DEEP LEARNING AND OBJECT DETECTION

What is Deep Learning?

Deep learning is a machine learning technique that teaches computers to do what comes naturally to humans. Deep learning is a key technology behind driverless cars enabling them to recognize a stop sign, or to distinguish a pedestrian from a lamppost.
Deep learning is one of the only methods by which we can overcome the challenges of feature extraction. This is because deep learning models are capable of learning to focus on the right features by themselves, requiring little guidance from the programmer. Basically, deep learning mimics the way our brain functions i.e. it learns from experience.Our brain is made up of billions of neurons that allows us to do amazing things. Even the brain of a one year old kid can solve complex problems which are very difficult to solve even using supercomputers.

Actually, our brain has sub-consciously trained itself to do such things over the years. Now, the question comes, how deep learning mimics the functionality of a brain? Well, deep learning uses the concept of artificial neurons that functions in a similar manner as the biological neurons present in our brain. Therefore, we can say that Deep Learning is a subfield of **machine learning** concerned with algorithms inspired by the structure and function of the brain called artificial neural networks.

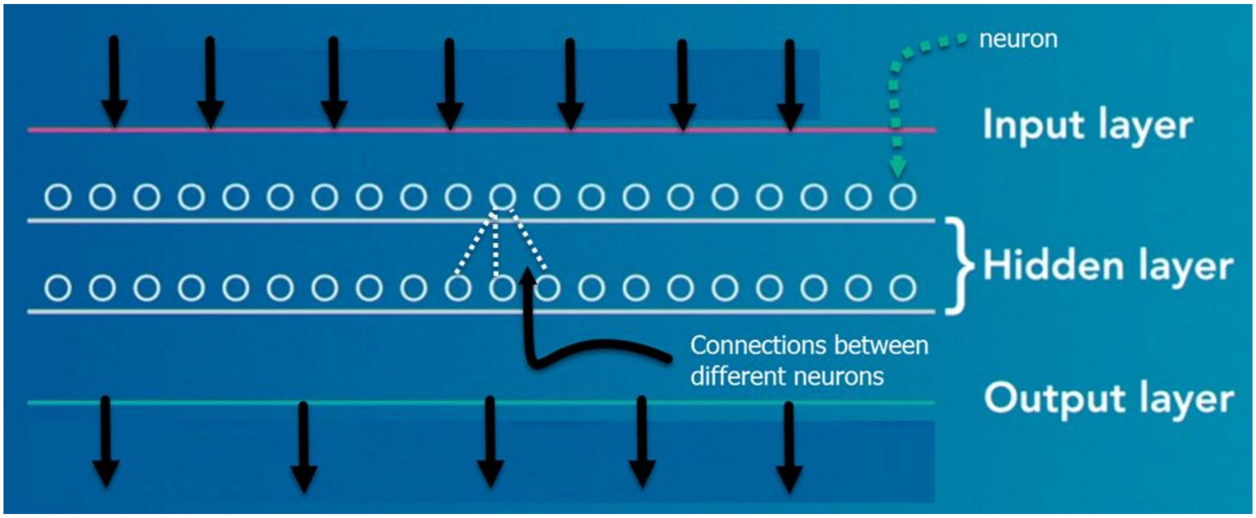
Now, deep learning takes this one step ahead. Deep learning automatically finds out the features which are important for classification because of deep neural networks, whereas in case of Machine Learning we had to manually define these features.

Deep learning algorithms are constructed with connected layers.

- The first layer is called the Input Layer
- The last layer is called the Output Layer
- All layers in between are called Hidden Layers. The word deep means the network join neurons in more than two layers

Each Hidden layer is composed of neurons. The neurons are connected to each other. The neuron will process and then propagate the input signal it receives the layer above it. The strength of the signal given the neuron in the next layer depends on the weight, bias and activation function.

The network consumes large amounts of input data and operates them through multiple layers; the network can learn increasingly complex features of the data at each layer.



Examples of Deep Learning at Work

Deep learning applications are used in industries from automated driving to medical devices.

- Automated Driving: Automotive researchers are using deep learning to automatically detect objects such as stop signs and traffic lights. In addition, deep learning is used to detect pedestrians, which helps decrease accidents.
- Aerospace and Defense: Deep learning is used to identify objects from satellites that locate areas of interest, and identify safe or unsafe zones for troops.
- Medical Research: Cancer researchers are using deep learning to automatically detect cancer cells. Teams at UCLA built an advanced microscope that yields a high-dimensional data set used to train a deep learning application to accurately identify cancer cells.
- Industrial Automation: Deep learning is helping to improve worker safety around heavy machinery by automatically detecting when people or objects are within an unsafe distance of machines.
- Electronics: Deep learning is being used in automated hearing and speech translation. For example, home assistance devices that respond to your voice and know your preferences are powered by deep learning applications.

