International Journal of Engineering Applied Sciences and Technology, 2021
Vol. 6, Issue 1, ISSN No. 2455-2143, Pages 249-254
Published Online May 2021 in IJEAST (http://www.ijeast.com)

ARTIFICIAL INTELLIGENCE IN COMPUTER VISION

Aryan Karn
Motilal Nehru National Institute of Technology Allahabad, Prayagraj
Department of Electronics and Communication Engineering

Abstract- **Computer vision is an area of research concerned with assisting computers in seeing. Computer vision issues aim to infer something about the world from observed picture data at the most abstract level. It is a multidisciplinary subject that may be loosely classified as a branch of artificial intelligence and machine learning, both of which may include using specific techniques and using general-purpose learning methods. As an interdisciplinary field of research, it may seem disorganized, with methods taken and reused from various engineering and computer science disciplines. While one specific vision issue may be readily solved with a hand-crafted statistical technique, another may need a vast and sophisticated ensemble of generic machine learning algorithms. Computer vision as a discipline is at the cutting edge of science. As with any frontier, it is thrilling and chaotic, with often no trustworthy authority to turn to. Numerous beneficial concepts lack a theoretical foundation, and some theories are rendered ineffective in reality; developed regions are widely dispersed, and often one seems totally unreachable from the other.**

Keywords—Computer vision, Artificial intelligence, Neural networks, CNN., Deep learning, machine learning

I. INTRODUCTION

Recently, computer vision has gained traction and popularity as a consequence of the many applications it has found in areas like health and medical, sports and entertainment, automaton design, and self-driving cars. Many of these applications rely on visual recognition tasks such as image order, restriction, and identification. Recent advances in Convolutional Neural Networks (CNNs) have resulted in an extraordinary performance in these best-in-class visual recognition assignments and frameworks, demonstrating the power of Convolutional Neural Networks. Consequently, convolutional neural networks (CNNs) have emerged as the basic building blocks of deep learning computations in computer vision.

Deep Neural Networks (DNN) is a kind of neural network that has better image identification skills and is often utilized in computer vision computations. Convolutional Neural Networks (CNN or ConvNet) is a subtype of Deep Neural Networks (DNNs) that are often employed in visual sign decoding. In addition, it is used in Computer Vision and Natural Language Processing to organize material (NLP). It is possible to construct a convolutional neural network using a variety of structural blocks. These structural blocks include convolution layers, pooling layers, and fully connected layers, all of which will be discussed briefly in this article. In the next sections, the author covers Deep Learning and the many neural network techniques lumped together. In addition, the book covers Convolutional Neural Networks, their construction, and their applications in several fields, including medicine and engineering.

II. LITERATURE SURVEY

A. Deep Learning and Neural Networks

Machine Learning is a subset of Deep Learning, a subset of Artificial Intelligence (AI). Machine learning uses algorithms and training data to automatically detect patterns and with little human intervention. Artificial Intelligence is a method for teaching computers to act like humans. At the same time, Deep Learning is inspired by the structure and function of the human brain, as represented symbolically by an artificial neural network. [12] While deep learning was originally proposed in the 1980s, it has shown significant benefits in recent years for two primary reasons:

A. This requires a significant level of knowledge. For instance, the development of autonomous vehicles necessitates the collection of many pictures and lengthy video recordings.

B. Deep learning requires a large capacity for recording. High-performance GPUs offer an efficient parallel design that is well-suited for deep learning. When used in conjunction with clusters or cloud computing, this significantly lowers the time required to train a deep learning network from weeks to hours or less. [11]. Deep learning may be used to solve a wide range of problems. For example, the author discusses autonomous driving, aerospace and military, medical research, industrial automation, and electronics in more detail in the closing part of the article.

In general, a Neural Network is a kind of algorithm that accepts certain input parameters and processes them using an Activation Function to get the desired Output. In this method, the input to output processing component is referred to as a neuron. Consider the fundamental example of calculating the purchase price of a home. Numerous factors must be taken into account, each of which has an impact on the Price component. For instance, the square footage of the room, the number of bedrooms, and the zip code. Thus, if we take price as an Output, the following Neural Network shows how a neural network could produce that Output using the parameters stated earlier as inputs.

