Alphabet Detection through Air Canvas Using Deep Learning and OpenCV

Aadesh[1], Hassan[2], Sahil[3], Dr H.S Guruprasad[4]

Student[1], Student[2], Student[3], Professor[4]
BMS college Of Engineering, Bangalore, India[1,2,3]

***Abstract*— Air Canvas is a hands-free digital drawing canvas that recognises and maps hand motions onto a PiTFT screen using a Raspberry Pi, a PiCamera, and OpenCV. Built-in buttons give the user the permission to replace the size and colour and their "brush." Very brush's direction is entirely controlled by open source OpenCV software, which has been adapted to following a calibrated screen to detect and register the person's hand coloring, mapping the index pointer onto the panel utilizing Pygame.**

Keywords — Alphabet Detection, gesture recognition, finger detection, python, openCV

I. INTRODUCTION

The conventional art of writing is being supplanted by digital art in the digital age. Digital art refers to art forms that are expressed and transmitted in a digital format. The digital manifestation is distinguished by its reliance on modern science and technology. Traditional art refers to work that was created prior to the advent of digital art. Pictorial art, auditory art, verbal art, and verbal inventive art can simply be divided into artwork, audio art, sound art, and sound imaginative art, which includes writing, paint, sculpting, architectural, musical, ballet, comedy, and other art works. The two types of art, traditional and digital, are closely intertwined and interconnected. Spite of the fact that societal growth is not the consequence of people's volition, the needs of human existence are the fundamental driving reason.
The very same thing occurs with art. Because art form and artistic styles are currently in a symbiotic relationship, we must properly understand the underlying idea of design of both.
Conventional learning tools include pencils, as well as blackboard and board. Digital art's major goal is to develop a hand motion recognizer that could be used to paint virtually. Digital art employs a variety of writing techniques, including the use of a computer, a tap surfaces, a computer pen, a pad, or electric protective gloves. Our solution, on the other hand, makes use of hand sign tracking, a deep learning technique, and Programming environment to allow artificial interaction

II. BACKGROUND

We started our effort by looking for open source type of motion detection application that used OpenCV and Python language together. As a result of this, the architecture of our project altered as we learned about different image processing methods. Hand gestures were used to control the colour and size variables in our early implementation. Initially, we had to make an image mask to separate the hand from the rest of the image. We successfully grabbed a picture, Gaussian blurred it, then applied a binary mask to dramatically contrast overall shape of palm from the background using OpenCV trial and error. This approach was picked from Izane's Finger Detection tutorial1 because it employs convexity detection, which involves detecting the valleys between the fingers.

III. RESEARCH AIMS AND APPROACH

In this Literature survey we have gone through other similar works that are implemented in the domain of gesture recognition and finger detection using deep learning procedures.
Our main objective is to survey various deep learning algorithms as well as suitable tools to implement the detection model. Main research aims were as follows: -

a) Current models and algorithms available and researched by researchers.
b) Data acquisition, gesture recognition representation, data environment and image processing
c) Performance of existing gesture recognition detection systems with its efficiency and output accuracy.

IV. LITERATURE SURVEY

M. Chen, G. AlRegib and B. Juang discussed that Air-writing is the practice of using hand or finger movements to write linguistic characters or words in a free area. The difference between air-writing and traditional handwriting is that the latter comprises the letters. While the latter has a circumscribed motion, the former does not. The order of events in the writing process. We'll look at how to recognize handwriting in the air. In a pair of related publications, there are several issues that need to be addressed. Part I dealt with recognition. Data from six-degree-of freedom hand motions is used to generate a set of characters or words.

M. Chen, G. AlRegib and B. Juang discussed that it follows the metaphor of pen-based writing, writing with a finger on a touch-based interface is intuitive. Recent advancements in tracking technology enable tracking of hand and finger motions without the use of user-worn sensors, and writing motion is no longer limited on a physical plane. When conventional input devices, such as a keyboard or a mouse, are not available or sufficient, air-writing provides a feasible alternative interface for text entry. Air-writing has the advantage of being "eye-free," needing minimal attention concentration [1], when compared to other unconventional input techniques like as typing using a virtual keyboard or similar systems.
When we utilize a controllerless tracking system like LEAP to track the finger movements while writing with a fingertip in the air.

