



A Survey Paper on Convolutional Neural Network

¹Ekta Upadhyay, ²Ranjeet Singh, ³Pallavi Upadhyay

^{1,1}Department of Information Technology,

^{2,1}Department of Computer Science Engineering

^{3,1}Department of Information Technology,

Buddha Institute of Technology, Gida, Gorakhpur, India

Abstract

In this era use of machines are growing promptly in every field such as pattern recognition, image, video processing project that can mimic like human cerebral network function and to achieve this Convolutional Neural Network of deep learning algorithm helps to train large datasets with millions of parameters of 2d image to provide desirable output using filters. Going through the convolutional layer then pooling layer and in last fully connected layer, Images becomes more effective how many times it filters become better than other.

In this article we are going through the basic of Convolution Neural Network and its working process.

Keywords-Deep Learning, Convolutional Neural Network, Handwritten digit recognition, MNIST, Pooling.

1- Introduction

In the world of technology deep learning has become one of the most aspect in the field of machine. Deep learning which is sub-field of Artificial Learning that focuses on creating large Neural Network Model which provides accuracy in the field of data processing decision. social media apps like Instagram, twitter Google, Microsoft and many other apps with million users having multiple features like face recognition, some apps for handwriting recognition. Which is machine learning problem to recognize clearly[3], as we know machines are man-made and machines does not have minds or visual cortex to understands or see the real word entity, so to understand the real world in 1960 human create a theorem named Convolutional neural network of deep learning algorithm that make machine much more understandable.

Convolutional Neural Network of deep learning is a type of feed forward Artificial Neural Network in which the connectivity pattern between its neurons is inspired by organization animal visual cortex. "Visual cortex processes information concerning from pattern and motion within functional maps that reflect the layout of the neural network circuits. [3]

In Deep learning Convolutional Neural Network is a class of deep neural network most commonly applied to analyze visual imagery. CNN uses a special technique called Convolution.

1.1 Convolution

Convolution is mathematical operation of function that produces a third function that expresses how one variable is dependent on another one. The Convolution refers to the result function and how to compute the function.

But in the context of a convolutional neural network a convolution is a linear operation that involves multiplications of a set of weights with the input much like a traditional neural network. Given that the

technique was designed for two-dimensional input, the multiplication is performed between an array of input and a two-dimensional array of weights, called a filter or a kernel.

The filter is smaller than the input data and the type of multiplication applied between a filter-sized patch of the input and the filter is a dot product. A dot product is the element-wise multiplication between the filter-sized patch of the input and filter, which is then summed.

1.2- Neural Network

Neural network is system of hardware and/or software patterned after the operation of neurons in the human brain.

2- History

In 1950- 1960 two scientists named David Hubel and Torsten Wiesel find that neurons inside the cat visual cortices response to small visual field. They find that the reason of visual space within the visual stimuli known as receptive field affect the attack of a single neurons. So the experiments starts with the biological field, Hubel and Wiesel proposed for visual pattern recognition living being have two types of cells one is *simple cell* and another is *Complex cell*. These cells become the the root of *deep neural network* or *convolutional neural network*.

- Complex cells which have larger field whose output is intensive to the exact position of the age in the field.
- simple cells which have whose output is maximized by state is having particular orientation within their respective field.[Wikipedia]

