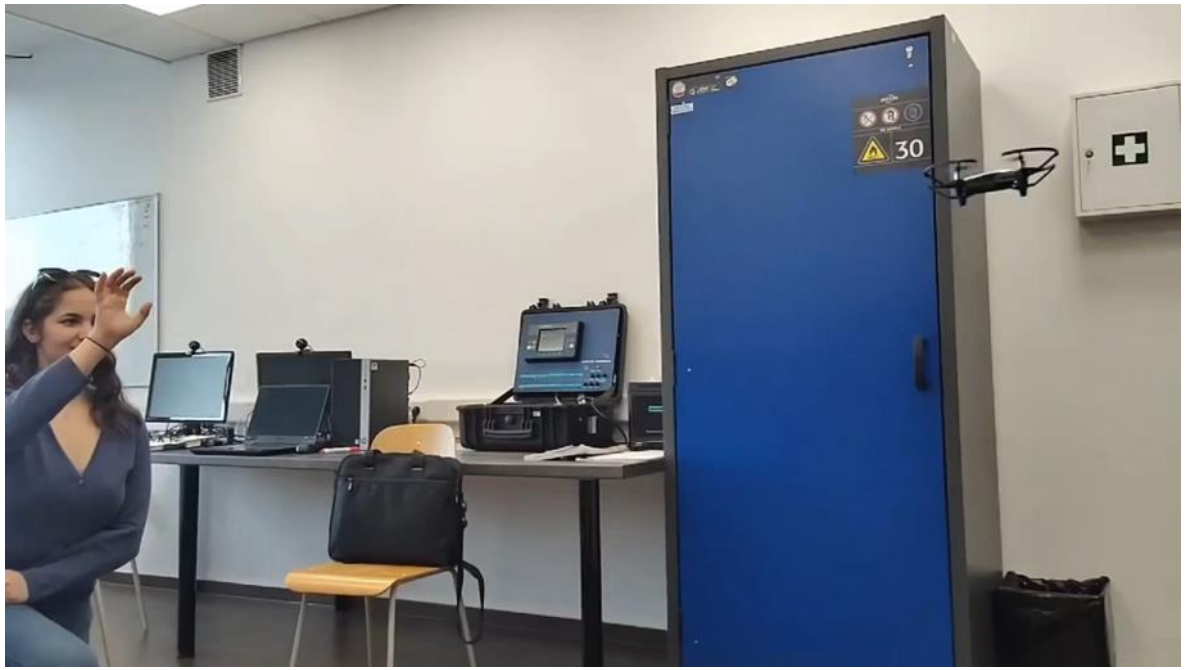
Bachelor Thesis

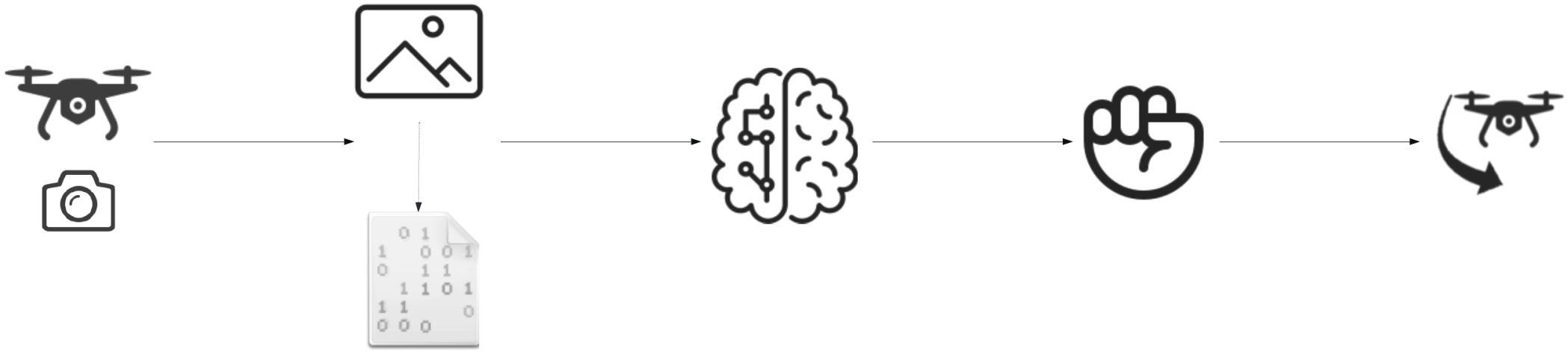
Institute of Information
Engineering, Automation, and
Mathematics