Google is trying to take their self-driving car initiative, known as WAYMO, to a whole new level of perfection using Deep Learning. Therefore, rather than using old hand-coded algorithms, they can now program system that can learn by themselves using data provided by different sensors. Deep learning is now the best approach to most perception tasks, as well as to many low-level control tasks. Hence, now even people who do not know how to drive or are disabled, can go ahead and take the ride without depending on anyone else

With the help of Deep Learning MIT is trying to predict the future.

How deep learning works?

Suppose we have three students each of them write down the digit nine on a piece of paper. Notably, they don’t all write it identically. The human brain can easily recognize the digits but what if a computer had to recognize them? That’s where deep learning comes in. A neural network could be trained to identify handwritten digits. Each number is represented as an image of 28×28 pixels. That amounts to a total of seven hundred and eighty-four pixels.

A neuron is the core entity of the network where the information processing takes place. Each of the 784 pixels is fed to a neuron in the first layer of our neural network. This forms the input layer.

On the other hand, we have the output layer with each neuron representing a digit with a hidden layer existing between them. The information is transferred from one layer to another over connecting channels. Each of these channels has a value attached to it and hence is called a weighted channel. All neurons have a unique number associated with it called bias. The bias is added to the weighted sum of inputs reaching to the neuron which is then applied to a function known as the activation function.

The result of the activation function determined if the neuron activated. Every activated neuron passes on information to the following layers. This continues until the second last layer. One neuron activated in the output layer corresponds to the input digit.

OBJECT DETECTION

Object Detection is the process of finding real-world object instances like car, bike, TV, flowers, and humans in still images or Videos. It allows for the recognition, localization, and detection of multiple objects within an image which provides us with a much better understanding of an image as a whole. It is commonly used in applications such as image retrieval, security, surveillance, and advanced driver assistance systems (ADAS).

Object Detection can be done via multiple ways:

- Feature-Based Object Detection
- Viola Jones Object Detection
- SVM Classifications with HOG Features
- Deep Learning Object Detection

Applications of Object Detection

Facial Recognition

A deep learning facial recognition system called the “DeepFace” has been developed by a group of researchers in the Facebook, which identifies human faces in a digital image very effectively. Google uses its own facial recognition system in Google Photos, which automatically segregates all the photos based on the person in the image. There are various components involved in Facial Recognition like the eyes, nose, mouth and eyebrows.

People Counting

Object detection can be also used for people counting, it is used for analyzing store performance or crowd statistics during festivals. These tend to be more difficult as people move out of the frame quickly. It is a very important application, as during crowd gathering this feature can be used for multiple purposes.

Industrial Quality Check

Object detection is also used in industrial processes to identify products. Finding a specific object through visual inspection is a basic task that is involved in multiple industrial processes like sorting, inventory management, machining, quality management, packaging etc.

Inventory management can be very tricky as items are hard to track in real time. Automatic object counting and localization allows improving inventory accuracy.

Self Driving Cars

Self-driving cars are the Future, there's no doubt in that. But the working behind it is very tricky as it combines a variety of techniques to perceive their surroundings, including radar, laser light, GPS, odometry, and computer vision. Advanced control systems interpret sensory information to identify appropriate navigation paths, as well as obstacles and once the image sensor detects any sign of a living being in its path, it automatically stops. This happens at a very fast rate and is a big step towards Driverless Cars.

Security

Object Detection plays a very important role in Security. Be it face ID of Apple or the retina scan used in all the sci-fi movies. It is also used by the government to access the security feed and match it with their existing database to find any criminals or to detect the robbers' vehicle.

