Convolutional Neural Network

S Narayanan

Contents

INTRODUCTION

- NEURAL NETWORK:

Neural networks are modeled after our brains. There are individual nodes that form the layers in the network, just like the neurons in our brains connect different areas.

There are three layers in a regular Neural Network:

- Input Layer: The layer in which we provide input to our model is known as the input layer. The entire number of features in our data (or the number of pixels in the case of a picture) is equal to the number of neurons in this layer.
- Hidden Layer: The hidden layer receives the input from the input layer. Depending on our model and the volume of the data, there may be numerous hidden levels. Each layer's output is calculated by multiplying the output of the layer below it by its learnable weights, adding learnable biases, and then computing the activation function, which makes the network nonlinear.
- Output Layer: After being passed into a logistic function like sigmoid or softmax, the output from the hidden layer is transformed into the probability score for each class.

Convolutional Neural Network

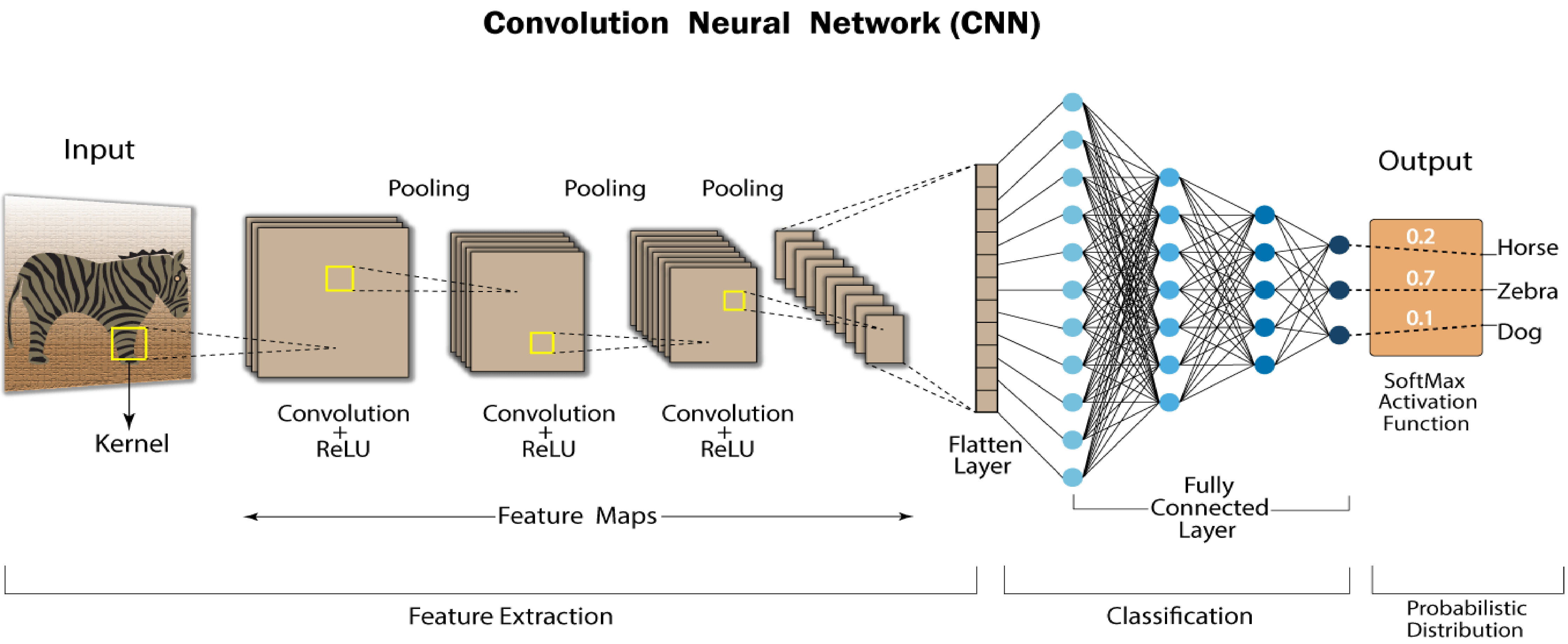
A Convolutional Neural Network (CNN) is a type of artificial neural network used primarily for image recognition and processing, due to its ability to recognize patterns in images.

