Received Date: 18/03/2021 Acceptance Date: 29/03/2021 Publication Date: 31/03/2021

IMAGE ACTIVITY TRACKING USING DEEP LEARNING: A Review

Akshay Kr. Singhal*1, Atima Jain*2, Bhavya Nigam*3, Deepak Sharma *4, Meenu Shukla *5
*1, 2, 3, 4 B. Tech Student, Dept. of Computer Science & Engineering, Krishna Engineering College, Mohan Nagar, Ghaziabad, Uttar Pradesh, India.
5 Asst. Professor, Dept. of Computer Science & Engineering, Krishna Engineering College, Mohan Nagar, Ghaziabad, Uttar Pradesh, India.

Abstract: *Deep Learning, over decade has become a burgeoning field for researchers since the technique has the proficiency to overcome the drawbacks of already existing traditional algorithms. Deep learning is a technology stimulated by functioning of human brain [2]. Deep learning technologies are fetching the major outlook for natural signal and information processing, like image classification, speech recognition. In deep learning, large datasets are analyzed automatically by network of artificial neurons, thus discovering the underlying patterns, without human intervention, deep learning identify patterns in unregulated data such as, Images, sound, video and text. Convolutional Neural Network (CNN) is mostly used for image classification in deep learning, performing better than many of the image classification subsets[2]. Numerous implementations, together with video surveillance systems, human-computer interaction, and robotics for human behavior characterization, require a multiple activity recognition system.[1]*

Key Words: *Deep Learning, Convolutional Neural Network.*

1. INTRODUCTION:

In this proposed review paper we aim to calibrate the information or knowledge we gained regarding the technologies we will be using in our project for tracking the activities going on in an image. This paper will majorly focus on the detailed explanations of computational neural networks (CNN) and deep learning also including our proposed methodology.

Deep Learning:

Deep learning is essential for a more extensive group of AI techniques dependent on learning portrayals of information. an observation (e.g., an image) can be addressed from various perspectives, for example, a vector of power or weight esteems per pixel, or in a more theoretical route as an assortment of edges, specific shape, and so forth. a few portrayals make it simpler to learn assignments (e.g., face acknowledgment or outward appearance acknowledgment) from models [1]. Deep neural networks, [2] convolutional deep neural networks, deep belief networks and recurrent neural networks are the various deep learning architectures that can be implemented in domains like computer vision, automatic speech recognition, natural language processing, audio recognition and biomedical and information where they have been shown to produce avant-garde results on multiple tasks.[8] It set-aparts as a class of machine learning algorithms that use an outpouring of numerous layers of nonlinear processing units for feature retrieval and mutation processes.[1] Each consecutive layer uses the output from the previous layer as input. The algorithms may be supervised or unsupervised and applications include pattern analysis i.e. unsupervised and classification i.e. supervised.

The figure illustrates the working of deep learning where an input image (of a car) passes through the networks, extracting features and producing the classification output.

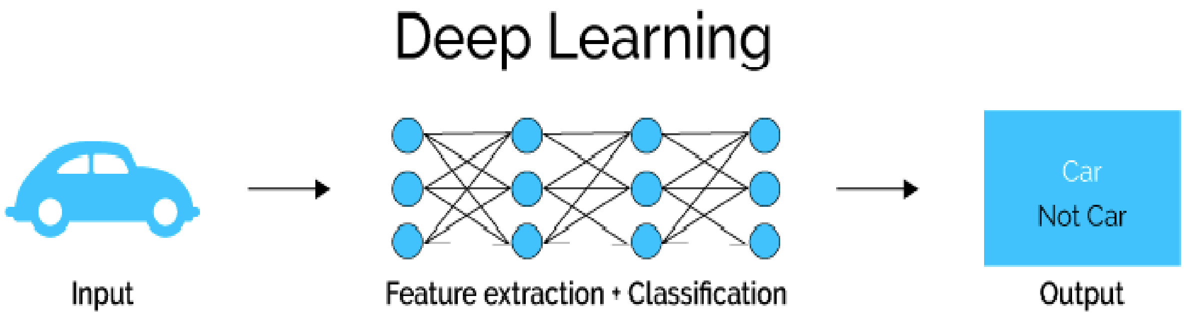

Fig: Representing the workflow Deep Learning.

Convolution Neural Network :

The recognition of object by machine learning covers a wide application area. In present time it has several domains like object detection, character recognition and expert system. The machine learning has become the pillar of many real time applications [7]. Machine learning contains many algorithms that trains and aids the computer to make real time decision and help machine to take action on desired environment condition. The growth of machine learning is very high in past years and is also influencing the new technologies in today's world. The today's technology like weather forecasting, self-driving system and AI [7].