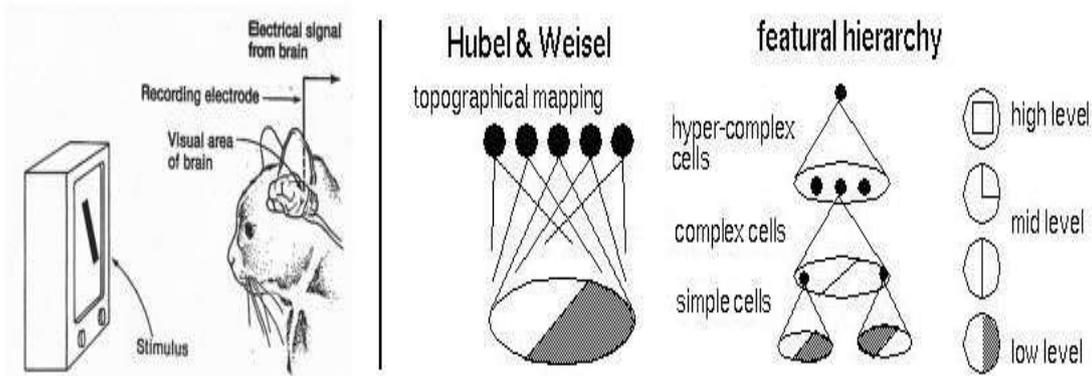


Fig 1- Image from Wikipedia

2.2- Kunihiko Fukushima's Neo-cognitron (Origin of the Convolutional Architecture)

After the Hubel and Wiesel discovery Kunihiko Fukushima inspired and introduced an artificial neural network that mimics the functioning of simple and complex cells. He proposed a network having an input layer with modular structure of 2 cascading layers of cells, in which the first layer of each model consists of “S-cells” shows essential quality similar to simple cell while the another is “C-cells” that is complex cell shows essential quality higher-order hypercomplex cell. A self-organized network that follow the path of “learning without supervisor” and hence has capability to learn incentive patterns based on the geometrical similarity of shapes without caring about their position.

So, the network work on unsupervised learning so it need not any supervisor or teacher to learn. And this was the beginning of Convolutional neural network [2] The conclusion of Fukushima’s experiment was that machine can learn without any problem having any distortion or shift in position of images. And introduced to the CNN.

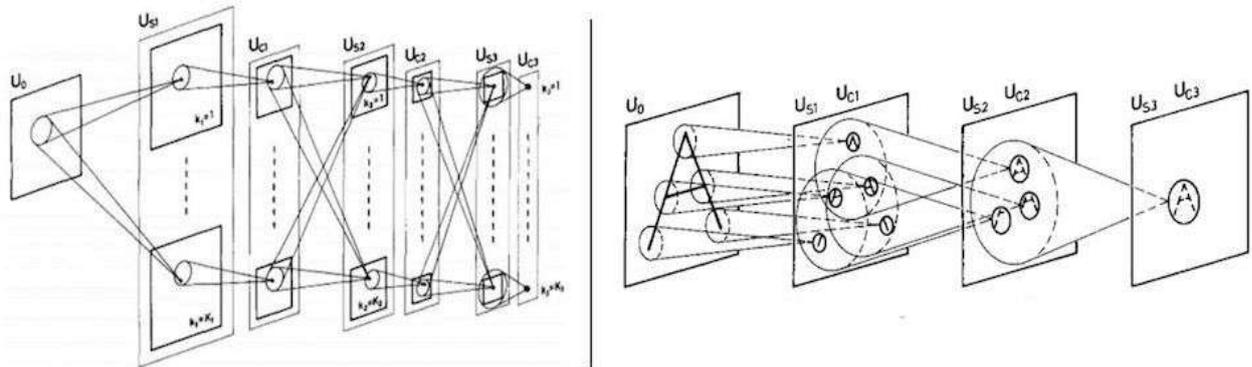


Fig 2-Image by Fukushima

2.3- The LeNet Aka CNN

While the Kunihiko Fukushima's Neocognitron was a successful model in the field of artificial intelligence, a newly modern application was introduced by Yann LeCun et al in 1990s, which recognize the handwritten digits. He used the methodology of MNIST (Modified National Institute of Standards and Technology Database) to train the CNN. As CNN need large datasets to train database so, MNIST contains 60,000 training images and Under 10,000 testing images. Half of the training set and half of the test set is taken from MNIST training data set. RGB codes of handwritten digits in figure-4 given below.[3]