A Convolutional Neural Network (CNN) is a type of deep learning algorithm that is particularly well-suited for image recognition and processing tasks. It is made up of multiple layers, including convolutional layers, pooling layers, and fully connected layers.

It has four layers:

- Convolutional layer
- Pooling layer
- The Activation layer
- Fully connected layer

Convolutional Neural Network



How CNN works?

1. The convolutional layers are the key component of a CNN
2. Filters are applied to the input image to extract features such as edges, textures, and shapes
3. The output of the convolutional layers is then passed through pooling layers
4. Used to down-sample the feature maps, reducing the spatial dimensions while retaining the most important information.
5. The output of the pooling layers is then passed through one or more fully connected layers, which are used to make a prediction or classify the image.

Layers in CNN

Convolutional layer:

- Its objective is to find a certain collection of features in the photographs supplied as input. Convolution filtering is used to do this.
- Convolutional neural networks' central element, the convolutional layer, is always the top layer.
- The basic idea is to "drag" a window representing the feature onto the image, calculate the convolution product between the feature and each section of the scanned image, and then apply the result to the entire image.
- The two concepts are equal in this context and a feature is then considered as a filter.

Layers in CNN

Pooling layer:

- The pooling operation involves shrinking the size of the photos while maintaining their crucial elements.
- This kind of layer, which receives many feature maps and applies the pooling operation on each of them, is frequently used between two layers of convolution.
- The image is divided into regular cells, and the maximum value is kept inside each cell. To prevent too much information loss, tiny square cells are frequently utilized in practice.
- The network's parameters and calculations are simplified by the pooling layer. By performing this, the network becomes more effective and overlearning is prevented.

Layers in CNN

Activation Layer:

- The "brain" of the CNN, or the activation function layer, is where the input is converted into a meaningful representation of the data. It is a necessary element.
- ReLU (Rectified Linear Units) refers to the real non-linear function defined by $ReLU(x)=max(0,x)$.
- The ReLU correction layer replaces all negative values received as inputs by zeros.
- It acts as an **activation function**.
- ReLU is the most widely activation function.
- Other examples are Sigmoid, Tanh and leaky ReLU

Layers in CNN

The fully-connected Layer:

- This kind of layer takes in a vector as input and creates a fresh vector as output. To accomplish this, it applies a linear combination to the input values received, followed optionally by an activation function.
- The association between the location of features in an image and a class is determined by the fully connected layer.
- The input table, which is the outcome of the preceding layer, corresponds to a feature map for a given feature: the high values represent the location of this feature in the image, which might be more or less exact depending on the pooling.
- A value in the table is given considerable weight if the location of a feature at a specific point in the image is indicative of a particular class.

Well-Known models of CNN

The most widely used CNN models:

- LeNet-5
- AlexNet
- VGG-16
- Inception-v1
- ResNet-50
- MobileNet

Well-Known models of CNN

ResNet:

- The issue of vanishing gradients, where the gradients become very small and training becomes s luggish, is one of the fundamental difficulties in training very deep neural networks.
- This issue is addressed by ResNet (Residual Neural Network), which Microsoft Research create d in 2015.
- ResNet uses the idea of "residual connections" to do so.
- A residual link is a shortcut connection that allows the gradient to flow directly to older layers b y ski-pping one or more layers.
- ResNet can now train far deeper networks than it could with earlier architectures without encou ntering the vanishing gradient issue.

