Smart Skies: Real-Time Gesture Detection in Miniature Drone

Priyanka Lal, Johnny Pham, Justin Wang
Advisors: Mozhgan Navardi, Daniel Mendat, Andreas Andreou

JOHNS HOPKINS
WHITING SCHOOL of ENGINEERING

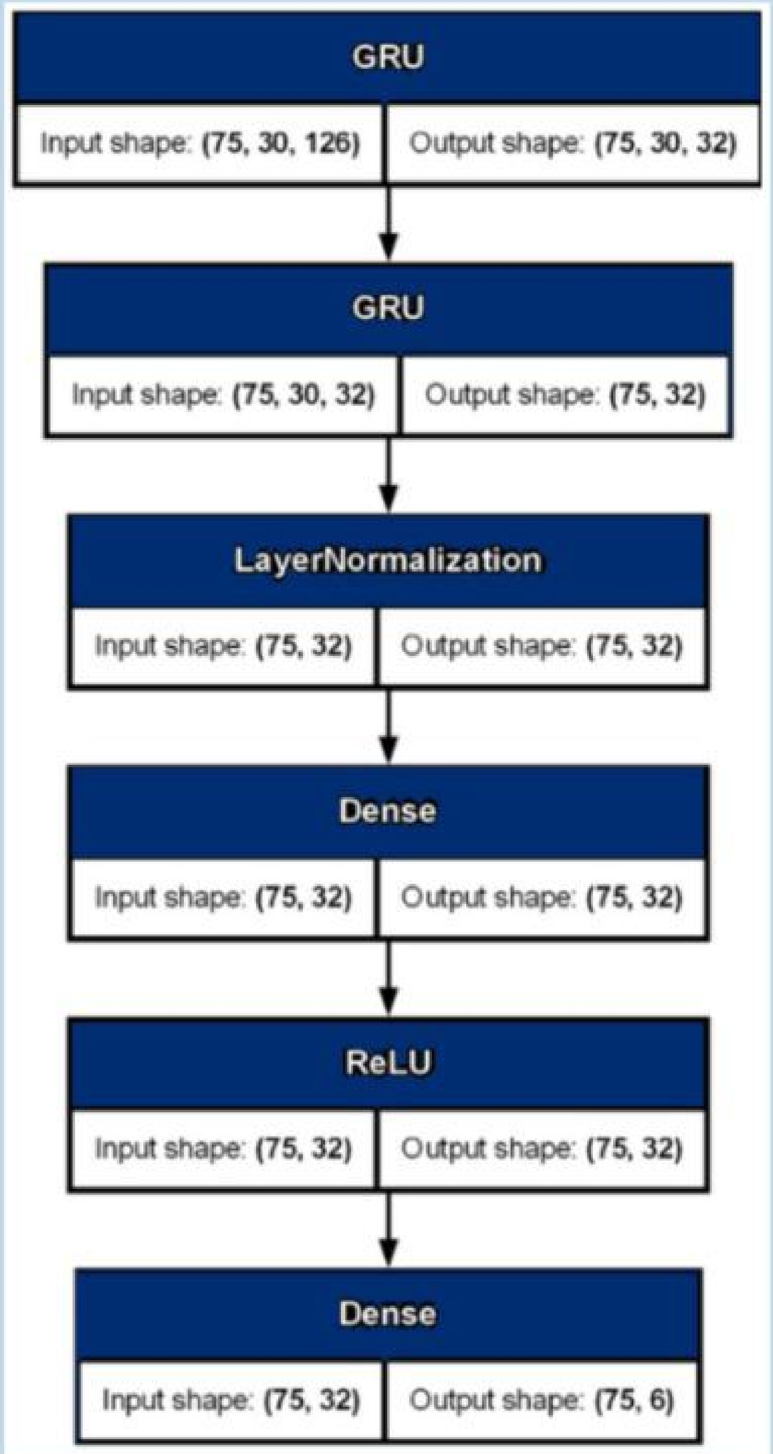
OUR VISION

AI at the Edge

AI at the edge enables real-time processing directly on devices like microcontrollers, reducing latency and dependency on cloud connectivity. This approach enhances privacy, lowers bandwidth usage, and allows true real-time inference and actions for autonomous machines.