CNN Architecture :

One of the subtypes of deep discriminative architecture is CNN which is comprised of one or more convolutional layer followed by one or more fully connected layer as it is defined in a standard multilayer neural network. The architecture of CNN shows the best performance by processing two dimensional data formats such as image, videos and so on. visual cortex is a complex arrangement of the cell and these cells are sensitive to a small region of the visual field called a receptive field. Since the animal visual cortex is the most powerful visual processing system, many different models exist in the literature such as Neocognitron, HMAX, and Lenet – 5[9]. CNN is differing from Neocognitron since it never requires a shared weight. General matrix multiplication and it is replaced in CNN by using Convolution layer, hence it is used to reduce the complexity of the network by reducing the number of weights, thus eliminates the need of separate feature extraction algorithm, which is widely used in standard learning algorithms. CNN's emerges spatially local correlation by implementing a local connectivity pattern between neurons of adjacent layers. We can illustrate this Like the architecture of standard Neural Network, CNN also having input, output and hidden layers in between. These hidden layers perform the function called feature detection and it performs three different type of operations on data called Convolution, Pooling and Rectifier Linear Unit (ReLu) [9]. The convolution layer is used to activate the certain feature of the image, by feeding the input image through a set of convolutional filters. Each filter enables different features from the image. The pooling layer helps to reduce the number of parameters irrelevant to the focused problem, by simplifying the output using nonlinear down sampling method. Rectified Linear unit make the network to learn faster and work more efficient by mapping negative values to zeros so that it will maintain only positive values [9]. These three different operations are repeatedly applied to ten or thousands of layers so that each layer can able to detect the different level of features [7].

Applications of convolutional neural network

A) Rice Disease Identification

Yang Lu et all [16], developed the CNN to diagnose the 10 different type of rice diseases such as rice blast, rice false smut, rice brown spot, rice bakanae disease,. The result obtained by them was 95.48% of accuracy. The work of Kawaski et all, proposed the CNN based system to identify the cucumber leaf disease and they got an accuracy of 94.9%. very few research works only done on automating the prediction of rice disease from these CNN shows more accuracy compare to other widely used digital image processing techniques and other traditional classification algorithm [7].

B) Human Activity Recognition

Human activity recognition (HAR) garnered a lot of attention since it has a high demand in the various application domain Earlier work of HAR in 1990, should be done using accelerometer . Very few works only carried for HAR using Deep learning technique. Currently facing problems in HAR are (i) Different people doing same activity in a different manner, hence it is translation invariant and fragment of an activity can manifest at different point in time. (ii) 1D time series signal is highly correlated in nature, but extracting sensor feature is very important for overall performance. All these issues are addressed effectively by the use of convolutional neural network, work done by Charissa Ann Ronao et all for Human activity recognition got an accuracy of 94.79% on raw data and interesting fact is that by employing fast Fourier transform for the HAR dataset to provide additional information return more accuracy of 95.75%[5][7].

2. LITERATURE SURVEY:

In Deep Learning [1] engulfs a distinct technique to building and training neural networks. Neural networks have been around since the 1950s, and like nuclear fusion, they've been an incredibly promising laboratory idea whose practical deployment has been beset by constant delays. They take a variety of numbers (that can represent pixels, audio waveforms, or words), run a progression of functions on that array, and output one or more numbers as outputs. The outputs are generally a forecast of certain properties you're attempting to figure from the input, for instance whether or not an image is a picture of a cat [2].

In Convolution Neural Network [3] order to provide an efficient learning method for 2D images, convolutional neural network (CNN) layers were incorporated [4]. By binding adjacent shifts of the same weights together through a pattern, in the same way as a filter sliding over an input vector, convolutional layers have the ability to compel feature learning with change invariance in the input vector. In doing so, they also greatly reduce the model complexity(as calculated by the number of free parameters in the layer's weight matrix) needed to represent equivalent shift-invariant features using completely connected layers, which simplifies SGD optimization and enhances generalization on suitable dataset [5].