Itaguchi Y, Yamada C, Yoshihara M discussed that Kusho is a sequential writing action that can be performed in the air or on a surface. When one uses Kusho behavior, he or she may monitor the finger movement at times but not at others; in other words, the activity may be carried out totally out of sight, such as in the space on the lateral side of the body or on the knee. This is a common occurrence in people from Kanji cultures, such as Chinese and Japanese. Most Kanji characters, which are made up of several straight and curved lines, have multiple pronunciations and meanings in Japanese (called strokes). When Japanese children grow up, they learn to utilize Kusho conduct, and even people from non-Kanji cultures do if they have learned how to write kanji characters.

Seosuk-dong, Dong-ku discussed Sign language is a pictorial communication used by deaf persons. The idea that symbol representations alter in tempo and structure in 3 (3D) space hampers gesture recognition understanding. We use Microsoft's Sensor module to generate 3D depth information from arm movements and a hierarchical conditional random field (CRF) to recognize hand signs from the motions in this study. A hierarchical CRF is used to find potential sign samples based on arm movements, and then a Boost Map anchoring approach is used to evaluate the divided signals' finger shapes.
Tests demonstrated the 90.4 percent of the cases, the presented approach accurately distinguished symbols from sign sentence data.

Amma, C., Gehrig, D. and Schultz discussed that we may develop specifications for a mobile wearable input device because we want one:
• inconspicuous (ideally, one should not notice wearing a computing device) • hands-free operation (no extra gadgets must be held) • reliable radio link and mechanical influence • long runtime without recharging batteries • independent of external factors such as temperature

Saoji, S.U., Dua, N. Choudhary discussed that Painting in the air has become one of the greatest intriguing and hard fields of computer vision applications in latest days. This could help the health interaction in a wide range of applications and contributes significantly to the progress of an iterative machine.
Numerous investigations have engaged in the development of novel tactics and techniques to reduce computation while boosting precision and recall.

Bach, B., Sicat, R., Pfister, H. and Quigley discussed that with AR-CANVAS, we introduce the notion of a virtual reality canvas for visualizations. This is different from the typical graphical canvas, which is normally vacant (white), rectangle, and plain. The AIR-CANVAS, but at the other side, refers to that part of a writer's visual field where true items are displayed in-situ for apparent and perhaps undetectable data visualization. Representations for wearable technology must be reconsidered reality due to the canvas's visuospatial complexity. As an example, we was using a library browsing scenario. Describe the AIR-CANVAS' essential features as well as the visualization design dimensions. We'll finish up with a short discussion. Designing visualizations on such a medium presents numerous challenges. The issue has been resolved.

Culjak, I., Abram, D., Pribanic discussed that the objective of this review is to triage and prioritization and educate an user with the foundations of OpenCV (Open Source Computer Vision) and requiring the learners to understand lengthy subsequent purchase or textbooks. Intel first released OpenCV, the open source toolbox for multimedia content analysis, and over a decade earlier. Since then, several programmers have committed to its most recent library improvements. First most latest big update (OpenCV 2) got launched in 2009, but it improved the C++ api significantly. Over 2500 enhanced methods are now included in the bundle. With over 2.5 million installations and over 4000 users, it is broadly applied from around globe.

Wu, Y. and Huang, T.S discussed that the necessity of motion as a natural gui has been driving factor behind analysis into gesture modelling, analysis, and recognition. A Human and Computer compiled intelligent machine, in particular, necessitates vision involved interaction. Many interdisciplinary studies are involved in gesture recognition. This paper provides an overview of contemporary vision-based gesture recognition algorithms. We'll go over how to recognize static hand position and temporal gestures. There are several gesture recognition application systems available. This is also discussed in this study. We'll wrap up with some views on further research avenues.

