CNN BASED NUMERAL RECOGNITION IN AIR WRITING

Priyanka Badhe[1], Kalyani Bhalerao[2], Arbaj Aga[3] , S. P. Dhanure[4,], P.K.Suryawanshi[5,], P.V.Igave[6,]
Dept. of E&TC, SKNCOE, Pune-411041, India

[1]priyankabadhe45@gmail.com
[2]kalyanid.bhalerao.kb@gmail.com
[3]Arbaj.aga786@gmail.com
[4]sudhir.dhanure_skncoe@sinhgad.com
preetisuryawanshi10@gmail.com
pradnya.renke@gmail.com

***Abstract*— India is a multilingual and multiscript country. Lot of work done on the character recognition of different languages but there is no much work towards the air writing character and multilingual string recognitionAir-writing is the practise of writing characters or words in free space without the use of a hand-held device, utilising finger or hand movements. Air writing varies from traditional handwriting in that the latter includes the pen-up-pen-down action, whereas the former lacks such a clearly defined sequence of occurrences. Due to the simple writing style, it has a significant advantage over the gesture-based system. However, it is a challenging task because of the non-uniform characters and different writing styles. This system proposes a robust and inexpensive system for recognizing multilingual numeral strings in an air writing environment, which uses only a generic device camera for input. The system is divided into two steps; air written multilingual numeral string segmentation and string recognition using deep learning algorithm. This system uses different Indian language for evaluation. CNN based system is used to classify the multilingual Indian numerals.**

Keywords—. **Air-writing,CNN, gesture-based, human-computer interaction,Vgg16**

I. Introduction

Air-writing systems render a form of gestural human-computer interaction. Such systems are beneficial for building advanced user interfaces that do not require traditional mechanisms of linguistic input such as pen-up-pen down motion, hardware input devices, or virtual keyboards by providing an interface for writing through hand gestures.

Different computer interfaces are used to give commands nowadays for the communication between humans and computers. Most of these are the particular devices that are designed for the human and machine fit. The development of computer vision technologies makes it possible to approach the interface problem from a human perspective, establishing communication between the computer and human more natural. The first task is to develop a system that will recognize hands for enabling real-time hand gesture recognition (HGR) via image. Image contains RGB values. From the hand movement, we draw a real-time handshape in the form of a graph. This graph can provide hand movements or change patterns. Recognizing the hand movement trajectory as an air written English alphabets and digits is the problem to solve.

Air writing may appear to be analogous to online handwriting recognition. The user can raise their hand from the touchpad when writing online. Though, when it comes to air writing, the system cannot tell which movements are parts of writing andwhich are not. As a result, multiple distinct additional strokes are mixed in the actual writing, complicating the classification process.Video-based writing recognition is unrestricted and natural. While in air writing, the hand may contact the face or the body, and the same hue may be confusing. Many studies have utilized distinctive marks around the writing finger to solve the problem. Writing on a surface (rather than a touchpad) feels more natural to people. In the proposed system, we will use a color stripe on the tip of the indexed finger.

Translating signals into the English alphabet is the Challenge in this system. When somebody writes a0, they write it as strokes. The optimal air writing algorithm should be able to distinguish between strokes and air movements. Though, when using air writing, many of the user's different movements coincide with perfect strokes and form part of the writing.

II. LITERATURE SURVEY

Bhattacharya and Chaudhuri [1] have proposed a multistage cascaded recognition scheme using wavelet-based multi-resolution representation to recognize English-Devanagari and English-Bangla mixed numerals. In this work, the authors have performed the similarly shaped numerals resulting in 17 and 16 classes, respectively.

Basu et al. [2] have proposed a method for recognizing postal code in a multi-script environment. They have performed the fusion of the similarly shaped numerals of four scripts, namely Latin (English), Devanagari, Bangla, and Arabic (Urdu), resulting in 25 classes. A quad-tree-based image partitioning is used for feature extraction, and an SVM classifier is used on these 25 classes for recognizing the handwritten numerals.

In [3], handwritten Arabic (Indian) numerals were classified using a two-stage classification technique with Nearest Mean (NMC), K-Nearest Neighbor (K-NNC), and Hidden Markov Models (HMMC) being used in the first stage and a Structural Classifier (SC) being applied in the second stage. They achieve an accuracy of 98% for numeral classification.