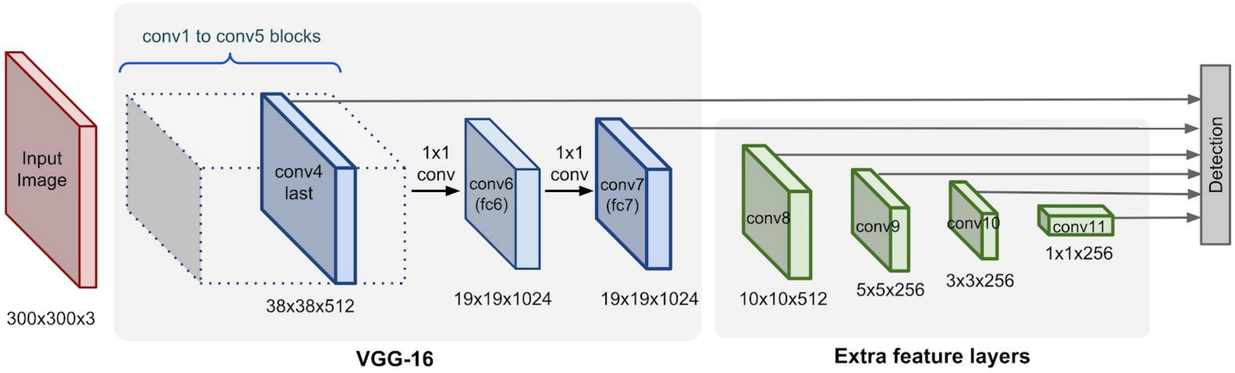
SSD (Single Shot MultiBox Detector) Model for Object Detection :

The SSD object detection composes of 2 parts:

1. Extract feature maps, and
2. Apply convolution filters to detect objects.

- SSD runs a convolutional network on input image only once and calculates a feature map. Now, we run a small 3×3 sized convolutional kernel on this feature map to predict the bounding boxes and classification probability. SSD also uses anchor boxes at various aspect ratio similar to Faster-RCNN and learns the off-set rather than learning the box. In order to handle the scale, SSD predicts bounding boxes after multiple convolutional layers. Since each convolutional layer operates at a different scale, it is able to detect objects of various scales.

- SSD uses the VGG-16 model pre-trained on ImageNet as its base model for extracting useful image features. On top of VGG16, SSD adds several conv feature layers of decreasing sizes. They can be seen as a pyramid representation of images at different scales. Intuitively large fine-grained feature maps at earlier levels are good at capturing small objects and small coarse-grained feature maps can detect large objects well. In SSD, the detection happens in every pyramidal layer, targeting at objects of various sizes.



Faster R-CNN

Faster R-CNN was developed by researchers at Microsoft. It is based on R-CNN which used a multi-phased approach to object detection. R-CNN used Selective search to determine region proposals, pushed these through a classification network and then used an SVM to classify the different regions.

Although it is a single unified model, the architecture is comprised of two modules:

- Module 1: Region Proposal Network. Convolutional neural network for proposing regions and the type of object to consider in the region.
- Module 2: Fast R-CNN. Convolutional neural network for extracting features from the proposed regions and outputting the bounding box and class labels.

Both modules operate on the same output of a deep CNN. The region proposal network acts as an attention mechanism for the Fast R-CNN network, informing the second network of where to look or pay attention.

YOLO(You Only Look Once)

The YOLO framework (You Only Look Once) on the other hand, deals with object detection in a different way. It takes the entire image in a single instance and predicts the bounding box coordinates and class probabilities for these boxes. The biggest advantage of using YOLO is its superb speed – it's incredibly fast and can process 45 frames per second. YOLO also understands generalized object representation.

YOLO divides up the image into a grid of 13 by 13 cells: Each of these cells is responsible for predicting 5 bounding boxes. A bounding box describes the rectangle that encloses an object. YOLO also outputs a confidence score that tells us how certain it is that the predicted bounding box actually encloses some object.

Prior detection systems repurpose classifiers or localizers to perform detection. They apply the model to an image at multiple locations and scales. High scoring regions of the image are considered detections.

YOLO uses a totally different approach. It applies a single neural network to the full image. This network divides the image into regions and predicts bounding boxes and probabilities for each region. These bounding boxes are weighted by the predicted probabilities.

Why GPU Matters in Deep Learning?

1. Every set of weights can be stored as a matrix (m,n).

2. GPUs are made to do common parallel problems fast. All similar calculations are done at the same time. This extremely boosts the performance in parallel computations.

What are the limitations of deep learning:

Deep learning has a vast scope but it faces some limitations.

- The first is as we discussed earlier is data. While deep learning is the most efficient way to deal with unstructured data, neural networks require a mass volume of data to train.
- Let’s assume we always have access to the necessary amount of data. Processing this is not within the capability of every machine and that brings us to our second limitation, computational power. Training a neural network requires graphical processing units which have thousands of cores as compared to CPUs and GPUs, of course, are more expensive.
- Finally, we come down to training time. Deep neural networks take hours or even months to train. The time increases with the amount of data and number of layers in the network

References:

https://www.csetutor.com/what-is-deep-learning-and-how-is-it-useful/
https://www.guru99.com/deep-learning-tutorial.html#2
https://skymind.ai/wiki/restricted-boltzmann-machine
https://dzone.com/articles/understanding-object-detection-using-yolo