IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

BASICS OF CONVOLUTION NEURAL NETWORK

Prachi Gadhire,Dr. Farhat Jummani
Research scholar, Assistant Professor
Computer department, [1]JJT university, Rajasthan ,India

ABSTRACT:

In recent years, deep learning has drawn the attention of scholars. Convolutional Neural Networks (CNN) are a deep learning technique that are frequently employed for resolving challenging issues. It gets around the restrictions of conventional machine learning techniques. This study was undertaken with the goal of educating readers on numerous CNN-related topics. In addition to describing CNN's three most used topologies and learning methods, this article also provides a conceptual explanation of the network. This study will encourage scholars to pursue this area of research by giving them a thorough understanding of CNN. For individuals who are interested in this subject, this study will serve as a resource and quick reference.

Keywords: Convolution Neural Network, Deep Neural Network

Introduction

In the past few years, deep learning has emerged as an area of interest to researchers due to the quickly expanding demand for learnable machines to solve numerous complicated issues. The subject of how humans learn is crucial because researchers often imitate human behaviour. The name "machine learning" was created because it refers to the process of incorporating learning, a fundamental human trait, into machines. By using three different learning methods, including learning under supervision, learning without supervision, and semi-supervised learning, machine learning claims to reduce the amount of work required. The requirement for traditional machine learning algorithms is feature extraction, which calls for a domain specialist. Additionally, choosing the right features for a particular situation is a difficult task. Deep learning techniques solve the feature selection issue by automatically extracting the important characteristics for a given problem from raw input rather than requiring pre-selected features. A deep learning model is

made up of several processing layers that may learn different data features at different degrees of abstraction. The network can learn several degrees to recognise different features. Deep learning has evolved as a method for achieving promising outcomes in a variety of applications, including facial recognition, subject classification, sentiment analysis, language translation, and small molecule bioactivity prediction. There are various deep learning architectures, including convolutional neural networks, recurrent neural networks, and deep belief networks.

Convolution Neural Network (CNN), Often referred to as ConvNet, it has a deep feed-forward architecture and an astoundingly superior capacity to generalise than networks with fully connected layers. CNN is defined as the idea of biologically inspired hierarchical feature detectors. It has the ability to learn extremely abstract traits and effectively recognise objects. The following are some of the significant justifications for why CNN is preferred to other traditional methods. The idea of adopting the concept of weight sharing, which significantly reduces the number of parameters that need training and improves generalisation, is the primary motivation for applying CNN. Less parameters provide for a smoother training process and prevent overfitting in CNN. Second, both the feature extraction step and the classification stage involve the learning process. Thirdly, creating big networks utilising general models of artificial neural network (ANN) is significantly more challenging than implementing in CNN. Due to their impressive performance in areas including image classification, object detection, face detection, speech recognition, vehicle recognition, diabetic retinopathy, facial expression recognition, and many more, CNNs are widely employed in a variety of fields. The purpose of this study is to develop a theoretical framework while enhancing knowledge and comprehension of CNN. The goal of this study is to summarise the fundamental concepts of CNN while also offering information on the general model, the three most popular designs, and learning algorithms. Additionally, the details of a novel learning method called ADAM have been revealed. Additionally, it calculates the learning rate for each distinct parameter. The sections are organised completely as follows.

2. General Model Of Convolution Neural Network

2.1Standard Model The conventional ANN model has numerous hidden layers in addition to a single input and output layer. A specific neuron receives input vector X and generates an output vector Y by applying some function F to it, which is represented by general equation (1) shown below. (1)

$$F(X, W) = Y \qquad (1)$$

where W stands for the weight vector, which symbolises how strongly neurons in two adjacent layers are connected. The weight vector that was created can now be utilised to classify images. The classification of images based on pixels has been extensively studied in the literature. Contextual information, such as the image's shape, gives better results or

outperforms, nonetheless. CNN is a model that is attracting interest due to its ability to classify data based on context. Figure 1 below describes the CNN model in general. Convolution layer (a), pooling layer (b), activation function (c), and fully connected layer (d) make up a standard CNN model. Below is an illustration of each component's functionality.

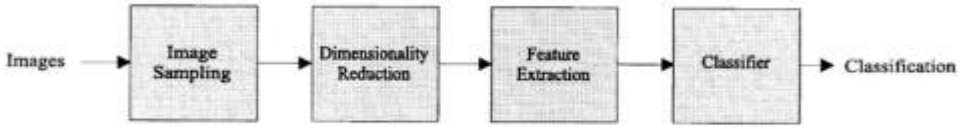

Figure 1: Elementary constituents of CNN [13]

2.2. Convolution Layer: The input layer receives a picture that needs to be classed, and the predicted class label is produced using features that were retrieved from the image. The local association between a single neuron in the subsequent layer and some neurons in the preceding layer is known as the receptive field. Utilizing receptive field, the local features from the input image are retrieved. A weight vector, which relates to the neurons in the next layer, is formed by the receptive field of a neuron connected with a specific location in the preceding layer. Similar features occurring at many points in the input data can be identified since the neurons in the plane share the same weights. This is illustrated in figure.