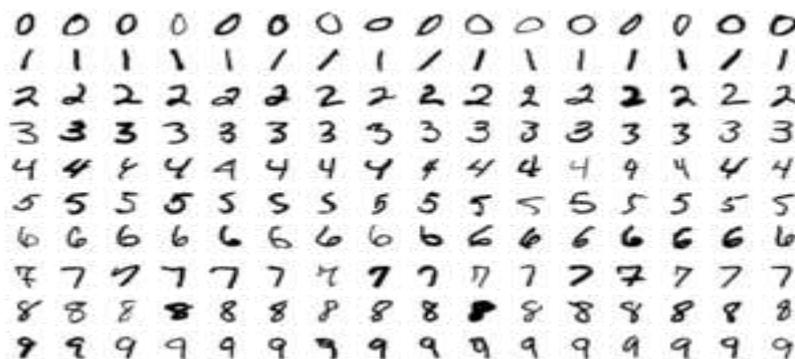


Fig 4- Example handwritten digits from the MNIST Datasets

Also in the original paper, the creator uses a support vector machine (SVM) to get an error rate of 0.8 %. An extended data set similar to a MNIST (pronounced "EM-nisst) has been published in 2017, Still at that time this technique provides low resolution of images.

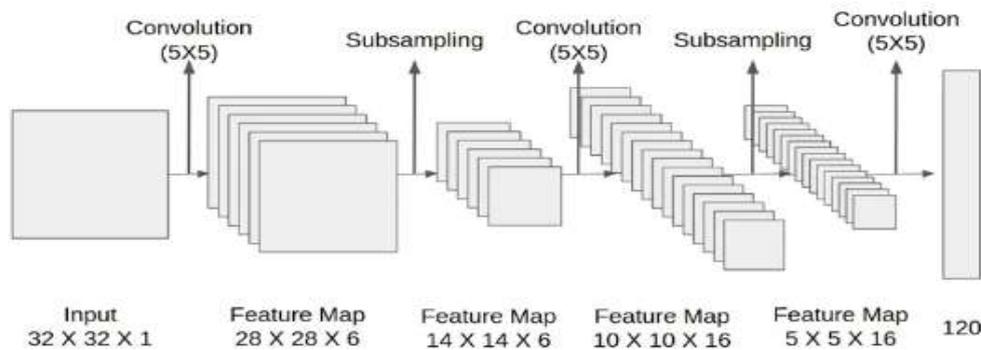


Fig 4- Image from Wikipedia

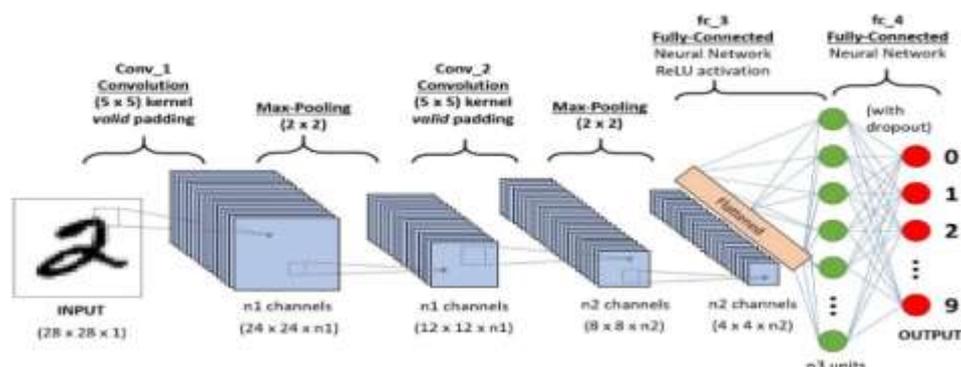
2.4- From The 1990s To Onwards

After the Kunihiko Fukushima's Neocognitro modern neural network the scientist or researchers start working further on the CNN model which provide a lot of enhancements and one day in the field of pattern and image recognition become successful.

After a time around 2012 Convolutional Neural Network was on another level a new deep convolutional neural network architecture known as AlexNex was introduced (according to an article) with 60% error rate that is 10% lower than the runner-up by utilizing GPS special neural network of AlexNex that hardly mimic the human version now over and over is CNN has become more popular and it continuous growing, now let us see how see works and its algorithm. Currently, ImageNet datasets starts from 2010 collaborating with the PSCAL VOC team contains 14,197,122 images in 1000 different object classes is using.