Choudhary et al. used a multilayered Perceptron (MLP) with one hidden layer in their work [4], which used a supervised learning technique based on the artificial neural network (ANN) for offline handwritten numeral identification. The tiny data set with few samples is a disadvantage of this study.

In [5], Convolutional Neural Networks (CNN) were proposed for Arabic numeral recognition and introduced a new dataset with 45,000 samples. This produced an average recognition rate of 95.7%.

In [6], the authors proposed a feature set calculated from the vertical and horizontal directions of the image along with the freeman chain code histogram (CCH). SVM was used in the classification phase. The results indicated that the recognition results were enhanced using this method. Deep Learning Algorithms have also been proposed for the recognition of Arabic and Persian numerals.

In [7], the authors developed a CNN to recognize multi-language numerals in the following languages (English, Arabic, Persian, Urdu, and Devanagari). The overall accuracy of the combined dataset was 99.26%, with a precision of 99.29%.

In [8], the authors propose a digit recognition method based on a simplified structural classification using a small set of primitive types and fuzzy memberships. The algorithm extracts five types of primitive segments of each image based on three types of feature points. An estimate of the likelihood of these primitives being close to the vertical boundaries of the image is calculated using the membership function. The classifier in their algorithm uses the primitives, extracted feature points, and fuzzy memberships to classify the digits. Using the NIST dataset for testing, they achieved a recognition rate of 87.33%-88.72%.

In [9], another algorithm is proposed that forms the fuzzy sets from the extracted features. The modified exponential membership function of type-2 is used to represent the fuzzy input sets. Fuzzy measure theory is used to manage the interaction between the fuzzy input sets. The algorithm was tested on the recognition of English and Devanagari digits as well as English characters. In all cases, the algorithm proved to achieve improved recognition over other methods.

In [10], the recognition of Hindi handwritten digit recognition is the target of the study. This algorithm is also based on the modified exponential membership function that is fitted to fuzzy sets. This is derived from the extracted features containing the normalized distances based on the box approach. The foraging model of E. Coli bacteria is used for optimization. The algorithm uses two window sizes: one for zero and one for the rest of the digits. Tested on a small dataset, the average recognition was 96%.

III. PROPOSED SYSTEM

The block diagram of the system is shown in Fig.1.

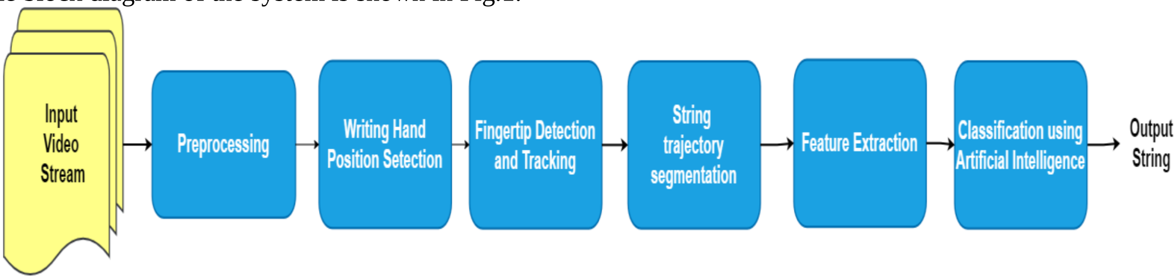


Fig.1. Block diagram of CNN based Numeral recognition in air writing

A. Data Collection

The English numeral dataset is collected from MNIST. The dataset consists of 60000 images with 6000 images of each class. It is well known dataset for digit classification.

B. Pre-processing

The captured image with mobile cameras are noisy; hence pre-processing is required to remove the unwanted noise from the image. The system utilizes a median filter to eradicate the salt and pepper noise. Median filtering is a valuable nonlinear process in reducing impulsive or salt-and-pepper noise. It is also helpful in preserving edges in an image while reducing random noise. Random bit errors in communication channels can cause impulsive or salt-and-pepper noise. A window glides along the picture in a median filter, and the median intensity value of the pixels within the window becomes the output intensity of the pixel being processed.

C. Fingertip detection and tracking

The colored strip placed on the fingertip within the hand area is detected in the HSV colorspace. First, find out the color ranges for the color strip using the thresholding technique and then track the selected tip using the tracking algorithms like optical flow, Kalman, etc. The tracked binary air written string is cropped and saved for further analysis.

