IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, December 2023

A Review Paper on Air Canvas Application and Text-to-Speech System using Deep Learning

Dumbre R. S[1], Gawade S. G[2], Dongare R. P.[3], Dr. Khatri A. A[4], Dr. Gunjal S. D.[5]
Students, Department of Computer Engineering[1,2,3]
Faculty, Department of Computer Engineering[4,5]
Jaihind College of Engineering, Kuran, India
rutujadumbre204@gmail.com, siddhigawade2002@gmail.com, dongarerenuka2151@gmail.com

Abstract: *Air Handwriting is one of the growing technologies in day by day life. Air handwriting enables user to write in air by using finger movements. The web camera detects the finger movements and converts it into readable format. This will be the natural way of communication with computer system. This will remove the need of physical input devices like keyboard, touchscreen and digital pen. This paper aims to provide an effective platform for both communication and practicing. The existing system has some disadvantages that are overcome in this paper. The existing system requires multiple fingers for writing. But using multiple fingers for different tasks like writing, changing color and erasing is complicated thing to remember. A strong, robust and efficient algorithm is proposed that will extract all the air writing trajectories or curves that are collected using a single web camera. The algorithm avoids restrictions on user's writing without using a delimiter and an imaginary box. The deep learning CNN algorithms are also used for converting hand written text into user readable text. Additionally, Optimizing algorithms for efficiency can help ensure smooth and responsive performance.*

Keywords: Air writing, OpenCv, Artificial Intelligence, optical character recognition, Text-to-speech, Handwritten text, color tracking

I. INTRODUCTION

With the fast growth of artificial intelligence technology, many intelligent applications have been developed such as smart TV and intelligent robots[1]. The most natural way for humans to communicate with these intelligent systems is dynamic gestures. The conventional text input method via the keyboard[4], mouse and touchscreen[1] is not as convenient as usual in some applications, such as using whiteboard, though it may greatly affect the user experience of smart devices, but for this we need some physical devices like digital pen. But what if you can give input text just through writing in the air? The user experience must be considerably improved in large amount. It is defined as writing alphanumeric and characters with hand or finger movements in a three dimensional (3D) free space[1]. Air writing is particularly useful for user interfaces that do not allow the user to type on the keyboard or write on the touchpad/touch screen or for text input for intelligent system control. The goal is to enhance accessibility for individuals with physical disabilities or limitations.

II. LITERATURE SURVEY

Air-Writing Recognition Based on Deep Convolutional Neural Networks[1]

The air handwriting can be carried out in three manners: isolated, connected, and overlapped air handwriting. In isolated writing, a letter is written in an imaginary bin with fixed height and width in this field of view of an image, one at a time.[1] In connected writing, multiple letters are written from left to right, which is same to writing on a paper. In the lterminal manner, one can write all kind of letters stacked contiguously one over another in the same imaginary box.[1] This system performs hand tracking only, avoiding the use of complicated procedures for finger tracking. This system deep CNN's for the recognition of air-handwriting digits and special direction symbols for smart-TV-like control.In this system skin and moving both features are combined to detect the moving skin region and then apply the Camshift algorithm to track the moving hand. The network complexity of this proposed neural networks is

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, December 2023

much lower than those of the popular methods, and this systems can operate in real time.[1] 2D camera- base systems often use for color markers on fingers to maximize tracing performance since finger pointing absence markers is challenging. . The two type of data are formed into trajectory datasets, which are used to learn CNN models in the offline training phase.[1] This work goal to develop a easy use yet effective system using a 1D or 2D network that utilizes only the writing trajectory data instead of images.[1] The system highlights the evolution of hand gesture recognition technology, primarily focusing on marker-based methods and their advancements.[1] It begins with 2D technology studies and discusses the steps involved in vision-based 2D hand gesture recognition, including hand/finger detection, tracking, feature extraction, and classification.[1]

Various studies are outlined: Oka et al.'s work using a com- plex device for fingertip tracking, Roy et al.'s simplification using a marker of fixed color for air-writing recognition, and Rahman et al.'s enhancement of marker tip tracking under varying lighting conditions.[1] Rahman et al. introduced a dual network configuration for noise elimination and digit recognition. Moreover, Misra et al. developed a hand gesture recognition scheme using a red marker on the finger, achieving a high recognition rate for various gestures. However, marker- based schemes impose constraints on user behavior, leading to a call for marker-free approaches as a more favorable option due to their flexibility.[1]