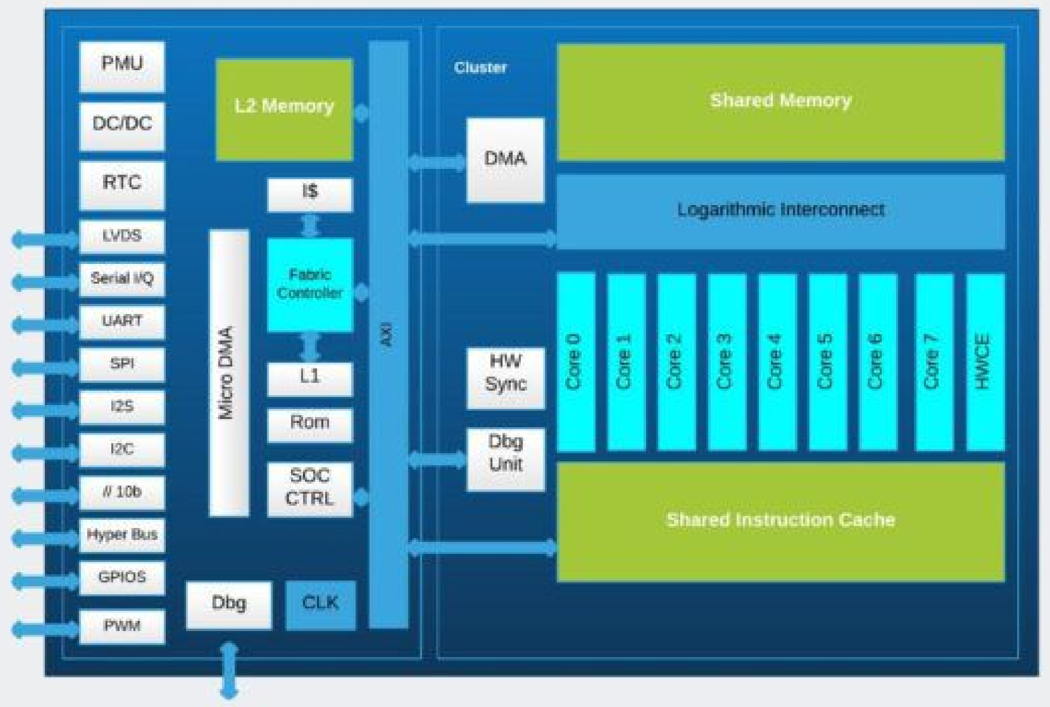
Camera-based Gesture Controlled Flight

In this project, we explore the steps and challenges in deploying a model on the Crazyflie drone to allow the drone's flight to be controlled through hand gestures. We believe this will open the path towards an all-in-one drone that does not require a physical controller or wireless connection, allowing for a truly modular and discrete personal drone.



The model processes sequences of hand gesture coordinates using two stacked GRU layers to capture temporal patterns in the motion. The normalization layer stabilizes the output, and the dense layer with ReLU activation enhances feature representation. The final dense layer maps the processed features to six gesture classes corresponding to different drone movements.

ONBOARD INTELLIGENCE: GAP8 AI DECK



The AI-deck onboarded on the Crazyflie consists of the GAP8 chip developed by Greenwaves Technologies, **a parallel, ultra-low-power and flexible RISC-V based IoT application processor** for the TinyML ecosystem, which includes MAC (multiply and accumulate) units for accelerating deep learning inference.