D. Classification using CNN and Vgg16 algorithm

In this approach, CNN and Vgg16 algorithms are used to classify signs. This section presents a detailed explanation of the CNN and Vgg16 algorithm.

1) Convolutional Neural Network (CNN)

CNN is a very effective algorithm for classification strategies. It is a feed-forward neural network including convolutional, pooling, flattening, and dense layers. The filter and kernels are used to process the image. Before beginning the training process, it is necessary to learn the fundamentals of CNN, which are detailed below. CNN's are a type of Neural Network that is exceptionally effective at image recognition and categorization CNNs, or large-layer feed-forward neural networks, are one type of large-layer feed-forward neural network.

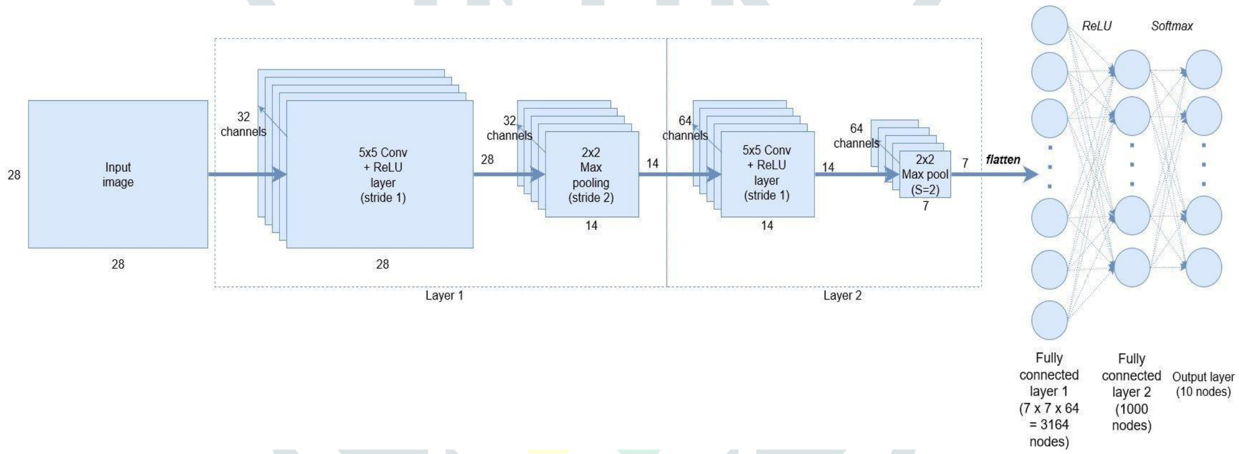


Fig.2. Architecture of CNN

- Convolutional Layer (CL):

A Convolutional Network's core component is the CL. CL's primary goal is to extract characteristics from the data it receives. Convolution preserves the spatial relationship between pixels by learning information from the input image's small kernel. A group of learnable neurons is used to hide the input image.

$$G[m,n] = (f * h)[m,n] = \sum_j \sum_k h[j,k] f[m-j, n-k] \quad (2)$$

- ReLU Layer

ReLU stands for rectified Linear units in a non-linear process. All non-positive feature map values are replaced with zero in a pixel-by-pixel way. To understand how the ReLU works, we'll assume the neuron input is x and the rectifier is given.

$$f(x) = max(0, x) \quad (3)$$

- Pooling Layer

The pooling layer keeps the most relevant information while reducing the complexity of each activation map. A sequence of non-overlapping rectangles is created from the supplied images. A non-linear technique, such as average or maximum, is used to down sample each region. This layer, frequently placed between CLs, improves generalization and convergence speed while also being resistant to translation and distortion.

- Flatten Layer

As the name implies, in this layer 2D pooled feature map is converted into one column array.

- Fully Connected Layer

So, you're going to snap images immediately. Examine each of the 24x24 windows. Apply 6000 different traits to it. Examine the image to find if it's a face or not. The image non-face region comprises the majority of the image. Instead, focus on regions where you could see a face. This allows us to spend more time checking probable areas.

2) Vgg16

The VGG16 model has 16 convolution layers and is a variation of the VGG model. VGGNet-16 has 16 convolutional layers and has a highly homogeneous architecture, which makes it quite appealing. It features only 3x3 convolutions, but several filters, similar to AlexNet.

