on sequence data. Unlike other neural networks, RNNs have a built-in memory that allows them to process inputs in a sequential manner. This makes them ideal for tasks such as language translation, and stock prediction, where the output depends on a sequence of inputs.

While Convolutional Neural Networks (CNNs) are great at capturing local patterns in data, they have a limited ability to model long-term dependencies. This is because CNNs only consider a fixed-size window of input data at a time, which restricts their ability to capture the full context of a sequence. In contrast, RNNs can capture dependencies between inputs that are far apart in time.

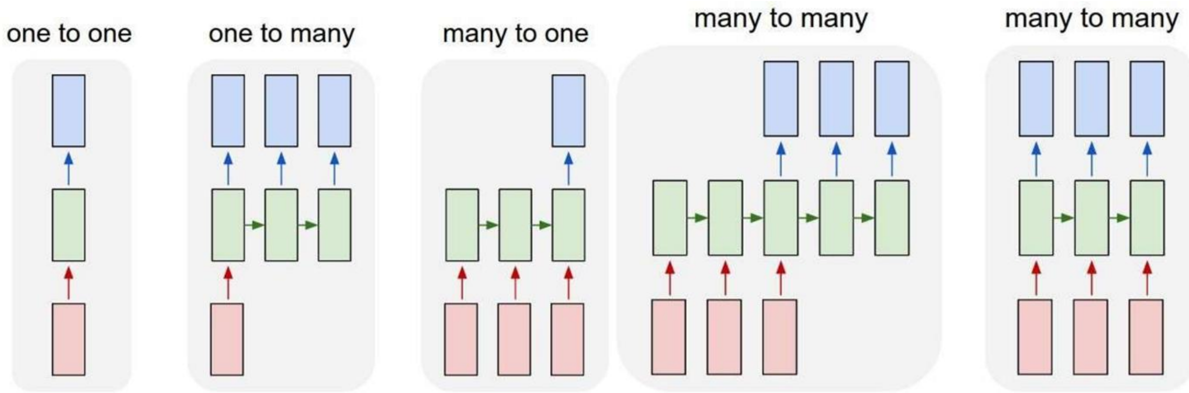

Figure 3- RNN

4. LSTM

One type of RNN is the Long Short-Term Memory (LSTM) network, which overcomes some of the weaknesses of traditional RNNs. One major weakness of RNNs is that they can suffer from the vanishing gradient problem. This means that RNNs can struggle to remember information from earlier in a sequence.

LSTMs address this issue by introducing a memory cell. The memory cell is divided into three parts: the forget gate, the input gate, and the output gate. The forget gate determines which information to discard from the memory cell, while the input gate decides which new information to store in the cell. The output gate controls how the information in the memory cell is used to generate the output. LSTMs are able to perform well on tasks that require memory, such as language translation and speech recognition.

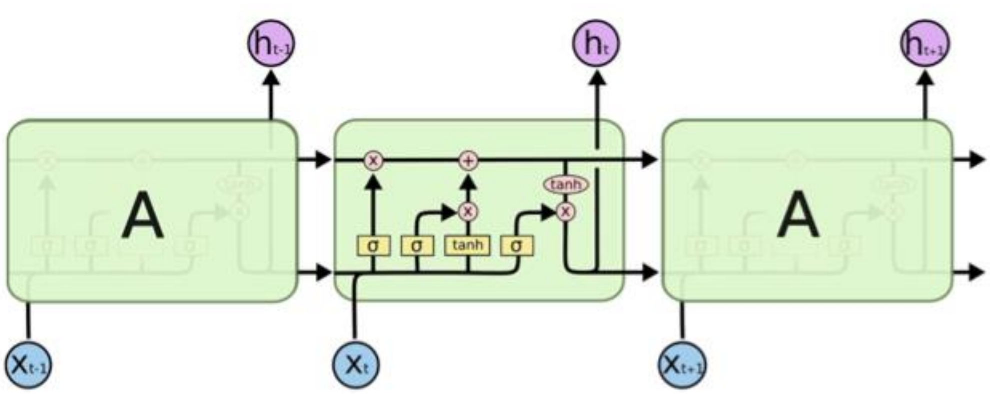

Figure 4- LSTM

5. GRU

Gated Recurrent Units (GRUs) are another type of recurrent neural network that share similarities with LSTMs, but have a simpler design. GRUs combine the forget and input gates into a single update gate, and merge the hidden and cell states into a single state vector. LSTMs and GRUs both provide solutions to the vanishing gradient problem and enable the modeling of long-term dependencies in sequence data such as language translation and speech recognition.