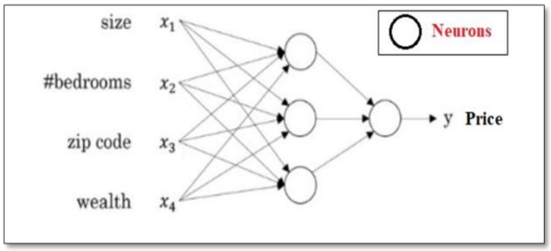


Fig 1: An example of Standard Neural Network [1]

Each circle represents a neuron that is given an Activation Function that computes the desired Output by combining distinct values for various input parameters. The Activation Function is determined by the algorithm's purpose/application. For instance, each circle represents a neuron that is given an Activation Function that computes the desired Output by combining distinct values for various input parameters. The Activation Function is determined by the algorithm's purpose/application. For instance, in the above example, the objective is to determine the maximum price of a house. For the sake of simplicity, let us suppose that the Output is solely dependent on two input variables, namely the size and number of bedrooms. In this instance, the bigger the house and the more bedrooms, the greater the price of the house. Thus, the Activation Function (Neuron) will be defined in such a manner that it will pick the greatest possible value for each input parameter and then compute the Output. Obviously, this seems to be very easy in this case, but when a large number of factors are involved, decision-making is not as straightforward as it appears based on maximum or minimum values alone. And here is where Data-Driven Machine Learning comes into play. The method takes advantage of data saved (learned!) from previous instances in order to calculate the optimal Output using the Activation Function. The above example shows a Standard Neural Network, which is often used to generate Output from statistical, numerical, and other quantitative data. The kind of Neural Network to employ is determined by the nature of the input data that the algorithm must handle. The following table summarizes the capabilities of different Neural Networks in processing various kinds of input data. [1] In the remainder of this article, the author will concentrate only on the Convolutional Neural Network method used in Deep Learning.

B. Deep Learning using Convolutional Neural Network for Computer Vision

In deep learning, a convolutional neural network (CNN), often known as a ConvNet, is a kind of deep neural network that is frequently used to analyze visual pictures. In certain areas, it is also referred to as a convolutional neural network (CNN). These artificial neural networks are referred to as shift-invariant artificial neural networks or space-invariant artificial neural networks due to their shared-weights architecture and translation invariance properties (SIANNs). Algorithms may be used to identify pictures and videos, create recommender systems, categorize images, do medical image analysis, and evaluate natural language. In the next part, the author discusses what Convolution is, how it extracts data from pictures, and the architecture and components of CNN, among other things. This will show how CNN examines the content of an image and processes the data in order to provide the intended result to the audience.

C. Architectural Overview

Convolution is a mathematical procedure that takes two functions and produces a third function that illustrates how the shape of one function is affected by the shape of the other. To complete the operation, the Convolution process requires the calculation of the Result function, as well as the initialization of the Result function. Convolution is a data processing technique that entails categorizing the components (content) of an image in order to assist Machine Learning and ultimately generate the desired Output through the algorithm. It is utilized in the processing of picture data. Deep Learning and Neural Networks are two different types of neural networks that are capable of analyzing image data. Deep Learning is a kind of neural network that enables data-driven learning. As indicated by the procedure's name, the convolution process separates the wheat from the chaff.

This structure may be seen as a three-dimensional volume of neurons in a cellular environment. A distinguishing feature of how CNNs have evolved from earlier feed-forward versions is their ability to improve computational efficiency via the addition of new layer types to their design. How about we take a closer look at the general design of CNNs right now? [4]

D. Basic CNN components

1. Convolutional Layer:

CNN, or convolutional neural network, is a kind of neural network model that is designed for dealing with two-dimensional image data, although it may also be used to deal with one-dimensional and three-dimensional data. Convolution is accomplished via the use of a channel (a small matrix whose size may be chosen). In this channel, which travels the whole picture network, the task is to reproduce the image's features by utilizing the pixel values that were first used. Each of these increases is added together to form a single number towards the end of the process. When doing a comparison action, the channel moves rightward by n units (this number may vary). After traversing over each place, a framework is produced that is much less in size than the information grid that was previously constructed. [7]