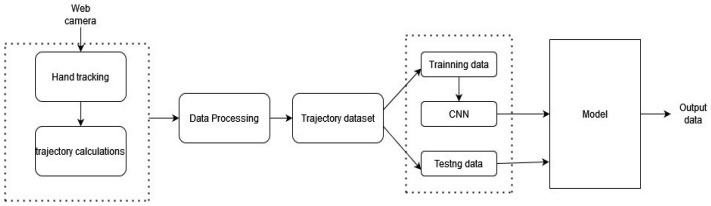

Fig. 1. Architecture of Air-Writing Recognition Based on Deep Convolutional Neural Networks

Effective Emotion Transplantation in an End-to-End Text- to-Speech System [2]

In this system, a successful method is used to transfer emotions in a speech generation framework for a specific speaker. This speaker's speech database only has a neutral style[2]. By updating a pre-trained expressive TTS model in two ways, they not only created the speaker's voice but also kept the expressive qualities of the original model[2]. So, this method effectively gave the ability to express emotions in the target speaker's voice[2]. They give the better result of this method through different measurements and evaluations.[2] The increasing importance of Text-to-Speech (TTS) technology in human-computer interaction systems and the substantial improvements in sound quality and naturalness of synthesized speech, largely attributed to advancements in deep learning. Specifically, end-to-end framework-based TTS models have shown success by directly inferring acoustic features from input character sequences without intricate feature engineering.[2] Researchers are now extending these models to synthesize more expressive speech, which presents variations in prosody compared to emotionally neutral speech.[2] The challenge lies in determining distinctive characteristics of different expressions and representing them using condition vectors to control expressive TTS models.[2] These condition vectors can be handcrafted or learned during the TTS model's training, with learned vectors, such as embedding vectors or latent variables, being mainly used in end-to-end TTS frameworks.[2]

Different methods, including supervised and unsupervised approaches, have been employed to train these embedding vectors for controlling expressive speech synthesis.[2] While supervised methods utilize labeled emotion data for training, unsupervised methods extract embedding vectors directly from speech waveforms during training without annotated labels. These approaches prove beneficial when obtaining labeled data is challenging or when the speech data contains ambiguous styles.[2]Despite the effectiveness of end-to-end TTS models with condition vectors, deploying them in real-world applications faces challenges due to issues related to high- quality expressive speech databases.[2] These databases play a crucial role in training models for synthesizing expressive speech in practical applications.[2] This system focuses on synthesizing expressive speech for a target speaker using only neutral speech data, discovering that as the model adapts, the synthesized speech loses expressiveness.[2] To address this issue, the paper proposes an effective emotion transplantation technique aiming to maintain expressiveness characteristics during

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, December 2023

model adaptation.[2] The technique involves a two-step update procedure for the pre-trained model: first, modifying the model's voice characteristics to match the target speaker's voice and, second, preserving its ability to generate expressive speech.[2] During adaptation, the proposed method synthe- sizes expressive speech from the adapted TTS model using the target speaker's voice and updates the model to minimize an "emotion loss" metric, measuring the distance between the extracted expressive condition vectors from the target speaker's synthesized expressive speech and the input expressive condition vectors used for expressive speech synthesis.[2] As the target speaker lacks expressive speech data, condition vectors are extracted from the source speaker's expressive speech. The model's update is driven to ensure that the emotional style of the synthesized expressive speech matches the emotional style included in the input condition vector. These alternating steps continue until model convergence.[2]

PHTI: Pashto Handwritten Text Image base for Deep Learning Applications [3]

In this system , they made the most complete and largest dataset in the Pashto language to date[3]. The dataset is created by 400 people with diverse characteristics such as gender, age, education, qualification, and profession[3]. It's named PHTI are includes 3,970 scanned pages. These pages are further divided into 36,082 text-line images, and each image has a corresponding ground-truth in UTF-8 codecs. They collect the data from 17 different sources, covering poetry, short stories, history, culture, news, health, and sports. This diverse collection ensures representation of various aspects of the language for comprehensive understanding[3].This system extract text and graphics from document images.[3] This process aims to reduce storage requirements and the high cost associated with exchanging image data. Converting document images allows for easier data exchange, analysis, and storage, thereby conserving resources like time and space.[3] In a standard DIA setup, Optical Character Recognition (OCR) is pivotal for digitizing image documents. However, despite the existence of effective OCR systems for other cursive scripts, the Pashto language lacks a robust system for several reasons, including a lack of sufficient real training data.[3] This paper highlights the crucial role of datasets in recognizing handwritten scripts and emphasizes their significance in the development, evaluation, and comparison of text recognition methods.[3] Creating datasets is labor-intensive, especially when dealing with languages in their early research stages. Researchers aim to construct comprehensive datasets covering diverse aspects of the targeted language.[3]