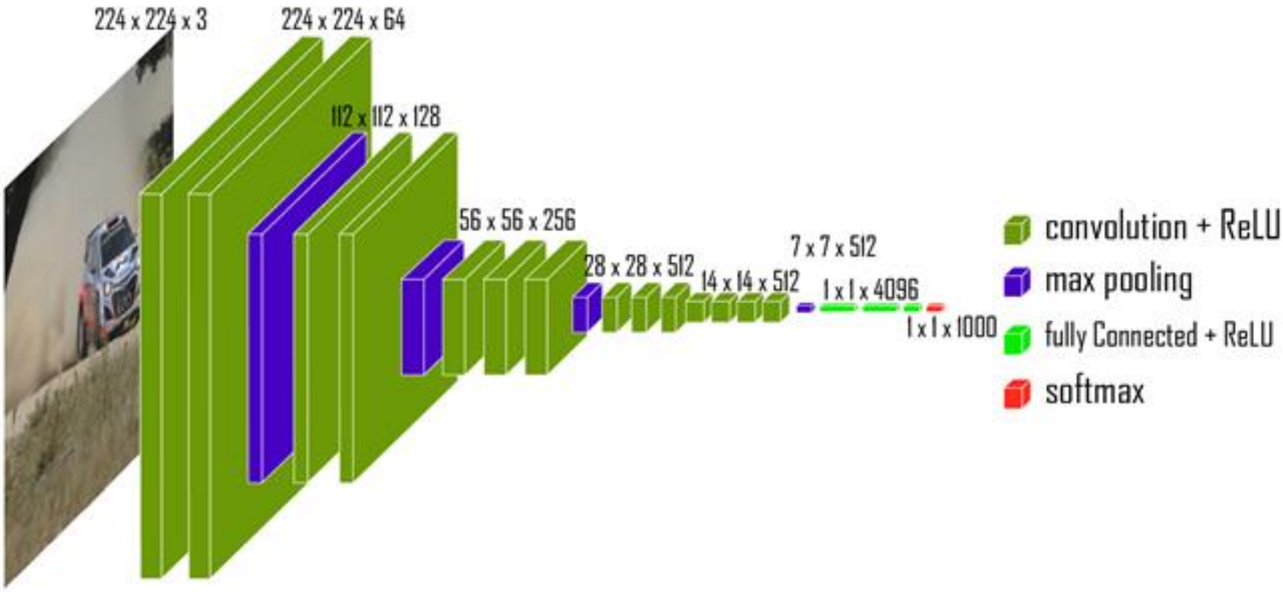


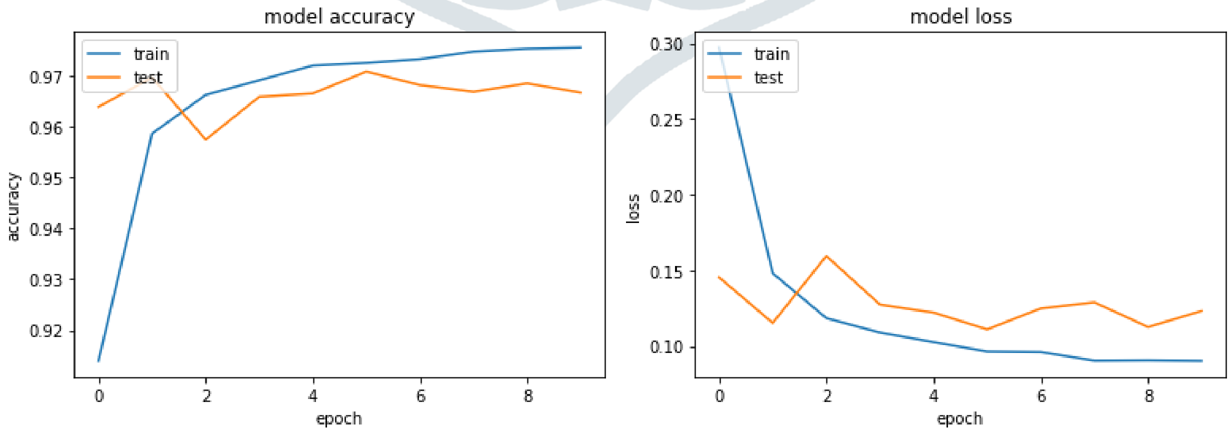
Fig.3. Architecture of Vgg16

VGG16 is a convolution neural net (CNN) architecture which was used to win ILSVR(Imagenet) competition in 2014 The most distinctive feature of VGG16 is that, rather than having a huge number of hyper-parameters, they focused on having 3x3 filter convolution layers with a stride 1 and always used the same padding and maxpool layer of 2x2 filter stride 2. Throughout the architecture, the convolution and max pool layers are arranged in the same way. It has two FC (completely connected layers) in the end, followed by a softmax for output.