A CNN's Architecture input is typically an order 3 tensor, such as an image with H rows, W columns, and 3 channels (R, G, B color channels) [6]. Higher order tensor inputs, on the other hand, can be treated in a similar way by CNN. After that, the input goes through a sequence of processing steps in order. A layer is a processing phase that can be a convolution layer, a pooling layer, a normalization layer, a completely connected layer, a loss layer, and so on. The specifics of these layers will be discussed later in this note.
For now, let us give an abstract description of the CNN structure first.

$$x^1 \longrightarrow w^1 \longrightarrow x^2 \longrightarrow \cdots \longrightarrow x^{L-1} \longrightarrow w^{L-1} \longrightarrow x^L \longrightarrow w^L \longrightarrow z$$

In a forward transfer, a CNN runs layer by layer, as seen in Equation . x 1 is the input, which is normally a picture (order 3 tensor). It goes through the entire process. The first sheet, which is the first package, is being processed. The parameters are denoted by as a tensor w1 involved in the processing of the first layer [6].

In a forward transfer, a CNN runs layer by layer, as seen in Equation 5. x 1 is the input, which is normally a picture [2](order 3 tensor). It goes through the entire process. The first sheet, which is the first package, is being processed. The parameters are denoted by as a tensor w1 involved in the processing of the first layer.

This processing continues until all layers in the CNN have been completed, yielding x L. Backward error propagation, a method for learning good parameter values in the CNN, is implemented as an additional layer. Assume you're working on an image classification issue with C groups. The performance of x L as a C dimensional vector, whose i-th entry encodes the prediction (the posterior probability of x 1 comes from the i-th class), is a typical strategy. Assume you're working on an image classification issue with C groups. The performance of x L as a C dimensional vector, whose i-th entry encodes the prediction (the posterior probability of x 1 comes from the i-th class), is a typical strategy. Let's pretend we're dealing with a C-class image classification issue. The performance of x L as a C dimensional vector with the i-th entry encoding the prediction (the posterior probability of x 1 comes from the i-th class) is a standard strategy[4].

The last layer is a loss layer. Let us suppose t is the corresponding target (ground-truth) value for the input x 1, then a cost or loss function can be used to measure the discrepancy between the CNN prediction x L and the target t. For example, a simple loss function could be although more complex loss functions are usually used.

$$z = 1/2 \, \|t - x^L\|^2$$

This squared `2 loss can be used in a regression problem. In a classification problem, the cross entropy loss is often used. The ground-truth in a classification problem is a categorical variable t. We first convert the categorical variable t to a C dimensional vector t (cf. the distance metric and data transformation note). Now both t and x L are probability mass functions, and the cross entropy loss measures the distance between them. Hence, we can minimize the cross entropy (cf. the information theory note.) Equation 5 explicitly models the loss function as a loss layer, whose processing is modeled as a box with parameters w L. Note that some layers may not have any parameters, that is, wi may be empty for some i. The softmax layer is one such example.