Despite the growing importance of handwritten text recognition in Document Image Analysis (DIA), there is limited attention given to creating datasets for Pashto handwritten text recognition.[3] This study introduces the Pashto Handwritten Text Image-base (PHTI) dataset, encompassing various genres of Pashto language, such as poetry, short stories, news, reli- gion, culture, jokes, and sports. This dataset includes 4,000 pages of handwritten Pashto materials, segmented into 36,082 text-line images with fully annotated ground-truth.[3] The subsequent sections of the paper review existing datasets related to Pashto in DIA, outline Pashto language features and character sets, detail the creation process of the PHTI dataset, discuss potential applications using supervised learning via the PHTI dataset, and conclude the overall work.[3]

Virtual Canvas for Interactive Learning using OpenCV [4]

This system is a straightforward, enjoyable, and useful tool that lets users write on a screen by easily waving a colored finger or object in the air[4]. While the concept of waving fingers to draw on a screen without touching it might seem unusual initially, it becomes practical when computer vision and Python are combined with color detection techniques for object tracking[4]. The primary aim of this system was to explore the use of computer vision in educational applications[4]. In recent years, air writing has emerged as a challenging and exciting area of research in image processing and pattern recognition. The project utilizes object tracking techniques to create a motion-to-text converter, which could serve as software in education, allowing students and teachers to write in the air. [4].In conclusion, the virtual canvas empowers users to effortlessly draw and write on the screen, fostering effective interaction. This technology challenges conventional writing methods and integrates various technologies to create a robust system.[4] While it thrives with a good camera and clear background, challenges arise in overly bright rooms or cluttered backgrounds, affecting object color detection. Nevertheless, despite these hurdles, the virtual canvas stands out as exceptional software for digital interaction.[4] This model has materialized successfully, paving the way for future enhancements, additional features, and heightened efficiency.[4]

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, December 2023

Air-Writing Recognition using Deep Convolutional and Recurrent Neural Network Architectures [5]

In this suggested system, deep learning architectures are used for recognizing air writing, where a person freely writes text in three-dimensional space[5]. The focus is on handwritten digits, specifically from 0 to 9, organized as multidimensional time-series captured from a Leap Motion Controller (LMC) sensor[5]. Both dynamic and static approaches to model the motion trajectory are examined. Several state-of- the-art convolutional and recurrent architectures are trained and compared. A Long Short-Term Memory (LSTM) network and its bidirectional counterpart (BLSTM) are employed to map the input sequence to a vector of fixed dimensionality. This vector is then passed to a dense layer for classifying the targeted air-written classes[5].

III. DATASET

Due to the nature of the collection process, there are more digital samples than character samples, and the number of samples in each class in the alphabetical part of the data set is almost equal to their frequency in the English language. As a result, four additional sets of these databases have been created to address these issues directly. Similarly, the EMNIST Digits class contains a limited set of Digital Database, which contains 28,000 samples per digit. Finally, the EMNIST-MNIST data set is designed to match exactly the size and specificity of the original MNIST database. It is intended to be a direct substitute for the first MNIST record containing the digits created by the conversion process described. Mainly used for verification.

IV. ALGORITHMS

4.1 Deep Convolution Network

CONVOLUTIONAL NURAL NETWORK A basic CNN is composed of several convolutional layers for feature extrac- tion, each of which is usually followed by a pooling layer. The last convolutional layer is also followed by one or more fully connected (dense) layers for classification. For the 1D and 2D trajectory data stated above, we design a 1D-CNN and 2D- CNN, respectively, to recognize the input digits (or directional symbols). The typical architectures of our proposed 1D-CNN and 2D-CNN [1] for recognizing digits are shown in FIGURE 5(a) and FIGURE 5(b), respectively, and consist of several 1D or 2D convolutional blocks. The architectures for directional symbols are similar; hence, they are neglected here. Each convolutional block contains convolution, maximal pooling, batch normalization, and activation function [1]. The CNN (1D or 2D) applies batch normalization after convolution and before activation because it helps to improve the performance and stability of neural networks[1].

