GestureBridge: Edge AI-Based Sign Language Recognition System for Communication Assistance

Yiming Zhu

Contact Information:

Yiming Zhu

Undergraduate Student, Department of Electrical and Computer Engineering

The Ohio State University

Email: zhu.3941@buckeyemail.osu.edu

Abstract:

This project, titled “GestureBridge”, aims to develop a low-power, real-time sign language recognition system based on Edge AI technology. Designed to run on resource-constrained microcontrollers, the system utilizes a convolutional neural network (CNN) trained with gesture image data and optimized using TinyML techniques for on-device inference. The goal is to empower communication for the deaf and hard-of-hearing community by enabling gesture-to-text or gesture-to-speech translation without relying on cloud computing.

Unlike existing cloud-based solutions, “GestureBridge” emphasizes privacy, latency reduction, and offline accessibility. The project will initially support static gesture recognition and later expand to pseudo-dynamic gesture sequences, with future plans for full 3D modeling and temporal gesture tracking. Through this phased development approach, the system is designed to

grow into a comprehensive, accessible, and socially impactful communication aid.

This work also serves as a research foundation aligning with broader interests in computer vision, embedded systems, and human-computer interaction, and is intended to integrate with advanced academic research in visual computing and real-time AI applications.

Background & Motivation：

Communication barriers pose significant challenges for deaf and hard-of-hearing individuals, limiting their access to information and social interaction. Sign language serves as a vital means of communication within this community, yet its use remains isolated from mainstream technology due to the lack of accessible, real-time translation systems.

Recent advances in computer vision and edge AI present new opportunities to develop portable, affordable, and privacy-preserving sign language recognition systems. However, existing solutions often rely on cloud computing or require expensive hardware, limiting their practical deployment.

Motivated by the need to bridge this accessibility gap, my project *GestureBridge* aims to leverage lightweight deep learning models deployed on embedded devices. This approach enables real-time, on-device sign language recognition, empowering users with seamless communication

tools. By integrating 3D dynamic gesture modeling and edge computing, the project seeks to advance assistive technologies that enhance social inclusion and improve quality of life for the deaf community.

Project Objectives:

- Develop a real-time, low-power hand gesture recognition system deployable on microcontrollers.
- Achieve reliable static gesture classification using TinyML for fully offline inference.
- Expand to dynamic gesture recognition via frame sequencing and temporal modeling.
- Integrate 3D modeling and multi-modal sensing for enhanced spatial accuracy.
- Enable semantic-level gesture-to-language translation using lightweight LLMs.
- Promote accessibility and inclusivity for deaf and hard-of-hearing individuals through deployable assistive technologies.

Technical Architecture:

GestureBridge adopts a progressive, multi-tier architecture designed to adapt to varying computational capabilities while maintaining core functionality across

deployment stages:

- Tier 1: Embedded Edge (Prototype Phase)
    - Platform: ESP32-CAM
    - Function: Static hand gesture recognition using TinyML
    - Limitations: Limited memory (≤ 512KB RAM), low compute power (240MHz), not suitable for dynamic gestures
    - Purpose: Proof-of-concept for ultra-low-power on-device inference
- Tier 2: Lightweight Edge–Cloud Hybrid (Transition Phase)
    - Architecture: ESP32-CAM captures frame, sends them via Wi-Fi/BLE to a smartphone or micro-edge processor
    - Phone: Performs temporal modeling (LSTM/CNN-seq), gesture sequence classification, optional TTS
    - Advantage: Improves temporal accuracy while preserving mobility and energy efficiency
    - Example: A mobile app receives gesture sequences and infers intent locally using lightweight LLM or offline NLP modules
- Tier 3: High-Capability Edge (Expansion Phase)
    - Platform: Jetson Nano / Raspberry Pi + USB camera or stereo camera

    - Function: Dynamic gesture classification, multi-frame inference, 3D modeling
    - Benefit: Enables richer interaction, real-time 3D hand mesh estimation, larger ML models
- Tier 4: Multi-Modal & Semantic AI Integration (Vision Phase)
    - Inputs: RGB + depth cameras, IMUs
    - Integration: Lightweight LLM for intent understanding, real-time speech generation
    - Outcome: True bidirectional communication—gesture-to-language + language-to-gesture in a standalone system

This architecture ensures long-term scalability while allowing development and testing to begin on low-cost devices. The goal is to balance affordability, performance, and user accessibility across hardware tiers.

Innovation & Differentiation:

- On-device Inference First: Unlike cloud-dependent systems, GestureBridge performs real-time inference directly on low-power devices like the ESP32-CAM, ensuring privacy, speed, and offline usability.
- Phased Architecture: The system adopts a staged deployment from static recognition to semantic-level communication, aligning with real-world

hardware constraints and scalability.

- 3D Hand Pose Integration: Unique use of CAD-based modeling and depth sensors enables accurate 3D gesture tracking in low-cost systems.
- Semantic Intelligence via LLMs: GestureBridge goes beyond classification by incorporating intent understanding and natural language generation, a novel feature for sign language systems at the edge.
- Accessible and Open: Focused on affordability and deployability in underserved regions, this system is built using open-source tools and consumer-grade hardware.