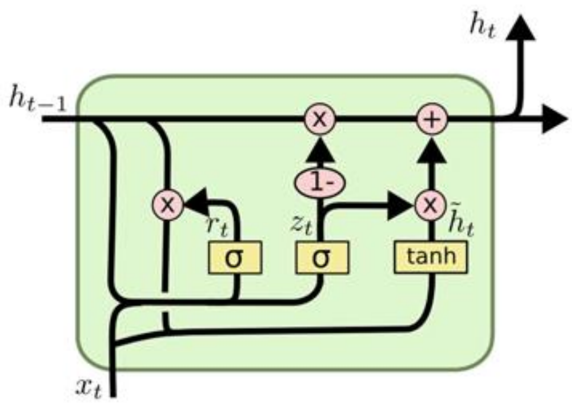

Figure 5- GRU

6. Transformer

The Transformer is a type of neural network architecture that addresses the problem of summarizing long input sequences using a single hidden state. In traditional recurrent neural networks, the last hidden state contains information about all previous inputs, which can lead to a loss of information and make it difficult to focus on the most relevant inputs.

The Transformer solves this problem by using a self-attention mechanism that allows the network to weight each input differently based on its relevance to the current output. This enables the network to focus more on the inputs that are most important for the output, and less on those that are less relevant. The main advantage of the Transformer over RNNs is that it is able to capture long-range dependencies in the input sequence without the need for recurrent connections. It is performance well in natural language processing problems, such as question answering.

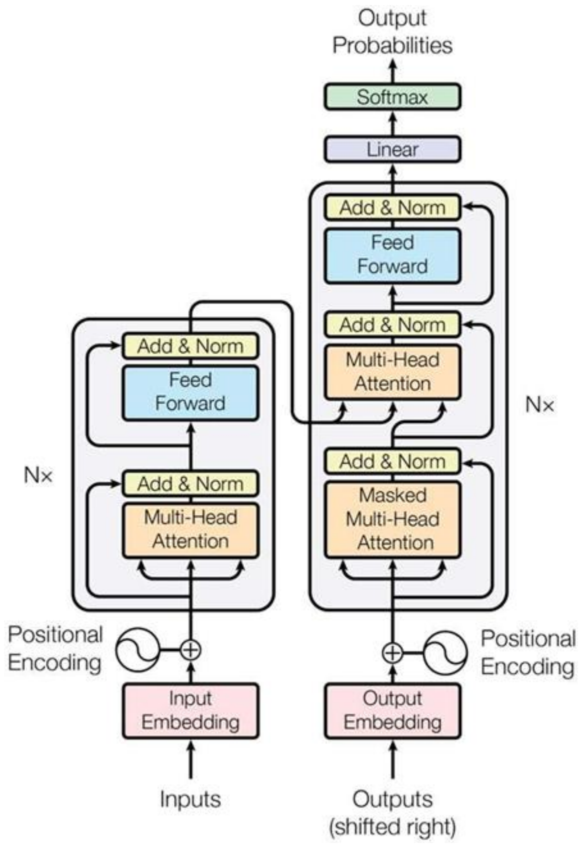

Figure 6- Transformer

7. Vision Transformer

The Vision Transformer is a type of neural network architecture that applies the Transformer's self-attention mechanism to image processing tasks. It divides a given input image into smaller patches, which are then fed into the Transformer encoder along with positional encodings.

The main advantage of Vision Transformer over convolutional neural networks (CNNs), it allows Vision Transformer to be more flexible and adaptable to different image sizes and resolutions, and can potentially reduce training time and computational resources required.

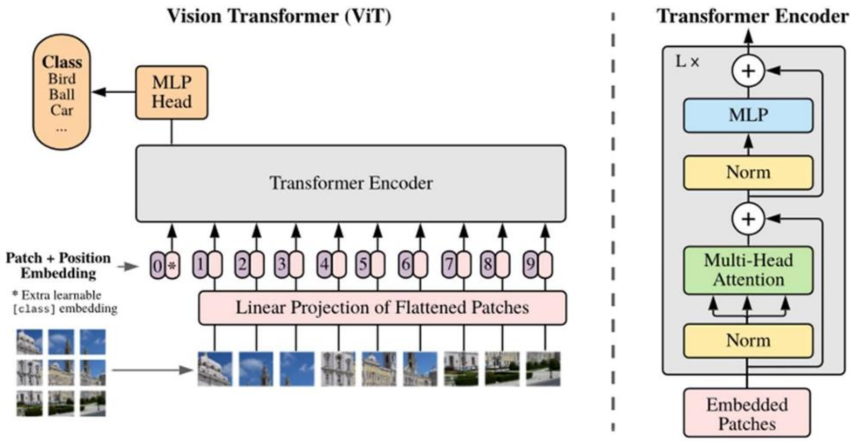

Figure 7- Vision Transformer

8. GCN

The Graph Convolutional Network (GCN) is a type of neural network architecture that is specifically designed to handle graph-structured data. GCN operates on graph-structured data by performing convolution-like operations that capture the relationships between nodes in the graph. Unlike traditional convolutional and recurrent neural networks, GCN is permutation invariant, meaning that the order of nodes in the input does not affect the output. On the other hand, MLP is not permutation nor translation invariant, and CNN is not permutation invariant nor rotation invariant. GCN has been applied successfully in many domains including social networks, biology, chemistry, and recommendation systems.