4.2 TEXT-TO-SPEECH

Speech synthesis is used by developers to create voice robots. Deep learning speech synthesis uses Deep Neural Networks (DNN) to produce artificial speech from text (text- to-speech) or spectrum [2]. The deep neural networks are trained using a large amount of recorded speech and, in the case of a text-to-speech system, the associated labels and/or input text.Each sentence can be pronounced differently depending on the meaning and emotional tone. To understand the right pronunciation, the system uses built-in dictionaries

V. APPLICATIONS

Air handwriting is a Gesture-Based input giving method. This can be simple and very effective tool of communication. Anyone can easily write without physically touching a screen. This technology can be used in various fields:

5.1 Human-Computer Interaction (HCI)

This technology focuses on motion-to-text converter. Gesture-based handwriting can improve the way people interact with computers and digital wearable devices. It can be used for text input, controlling applications, and navigating user interfaces using natural hand movements.

5.2 Virtual Reality (VR) and Augmented Reality (AR)

Air handwriting can be used in VR and AR for giving input and draw diagrams. This can be used for gaming to control the movements of game play virtually.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, December 2023

5.3 Accessibility and Assistive Technology:
Air handwriting can profitable for individuals with physical disabilities who may face difficulties using traditional input devices like keyboards or touchscreens. It allows them to input text and interact with devices using hand or finger gestures.

5.4 Presentation and Collaboration
During presentations or collaborative meetings, air hand- writing can be used to draw diagrams, annotate slides, or write notes in the air, which can then be displayed on a screen for all participants to see. The work can be saved for future use.

5.5 Healthcare
In healthcare settings, air handwriting can be used by doctors and surgeons to annotate medical images, take notes, or navigate through digital patient records in a hands-free manner, reducing the risk of being contaminated.

5.6 Education
Air handwriting can be used in educational applications to teach handwriting skills to children, especially in environments where physical contact with surfaces is discouraged, such as during a pandemic.

5.7 Art and Design
Artists and designers can use air handwriting as a digital sketching and drawing tool, allowing for more natural and expressive creative processes.

VI. CONCLUSION

In this system, we have proposed deep CNNs for the recognition of digits, characters and display on screen. The MINIST dataset is use to match the handwritten text with characters. the Text-to-Speech algorithm is used to convert the plane text into speech. Main goal is to eliminate the need for traditional input devices like keyboards and touchscreens by allowing users to write in the air using hand movements. A robust air-writing detection algorithm based on a web camera is developed that performs hand tracking only, avoiding the use of complicated procedures for finger tracking. By using Deep CNN we aim to improve the accuracy and robustness of the technology, making it more reliable and responsive.

REFERENCES

[1]. CHAUR-HEH HSIEH 1,YOU-SHEN LO2,JEN-YANG CHEN AND SHENG-KAI TANG2, ”Air-Writing Recognition Based on Deep Convolutional Neural Networks” October 21, 2021
[2]. YOUNG-SUN JOO 1,2, (Member, IEEE), HANBIN BAE1 , YOUNG- IK KIM1 , HOON-YOUNG CHO1 , AND HONG-GOO KANG 2 ,
[3]. (Member, IEEE)”Effective Emotion Transplantation in an End-to-End Text-to-Speech System ” September 4, 2020.
[4]. IBRAR HUSSAIN 1,2, RIAZ AHMAD2 , SIRAJ MUHAMMAD2 , KHALIL ULLAH 3 , HABIB SHAH 4 , AND ABDALLAH NAMOUN
[5]. 5 , (Member, IEEE) ”PHTI: Pashto Handwritten Text Imagebase for Deep Learning Applications”,25 October 2022.
[6]. Palak Rai ,Reeya Gupta ,Vinicia Dsouza ,Dipti Jadhav , “Virtual Canvas for Interactive Learning using OpenCV ”, Oct 7-9, 2022
[7]. Grigoris Bastas, Kosmas Kritsis and Vassilis Katsouros, ‘Air-Writing Recognition using Deep Convolutional and Recurrent Neural Network Architectures” 2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR).
[8]. Pranavi Srungavarapu, Eswar Pavan Maganti, Srilekkha Sakhamuri, Sai Pavan Kalyan Veerada, Anuradha Chinta, ”Virtual Sketch using Open CV ”, IJITEE, Volume-10 Issue-8, June 2021