Air-Typing Virtual Keyboard on NXP i.MX 8M Plus

Nejneriu Mihai
Technical University “Gheorghe Asachi” of Iasi, Faculty of Automatic and Computer Engineering

Abstract

The Air-Typing Virtual Keyboard project presents a touchless interface designed for the i.MX 8M Plus EVK, addressing the limitations of traditional physical peripherals in sterile or space-constrained environments. The system utilizes MediaPipe for real-time hand landmark detection and a custom neural network trained on skeletal coordinates to classify right-hand gesture intent. By implementing a multi-threaded architecture and specific collision logic for the left hand, the project enables high-accuracy typing and navigation without physical contact, ensuring privacy and low-latency operation.

Introduction

Traditional keyboard setups are suboptimal in scenarios where hygiene, space, or accessibility are primary concerns. This project proposes a solution using a standard camera connected to an embedded platform to detect hand gestures and map fingertip positions to virtual keypresses. The scope focuses on providing immediate feedback through a visual overlay, ensuring local processing to maintain user privacy and eliminate cloud-based latency.

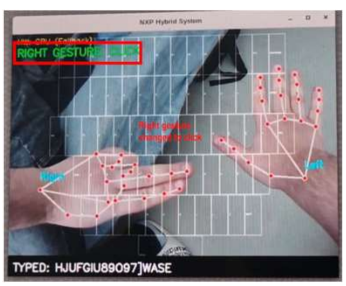

Figure 3. Left Click

Methods and Materials

The system architecture centers on a multi-threaded Python implementation where one thread captures live frames while a second thread processes hand landmarks every few milliseconds to optimize CPU load. Data Collection: A dedicated collector script was used to record approximately 600 samples (200 per class) of 21 hand landmarks, providing the 63 inputs required for training. Neural Network: A three-layer model was trained to classify the right hand into three distinct outputs with high accuracy. Typing Logic: For the left hand, a pressing action is detected by calculating the Euclidean distance between internal finger points; a key is triggered only when these points are sufficiently close. Input Validation: The system incorporates a "leeway" mechanism requiring a key to be held for a set duration to be registered and a release-before-repeat rule to prevent accidental duplicate inputs.

Results

The final implementation achieves excellent accuracy in gesture recognition and precise fingertip-to-key collision detection within the 640x480 resolution camera view. An interactive keyboard overlay with enlarged text ensures high readability on the external monitor, and a dedicated visual console bar at the bottom of the frame displays the typed text string in real-time. The neural network successfully identifies the three navigational labels for the right hand, allowing the user to switch modes without interfering with the typing process. The multi-threaded approach allows the image to be displayed live while maintaining a controlled processing rate for the AI logic, preventing CPU saturation on the NXP board. Precise interaction is further ensured by a geometric typing system for the left hand that calculates Euclidean distances between finger joints to distinguish intentional presses from hovering motions. This stability is reinforced by a press leeway mechanism and a release-to-repeat rule that together prevent accidental triggers and unintended character streaming during high-speed use.

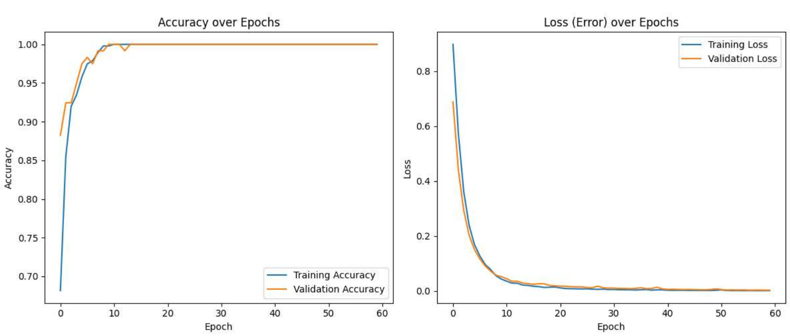

Figure 4. Results

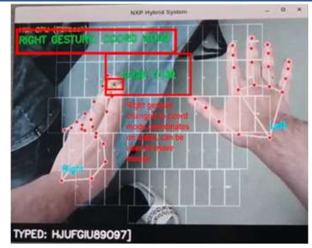

Figure 1. Coord Mode

Figure 2. Right Click

Discussion

The technical implementation focuses on combining geometric analysis with machine learning to ensure high input precision. By calculating Euclidean distances between finger joints, the system effectively distinguishes intentional presses from incidental hovering motions. This stability is reinforced by a multi-threaded framework that prevents CPU saturation while maintaining a fluid live display on the external monitor. Specific mechanisms like press leeway and release-to-repeat further filter out false triggers, creating a reliable and responsive interaction model.

Conclusions

This project proves that sophisticated touchless interfaces can be successfully executed on embedded hardware using only standard webcam data and optimized software. By processing skeletal landmarks locally, the system offers a high-performance alternative to physical peripherals while ensuring user privacy and low latency. The final implementation serves as a robust proof of concept for contact-free technologies in environments where hygiene and space constraints are critical.

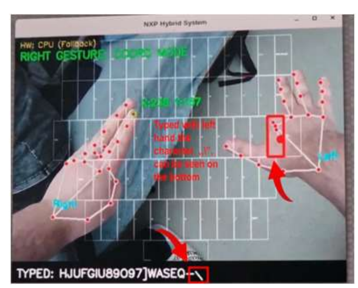

Figure 5. Typing with left hand

Contact

Nejneriu Mihai

References

1. https://arxiv.org/abs/2006.10214
2. https://research.google/blog/on-device-real-time-hand-tracking-with-mediapipe/
3. https://www.mouser.com/pdfDocs/AN12964.pdf
4. https://www.nxp.com.cn/docs/en/user-guide/IMX_ANDROID_TENSORFLOWLITE_USERS_GUIDE.pdf
5. https://www.nxp.com/docs/en/user-guide/UG10338.pdf