Risks and Challenges

1. Hardware Constraints

   Deploying real-time gesture recognition on low-power embedded devices such as ESP32-CAM presents inherent trade-offs between computational capability and power consumption. Achieving sufficient inference speed and model complexity without increasing device size or heat dissipation remains a key challenge. Careful optimization and lightweight model design are required to balance accuracy and resource usage.

2. Data Collection and Representation

   Most publicly available hand gesture datasets are captured from a third-person perspective, focusing on outward-facing hand images. In contrast,

our system aims to support wearable, user-facing devices (e.g., necklace or neck-mounted cameras), resulting in limited availability of training data from this novel viewpoint. This scarcity necessitates the integration of 3D hand modeling and synthetic data generation to augment datasets and improve model generalization. Developing robust 3D hand pose reconstruction methods will be critical to overcoming this barrier.

3. Model Accuracy and Generalization

   While initial static gesture recognition models are expected to achieve reasonable accuracy, extending to dynamic and user-specific gestures under diverse environmental conditions will require further refinement. Incorporating multi-modal sensor data and advanced temporal modeling techniques will be explored to enhance robustness and performance.

Phases：

Phase 1: Static Gesture Recognition on ESP32-CAM using TinyML

This phase focuses on developing a lightweight convolutional neural network (CNN) model for static hand gesture recognition on the ESP32-CAM platform. The model will be trained and optimized using TinyML techniques to enable fully on-device inference under strict memory and power constraints. This stage establishes the foundation for efficient, edge-based visual recognition in real-time.

Phase 2: Continuous Gesture Recognition via Lightweight Offloading

To support dynamic gesture interpretation, this stage introduces temporal modeling by increasing frame rate and enabling frame sequence analysis. While the ESP32-CAM remains the primary sensor, heavier computational tasks are offloaded to lightweight companion devices, such as smartphones or portable processors. This phase ensures improved temporal resolution while maintaining portability and thermal efficiency.

Phase 3: Advanced Embedded Inference on More Capable Edge Platforms

With more sophisticated models, the system is migrated to more powerful embedded platforms such as Jetson Nano or Raspberry Pi. The goal is to enhance the capacity for dynamic, multi-frame inference while preserving the form factor necessary for user-deployed assistive systems.

Phase 4: Integration with 3D Modeling and Multi-Modal Sensing

In the final phase, the project incorporates 3D hand gesture modeling using hybrid sensor inputs,such as depth cameras to enhance spatial accuracy. Leveraging experience with Onshape and CAD design, this phase enables the development of realistic, adaptable 3D hand models and potentially physical assistive devices. The integration of hybrid 3D modeling aligns the system with real-world deployment requirements and supports future interdisciplinary expansion.

Phase 5: Semantic Gesture Understanding and Language Generation via LLMs

The final phase aims to empower the system with semantic understanding by interpreting the user's intent behind dynamic gestures. By integrating lightweight large language models (LLMs) via mobile devices, the system can translate recognized gestures into meaningful spoken language. This enables the device to function as a bidirectional communication assistant, expressing the user's intended message naturally via voice synthesis or text on-screen—further closing the gap between gestural communication and spoken language.

Skills and Technical Readiness：

I have practical experience in Python programming and basic familiarity with C++, which supports my development of machine learning models and embedded applications. I have used OpenCV and PyTorch in simple image recognition projects, gaining foundational knowledge in computer vision and deep learning.

My skills in 3D modeling with Onshape and CAD tools assist in designing custom hand pose representations.

Though I am still building my understanding of photogrammetry and 3D reconstruction, I am highly motivated to deepen my expertise. My academic background in electrical and computer engineering, combined with hands-on practice, has prepared me to take on challenges in real-time gesture recognition and edge AI deployment.

I am eager to learn and collaborate across disciplines to strengthen my technical capabilities.

Research Relevance and Fit:

The proposed project, *GestureBridge*, focuses on developing an edge AI-based sign language recognition system for deaf and hard-of-hearing users, utilizing computer vision, gesture tracking, and efficient on-device inference (TinyML), with planned extensions to dynamic gesture recognition and 3D modeling.

This work aligns with the lab's emphasis on scene modeling and understanding, particularly the transformation of raw data into usable 3D representations such as meshes and level-of-detail models. My experience with Onshape and CAD tools supports the design and prototyping of 3D hand pose representations and assistive devices, bridging gesture recognition with practical 3D modeling.

Furthermore, the project's goal of dynamic 3D gesture modeling resonates with the lab's research on optimizing Structure-from-Motion and photogrammetry workflows. Addressing the computational constraints of embedded devices, it reflects the lab's focus on efficient, accurate, and scalable 3D reconstruction. Integration of multi-modal sensor data, including depth and LiDAR, further complements the lab's hybrid 3D modeling expertise.

Finally, *GestureBridge* exemplifies interdisciplinary application by combining computer vision, edge AI, and assistive technology to mitigate communication

barriers for the Deaf community. This aligns with the lab's collaborative approach to solving real-world problems across diverse domains, making the project a pertinent extension of the lab's interdisciplinary research efforts.

Appendix：