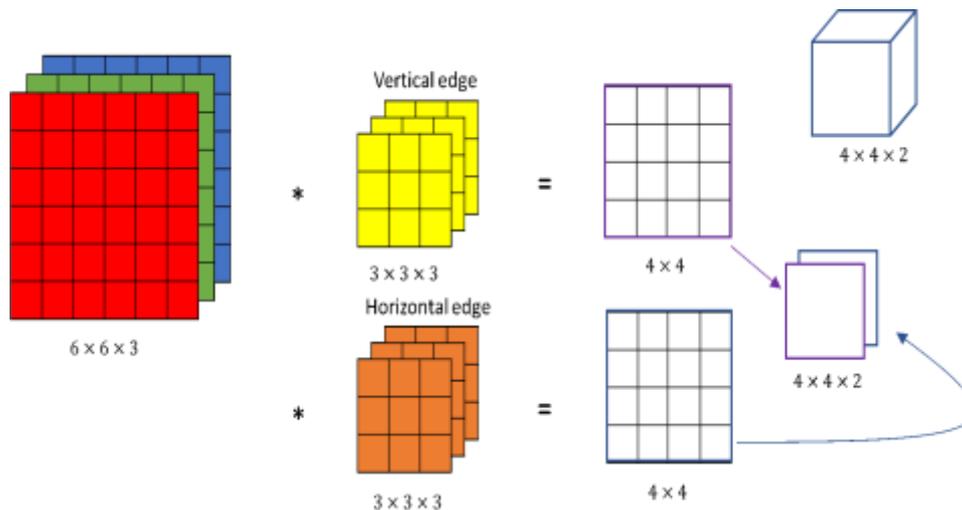
3- Architecture of Convolutional Neural Network

There are three type of layer that make the convolutional neural network which are convolutional layer, pooling layer, and fully connected layer . When these layers are stack together a CNN architecture will be formed. Convolutional neural network first takes inputs which are images goes through the filters which consists of multiple layers of neurons. In addition to these three layers, there are two more important parameters which are the dropout layer and the activation function which perform in last.



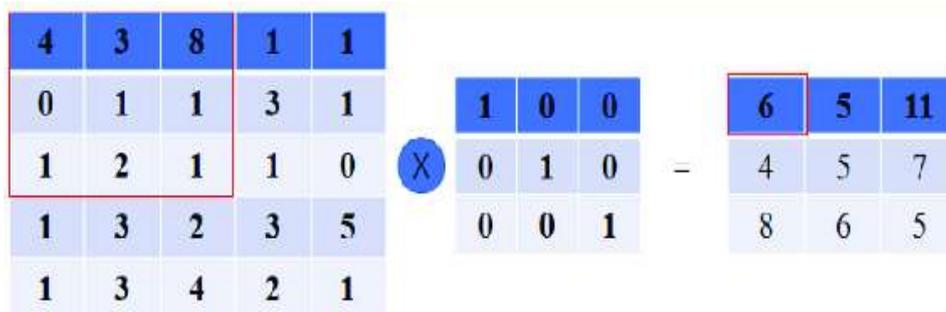
3.1- Input Layer

The input to a CNN is mostly an image which contain RGB as shown in the figure...



3.2- Convolution Layer

Here, we basically define filters and we compute the convolution between the defined filters and each of the 3 images.



In above figure, we are *building another image* by applying the filter to our input image. Note that depending on the filter we apply we will get a different image.

In the same way we apply to remaining (above is for red image, then we do same for green and blue) images. We can apply more than one filter. More filters we use, we can preserve spatial dimensions better.

We use convolution instead of considering flatten image as input as we will end up with a massive number of parameters that will need to be optimized and computationally expensive.