The 16 in VGG16 alludes to the fact that it contains 16 layers with different weights. This network is quite huge, with approximately 138 million (estimated) parameters.

IV. Result and Discussion

In this approach, for classification of English digits, CNN and Vgg16 algorithms are used. The training progresses.



(a) (b)

Fig.4. Training progress graph of CNN algorithm (a) Accuracy (b) Loss

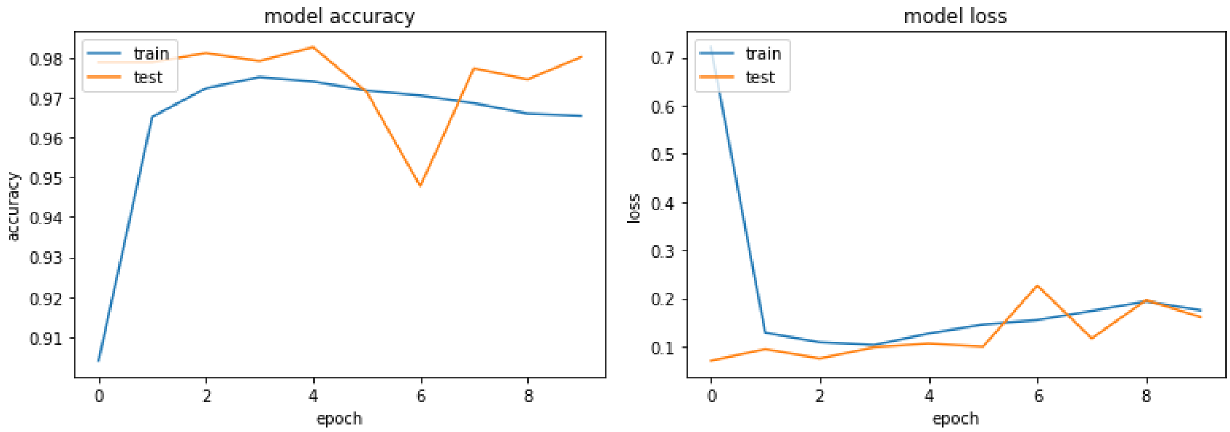


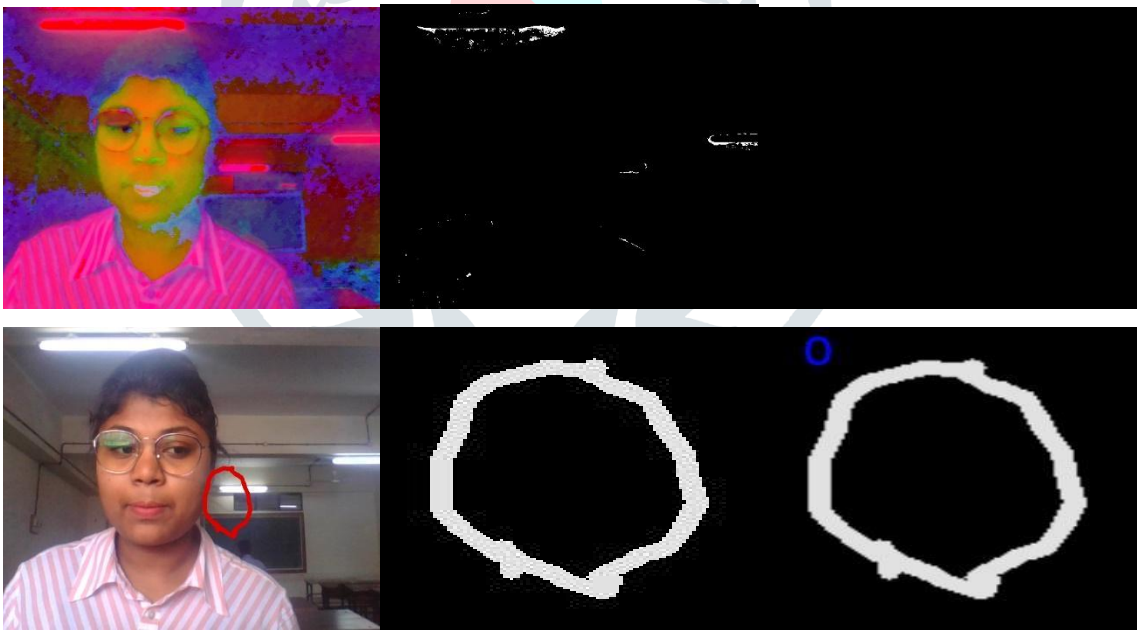
Fig.5. Training progress graph of Vgg16 algorithm (a) Accuracy (b) Loss