Well-Known models of CNN

VGGNet:

- The Visual Geometry Group (VGG) at the University of Oxford created the convolutional neural network architecture known as VGGNet. It was first discussed by Karen Simonyan and Andrew Zisserman in a 2014 paper, and it was employed to triumph in the ILSVRC-2014 competition.
- VGGNet is renowned for being straightforward while still doing well in picture classification tasks. Convolutional and max-pooling layers are stacked on top of fully connected layers.
- The use of very tiny convolutional filters (3x3) with a very deep architecture (up to 19 layers) is the distinguishing feature of VGGNet.
- Later, this architecture underwent modifications and served as the foundation for a number of computer vision applications, including picture segmentation and object detection.

Well-Known models of CNN

MobileNet:

- Google created MobileNet in 2017 with the express purpose of being used on mobile and embedded devices with constrained processing capabilities.
- In order to decrease the number of parameters and computational expense while retaining high accuracy, MobileNet uses depth-wise separable convolutions.
- A depth-wise separable convolution combines the results of a single pointwise convolution (1x1 convolution) plus a single convolutional filter applied to each input channel. Compared to traditional convolutional layers, this method requires less parameters and computational resources.
- The lightweight and efficient design of the MobileNet architecture makes it ideal for deployment on mobile and embedded devices with constrained memory and processing power.

Why CNN?

Parameters:

- The number of parameters in a neural network grows rapidly with the increase in the number of layers.
- This can make training for a model computationally heavy (and sometimes not feasible).
- Tuning so many of parameters can be a very huge task.
- The time taken for tuning these parameters is diminished by CNNs.

Network:

- CNNs are fully connected feed forward neural networks.
- CNNs are very effective in reducing the number of parameters without losing on the quality of models.
- Images have high dimensionality (as each pixel is considered as a feature) which suits the above described abilities of CNNs.

Application of CNN

- Facial Recognition:

  A convolutional neural network separates facial recognition into the following key elements:

  - Recognizing each person in the photo despite other influences like light, angle, stance, etc., concentrating on each face.
  - Finding distinctive qualities
  - Matching a face with a name by comparing all the obtained data to data that is already present in the database.

Application of CNN

- Understanding Climate:
  - CNNs can be very useful in the fight against climate change, especially in figuring out why th ere are such significant shifts and what we might try to do about it.
  - It is claimed that the information in these collections of natural history can also help researche rs get deeper social and scientific insights, however this would call for experts to physically vi sit these arc-ives, such as researchers.
  - In order to conduct more extensive experiments in this area, more personnel are required.
- Advertising:
  - With the introduction of programmatic buying and data-driven personalized advertising, CNNs have already made a significant effect in the advertising industry.

Recurrent Neural Network(RNN)

- Recurrent neural networks (RNNs) are a type of neural network in which the results of one step are used as inputs for subsequent steps.
- The Hidden state of RNN, which retains some information about a sequence, is its primary and most significant characteristic. The state, which recalls the previous input to the network, is also known as memory state.