Van Der Walt, S., Colbert, S.C discussed that in Python, NumPy matrices are the default container for quantitative form. We show how well these arrays improve the accuracy of arithmetic operations in an elevated vocabulary. Three ways are utilised to improve effectiveness: vectorizing computations, minimize data copying in memory, and lowering operation counts.The NumPy matrix format is first presented, followed by demonstrations of how to use it for rapid analysis or how to integrate array information with other tools.

Suarez, J. and Murphy, R.R discussed that This study presented a non-interactive while non-receipt e - voting based on smart contracts, one-time ring signatures, and encryption techniques. During the voting process, the contract is utilised to record, manage, calculate, and check; For ensuring that the enrollment and voting are private, ring signature is used to respect the confidentiality of the digital voting scheme. They employ DPOS for consensus process and the hacker's node to randomize the choices and collect the results for huge voting on the network.

Zabulis, X., Baltzakis, H. and Argyros discussed that People are becoming more aware of the relevance of the voting system as a growing number of votes materialize in real life. The majority of plans are currently centralized.

Howse, J discussed that Images are required for all CV applications. The majority of them must also generate photos as output. A camera as an input source and a window as an output destination can be required for an interactive CV application. Image files, video files, and raw bytes are among the various available sources and destinations. Raw bytes, for example, could be received/sent over a network connection or created by an algorithm if we're using generative graphics in our program.

Gupta, N., Mittal, P., Roy, S.D discussed that a gesture-based interface entails tracking a flowing hand between frames and pulling the gesture's semantic interpretation. This is a difficult task because the hand's position as well as appearance change. Furthermore, such a system should be resistant to changes in gesture pace. This study describes an innovative approach to creating a gesture-based interface. For the moving hand, we present an on-line predictive Eigen Tracker. The eigenspace of our tracker can be learned on the fly. Based on the eigenspace reconstruction error, we offer a new state-based encoding technique for hand motions. As a result, the system is unaffected by the speed of the gesture performed. Learning is used to adapt the gesture recognition system.

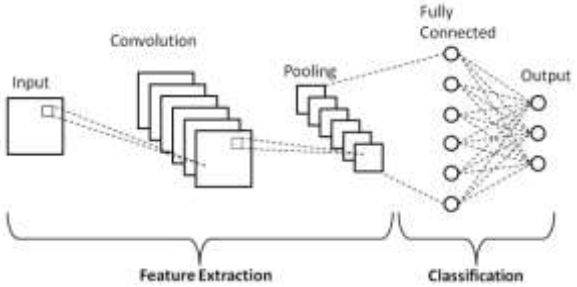
SINGH, A.K discussed that It could be beneficial to be able to keep track of daily schedules and information notes more quickly and simply. I propose the "Text Corner" program, which uses computer webcams to detect and track a pen utilizing Contour-based tracking. The pen's movement is also employed to decipher the user's writings. The user should be able to take short notes, type short messages, and draw simple sketches with their free hand while using this application. In the video frame, the pen's path is traced, and this path is utilized to draw small sketches, type short messages, and sign documents.
The image file can then be saved to a hard drive or emailed. The OpenCV library was used to create this program, which was written in Python. Because as years progress toward mechanization and a better destiny, an automated production system capable of composing text, simple doodles, and brief greetings and sending them via e-mail or storing them in writing or png format has now become a need. Consider the case below: Ram is at business, and he knows to pay the management fee and phone a friend named "Rahul" when he comes home in the evening. It was as simple as opening the app, swinging the pen in the air, and scribbling notes like "pay for things" and "call Rahul." The application of the program.

V. PROPOSED METHOD

Convolutional Neural Network:

A convolutional neural network (CNN) is a type of artificial neural network used in image recognition and processing that is specifically designed to process pixel data.
CNNs are powerful image processing, artificial intelligence (AI) that use deep learning to perform both generative and descriptive tasks, often using machine vison that includes image and video recognition, along with recommender systems and natural language processing (NLP).
The layers of a CNN consist of an input layer, an output layer and a hidden layer that includes multiple convolutional layers, pooling layers, fully connected layers and normalization layers. The removal of limitations and increase in efficiency for image processing results in a system that is far more effective, simpler to trains limited for image processing and natural language processing.