It builds on open source components from the PULP (Parallel Ultra-Low-Power Processing-Platform) project, an open source multi-core computing platform developed by ETH Zurich and the University of Bologna, achieving leading-edge energy-efficiency and featuring widely-tunable performance. It also leverages the flexibility and openness the RISC-V ISA, a rich set of peripherals, secure execution, associated with a powerful parallel processing engine for flexible multisensor (image, audio, inertial) data analysis and fusion.

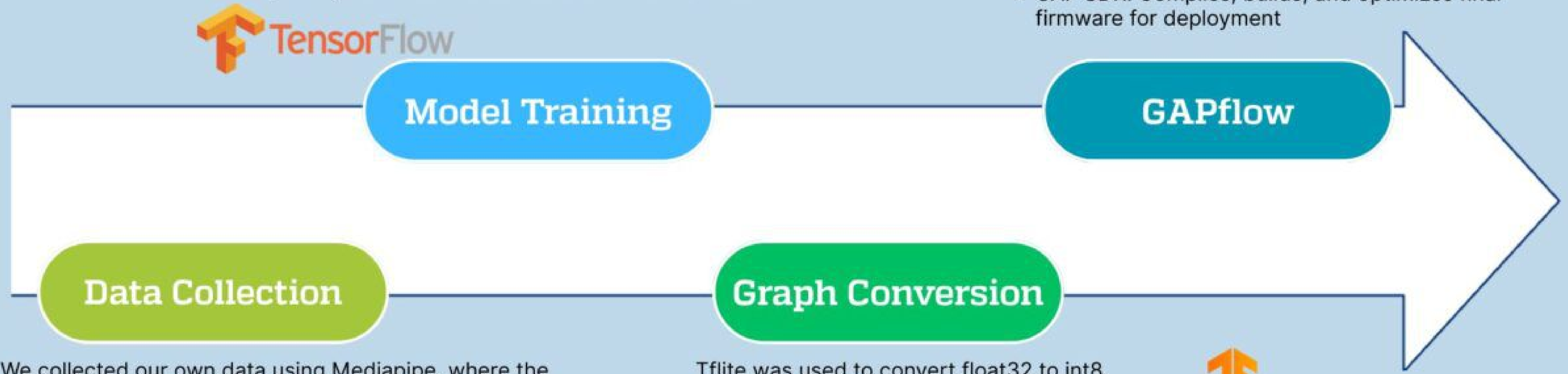
When the cluster runs CNN-based applications, it can offload convolutional layers computation to a dedicated accelerator, the Hardware Convolution Engine (HWCE). This block can evaluate a 5×5 convolution or three 3×3 convolutions on 4-, 8- or 16-bit operands in a single cycle.

METHODOLOGY

Given that the data was sequential data, which was beneficial for gesture recognition, the GRU was used for the model. It provided a good balance of performance and memory usage, which is crucial in such limited resource environments like the Crazyflie. Normalization and dropouts were used within the model to prevent overfitting, especially for such a small model and small dataset.

Using GreenWaves' GAPflow toolchain, the .tflite can be transformed into GAP8-optimized C code, ready for flashing to the AI deck through tools like:

- NNTool: Converts model layers into C code
- AutoTiler: Splits computation and memory usage for GAP8's clustered architecture
- GAP SDK: Compiles, builds, and optimizes final firmware for deployment



We collected our own data using Mediapipe, where the 21 key features detected from the hand movement wre transferred into numpy arrays. There are a total of 30 frames comprising of 5 static and 1 dynamic gestures, each with 18 data points.