Touchless Drone Navigation via Hand Gestures

Ivana Dukayová
Thesis supervisor: doc. Ing. MSc. Martin Klaučo, PhD.
Thesis consultant: Ing. Patrik Valábek
June 18, 2024



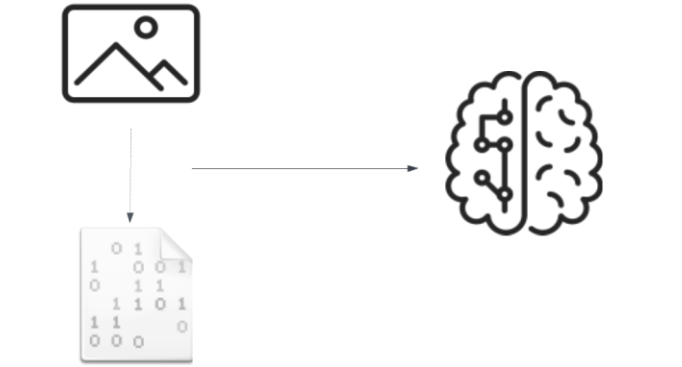
Our Goals



Obtaining the image | Data processing | Gesture prediction | Command conversion | Command execution

Our Goals

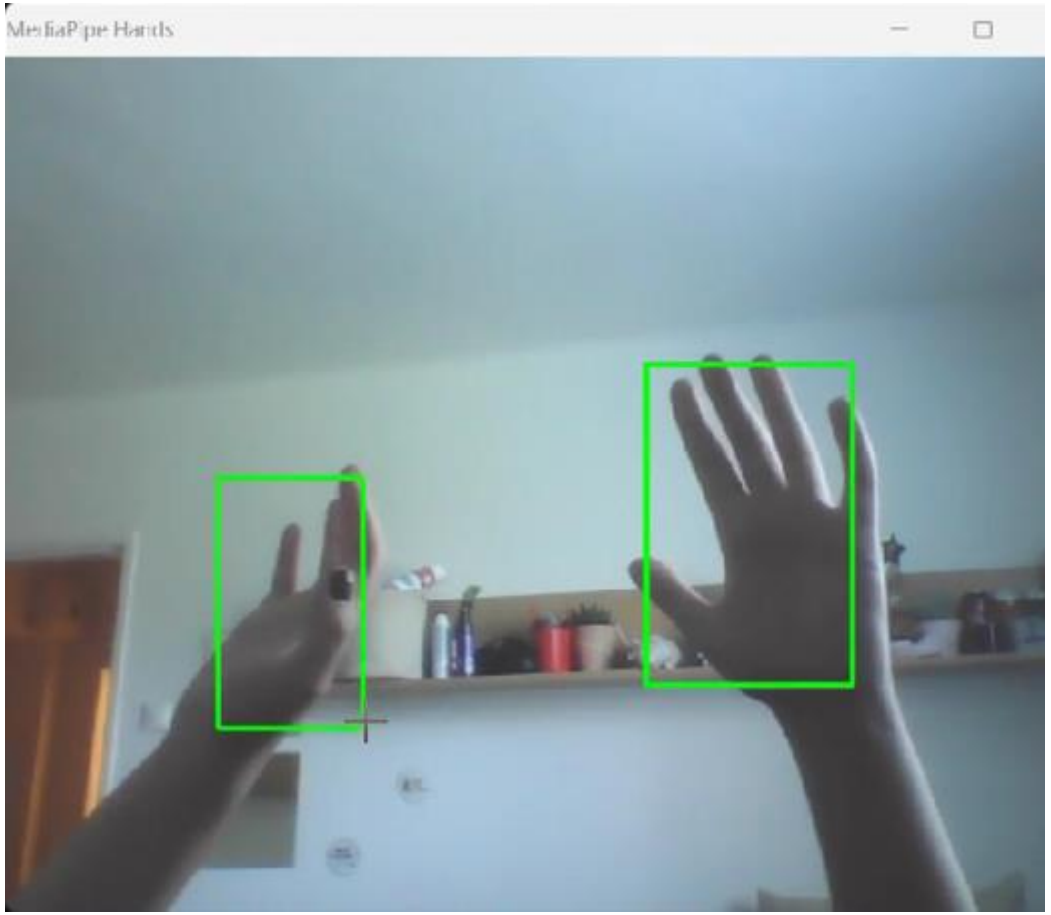
MediaPipe