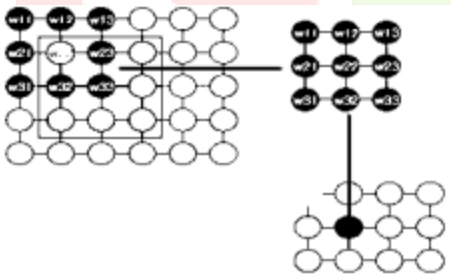

Figure 2: Receptive field of particular neuron in the next layer

The feature map is produced by sliding the weight vector, sometimes referred to as the filter or kernel, across the input vector. Convolution operation is the term used to describe the process of sliding the filter both horizontally and vertically. This procedure creates N filters and N feature maps by extracting N features from the input image and placing them on a single layer that represents each different feature. There are much fewer trainable parameters because of the local receptive field phenomena. Following the convolution procedure, the output aij in the following layer for position (i,j) is calculated using the formula indicated below:

$$a_{ij} = \sigma((W * X)_{ij} + b) \qquad (2)$$

where X represents the input sent to the layer, W represents a filter or kernel that slides over the input, b represents the bias, * represents the convolution operation, and represents the nonlinearity introduced in the network.

Layering Pools Once a feature has been identified, its precise placement is less important. Thus, the pooling or sub-sampling layer comes after the convolution layer. Utilizing the pooling strategy has the significant advantages of introducing translation invariance and drastically reducing the number of trainable parameters. Figure 3 illustrates how to perform a pooling operation by selecting a window and then passing the input components contained within that window via a pooling function.

Figure 3: Pooling operation performed by choosing a 2 x 2 window

Another output vector is produced by the pooling function. There are only a few pooling methods, such as average pooling and max-pooling, however max-pooling is the most used method and considerably reduces map size. Because it does not participate in the forward flow while computing mistakes, the error does not back-propagate to the winning unit.

2.4. Fully Connected Layer

The fully connected network in traditional models is analogous to the fully connected layer. The fully connected layer receives the result of the first phase, which comprises repetitive convolution and pooling, and computes the dot product of the weight vector and input vector to produce the final output. Gradient descent, sometimes referred to as batch mode learning or the offline approach, lowers the cost function by estimating the cost throughout the full training dataset. It changes the parameters only once every epoch, or complete traversal of the training dataset. Although it produces global minima, the size of the training dataset has a significant impact on how long it takes to train the network. Stochastic gradient descent was used to replace this method of decreasing the cost function.

2.5. Activation Function

The usage of the sigmoid activation function in traditional machine learning methods is extensively documented in the literature. Rectified Linear Unit (ReLU) use has proven to be superior to the former when it comes to introducing non-linearity for two main reasons. First off, it is simple to calculate the partial derivative of ReLU. However, a big gradient that is flowing through the network reduces ReLU efficiency, and an update in weight prevents activation of the neuron, which results in the Dying ReLU problem, a significant difficulty that is frequently experienced. Leaky

ReLU can handle this problem; if x>0, the function activates as f(x)=x and if x0, the function deactivates.

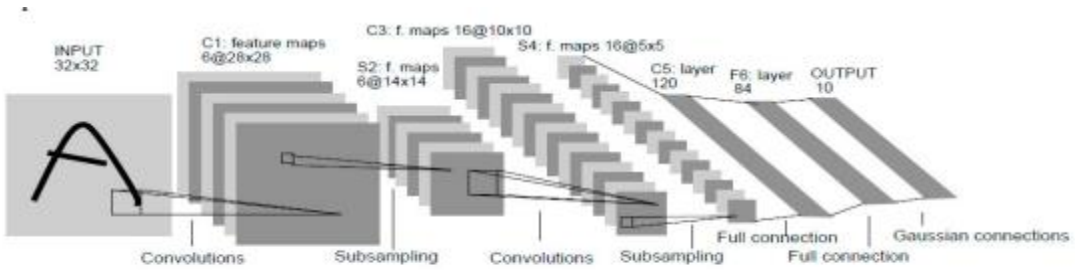

Figure 4 : Architecture of LeNet5, a CNN where each box represents a different feature map.[16]

The adjacent units in the feature map are the outcome of the adjacent units in the preceding layer. Contiguous receptive fields overlap as a result of this. Convolution layer is the first layer, and it is made up of neurons that produce sigmoid activation when applied to the weighted sum. As seen in figure 4, if a 5 5 area is selected as an input and a horizontal shift on the area is carried out, it will result in the overlap of four rows and five columns while computing contiguous units in the feature map. Different feature maps are produced as a result of applying various weight vectors to the same input image. The produced feature maps can be used to extract various features. CNN's crucial characteristic is that small change . The fundamental characteristic of CNN is that a small change in the input has no impact on the feature map. Since it is not important to have a precise position for a feature in an image, subsampling is done to lower the precision value. Figure 4 illustrates how sub-sampling is represented in the second layer. The amount of feature maps obtained through subsampling is equal to that of feature maps obtained through convolution. Here, the average of the four inputs has been computed for the sub-sampling layer 2 2 area, multiplied by the trainable coefficient, added the trainable bias, and then sent to the sigmoid function. As the spatial resolution is reduced layer by layer, an increase in the number of feature maps may be seen. The back propagation approach is used to carry out the learning.