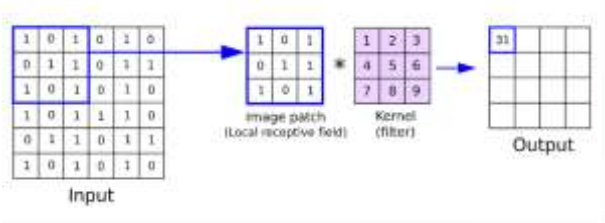
There are two main parts to a CNN architecture:
1. A convolution tool that separates and identifies the various features of the image for analysis in a process called as Feature Extraction
2. A fully connected layer that utilizes the output from the convolution process and predicts the class of the image based on the features extracted in previous stages.



This paper proposes a unique Alphabet Detection system idea within seven phases which are elucidated below:

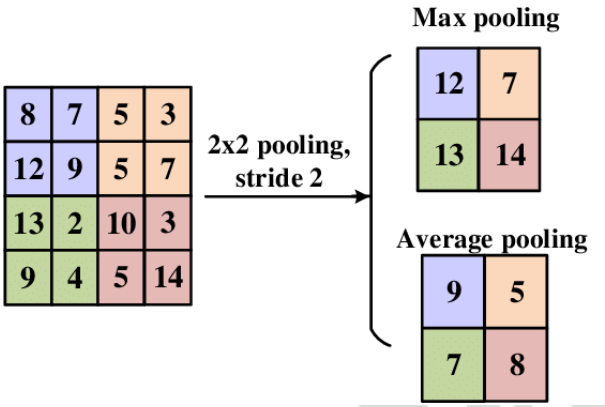
A. Convolutional Layer:

This layer is the first layer that is used to extract the various features from the input images. In this layer, the mathematical operation of convolution is performed between the input image and a filter of a particular size MxM. By sliding the filter over the input image, the dot product is taken between the filter and the parts of the input image with respect to the size of the filter (MxM). The output is termed as the Feature map which gives us information about the image such as the corners and edges. Later, this feature map is fed to other layers to learn several other features of the input image.



B. Pooling Layer:

In most cases, a Convolutional Layer is followed by a Pooling Layer. The primary aim of this layer is to decrease the size of the convolved feature map to reduce the computational costs. This is performed by decreasing the connections between layers and independently operates on each feature map. Depending upon method used, there are several types of Pooling operations.
In Max Pooling, the largest element is taken from feature map. Average Pooling calculates the average of the elements in a predefined sized Image section. The total sum of the elements in the predefined section is computed in Sum Pooling. The Pooling Layer usually serves as a bridge between the Convolutional Layer and the FC Layer.



C. Dense Layer :

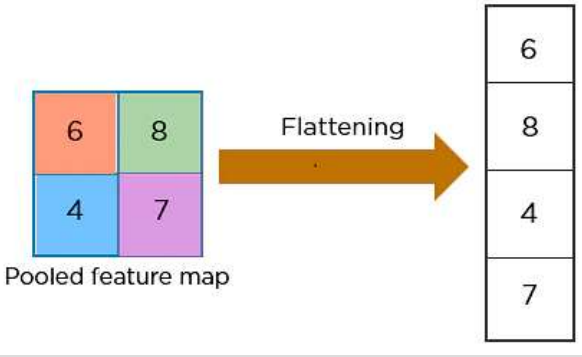
In any neural network, a dense layer is a layer that is deeply connected with its preceding layer which means the neurons of the layer are connected to every neuron of its preceding layer. This layer is the most commonly used layer in artificial neural network networks.
The dense layer's neuron in a model receives output from every neuron of its preceding layer, where neurons of the dense layer perform matrix-vector multiplication. Matrix vector multiplication is a procedure where the row vector of the output from the preceding layers is equal to the column vector of the dense layer. The general rule of matrix-vector multiplication is that the row vector must have as many columns like the column vector.