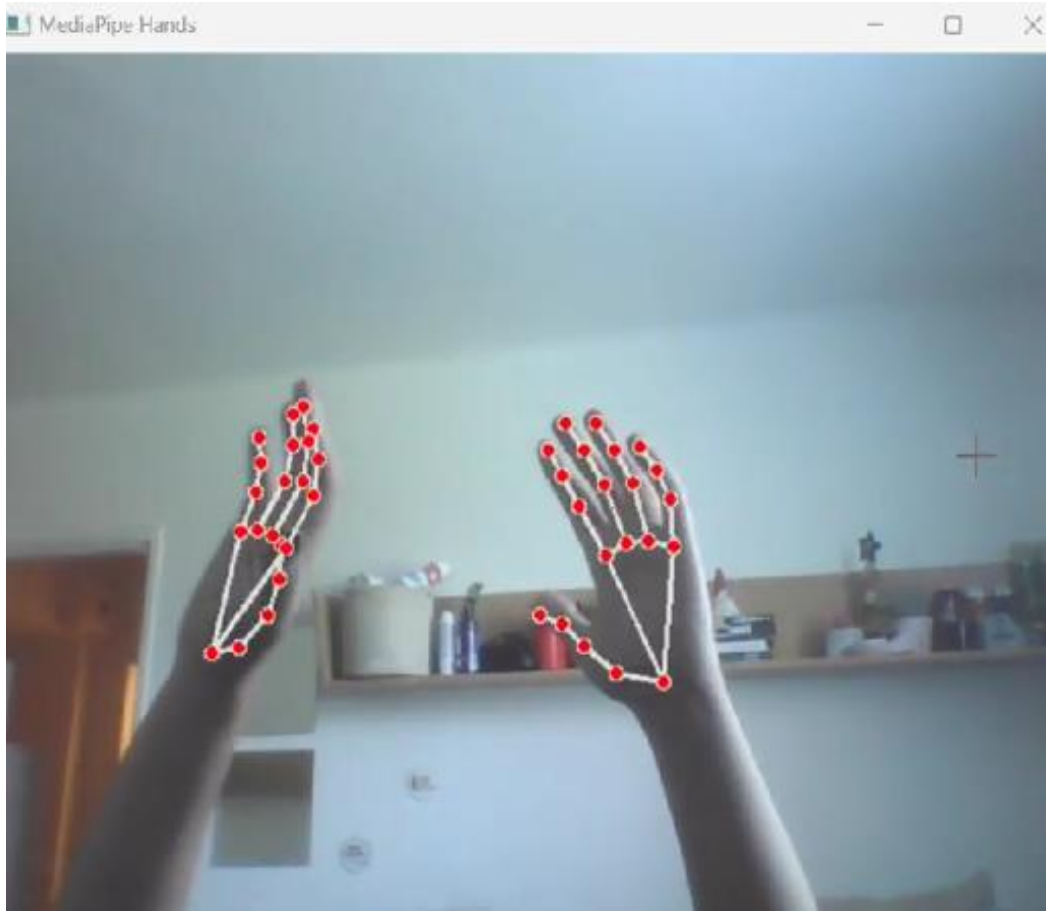
Obtaining the image

Data processing

Data Collection and Preparation

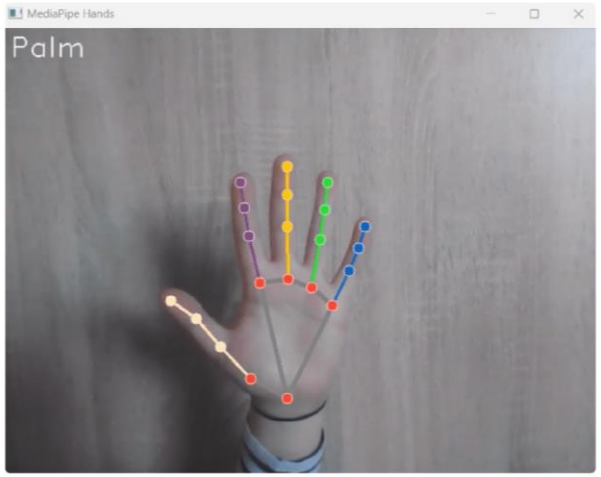


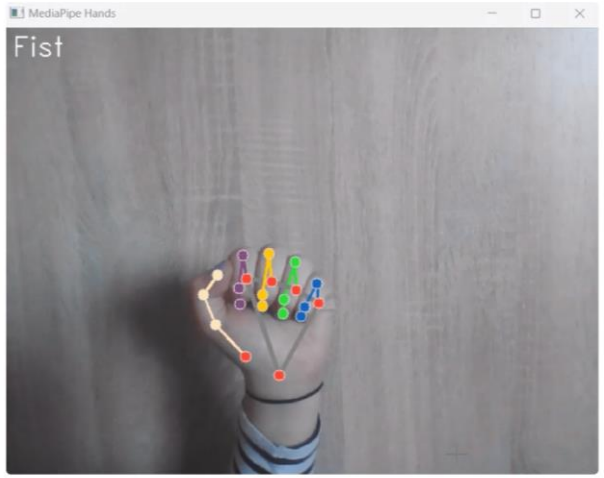