3.2. AlexNet Architecture This piece provides a succinct description of the AlexNet architecture, a modified version of LeNet introduced by [12]. It has five convolution layers, three fully connected layers, and three partial connected layers. The outputs are sent to a 1000-way softmax, which divides 1.2 million high resolution images into 1000 different classes.

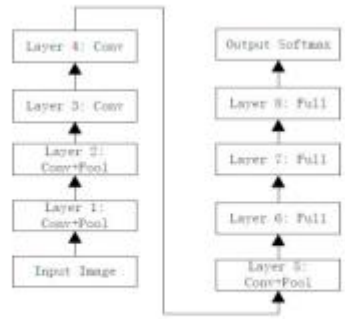

Figure 5: AlexNet architecture [19]

To train the network more quickly, non-saturating neurons and effective GPU implementation are used. As a result, a huge network is needed to enable the network to classify objects from millions of photos, which may ultimately result in a high demand for training a very large number of weights and the overfitting issue. This issue was solved by using the dropout method. In this method, the neurons that have a probability of 0.5 are damped and do not participate in forward or backward propagation. Overfitting is significantly reduced since the neurons that depend on these damped neurons must learn the most robust features entirely on their own. The dropout method doubles the amount of iterations needed to converge. It takes five to six days to train the network using two GTX 580 3GB GPUs. The main characteristics of this design included the addition of ReLU non-linearity in CNN, which caused the convergence rate to rise quickly.

3.3. GoogleNet Architecture It was suggested to use the GoogleNet paradigm. After winning the ILSVRC14 competition, it became enlightened. The main objective was to create a model with a smaller budget that could consume less memory, power, and trainable parameters. The number of trainable parameters employed in the network was dramatically decreased by the model. The following is a general description of the architecture. In essence, it employs 12 million fewer parameters than the model put forward by. The architecture attempted to create a network that could more accurately identify the items in an image. . This might be accomplished by expanding the network's size, which would increase the number of layers, but a key negative of this idea was that doing so would increase the number of parameters that would need to be trained, which would create the issue of overfitting. Another significant drawback is that as the number of filters increases, so does the computation, which raises overhead. Implementing a sparse matrix was the suggested approach. In order to create the best network topology, the highly correlated units join to form a cluster in the previous layer and send input to the next layer.. . Even though the computations are sped up by 100 times, the overhead of cache misses still exists when using the non-uniform sparse matrix. Using highly optimised numerical libraries to achieve faster computations is also ineffective. As a result, the state of the art relies on uniform sparse matrices.

Conclusion

In comparison to shallow networks, deep learning has the major advantage of being able to independently identify pertinent characteristics in high dimensional data. There is enough information available on several deep learning approaches, including CNN, deep belief networks, and recurrent neural networks. This study has clarified the fundamental principles of CNN, a deep learning method used to resolve several challenging issues. The general CNN model, different architectures, and two significant learning methods have all been covered in this paper. CNN has become a well-known method for classification based on contextual data. It has a tremendous capacity for learning contextual cues, and by doing so, has solved the difficulties associated with pixel-by-pixel categorization. It drastically minimises the number of parameters needed. In remote sensing, ocean front recognition, high-resolution data, traffic sign recognition, audio scene, and segmenting MR brain pictures, CNN is frequently utilised for classification. Researchers who are interested in exploring this area will benefit greatly from this work. It will serve as a resource for students, researchers, and anybody with an interest in this area.

References

[1] Arel, I., Rose, D. C., and Karnowski, T. P. (2010) “Deep machine learning-a new frontier in artificial intelligence research [research frontier].” IEEE computational intelligence magazine 5 (4): 13-18.

[2] Carbonell, J. G., Michalski, R. S., and Mitchell, T. M. (1983) “An overview of machine learning. In Machine learning.” Springer Berlin Heidelberg (pp. 3-23).

[3] Chen, H., Tang, Y., Li, L., Yuan, Y., Li, X., & Tang, Y. (2013) "Error analysis of stochastic gradient descent ranking." IEEE transactions on cybernetics 43 (3): 898-909.

[4] Dharmadhikari, S. C., Ingle, M., and Kulkarni, P. (2011) “Empirical studies on machine learning based text classification algorithms.” Advanced Computing 2 (6): 161.

[5] Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., and Darrell, T. (2014) “Decaf: A deep convolutional activation feature for generic visual recognition.” In International conference on machine learning (pp. 647-655).