The comparative analysis of the CNN and vgg16 algorithm for English numerals recognition system is as tabulated in Table 5.1.

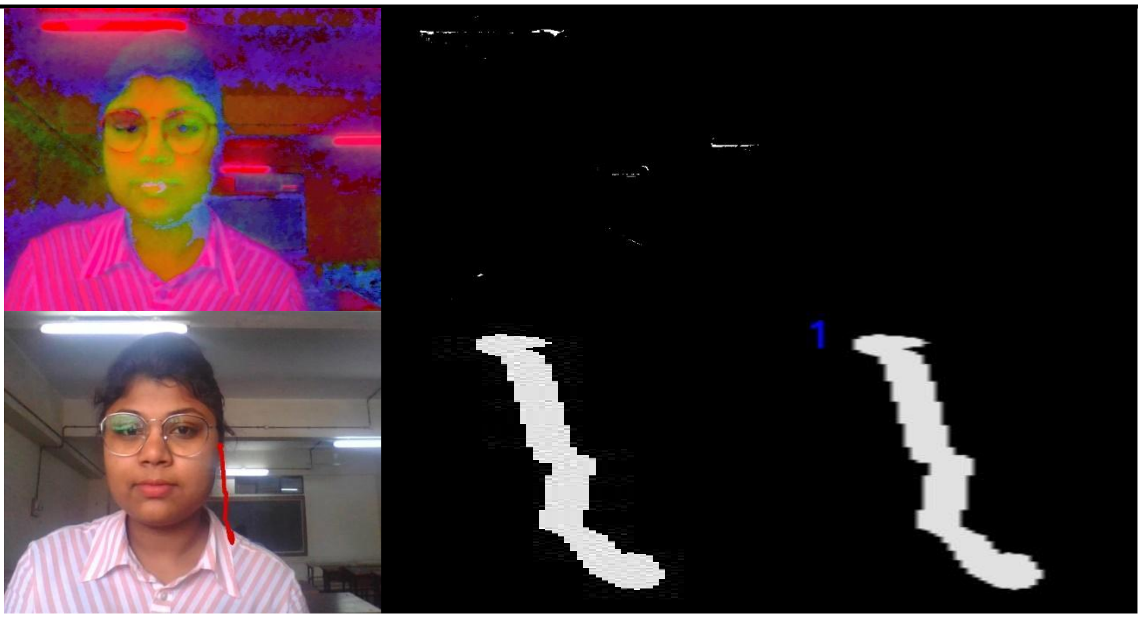
TABLE I
Accuracy of the system

Algorithm	**Training Accuracy**	**Validation Accuracy**	**Training Loss**	**Validation Loss**
CNN	0.9756	0.9668	0.0903	0.123
Vgg16	0.9654	0.9802	0.1751	0.1611

From Table 5 it is observed that the results of the Vgg16 algorithm shows the promising results than the CNN algorithm. The qualitative analysis of the proposed system is as shown in Fig. 6.



(a)



(b)

Fig.6.4. The qualitative analysis of the system in real time

The qualitative analysis of the system shows the promising results in real time. The system is somewhat affected by light condition.