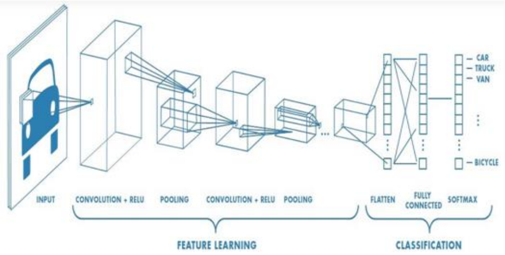


Fig 3: Architecture of a CNN [6]

Edge Detection Example:

In Figures 4 and 5 below, to detect the horizontal and vertical images with the help of a matrix, let's consider a greyscale image, with a 6 x 6 matrix and a filter of 3x3 applied to it.[14]

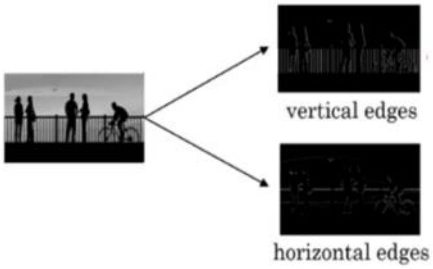
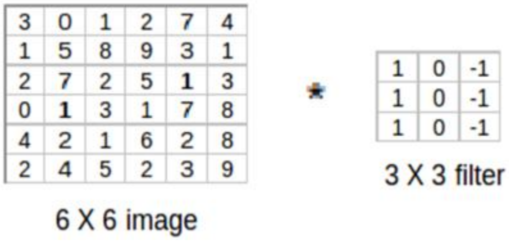


Fig 4: Identifying the edges [14]



3	0	1	2	7	4
1	5	8	9	3	1
2	7	2	5	1	3
0	1	3	1	7	8
4	2	1	6	2	8
2	4	5	2	3	9

6 X 6 image

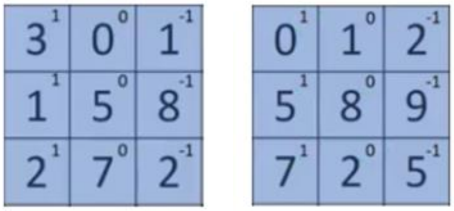
*

1	0	-1
1	0	-1
1	0	-1

3 X 3 filter

Fig 5: Scanning of the image pattern [14]

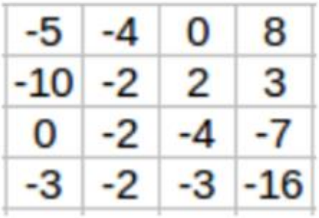
After the above calculations of the matrix, we get the matrix as shown in fig:6. To calculate it, we take the initial 3 X 3 framework from the 6 X 6 picture and increase it with the channel. Let's consider the following matrix of 4x4 order and the calculation takes place as: for example, 3*1 + 0 + 1*-1 + 1*1 + 5*0 + 8*-1 + 2*1 + 7*0 + 2*-1 = - 5. To compute the second component of the 4 X 4 order, we will move the channel one step ahead to the right side of the original Greyscale matrix and so on: [14]



3^{1}	0^{0}	1^{-1}
1^{1}	5^{0}	8^{-1}
2^{1}	7^{0}	2^{-1}

0^{1}	1^{0}	2^{-1}
5^{1}	8^{0}	9^{-1}
7^{1}	2^{0}	5^{-1}

Fig 6: Matrix calculation in convolutional layer [14]



-5	-4	0	8
-10	-2	2	3
0	-2	-4	-7
-3	-2	-3	-16

Fig 7: Convolving over the entire image [14]

The way to detect the vertical edge in the image is to look for the pixel values as, if the pixel values are greater, then brightness at that part of the image will be more, and if the value is less, it will be dark. [14]

2. Pooling Layer:

Spatial pooling (alternatively referred to as subsampling or down sampling) lowers the dimensionality of each element map while preserving the most critical data. Spatial pooling may occur in a number of ways: Quantifiers include the terms maximum, average, and total. If Max Pooling occurs, we define a spatial neighborhood (for example, a 22-window neighborhood) and choose the biggest component from the redressed highlight map contained inside that neighborhood. Rather than choosing the biggest component, we might choose the average (Average Pooling) or a total of all components included inside that window. Max Pooling has been shown to be increasingly effective with time. [8] Max pooling, as shown below, chooses the component with the largest size from the rectified feature map. Choosing the biggest component is equal to using the conventional pooling method. The phrase "sum pooling" refers to the gathering of all components in an element map. [8]

