Journal of Harbin Engineering University
ISSN: 1006-7043

Vol 46 No. 06
June 2025

AIR HANDWRITING BY USING CNN MODEL

Kalyani Sampat Gawande, Dr. D. M. Yadav

Department of Computer Engineering

S.N.D College of Engineering & Reserch Center Nashik, Maharashtra

Principal
ME (Electronics), Ph.D (Electronics)

Department of Electronics Engineering

S.N.D College of Engineering & Reserch Center Nashik, Maharashtra

Abstract*:* Gesture recognition has gained significant attention due to the growth of IoT and smart device technologies. Among its various challenges, air-writing—writing characters in mid-air using hand movements—stands out as a complex yet important task. This paper presents a wearable air-writing system that allows users to write English letters freely in three-dimensional space, without requiring strict adherence to writing conventions. The system is based on data from an Inertial Measurement Unit (IMU) and uses Dynamic Time Warping (DTW) as the core algorithm for gesture recognition. To improve accuracy and optimize the combination of IMU data with DTW, we propose a novel adjustment method that enhances system performance. Experimental evaluations show that the system achieves a recognition accuracy of 84.6%. Moreover, the results reveal that the effectiveness of DTW-based recognition varies between users, highlighting the importance of adapting the system to individual usage patterns.

Keywords—air-writing, inertial measurement unit, dynamic time warping, gesture recognition.

1. INTRODUCTION

The air writing recognition system captures words written in the air by the user using a PC's webcam and employs a convolutional neural network (CNN) to classify the captured input into predefined categories. While many existing systems rely on expensive and sophisticated tracking setups for gesture recognition, this approach aims to achieve similar performance using a more affordable and accessible setup.Air writing is a full-body activity, especially beneficial for helping children practice writing uppercase and lowercase letters in an engaging, physical manner. The process involves standing upright, extending the arm straight with the elbow locked, and tracing letters in the air with the index finger.In modern technology, air writing serves as a natural human-computer interaction method where users form characters or words through hand or finger gestures in free space. Despite its intuitiveness, recognizing these gestures accurately is difficult due to the variability in individual writing styles and differences in gesture speed, which impact the consistency of the captured data.To address these challenges, the proposed system uses a standard webcam to capture air-writing inputs and supports recognition of both alphabets and digits. For performance evaluation, two datasets were created: one for alphabet characters and another for numeric digits. Data were collected from 17 participants who wrote each letter (A–Z) and digit (0–9) multiple times, ranging from 5 to 10 repetitions.During data acquisition, fingertip positions were tracked using MediaPipe technology. The visual data was processed through a CNN to extract spatial features, while the temporal sequences of movements were analyzed by a Bidirectional Long Short-Term Memory (BiLSTM) network to understand time-based patterns. The results from both networks were combined and further refined through 5-fold cross-validation, leading to improved accuracy in recognizing air-written characters.

2. RELATED WORKS

Air handwriting, also known as air-writing, utilizes gesture recognition and augmented reality technologies to enable users to draw or write in three-dimensional space without traditional tools.

This study presents a novel approach using a **Convolutional Neural Network (CNN)** to interpret air-drawn gestures, offering a seamless and interactive method for engaging with digital systems without physical contact.The air-writing system proposed in this work begins by collecting gesture data that represent letters, symbols, and numbers through hand movements. These gestures serve as training input for a CNN designed to identify and classify air-written content. To ensure the model performs reliably across various handwriting styles and environments, data augmentation techniques are applied—introducing variations in speed, orientation, and angle of gestures. This not only increases the dataset size but also strengthens the model's ability to generalize across users.Another related approach involves recognizing air-written alphabets using **wearable devices** integrated with **Inertial Measurement Units (IMUs)**. In such systems, **Dynamic Time Warping (DTW)** is employed to match gesture trajectories over time. While effective, these methods often depend heavily on individual user behavior, making them highly personalized solutions.Meanwhile, the advancement of sensor technologies has made it easier to capture finger and joint movements accurately in 3D space. Building on this progress, researchers have developed **trajectory-based recognition systems** using CNNs, focusing on identifying characters drawn in the air based on movement paths.Expanding on this idea, some recent models combine **CNN and Long Short-Term Memory (LSTM)** networks—forming a hybrid architecture referred to as **CNN-LSTM**. These models have been tested on publicly available datasets such as the RealSense Trajectory Digit (RTD) and RealSense Trajectory Character (RTC) datasets, which include tens of thousands of digit and character samples. The CNN-LSTM framework typically includes stacked convolutional layers with pooling, followed by LSTM layers to capture temporal patterns, and dense layers for classification. To mitigate overfitting, a dropout rate (e.g., 0.4) is applied in the final dense layers.

