GLIMPSE -Journal of Computer Science •Vol. 1(1) , JANUARY-JUNE 2022, pp. 7-10

COMPUTER VISION AND ITS APPLICATION AREAS

[1]Neha Bagawari , [2] Tarushi Khanna

[1]Assistant Professor, Ajay Kumar Garg Engineering College, Ghaziabad, UP, India

BTech , CSE, III rd Year Student, Ajay Kumar Garg Engineering College, Ghaziabad, UP, India

[1]bagawarineha@akgec.ac.in, [2] tarushikhushi@gmail.com

Abstract: **For many years, creating machines that possessed the characteristics of human intelligence, for example, the skill to "see", and interpret the acts of their surroundings, was a dream for many. But now, the figments of yesterday have become the facts of today. [1]. Computer vision technology is taking a major part in our lives steadily due to the growth of artificial intelligence and the easy availability of computational power.**

As the term consists of two components, "Computer" that is the manmade electronic machine built for performing various processes, operations, and calculations, from sets of instructions directed by software or hardware, and "Vision", the act or power of sensing with the eyes to conceive something, Computer vision can be defined as the visual perception provided to the machines via which machines try to understand what they see to achieve a goal. [2]

In simpler words, Computer Vision, a specialized area of study of artificial intelligence, deep learning, and machine learning, that uses specialized methods and makes use of general learning algorithms. The goal of computer vision is to understand the contents of digital images. Basically, this involves the methods of development that attempt to reproduce the capability of human vision.

Example of Computer-Vision: Google Translation App, Automatic Cars, Real-time sports tracking, etc.

Keywords: **Computer Vision, CNN, ANN, DIP**

I. INTRODUCTION

Computer vision is a branch of artificial intelligence that imitates parts of the complexhuman vision system and enables the computers and systems to visualize and derive meaningful information from the visual world that includes digital images, videos, and other inputs. So, it can be stated that AI enables computers to think, whereas Computer Vision enables them to visualize, observe and understand. It trains computers to apply interpretation and understanding to the visual world. Using videos, digital images from cameras, and applying deep learning models to it, accurate identification and classification of objects could be done by machinesincluding the reaction to what they see. It is one of the main technologies that enable the intreratcion of the digital world with the physical world. [2]

II. HISTORY OF COMPUTER VISION

The experimentation with Computer-Vision started back in the 1950s. For the first time, it was used to interpret handwritten digits and alphabets.

About the same time, neurophysiologists attempted to correlate a response in a cat's brain by displaying an array of images to it. They discovered that it responded first to the harder lines or edges and in the view of science, this meant that image processing begins with a simpler portion of the images like edges. [2]

1. Evolution of Computer-Vision

Before the advent of deep learning,to accomplish computer vision task manual coding and lot of extra work was required to be done by human operators and developers.For instance, for performing simple facial recognition, the following steps would be required:

- Create a database: To store individual images of the object that were to be tracked in a particular format.
- Images Annotation: For each image, several points, such as the width of the nose, measure of distance between both the eyes, and various other features or measurements that help to identify the identity of every person should be entered.
- Capture new images: Then, to capture new images, the entire measurement process again should be done again, with the key features of the given images as done before.

After this manual work, the application does the comparison between the measurements of the pre-stored images in the database to the newly entered images and tells if the image corresponds with any of the profiles tracked. In fact, most of the work was being done manually and there was very little

automation involved. The amount of random sampling error in the results is still very large.

Computer vision problems could be solved much easily with machine learning approach. Machine learning helped to reduce the manual coding to a greater extent. Specific patterns in images could be detected with programs of much smaller length.Later they used mathematical and statistical learning algorithms for eg, support vector machine, logistic regression, linear regression, decision trees, etc for the same.

Few problems which always used to be a challenge and could not be solved by basic software development tools were very easily solved by ML (Machine Learning)tools. For example, back then ML engineers created a software that predicted survival windows of breast cancer better than human experts. But for building the software a lot of effort and time was needed from the engineers and breast cancer experts.

Deep learning provided analtogether different approach by using neural networks for working on machine learning models, which would be a general-purpose function and capable of solving any problem representable through instances or examples. When a neural network is provided with various labeled examples of a particular data, it will be able to extract common patterns between those examples and transform them into a mathematical equation that eventually helps solve the problem further and classify better.

For example, a facial recognition application that uses deep learning only requires a developed preconstructed algorithm and training the algorithm with the examples of the faces of the persons it intends to detect. If enough examples are provided, the neural network generally detects the faces without the need for any further instructions or measurements.