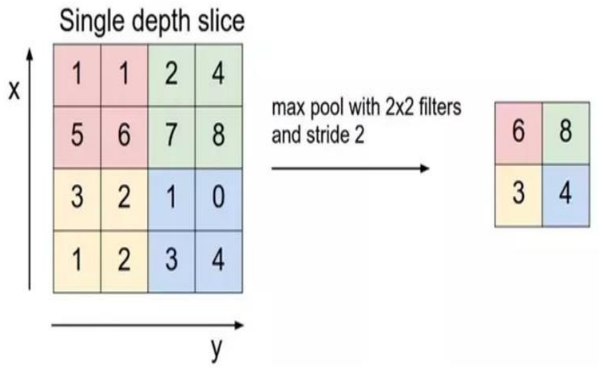


Fig 8: Max Pooling [8]

3. Fully Connected Layers:

Fully Linked layers have every neuron in the layer above it connected to every neuron in the layer below it. To put it simply, FC works in the same manner as a conventional neural network, such as a Multi-Layer Perceptron, does (MLP). The main distinction is that information sources would be molded and organized in the manner defined by earlier phases of a CNN, rather than the other way around. [7]. As illustrated in the diagram below, the feature map matrix is converted into a vector as(x1,x2,...xn) by utilizing the FC layer, and the resulting vectors are merged to create a model. Then, using the activation function, we can classify the Output into different categories.

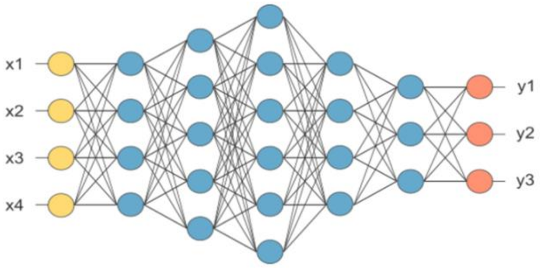


Fig 9: Fully connected Layer [7]

III APPLICATIONS

A. HealthCare:

Computer vision is extensively employed in the diagnosis of diseases by analyzing X-rays, magnetic resonance imaging (MRI), and other medical pictures. It has been shown to be just as convincing as traditional human specialists in the area when it comes to accuracy. On a regular basis, Computer Vision is effectively diagnosing pneumonia, cerebrum tumor's, diabetes, Parkinson's illness, malignant uterine growth, and a host of other medical issues, and the technology is getting more advanced. With the use of best-in-class image processing technology and computer vision methods, early identification of any potential diseases will be feasible. In this manner, treatment may be administered at an inconvenient time during the disease or, in any event, the likelihood of their recurring is decreased [2].

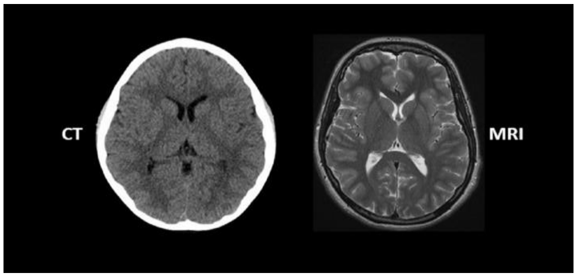


Fig 10: Cross-section of the 3D image of CT Scan and MRI [2]