LITERATURE SURVEY

1. **CNN-Based Framework for Air-Writing Numeral Recognition**
Air-writing involves forming characters in mid-air using hand movements with full spatial freedom (six degrees of freedom). This work presents a CNN-powered recognition framework that leverages a standard video camera to interpret unistroke numerals. A colored marker is used by the user to write in the air, and the system applies color-based image segmentation to isolate and track the marker tip's movement, effectively capturing the gesture trajectory for recognition.

2. **Wearable System for Air-Writing Recognition Using DTW**
With the growing integration of IoT and smart devices, gesture recognition—particularly air-writing—has emerged as a significant research focus. This study introduces a wearable system that enables users to draw English letters in 3D space without following strict handwriting rules. The system utilizes an Inertial Measurement Unit (IMU) to track motion data, and **Dynamic Time Warping (DTW)** is employed to compare gesture patterns for effective recognition.

3. **Trajectory-Based Character Identification Using CNN**
Air-writing can be described as writing letters or numbers in space using finger or handheld markers. Unlike conventional handwriting, this method relies on tracking hand or finger motion in a 3D plane. Advances in sensor technology have significantly improved the accuracy of capturing such gestures. This research proposes a trajectory-based character recognition approach that utilizes a **Convolutional Neural Network (CNN)** to analyze and classify these gesture patterns.

4. **Air-Writing Recognition via Hybrid CNN-LSTM Network**
This study presents an air-writing recognition system that combines CNN with **Long Short-Term Memory (LSTM)** networks to capture both spatial and temporal features of handwritten characters. Two public datasets—RealSense Trajectory Digit (RTD) and Character (RTC)—containing 20,000 digits and 30,000 characters respectively, were used for training. The architecture consists of two CNN layers with pooling, followed by two LSTM layers and two dense layers. A dropout rate of 0.4 was applied to avoid overfitting. The system achieved high recognition accuracy (99.63% for digits, 98.74% for characters), with each character processed in approximately 14 milliseconds, making it suitable for real-time applications.

5. **Enhanced Digit Recognition with CNN Models**
Traditional handwriting recognition systems relied heavily on manually designed features and domain-specific knowledge, making them complex to develop and less adaptable. Recent progress in deep learning, particularly with CNNs, has revolutionized the field. This work focuses on enhancing handwritten digit recognition by leveraging CNN architectures, taking advantage of modern computational power and large-scale handwritten datasets to boost both accuracy and scalability.

6. **Air-Writing Implementation Using Python** Air-writing has become a captivating and complex research domain within image processing and pattern recognition. It plays a vital role in advancing automation and improving human-computer interaction across various sectors. This study focuses on developing an air-writing system using Python, aiming to elevate the efficiency and intuitiveness of gesture-based interfaces for real-world applications.

3. Proposed System & System Architecture

Air handwriting combines gesture recognition and augmented reality to enable users to write or sketch in three-dimensional space, eliminating the need for traditional input devices. This research presents a novel approach that utilizes a **Convolutional Neural Network (CNN)** to interpret these air-drawn gestures, offering a natural and efficient method for human-computer interaction.The system begins by capturing hand motion data representing letters, digits, and symbols. This gesture data is used to train a CNN specifically designed to recognize various writing styles and adapt to different user behaviors and movement patterns. To improve generalization and robustness, data augmentation techniques are applied—introducing variations in orientation, speed, and direction for each gesture sample.The proposed deep CNN model consists of multiple layers—including convolutional, pooling, and fully connected layers—that collaboratively extract and process features from the input gesture sequences. **Convolutional layers** focus on preserving spatial characteristics of the movement, while **pooling layers** help in reducing dimensionality and computational load, making the system efficient for real-time processing.The final classification is performed using a **Softmax activation function**, which enables the model to accurately categorize input gestures into multiple character classes. This layered architecture ensures that spatial variations are effectively learned and mapped to their corresponding symbols or letters, enabling high recognition accuracy in diverse and dynamic conditions.