$$\mathbf{Ax} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix}.$$

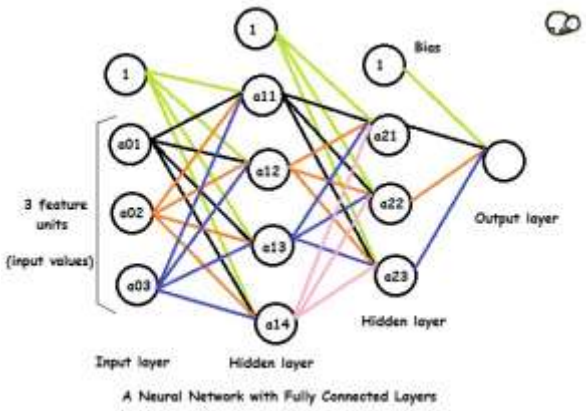
D. Flattening:

Flattening is used to convert all the resultant 2-Dimensional arrays from pooled feature maps into a single long continuous linear vector.
The flattened matrix is fed as input to the fully connected layer to classify the image.



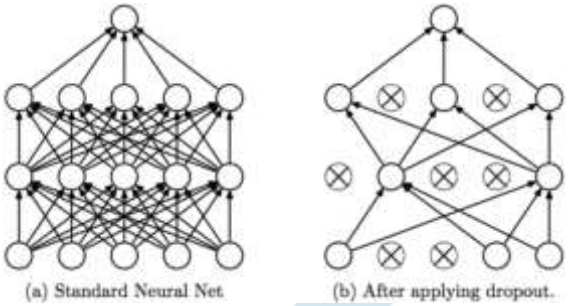
E. **Fully Connected Layer :**

The Fully Connected (FC) layer consists of the weights and biases along with the neurons and is used to connect the neurons between two different layers. These layers are usually placed before the output layer and form the last few layers of a CNN Architecture.
In this, the input image from the previous layers are flattened and fed to the FC layer. The flattened vector then undergoes few more FC layers where the mathematical functions operations usually take place. In this stage, the classification process begins to take place.



F. Dropout:

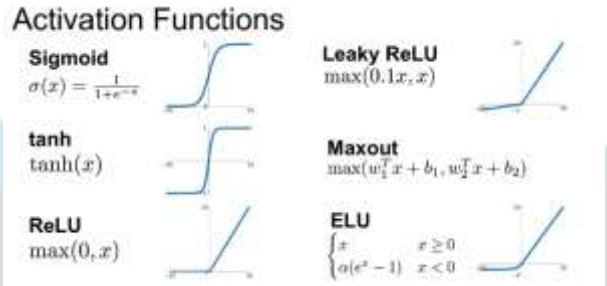
Usually, when all the features are connected to the FC layer, it can cause overfitting in the training dataset. Overfitting occurs when a particular model works so well on the training data causing a negative impact in the model's performance when used on a new data.
To overcome this problem, a dropout layer is utilised wherein a few neurons are dropped from the neural network during training process resulting in reduced size of the model. On passing a dropout of 0.3, 30% of the nodes are dropped out randomly from the neural network.



(a) Standard Neural Net (b) After applying dropout.

G. Activation Functions:

Finally, one of the most important parameters of the CNN model is the activation function. They are used to learn and approximate any kind of continuous and complex relationship between variables of the network. In simple words, it decides which information of the model should fire in the forward direction and which ones should not at the end of the network.
It adds non-linearity to the network. There are several commonly used activation functions such as the **ReLU**, **Softmax**, **tanH** and the **Sigmoid** functions. Each of these functions have a specific usage. For a binary classification CNN model, sigmoid and softmax functions are preferred an for a multi-class classification, generally softmax us used.