3. METHODOLOGY:

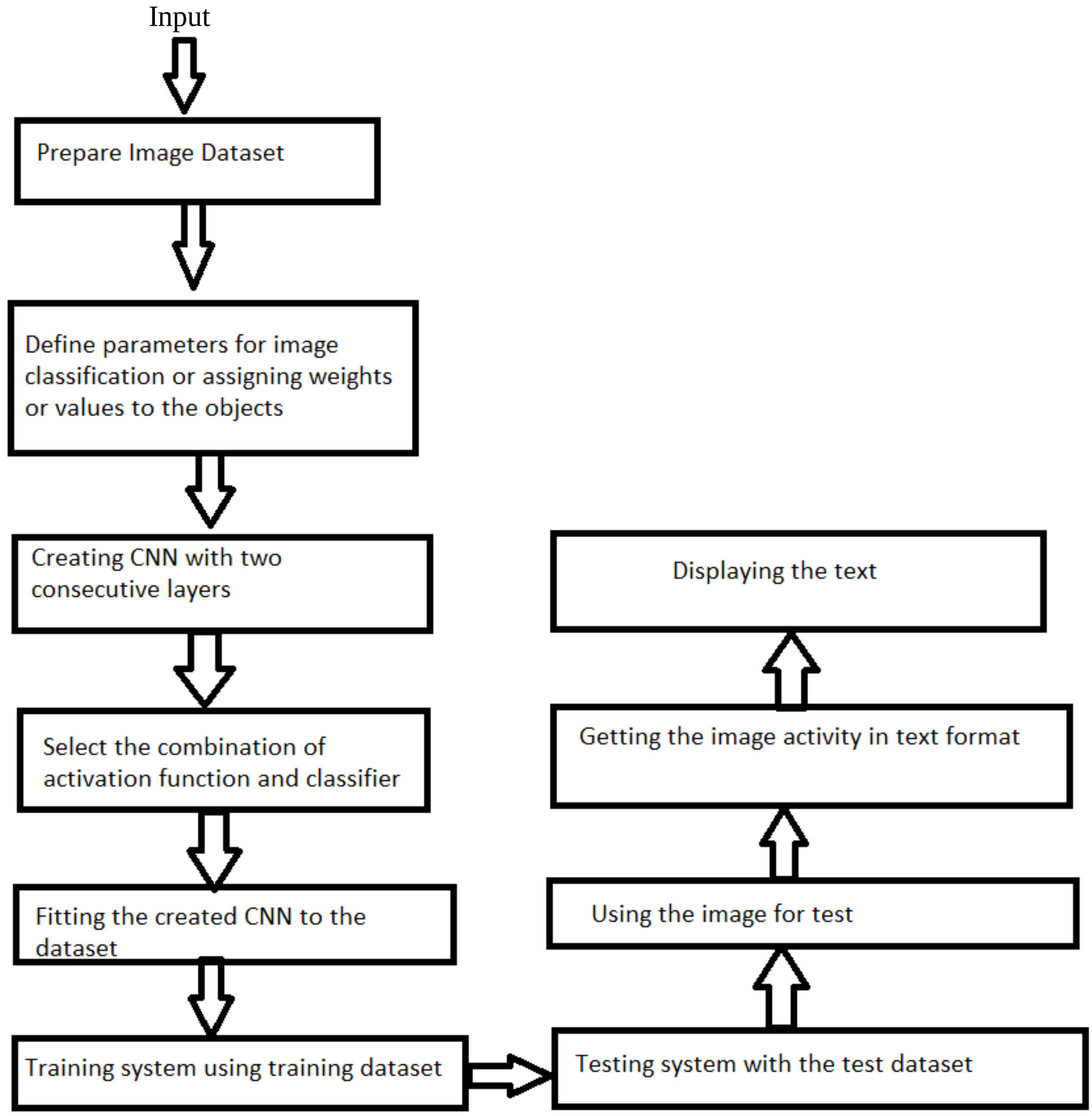

Fig: Flowchart representing the workflow of image tracking using CNN

4. CONCLUSION:

This paper concludes that deep learning and convolution neural networks are the important approaches used for image tracking and computer vision, for building expert machines that can visualize, analyze and interpret the digital images by training and testing the system on the dataset. Activates certain features of the image by passing them through a set of filters and ignoring the irrelevant features.

5. FUTURE SCOPE:

The most important reason for high scope of deep learning is that it doesn't requires feature engineering. Means that there is no need of providing it the features that we have extracted from the data. Deep learning extract the data features itself. Because of the reason stated above, it is more appropriate for generating generalized models. In future, these approaches will more will be able to work more effectively on unstructured data and fuzzy or blur images in case of CNN. Will be able to provide more accurate and precise results in real time.

REFERENCES:

1. Ben-Bright Benuwa, Yongzhao Zhan, "A Review of Deep Machine Learning: International Journal of Engineering Research in Africa", Vol. 24, April 2016, pp 124-136.

2. Karan Chauhan, Shrwan Ram, "Image Classification with Deep Learning and Comparison between Different Convolutional Neural Network Structures using Tensorflow and Keras : International Journal of Advance Engineering and Research Development", Volume 5, Feburary/2018.
3. M Manoj krishna, M Neelima and all, "Image classification using Deep learning: International Journal of Engineering & Technology", 2018, pp 614-617.
4. Ayush Singha, Pradeep Sinha and all, "Use of Deep Learning in Modern Recommendation System: A Summary of Recent Works : International Journal of Computer Applications", Volume 180, December 2017.
5. Jianxin Wu, "Introduction to Convolutional Neural Networks : National Key Lab for Novel Software Technology Nanjing University, China" , May 2017.
6. Xizhao Wang, Yanxia Zhao, Farhad Pourpanah, "Recent advances in deep learning: International Journal of Machine Learning and Cybernetics", February 2020, pp 747-750.
7. Zhuwei Qin, Fuxun Yu and others, "How convolutional neural network see the world - A survey of convolutional neural network visualization methods: Mathematical Foundations of Computing", Volume 1, May 2018, pp 149-180.
8. Asifullah_Khan, Anabia_Sohail, "A Survey of the Recent Architectures of Deep Convolutional Neural Networks", May 2020.
9. M.Sornam, Kavitha muthusubash, "A Survey on Image Classification and Activity Recognition using Deep Convolutional Neural Network: 2017 Ninth International Conference on Advanced Computing (ICoAC)" , 2017.