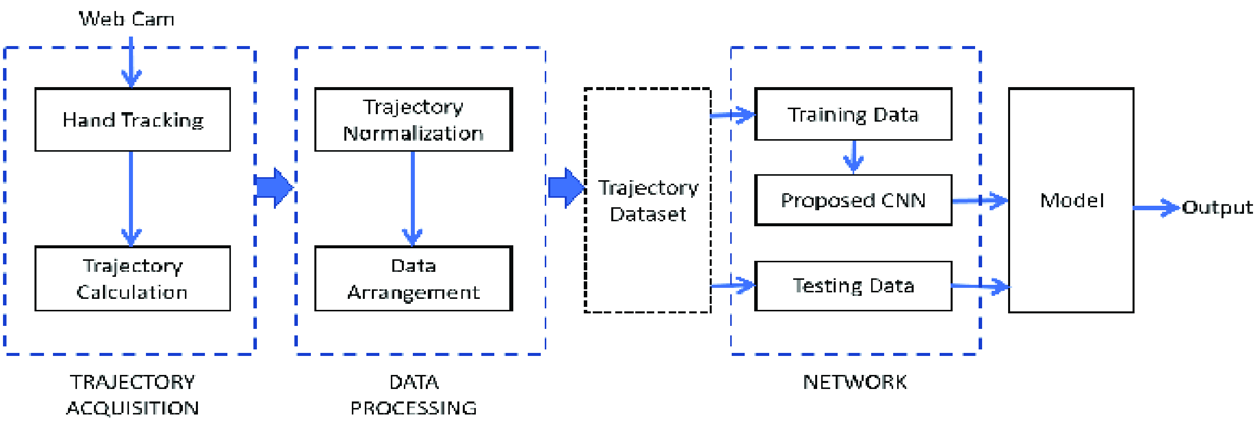

Fig 1. proposed air-writing recognition system.

This system addresses the air-writing challenge, which traditionally depends on depth-sensing technologies like Kinect and LEAP Motion, or on wearable motion-detection devices such as the Myo armband. In some cases, multi-camera systems are also used to derive depth-related data. Although these methods typically provide accurate tracking and reliable results, their reliance on specialized external hardware makes them less practical and cost-efficient for widespread or everyday use.

Journal of Harbin Engineering University
ISSN: 1006-7043
Vol 46 No. 06
June 2025

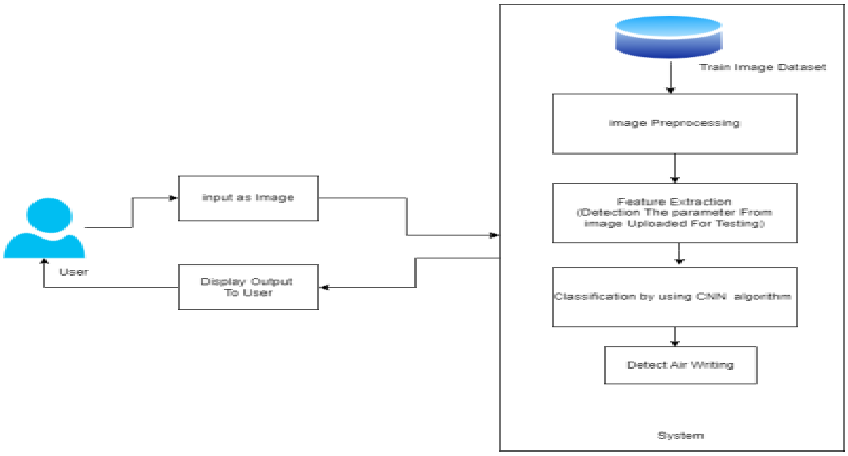

Fig 2. System Architecture

Image Processing:

Image processing refers to the technique of performing operations on images to enhance them or to extract meaningful information. It falls under the broader category of signal processing, where the input is an image, and the output might be a transformed image or data derived from the image, such as patterns, features, or other visual characteristics.