VI. CONCLUSION AND FUTURE WORK

The approach has the ability to put standard writing methods to the test. It eliminates the need to write down notes on a mobile phone by giving a convenient on-the-go method of doing so. It will also be very useful in assisting people with disabilities to communicate more freely. Even senior persons or those who have difficulty using keyboards will be able to operate the system with ease. Extending the capabilities, the system will be able to operate IoT devices in the near future. It is also feasible to draw in the air. The system will be an outstanding program for smart wearables, allowing individuals to interact with the digital world more effectively. Text can be brought to life via augmented reality. There are certain flaws in the system that could be addressed. There are several shortcomings in the system that can be addressed in the future. To begin with, employing a handwriting recognizer instead of a character recognizer allows the user to write word by word, which speeds up the writing process. Second, instead of employing the number of fingertips, hand-gestures with a pause can be utilized to control the real-time system, as demonstrated. Finally, our technology occasionally detects and modifies the condition of fingertips in the background. Air-writing systems should solely follow their master's control gestures and not be led astray by others. We also used the EMNIST dataset, which isn't an actual air-character dataset. The accuracy and speed of fingertip recognition can be improved with upcoming object detection techniques like YOLO v3. Artificial Intelligence (AI) will continue to advance in the future.

REFERENCES

[1] Chen, M., AlRegib, G. and Juang, B., 2016. Air-Writing Recognition—Part I: Modeling and Recognition of Characters, Words, and Connecting Motions. IEEE Transactions on Human-Machine Systems, 46(3), pp.403-413.
[2] Chen, M., AlRegib, G. and Juang, B., 2016. Air-Writing Recognition—Part II: Detection and Recognition of Writing Activity in Continuous Stream of Motion Data. IEEE Transactions on Human-Machine Systems, 46(3), pp.436-444.
[3] Itaguchi, Y., Yamada, C., Yoshihara, M. and Fukuzawa, K., 2017. Writing in the air: A visualization tool for written languages. PLOS ONE, 12(6), p.e0178735.

[4] B. Bach, R. Sicat, J. Beyer, M. Cordeil, and H. Pfister. The hologram in my hand: How effective is interactive exploration of 3d visualizations in augmented reality? In IEEE TVCG, 2018.
[5] Yang, H., 2014. Sign Language Recognition with the Kinect Sensor Based on Conditional Random Fields. Sensors, 15(1), pp.135-147.
[6] Aloysius, N. and Geetha, M., 2020. A scale space model of weighted average CNN ensemble for ASL fingerspelling recognition. International Journal of Computational Science and Engineering, 22(1), p.154.
[7] Yurtman, A. and Barshan, B., 2017. Activity Recognition Invariant to Sensor Orientation with Wearable Motion Sensors. Sensors, 17(8), p.1838
[8] Saoji, S.U., Dua, N., Choudhary, A.K. and Phogat, B., 2021. Air Canvas Application Using OpenCV and Numpy in Python. IRJET, 8(08).
[9] Culjak, I., Abram, D., Pribanic, T., Dzapo, H. and Cifrek, M., 2012, May. A brief introduction to OpenCV. In 2012 proceedings of the 35th international convention MIPRO (pp. 1725-1730). IEEE.
[10] Van Der Walt, S., Colbert, S.C. and Varoquaux, G., 2011. The NumPy array: a structure for efficient numerical computation. Computing in science & engineering, 13(2), pp.22-30.
[11] Suarez, J. and Murphy, R.R., 2012, September. Hand gesture recognition with depth images: A review. In 2012 IEEE RO-MAN: the 21st IEEE international symposium on robot and human interactive communication (pp. 411-417). IEEE.
[12] S.Ghotkar, A. and K. Kharate, G., 2013. Vision based Real Time Hand Gesture Recognition Techniques for Human Computer Interaction. International Journal of Computer Applications, 70(16), pp.1-8.
[13] Howse, J., 2013. OpenCV computer vision with python. Birmingham: Packt Publishing.
[14] Gupta, N., Mittal, P., Roy, S.D., Chaudhury, S. and Banerjee, S., 2002. Developing a gesture-based interface. IETE Journal of Research, 48(3-4), pp.237-244.
[15] SINGH, A.K., 2021. Sketch and Signature using Object Tracking.