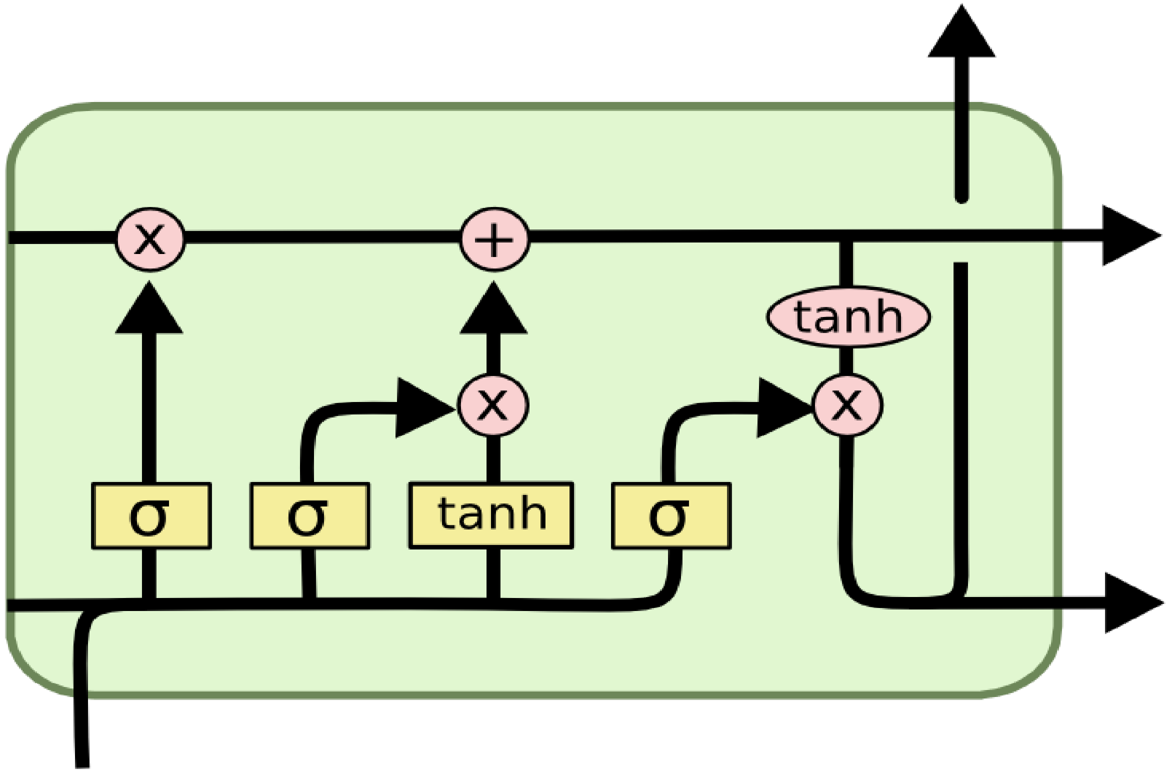
How it works?

  - The network receives the input in a single step.
  - Then, using a set of current inputs and the previous state, determine its present state.
  - For the following time step, the current time becomes time-1.Depending on the issue, one can travel back as many time steps and combine the data from all the prior states.
  - The final current state is used to determine the output after all the time steps have been finished.
  - The error is then generated once the output is compared to the goal output, which is the actual output.
  - The network (RNN) is trained using backpropagation(BPTT) through time after the mistake is back-propagated to it in order to update the weights.

Long Short-Term Memory(LSTM)

- Long Short-Term Memory (LSTM) are RNN’s that are especially made to handle sequential data, including time series, speech, and text. Sequential data can be used to teach LSTM networks long-term dependencies.
- It tackled the problem of long-term dependencies of RNN in which the RNN cannot predict the word stored in the long-term memory but can give more accurate predictions from the recent information.
- RNN does not give an efficient performance. LSTM can by default retain the information for a long period of time.
- There are three gates that help in information retention and memory manipulation:
    - **Forget Gate:** The information that is no longer useful in the cell state is removed with the forget gate
    - **Input gate:** The addition of useful information to the cell state is done by the input gate.
    - **Output gate:** The task of extracting useful information from the current cell state to be presented as output is done by the output gate.

Long Short-Term Memory(LSTM)



RNN vs LSTM

RNN	LSTM
Standard RNNs might be difficult to train to address issues that call for learning long-term temporal dependencies.	A "memory cell" found in LSTM units is capable of storing data in memory for extended periods of time. They can learn longer-term dependencies thanks to this memory cell.
Information isn't kept in the memory of an RNN for very long.	Information is kept in the memory for a very long time by LSTM.
The overall training speed of RNNs is relatively low compared to feedforward networks.	LSTM is comparatively faster than RNN
Training RNNs on very long sequences is challenging, especially when using ReLU or tanh activations.	LSTM was designed to have better performances on long sequences problems.
RNN is much more suitable for short sequences.	LSTM is suitable for long sequences and performance dips for short sequences.