CNN Algorithm:

Convolutional Neural Networks (CNNs) are a specialized architecture within deep learning, primarily designed to handle visual inputs like images. While there are several types of neural networks in deep learning, CNNs are particularly effective for tasks involving image recognition and pixel-level analysis. These networks are modeled after the visual processing structure of the human brain and are widely utilized in fields such as image classification, object detection, and image segmentation. A CNN typically comprises several layers, including convolutional layers that extract spatial features, pooling layers that reduce dimensionality, and fully connected layers that perform classification. Figure 3 illustrates a standard CNN architecture.

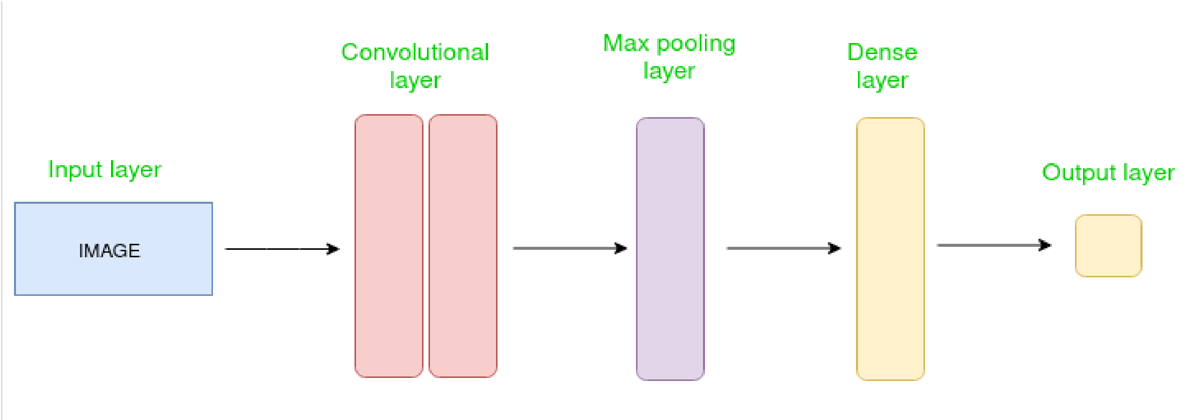

Fig 2. CNN Architecture

4. METHODOLOGY

The creation of an air handwriting system powered by Convolutional Neural Networks (CNNs) involves a structured methodology to ensure reliable gesture interpretation, classification accuracy, and responsiveness for real-time applications. The development process includes several stages—from data acquisition and preprocessing to model design, training, and deployment.

A. Data Collection and Preparation

1. Gesture Capture:

The first step involves recording hand gestures using either sensor-based or camera-based setups that track movements in three-dimensional space. In this project, a high-definition camera was utilized to collect gesture sequences that correspond to various alphabets, digits, and special characters.

2. Annotation and Augmentation:

Each recorded gesture is tagged with the appropriate label to denote the character it represents. To improve the model's adaptability across different handwriting styles, the dataset is enhanced using augmentation techniques such as rotating, scaling, and shifting.

3. Preprocessing:

To prepare the gestures for CNN input, the image data undergoes preprocessing steps including resizing, converting to grayscale, and normalizing pixel values. This ensures consistency in input dimensions and improves the training process.

B. CNN Model Design

1. Model Selection:

A CNN was selected for its proven ability to recognize spatial patterns in visual data. The model architecture consists of several layers that work together to extract and interpret features from input gestures.

2. Layer Configuration:

The network is organized into three key components:

- **Convolutional Layers** – Capture spatial features and edges from the input.
- **Pooling Layers** – Reduce data dimensionality while retaining key features.
- **Fully Connected Layers** – Final decision-making layers for classification.