B. Automobile:

With the expanded publicity of oneself driving autos, car businesses are vigorously subject to Computer Vision since it is intended for understanding the driving condition, including identifying impediments, people on footpaths, and conceivable crash ways. Self-driving autos are gradually advancing into the market, with more organizations searching for imaginative approaches to bring progressively electric vehicles onto the street. Computer Vision innovation enables these self-driving vehicles 'to see' the earth while AI calculations make the "minds" that help that Computer Vision translate the items around the vehicle. Self-driving autos are furnished with numerous cameras to give a total 360-degree perspective on nature inside the scope of several meters. Tesla vehicles, for example, utilize something like 8 encompasses cameras to accomplish this accomplishment. Twelve ultrasonic sensors for identifying hard and delicate articles out and about and a front-oriented radar that empowers the identification of different vehicles even through downpour or mist are additionally introduced to supplement the cameras. With a lot of information being encouraged into the vehicle, a basic PC won't be sufficient to deal with the inundation of data. This is the reason all self-driving autos have a locally available PC with Computer Vision highlights made through AI. The cameras and sensors are entrusted to both recognize and group protests in nature - like people on foot. The area, thickness, shape, and profundity of the items must be considered quickly to empower the remainder of the driving framework to settle on proper choices. Every one of these calculations is just conceivable through the incorporation of AI and deep neural systems, which results in highlights like the person on foot recognition [15].

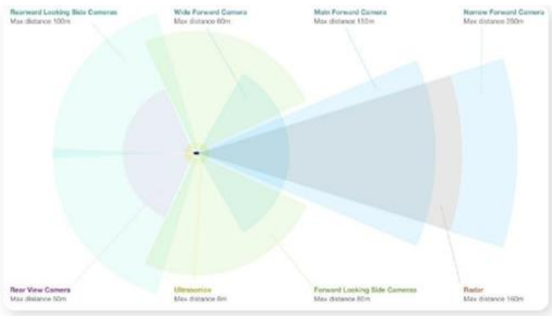


Fig 11: Tesla Car's Vision, Source: Tesla [15]

C. Astronomy:

Our full understanding of the universe is based on photon estimations, which are mostly composed of pictures of the universe. This opens the door to the potential of utilizing Computer Vision in astronomy since our universe is so enormous, and our universe's lone natural rule predicts that the data gathered will be just as large. It will be impossible for the stargazer, or for anybody else, to physically contemplate this information in its entirety. We can decipher all of the data in a short period of time, thanks to Computer Vision. To put it another way, computer vision is currently being utilized to find new planets and big bodies, with applications such as exoplanet imaging, star and cosmic system grouping, and other similar tasks [3].

D. Industrial:

In Industries, Computer Vision is utilized on the mechanical production systems for checking groups, identifying harmed parts, for the examination of the completed merchandise. Here, Machine Vision apparatuses help in discovering infinitesimal level surrenders in items that basically can't be distinguished through human vision. In assembling undertakings, perusing scanner tags or QR code are fundamental as they give a one of a kind recognizable proof to an item. Perusing a great many standardized identifications in a day isn't a simple errand for people; at the same time, it very well may be done effectively in minutes through Computer Vision [3].

IV ConvNet ARCHITECTURE

Convolutional Neural Networks (CNNs) is a kind of neural network that has been around since the mid-90s. You'll discover some more visually arresting designs in the section below [9].

A convolutional neural network was in the process of being created from the late 1990s to the middle of the 2010s, and it was known as LeNet during that time period. The tasks that convolutional neural networks were capable of doing grew increasingly fascinating as an ever-increasing quantity of information and processing power became accessible.

(2).AlexNet (2012) – In 2012, Alex Krizhevsky (together with others) published AlexNet, which was a more in-depth and much more complete version of the LeNet. AlexNet was the clear winner of the inaugural ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012, outperforming the competition by a wide margin. This research represented a major advancement over prior techniques, and the widespread use of CNNs today may be linked back to the findings of this study.

A Convolutional Network created by Matthew Zeiler and Rob Fergus was presented at the ILSVRC 2013 as part of the 3.ZF Net (2013) session. The ZDNet was the moniker that was given to this network (short for Zeiler and Fergus Net). It was feasible to make improvements to AlexNet by altering the design hyperparameters used in its creation. When Szegedy and colleagues from Google presented a Convolutional Network at the ILSVRC 2014 conference, it was given the moniker "GoogleLeNet" (2014). This organization's main goal was the creation of an Inception Module that significantly decreased the number of parameters in the system (4M, contrasted with AlexNet with 60M).