3.3- Pooling

Similar to the Convolutional Layer, the Pooling layer is responsible for reducing the spatial size of the Convolved Feature. This is to **decrease the computational power required to process the data** through dimensionality reduction. Furthermore, it is useful for **extracting dominant features** which are rotational and positional invariant, thus maintaining the process of effectively training of the model.[4]

There are two types of Pooling:

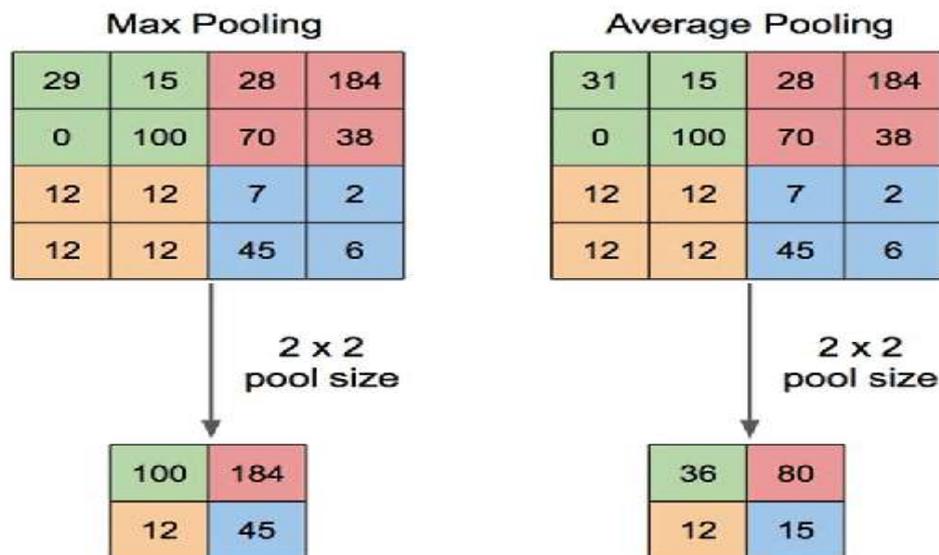
3.1 Max Pooling: Max Pooling returns the **maximum value** from the portion of the image covered by the Kernel.

3.2 Average Pooling: Average Pooling returns the **average of all the values** from the portion of the image covered by the Kernel.

Max Pooling also performs as a **Noise Suppressant**. It discards the noisy activation altogether and also performs DE-noising along with dimensionality reduction. On the other hand, Average Pooling simply performs dimensionality reduction as a noise suppressing mechanism. Hence, we can say that **Max Pooling performs a lot better than Average Pooling**.

The Convolutional Layer and the Pooling Layer, together form the i-th layer of a Convolutional Neural Network. Depending on the complexities in the images, the number of such layers may be increased for capturing low-levels details even further, but at the cost of more computational power.

After going through the above process, we have successfully enabled the model to understand the features. Moving on, we are going to flatten the final output and feed it to a regular Neural Network for classification purposes.



When these processes were stacked together going through activation functions provides the desired output.

3.4- Fully Connected Layer

After pooling layer fully connected layer takes input as output of pooling layer.

5- Conclusion

The new beginning in the field of deep neural network has started with the minor stage facing higher challenges growing through step-by-step with Convolutional Neural Network has become highly successful. Conventionally, the starting CovLayer was responsible for capturing low level features such as edges, color, gradient decent etc., With added layers, the architecture adapts to higher-Level features as well giving us a network, which has the wholesome understanding of images in the datasets.

The experiment started with the Cells that is S-cells and C-cells for Handwritten recognition and goes through MNIST (Modified National Institute of Standards and Technology Database) database having training images and testing images which improves purity level to 73% and accordingly increases.



6- References

- 1- LeNet by Rachel Lea Ballantyne Drraelos "History of convolutional neural network" in Apr 13, 2019
- 2- FukuFukush's Neocognitton by Kunihiko Fukushima "Neocognitron: A self-organizing Neural Network Model for a Machan8sm of Pattern Recognition Unaffected by Shift in position.
- 3- MNIST and History by Wikipediaof "Convolutional Neural Network"
- 4- Architecture of CNN by" K Keiron O'Shed and RyanNash "A Introduction to convolutional Neural Network"