Deep learning is the most popular and very effective method to perform computer vision problems. In most cases, building an advanced deep learning algorithm starts from gathering a large amount of labeled training data and tuning its parameters such as the layers of neural networks and training epochs. Compared to previous approaches of machine learning algorithms, deep learning is both faster and easier to develop and computing resources.[4]

III. APPLICATION OF COMPUTER VISION

A few examples of tasks established using computer vision are :

Image classification - A technique that is used to visualize an image and can be used to classify it (eg -a cat, an orange, a person's face). More precisely speaking, it is capable enough to accurately classify the given image. For eg., Facebook uses this technology to automatically tag people in given photos, google photos use this technology to categorize the images. Similarly, this technology could be used to automaticallyidentify and segregate the images uploaded by users that are objectionable according to the social media company guidelines.

Object tracking - Used to follow or track an object after it is detected. This is often used with objects and their movement detection or in real-time video inputs. For example, Autonomous vehicles detect and classify obstacles while driving such as pedestrians, other vehicles, and their respective motions to avoid accidents, collisions and obey traffic laws.[7]

Object detection - Image classification is also used to classify the image and then detect its presence in a video or image. A few examples include damage detection on an assembly line that requires maintenance.

Content-based image retrieval - Computer vision is used to identify, retrieve, and lookup for the images stored in large databases, based on their content rather than metadata tags associated with them. This technology can be used for performing digital asset management and can increase the accuracy of search and retrieval.[2]

IV. HOW DOES COMPUTER VISION WORKS.

Computer vision is a technology that works by imitation of the human brain. Similar to a human brain, computer vision systems also rely on patterns to decode individual objects.[5] [1]

A Computer vision algorithm interprets an image as a series of pixels, with each pixel having its own set of color values. A pixel is basically defined as a unit that consists of a combination of 3 additive primary colors ie, RGB(Red Green Blue). This combination may vary in intensities to represent different colors. Colors are stored inside pixels. Computer Vision models are then trained with a large amount of data consisting of many many pixels — computers process images, add labels to objects on them, and find related patterns in the respective objects. For example, if we send a million images of different cars as input, the computer will analyze and identify patterns that are similar to all cars, and then at the end of this process, a model is created that results in accurate detection of whether a particular image is a car every time we send them pictures. [2] [1]

This task is accomplished by, the two most basic terminologies used, those are (1). CNN (Convolutional Neural Network) and (2). Deep Learning.

Deep Learning: Advances in machine learning had a great impact on the destiny of computer vision technology. Deep learning, in particular, had a very powerful impact on computer vision. It made computer vision algorithms highly efficient

in practical life. The emergence of the convolutional neural network made computer vision possible for general commercial or industrial applications and cemented the technology as a worthy investment for companies looking to automate tasks. [8]

The machine is taught by the deep neural network training process by using a lot of data sets and countless training cycles. It follows the bottom-up approach. During this training process, the algorithm automatically extracts the relevant features in general, for eg - of 'cars'. This process produces a model that is applied to previously unseen images to produce an accurate classification of the same.

Convolutional Neural Networks -A CNN provides the vision to a machine learning or deep learning model by breaking or converting the images into its unitary form ie -labeled and tagged pixels. Labels are used to conduct mathematical operations on two functions to produce another, a third function is known as convolutions and make predictions about what is been seen. The algorithm is inspired by the neurons in our brain and is designed to recognize the patterns in complex input data and often performs the best when patterns are recognized in audio, images, or videos. It basically helps you to build predictive models using huge databases. Some of the main applications of CNN consist of sentence classification, face recognition, text recognition, object detection, localization image characterization, etc. [2]

As mentioned, the algorithm takes inspiration from neurons in our brain by using neural networks. A neural network is a group of connected I/O units called neurons which consists of weights and some biases, that are the features of connections between the neurons. Weight is basically a parameter that represents the strength of the connection between units (neurons) and bias is an indication that tells whether neurons are active or inactive. (Weights and bias)

Instead of general matrix multiplication, CNN uses a mathematical linear operation called convolution in its layers. It contains at least one fully connected layer and a convolutional layer as in a standard multi-layer neural network.[5]
Some of the basic works of CNN are -

- Object detection
- Faster r-CNN
- YOLO- you only look once
- Single-shot detector (SSD)
- Retina net

V. CNN ARCHITECTURE

They are the most prominent and most widely used category of neural networks. It contains a multi-layer neural network architecture, which contains convolutional layers pursued by fully connected layers. Few ptional layers,ie Subsampling layers can exist between these two layers. They results the best of DNN models. Its architects include :

Convolutional layer
It is the basic building block of a convolutional neural network. It determines the output to the given inputs, accordingly. This is achieved through special filters that are used to extract information from the input pictures, known as the kernels. This helps in calculating the dot product between the input and filter values, which in a result builds a 2-Dimensional activation map for that filter. CNN then learns the filters that are activated only when a particular type of feature at some spatial position of the input is under observation.

Nonlinearity layer
Nonlinear functions have a curvature When they are plotted and degrees higher than one. The main purpose of this layer is to transform the input signal into the output signal which will be used as an input to the next layer.
Pooling layer

CNN has layers that take the outputs from a neuron at one layer and add it into another individual neuron in the next layer. Its main purpose is to scale down the size of the representation to reduce the number of calculations and parameters in the given model.