9. Autoencoder (AE)

The autoencoder is a type of neural network architecture that is used for **unsupervised** learning (no label there). It consists of two main components: an encoder and a decoder that reconstructs the original input from this representation. The goal of the autoencoder is to learn a compressed representation of the input that captures the most important features, while minimizing the difference between the original input and the reconstructed output.

Autoencoders have been successfully applied to a variety of data types, including images, text, and audio. In computer vision, autoencoders have been used for tasks such as image denoising, image compression, and image generation.

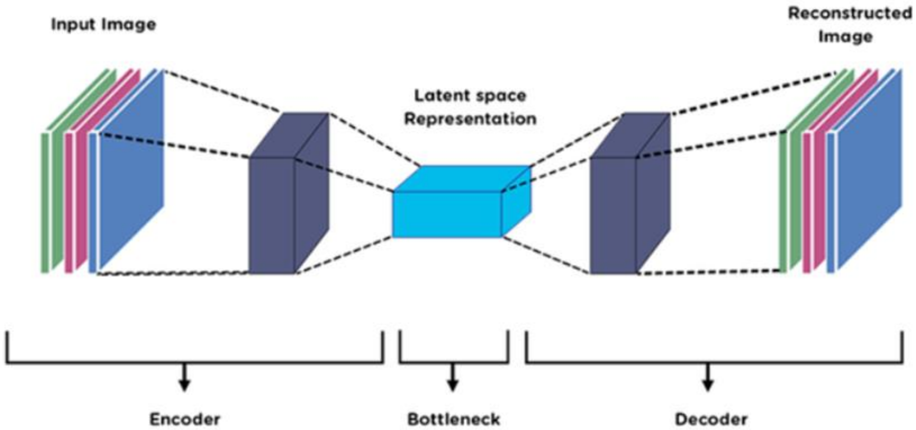

Figure 8- Autoencoder

10. VAE

Variational Autoencoders (VAEs) are a type of neural network architecture used for unsupervised learning (without labels on the data). An improper choice of input to the decoder can result in incorrect generation of data, which is a common problem in autoencoder-based generative models. The main advantage of autoencoders over other generative models is their ability to generate new data that is similar to the input data, while also minimizing the loss in the latent space and the error from the Gaussian mixture. Variational Autoencoders have been used for a wide range of applications, including image and speech recognition.

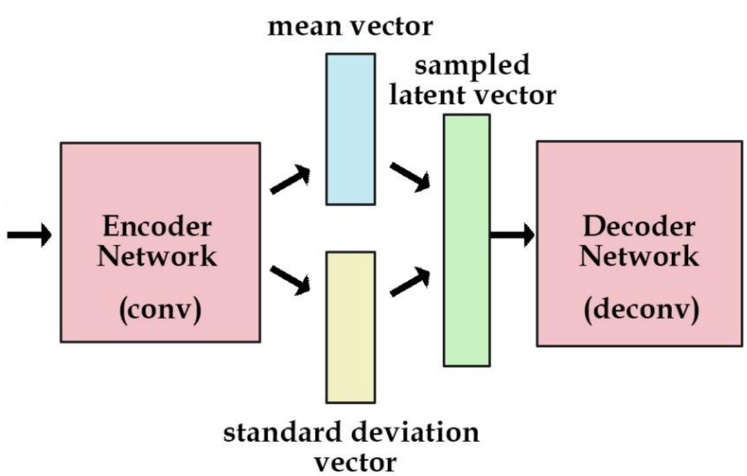

Figure 9- Variational Autoencoders

11. GAN

Generative Adversarial Networks (GANs) are a type of neural network architecture used for generative modeling. GANs use a two-part architecture consisting of a generator network and a discriminator network. The generator network generates synthetic data, while the discriminator network tries to distinguish between the synthetic and real data.

The main advantage of GANs over other generative models, such as Variational Autoencoders (VAEs), is their ability to generate more realistic data. GANs have been used for many types of the data such as test, image and so on.

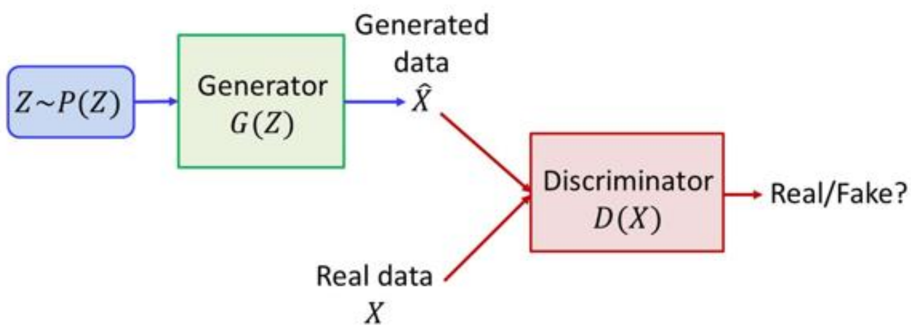

Figure 10- Generative Adversarial Networks