Tflite was used to convert float32 to int8, which reduced the size of the model even further without majorly impacting the performance of the model

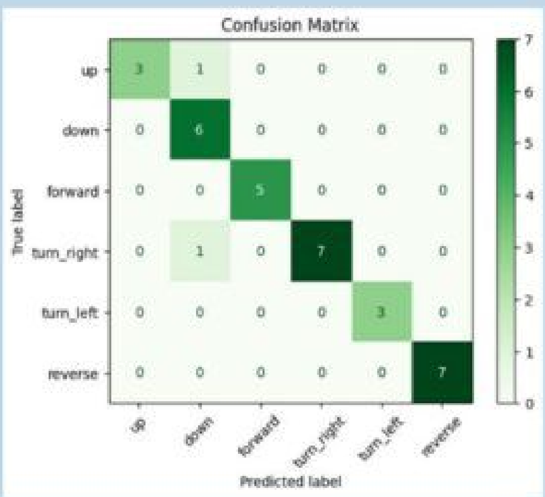
TensorFlow Lite

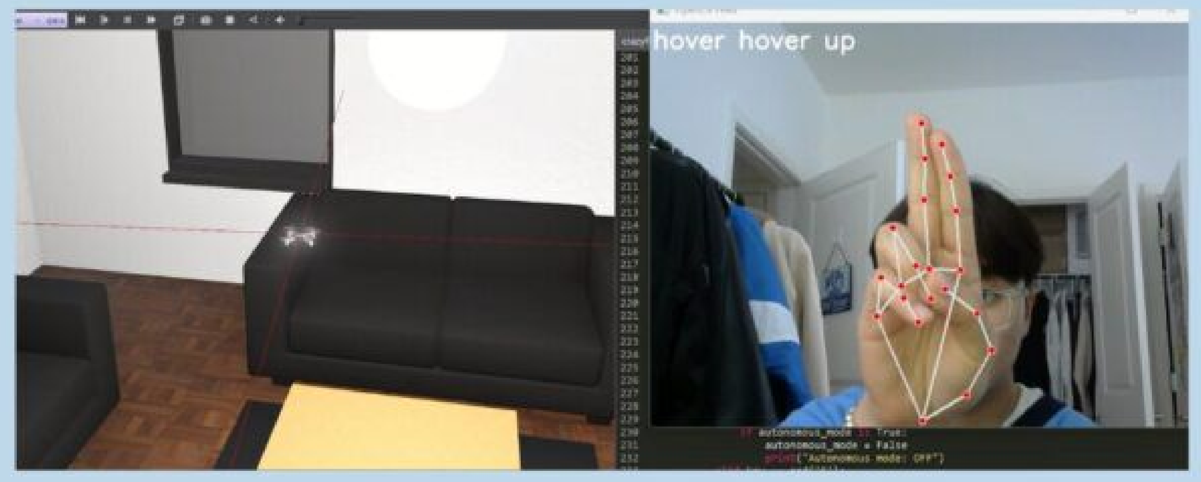
RESULTS

We successfully deployed the model onto a Crazyflie in a simulated environment within Webots, and was able to achieve real-time control of the drone through exclusive use of a webcam.

Accuracy: 0. 9393
Quantized Model Size: 214kB
Memory Savings: 100kB

The confusion matrix on the right summarizes the performance of gesture classification model used to control drone movements.





The left half of this screen capture image is the drone in a simulated apartment setting, the right half is the video from the webcam, with the hand features and skeleton overlaid atop along with the predicted gesture.

CONCLUSION

In this project, we demonstrate the feasibility of controlling drones solely with gestures which opens up a new angle for human-robot interaction. Our system can be easily deployed and used by many, as it enables intuitive, hands-free drone control without any external setup. Its edge ML design ensures real-time responsiveness without relying on cloud or bulky controllers. This makes it ideal for applications in search and rescue, assistive tech, warehouse automation, and interactive education.

REFERENCES

[1] https://www.bitcraze.io/documentation/tutorials/getting-started-with-simulation/
[2] https://cms.tinyml.org/wp-content/uploads/talks2020/tinyML_Talks_Manuele_Rusci_201027.pdf
[3] https://www.tensorflow.org/tutorials
[4] https://www.hackster.io/news/getting-started-with-edge-ai-using-the-gap8-processor-f207b511a00c
[5] https://iis-projects.ee.ethz.ch/index.php/PULP