Fully connected layer
Fully Connected layers are standard deep Neural Networks, whose objective is to build the predictions arrived from the activations, to be used for regression or classification. It has a similar principle as the conventional multi-layer perceptron neural system (MLP).

VI. BASIC APPLICATION OF OPEN CV IN OBJECT DETECTION

Object detection
It is a very demanding issue as there are many factors that must be considered, for example, lighting conditions, object's shapes, and colors, occlusion, and many more. A few examples that use this technique include - arms detection, wild animal detection, miscellaneous object detection, and human being detection–

Wild animal detection includes research that is based on Animal detection and is helpful for research that is related to the locomotive behavior of the intended animal, the applications which are used in real life, the applications which are used in everyday life, and to spot the dangerous animal in the public residentiary zone.

Human being detection is recognition is a method based on human activities to recognize various activities via the outer

sensors, for example, the video sensors to improve the fields of human health and well-being services, by allowing the automatic checking activities of the patients.

Small arm (gun) detection Most criminal or illegal actions occur by utilization of hand held arms such as guns, revolvers, and knives. A few observations state that these arms are also utilized for differing illegal actions like assault, thievery, etc. Automatic strategies for weapon identification are used lately to avoid and reduce criminal acts.

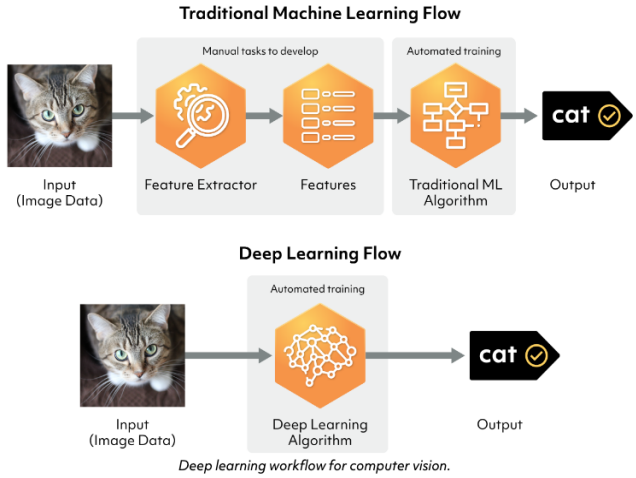

Fig 1: Machine learning and deep learning workflow

VII. CONCLUSION

Thus, we have seen that with the advent of deep learning techniques we have a come long way in computer vision, and have huge application areas where these deep learning techniques can be used for the purpose of image classification, object localization, object detection, object tracking, performing transformation analysis, etc.

REFERENCES

[1] https://xd.adobe.com/ideas/principles/emerging-technology/what-is-computer-vision-how-does-it-work/
[2] https://www.ibm.com/in-en/topics/computer-vision#:~:text=Computer%20vision%20is%20a%20field,recommendations%20based%20on%20that%20information.
[3] https://www.verdict.co.uk/computer-vision-timeline/
[4] https://towardsdatascience.com/everything-you-ever-wanted-to-know-about-computer-vision-heres-a-look-why-it-s-so-awesome-e8a58dfb641e
[5] Dhillon, Anamika, and Gyanendra K. Verma. "Convolutional neural network: a review of models, methodologies and applications to object detection." *Progress in Artificial Intelligence* 9.2 (2020): 85-112.
[6] https://www.v7labs.com/blog/computer-vision-applications
[7] https://viso.ai/computer-vision/the-most-popular-computer-vision-tools/
[8] https://www.dynam.ai/what-is-computer-vision-technology/
[9] https://www.analyticsvidhya.com/blog/2021/06/everything-happening-in-computer-vision-that-you-should-know/

ABOUT THE AUTHORS

Ms. Neha Bagwari has completed her graduation from Government College of Engineering, Amravati, Maharashtra. Post-graduation she has worked with Cognizant Technology Solutions and HSBC GLT India. Combined she has work experience in Manufacturing and Logistics, Life Sciences, and Banking Domain. Her work majorly was around Java, J2EE technologies. She did her MTech from ABES Engineering College, Ghaziabad affiliated with Dr. APJAK Technical University, Lucknow. Her research subject was "Big Data". She has experience working with the Noida Institute of Engineering and Technology in the education field. She has published one research article in International Conferences. She is an Assistant Professor in the Department of Computer Science and Engineering of AKGEC Ghaziabad affiliated with Dr. APJAK Technical University, Lucknow. Her present research interests include investigating the different application areas of Big Data and the technologies providing solution to it.

Ms. Tarushi Khanna is a student in B.Tech Third year, CSE Department, AKGEC. Her interest areas are Data Analytics, Machine Learning, Deep Learning, and Artificial Intelligence. She is looking forward to an opportunity to start her career as a Data Scientist and Machine Learning Engineer along with Computer Vision.