→ Palm detector

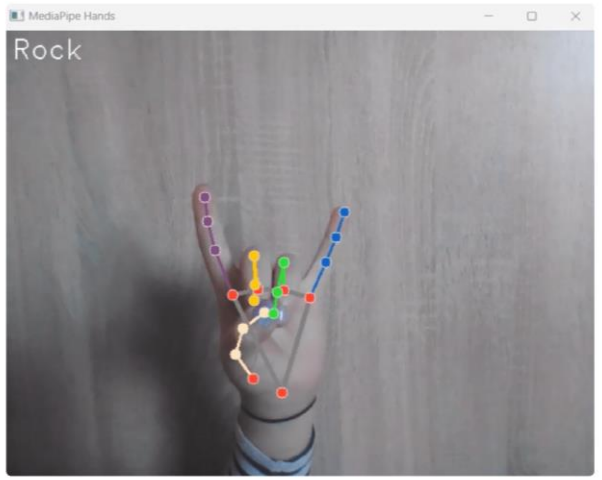


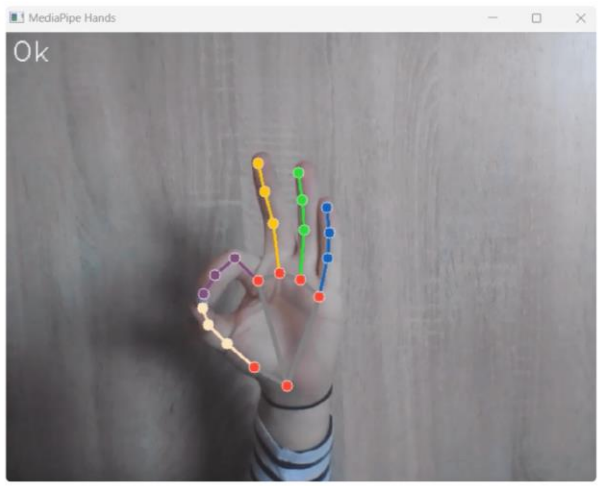
→ Hand landmark locator

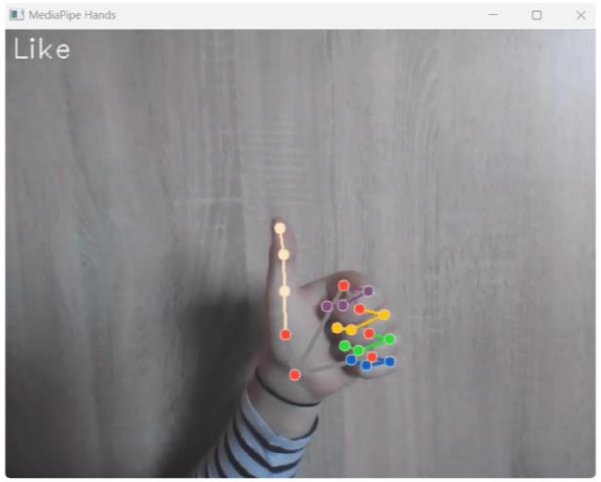
MediaPipe

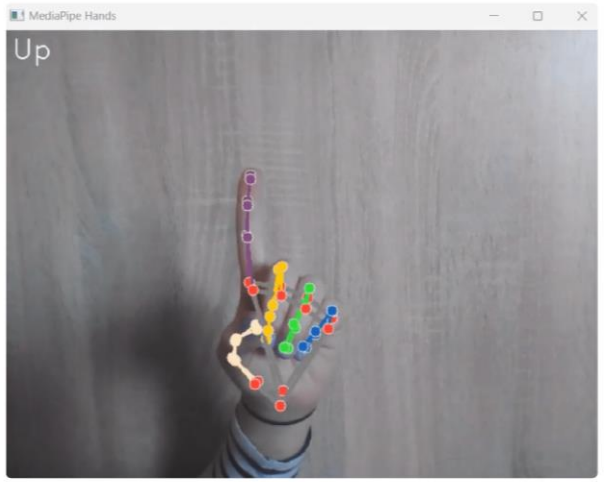