3. Activation and Loss Functions:

Rectified Linear Unit (ReLU) functions are used in hidden layers to introduce non-linearity. The output layer employs the Softmax function to perform multi-class classification. Cross-entropy loss is used to measure model error during training.

C. Integration for Real-Time Use

1. Real-Time Gesture Detection:

Once the model is trained, it is deployed in a real-time system where a live camera captures hand motions. The CNN processes this input and identifies the characters or symbols being written in the air.

2. Output Visualization:

Recognized characters are shown on a display or projected onto a surface, allowing users to see the system's interpretation of their input immediately.

3. Feedback and Correction:

To improve recognition accuracy over time, the system incorporates a feedback mechanism. If a gesture is misclassified, the user can correct it, enabling the model to adjust and learn from errors continuously.

ALGORITHM & FLOWCHART DETAILED

To implement air handwriting recognition effectively, we designed an algorithm based on a Convolutional Neural Network (CNN) tailored to identify and classify hand gestures in real time. This method encompasses all critical stages—from capturing hand motion to translating these into digital handwriting that can be shown on a display or projected surface. The approach relies on a structured data pipeline and a CNN architecture optimized to extract spatial features from gesture sequences.

A. Algorithm for Air Handwriting Recognition Using CNN

1. Data Collection:

- Utilize a high-resolution camera to record hand gestures that correspond to letters, numbers, and symbols.
- Gather gesture data either live or from a pre-labeled dataset representing each character or symbol.

2. Data Preparation:

- Convert images of gestures to grayscale to reduce processing demands.
- Resize images uniformly (e.g., 64x64 pixels) to maintain consistent input dimensions for the CNN.
- Normalize pixel values to a 0–1 scale, enhancing model training stability.

3. Data Augmentation:

- Apply transformations such as rotation, scaling, and shifting to mimic diverse handwriting styles.
- Keep labels consistent for all augmented data to preserve dataset accuracy.

4. CNN Model Setup:

- Build a CNN comprising multiple convolutional layers, pooling layers, and fully connected layers.
- Match the input layer size with the preprocessed image dimensions (e.g., 64x64x1).
- Use a Softmax activation function in the output layer to classify input into relevant categories (e.g., alphabets A–Z, digits 0–9).

5. Model Training:

- Divide the dataset into training, validation, and test subsets, typically around 70%, 15%, and 15%.
- Train the model with an appropriate batch size and learning rate, optimized through experimentation.
- Compute cross-entropy loss for each epoch and update weights using the Adam optimizer.
- Evaluate performance on the validation set periodically and adjust hyperparameters to avoid overfitting.

6. Model Testing:

- Assess the model's effectiveness using the test dataset by measuring accuracy, precision, and recall.
- Verify processing speed to ensure suitability for real-time gesture recognition.

7. Real-Time Recognition:

- Embed the trained CNN within a real-time system that continuously captures hand gestures.
- Preprocess each new gesture frame and predict the corresponding character or symbol.
- Immediately display the prediction to provide user feedback.

8. Feedback and Model Improvement:

- Include a mechanism for users to correct misclassified gestures.
- Periodically retrain or fine-tune the model based on user corrections to personalize and improve accuracy.

B. Algorithm Flowchart Steps

1. Begin Air Handwriting Process
2. Capture Hand Gesture Data
3. Preprocess Images
4. Perform Data Augmentation
5. Initialize CNN Architecture
6. Train CNN with Dataset
7. Evaluate Model Performance
8. Preprocess Input and Predict
9. Display Recognized Output
10. Receive User Feedback & Update Model
11. End Process

RESULTS & OUTPUTS

Figure 5: Alphabetical Result

Figure 5: Numerical Result

5. CONCLUSION

In conclusion, this paper introduces an innovative air-writing system driven by an inertial measurement unit (IMU) that enables users to write in the air using gesturesThe system utilizes Dynamic Time Warping (DTW) as the core algorithm for recognizing gestures, enabling effective comparison of time-series data and enhancing the precision of character recognition. To boost the system's accuracy and reduce the processing time for each gesture, an adaptive adjustment mechanism has been introduced. This is especially important since users interact with such systems in diverse and changing environments.