V. Conclusion

In this method, recognition of air writing for English digits has been presented. The system used MNIST English numeral dataset for development of the method. The digit is drawn in the air with the help of blue tip pen. The system converts RGB image into HSV format. The Mask for blue color is applied over the video frame to segment out the blue color. This blue color is track and make a trajectory. This trajectory then segmented and cropped. The cropped air written character is then classified using deep learning algorithms i.e., CNN and Vgg16.

The training and validation performance shows that the system shows accurate and valid results for both algorithms. In this system input will take by camera and output will be displayed on PC. In this system CNN and Vgg16 algorithms are used to train and test the air written English Numerals. The qualitative and quantitative analysis of the system shows that the vgg16 algorithm outperforms the CNN algorithm.

The CNN algorithm achieved a training and validation accuracy of 97.56% and 96.68%, and loss of 0.0903 and 0.1232 while Vgg16 achieved a training and validation accuracy of 96.54% and 98.02% and loss of 0.1751 and 0.1611.

VI. Acknowledgments

It gives us great pleasure in presenting the preliminary project report on ‘Convolutional Neural Network based Method for Numeral Recognition in Air Writing.

I would like to take this opportunity to thank my internal **Guide Mr. S. P. Dhanure and Co-Guide Ms. P. K. Suryawanshi** Department of Electronics and Telecommunication Engineering, SMT. KASHIBAI NAVALE COLLEGE OF ENGINEERING, for his unconditional guidance. I am re- ally grateful to them for their kind support.

We are highly grateful to Dr. S. K. Jagtap Head of Department, Electronics and Telecommunication Engineering, SMT. KASHIBAI NAVALECOLLEGE OF ENGINEERING, for providing necessary facilities during the course of the work.

We admit thanks to project coordinator, Department of Electronics and Telecommunication Engineering, for giving us such an opportunity to carry on such mind stimulating and innovative Project.

References

[1] U. Bhattacharya and B. B. Chaudhuri, "Databases for research on recognition of handwritten characters of Indian scripts," *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*, Vol. 2, 2005, pp. 789-793.

[2] Basu, S., Das, N., Sarkar, R., Kundu, M., Nasipuri, M., &Basu, D. K. "A novel framework for automatic sorting of postal documents with multi-script address blocks. Pattern Recognition," Vol. 43, Issue 10, 2010, pp. 3507–3521.

[3] A. Sabri, M. Marwan, H. Abu-Amara, "Recognition of handwritten Arabic (Indian) Numerals Using Radon-Fourier-based Features," In Proceedings of the 9th WSEAS International Conference on Signal Processing, Robotics and Automation (ISPRA'10), Cambridge, UK, 20–22 February 2010, pp. 158–163.

[4] Choudhary, Amit & Rishi, Rahul & Savita, Ahlawat, "Handwritten Numeral Recognition Using Modified BP ANN Structure," Communications in Computer and Information Science, 2011,

[5] H. A. Alwzwazy, H. M. Albehadili, Y. S. Alwan, and N. E. Islam, "Handwritten digit recognition using convolutional neural networks," International Journal of Innovative Research in Computer and Communication Engineering, vol. 4, no. 2, 2016, pp. 1101-1106.

[6] A. Boukharouba and A. Bennia, "Novel feature extraction technique for the recognition of handwritten digits," Applied Computing and Informatics, vol. 13, no. 1, 2015, pp. 19-26.

[7] G. Latif, J. Alghazo, L. Alzubaidi, M. M. Naseer and Y. Alghazo, "Deep Convolutional Neural Network for Recognition of Unified Multi-Language Handwritten Numerals," 2018 IEEE 2nd International Workshop on Arabic and Derived Script Analysis and Recognition (ASAR), 2018, pp. 90-95.

[8] C. Jou and H. C. Lee, "Handwritten numeral recognition based on simplified structural classification and fuzzy memberships," Expert Systems with Applications, Vol. 36, no. 9, 2009, pp. 11858-11863.

[9] O. V. R. Murthy and M. Hanmandlu, "Interactive fuzzy model based recognition of handwritten characters," Journal of Pattern Recognition Research, vol. 2, pp. 154-165, May 2011.

[10] M. Hanmandlu, A. V. Nath, A. C. Mishra, and A. C. Madasu, "Fuzzy model based recognition of handwritten Hindi numerals using bacterial foraging," in Proc. Int. Conf. Comput. Inf. Sci., 2007, pp. 309-314.