In the 2014 International Laser Scanning and Vision Research Conference (ILSVRC), a system that became known as the VGGNet was the first to cross the finish line. In particular, it aimed at demonstrating how important it is for efficient execution to have a system with sufficient depth (i.e., layers). It was ResNets (2015), a Residual Network created by Kaiming He (and others), that was awarded sixth place in the ILSVRC 2015. ResNets (2015) was the winner of the ILSVRC 2015. Convolutional Neural Network models such as ResNets are currently by a wide margin the best-in-class models, and they will continue to be the default option for utilizing ConvNets for the foreseeable future (as of May 2016).

The seventh source is DenseNet, which was launched in August 2016. A network of nodes that are closely packed together. This densely linked convolutional network, developed by Gao Huang (and others) and published recently, has each layer directly connected to every other layer in a feed-forward architecture, with each layer being straightforwardly correlated with each other layer. Following the completion of five highly concentrated article acknowledgment benchmark assignments, the DenseNet was found to have gained substantial gains over prior best-in-class architectures, results revealed. View this video to see exactly how the Torch was carried out.

V. CONCLUSION

At the beginning of the paper, we discussed the overview of deep learning and how Neural networks in dep learning are deployed to process various inputs to gain desired outputs. In the later part, the author has focused on Convolutional Neural

Network and explained in detail a convolution operation, the system architecture of CNN, and how the layers of CNN work in coordination to identify the highlights and the patterns of an image. Using these algorithms author has described how CNN can be applied in various industries. Through this paper, it can be concluded that CNN has become a very powerful tool in machine learning. By providing various images as input data at the machine learning phase can facilitate the learning process faster, and the data can be deployed for multiple-output functions, which is a major advantage of CNN. Apart from the application, the author mentioned in the earlier section that CNN is now also being considered for IoT, Commercial, and domestic security systems. Thus, CNN has gained a very prominent place in Data Engineering and still is gaining.

VI REFERENCES

[1] Pulkit Sharma, Analytics Vidhya, An Introductory Guide to Deep Learning and Neural Networks, 22 Oct. 2018, https://www.analyticsvidhya.com/blog/2018/10/introduction-neural-networks-deep-learning/

[2] Khan, S. A Guide to Convolutional Neural Networks for Computer Vision. Morgan &amp; Claypool, 2018.

[3] Verma, Shiva. "Understanding 1D and 3D Convolution Neural Network: Keras." Medium, Towards Data Science, 1 Oct. 2019,

towardsdatascience.com/understanding-1d-and-3d-convolution-neural-network-keras-9d8f76e29610.

[4] Upadhyay, Yash. "Computer Vision: A Study on Different CNN Architectures and Their Applications." Medium, AlumnAI Academy, 29 Mar. 2019, medium.com/alumnaiacademy/introduction-to-computer-vision-4fc2a2ba9dc

[5]CS231n Convolutional Neural Networks for Visual Recognition, cs231n.github.io/convolutional-networks/.

[6] "Autopilot." Tesla, Inc, www.tesla.com/autopilot.

[7] Deshpande, Adit. "A Beginner's Guide To Understanding Convolutional Neural Networks."

A Beginner's Guide To Understanding Convolutional Neural Networks – Adit Deshpande – Engineering at forwarding | UCLA CS '19,

adeshpande3.github.io/adeshpande3.github.io/A-Beginner's-Guide-To-Understanding-Convolutional-Neural-Networks/.

[8] Upadhyay, Yash. "Computer Vision: A Study On Different CNN Architectures and Their Applications." Medium, AlumnAI Academy, 29 Mar. 2019, medium.com/alumnaiacademy/introduction-to-computer-vision-4fc2a2ba9dc.

[9] Ujjwalkarn. "An Intuitive Explanation of Convolutional Neural Networks." The Data Science Blog, 29 May 2017, https://www.ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/.

[10] Gibson, Adam, and Josh Patterson. "Deep Learning." O'Reilly | Safari, O'Reilly Media, Inc.,

www.oreilly.com/library/view/deep-learning/9781491924570/ch04.html

[11] "What Is Deep Learning?: How It Works, Techniques &amp; Applications." How It Works, Techniques &amp; Applications - MATLAB &amp; Simulink, www.mathworks.com/discovery/deep-learning.html.

[12] Brownlee, Jason. "What Is Deep Learning?" Machine Learning Mastery, 31 Oct. 2019, machinelearningmastery.com/what-is-deep-learning/.