A notable aspect of this adjustment mechanism is the integration of an idle-cutter technique. This feature tackles a common issue in gesture recognition: involuntary hand tremors or unintended movements caused by lack of user control or unconscious actions. These disturbances often compromise recognition accuracy, and the idle-cutter offers a practical solution to minimize their impact.

Furthermore, the study proposes a multi-template strategy that innovatively optimizes the use of DTW for gesture recognition. This method is flexible and applicable beyond merely analyzing hand angles, indicating its potential for broader use in various gesture-related technologies.

Experimental results show that in user-specific settings—where the system is customized to an individual's gestures—the air-writing system achieves high accuracy and consistency. This outcome highlights the effectiveness of the proposed approach and its potential to advance gesture recognition technologies.

6. FUTUREWORK

The future potential of air handwriting recognition is extensive and could significantly impact various sectors, especially with the ongoing progress in augmented reality (AR) and virtual reality (VR) technologies. In the field of education, this technology has the ability to transform how teachers and students interact by enabling collaboration in virtual environments without the need for physical writing instruments. It can greatly enhance remote learning by allowing participants to write, draw, and annotate in a shared digital space, making learning more engaging and immersive.

Furthermore, integrating air handwriting into virtual teamwork platforms can improve brainstorming, project management, and creative processes, facilitating smooth interaction regardless of participants' physical locations. The technology also shows great promise for accessibility and assistive applications. For people with physical disabilities, air handwriting can serve as an empowering communication tool, enabling interaction with digital devices through simple hand gestures.

As gesture recognition continues to improve and models become more customizable to individual users, air handwriting could become a natural and intuitive interface in smart homes, healthcare, and Internet of Things (IoT) environments. With advancements in sophisticated algorithms, including deep learning and reinforcement learning techniques, future air handwriting systems may evolve into context-aware platforms capable of interpreting not only isolated characters but also complex commands and expressions in real time.

REFERENCES

[1] Yuqi Luo, Jiang Liu," Wearable Air-writing Recognition System employing Dynamic Time Warping" Shigeru Shimamoto School of Fundamental Science and Engineering Waseda University Tokyo, Japan luoyuqi1997@toki.waseda.jp, jiang@waseda.jp, shima@waseda.jp.

[2] Aradhana Kumari Singh a , Deepika Koundal " A Temporal Convolutional Network for modeling raw 3D sequences and air-writing recognition" Received 1 October 2023; Received inrevised form 20 November 2023; Accepted 25 November 2023 Available online 28 November 2023.

[3] Prasun Roy, Subhankar Ghosh, and Umapada Pal "A CNN Based Framework for Unistroke Numeral Recognition in Air-Writing" 2018 16th International Conference on Frontiers in Handwriting Recognition Computer Vision and Pattern Recognition Unit, Indian Statistical Institute 203, B. T. Road, Kolkata, WB - 700108, India.

[4] Md. Shahinur Alam1, Ki-Chul Kwon, and Nam Kim2 "Trajectory-based Air-writing Character Recognition Using ConvolutionalNeuralNetwork" 2019 4th International Conference on Control, Robotics and Cybernetics (CRC Dept. of Computer and Communication Engineering Chungbuk National University.

[5] Z. -T. Liu, D. P. Y. Wong and P. H. Chou,"An Imu-Base WearableRing for On-Surface Handwriting Recognition," 2020 InternationalSymposium on VLSI Design, Automation and Test (VLSI-DAT), 2020,pp. 1-4, doi: 10.1109/VLSIDAT49148.2020.9196479.

[6] Savita Ahlawat 1, Amit Choudhary 2, Anand Nayyar 3, Saurabh Singh 4 and Byungun Yoon 4, "Improved Handwritten Digit Recognition Using Convolutional Neural Networks (CNN)" Received: 25 May 2020; Accepted: 9 June 2020; Published: 12 June 2020

.

.