Deep Learning is a subset of Machine Learning which is essentially a neural network with several layers. Deep Learning seeks to reproduce the functioning of a human brain by making the algorithm learn through a large amount of data. [1]

Convolutional Neural Network

Convolutional neural networks (CNN) are constituted by several layers and are mainly used for image classification and object recognition tasks. The layers of CNN treat and extract the data characteristics. CNN have three main types of layers: convolutional layer, pooling layer, and fully-connected layer. [2]

The convolutional layer is the first layer of the CNN. It contains filters that perform the convolution operation on the input data. The convolution operation consists in moving the filter across the receptive fields of the image to check if the feature is present. The filter is only applied to an area of the image. Then, a dot product is calculated between the input pixels and the filter and gives an output array. The process is repeating until the filter has swept across the entire image. The final output obtains from the series of dot products is called the feature map.

This map supplies the pooling layer. Pooling is a subsampling operation that reduces the dimensions of the feature map. Then, the clustering layer converts the resulting two-dimensional arrays from the clustered feature map into a single, long, continuous linear vector by flattening them.

The pooling operation sweeps a filter across the entire input. This filter applies an aggregation function to the values within the receptive field, filling the output array. They are two main types of pooling. In the max pooling, the filter moves across the input and selects the pixel with the maximum value to send to the output array. While in the average pooling, the filter calculates the average value within the receptive field to send to the output array.

The fully-connected layer performs the task of classification based on the features extracted through the previous layers. Each input is connected to the neurons allowing images to be classified and identified. The full-connected layer usually uses a softmax activation function to classify inputs by producing a probability from 0 to 1.

With each layer, the CNN increases in its complexity, identifying greater portions of the image. Earlier layers focus on simple features such as colors and edges for example.

As the image data progresses through the layers of the CNN, it starts to recognize larger elements or shapes of the object until it finally identifies the intended object. [3]

YOLO Algorithm

You Only Look Once (YOLO) is a Deep Learning algorithm that provides real-time object detection. It is popular because of its speed, accuracy, and its learning capabilities. It employs convolutional neural networks (CNN) to detect and recognizes the objects. As the name suggests, the algorithm requires only a single forward. [4]

The YOLO algorithm uses the Residual blocks technique. The image is divided into various grid. The grid cells have equal dimension and detect objects that appear within them. Then, the algorithm creates bounding boxes to highlight different objects in an image. Each bounding box has attributes: the width, the height, the class (for example, pedestrian, car …) and the bounding box center.

Another technique used by Yolo algorithm is the intersection over union (IOU). IOU describe how boxes overlap and provide an output box that surrounds the objects perfectly to Yolo.

Each grid cell allows to predict the bounding boxes and their confidence scores. The IOU is equal to 1 if the predicted bounding box is the same as the real one and it eliminates the bounding boxes that are not equal to the real box. [5]

Une image contenant texte, intérieur, différent, arrangé

Description générée automatiquement

[1] What Is Deep Learning?

<https://www.ibm.com/cloud/learn/deep-learning>

[2] « Quels sont les algorithmes de deep learning ? » Mobiskill, 26 mai 2021, <https://mobiskill.fr/blog/conseils-emploi-tech/quels-sont-les-algorithmes-de-deep-learning/>

[3] What Are Convolutional Neural Networks?

<https://www.ibm.com/cloud/learn/convolutional-neural-networks>

[4] Gandhi, Rohith. « R-CNN, Fast R-CNN, Faster R-CNN, YOLO — Object Detection Algorithms ». Medium, 9 juillet 2018, <https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e>

[5] « Introduction to YOLO Algorithm for Object Detection ». Engineering Education (EngEd) Program | Section, <https://www.section.io/engineering-education/introduction-to-yolo-algorithm-for-object-detection/>