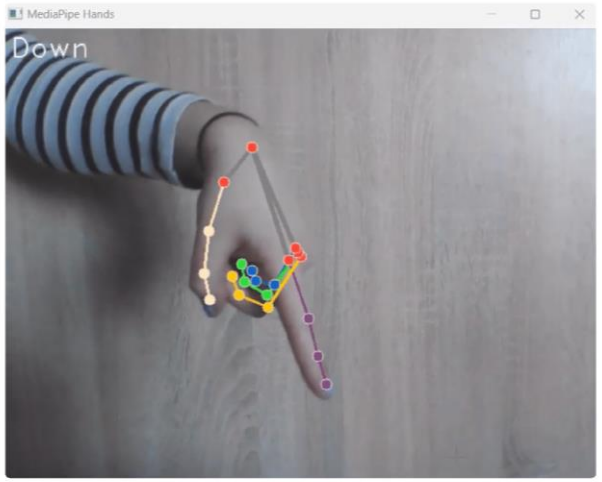


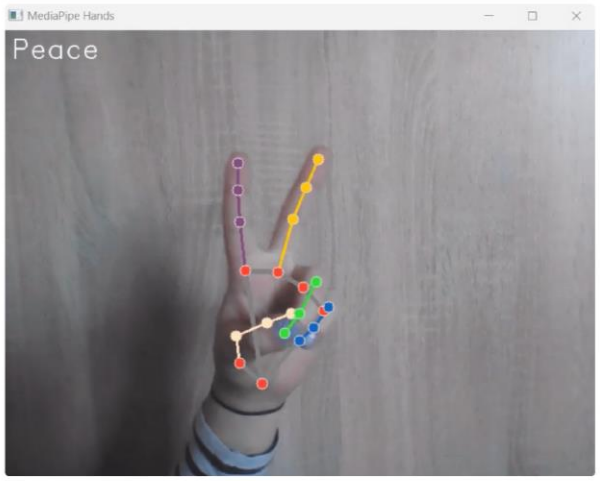




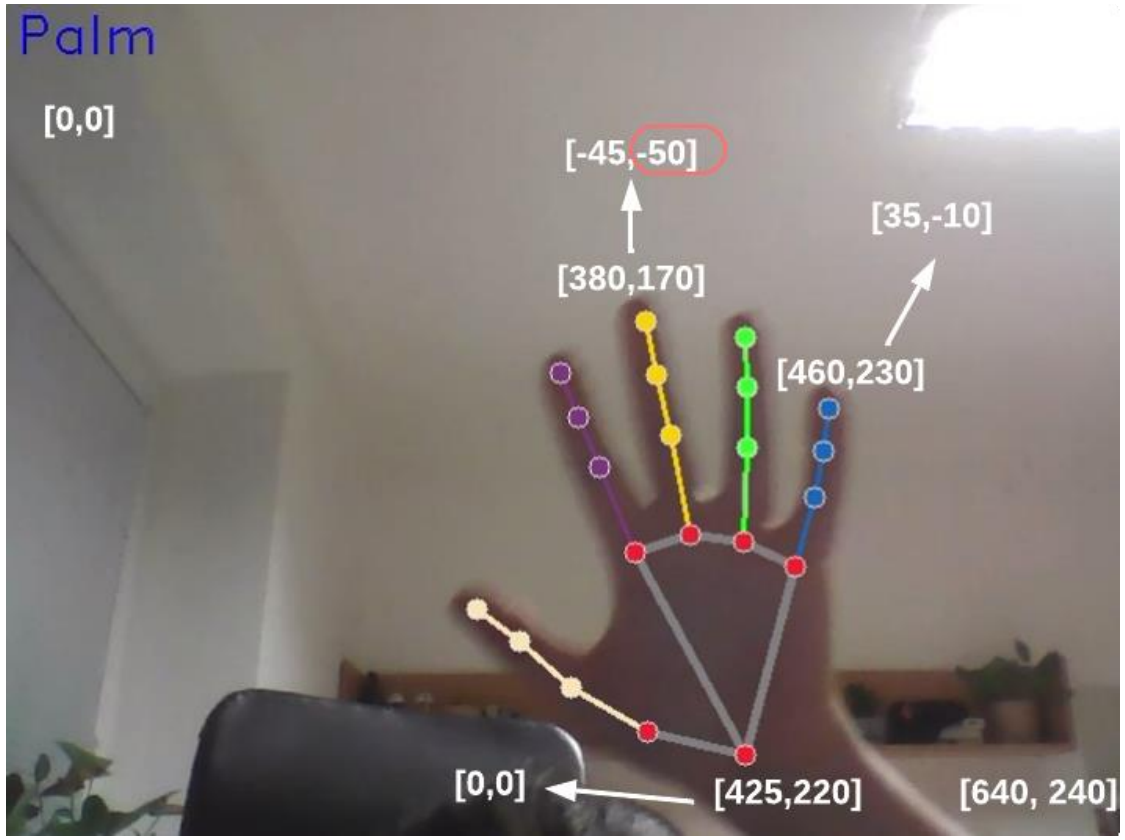


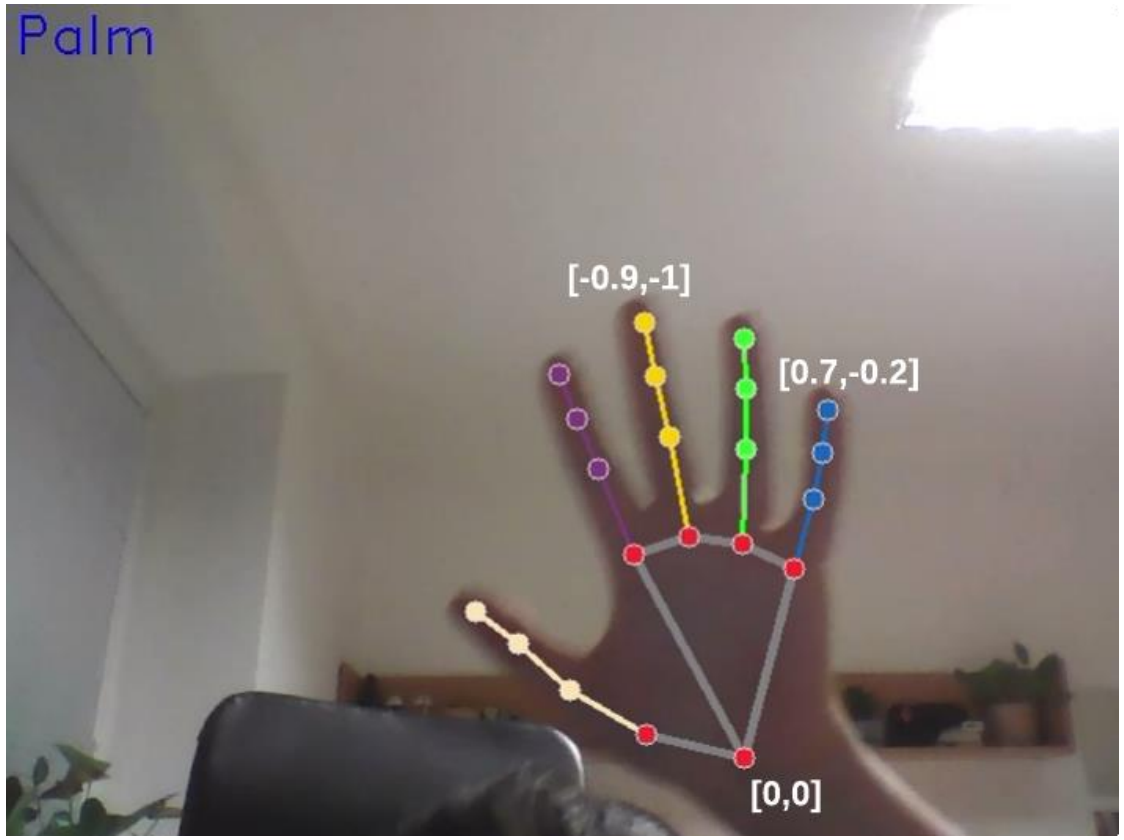




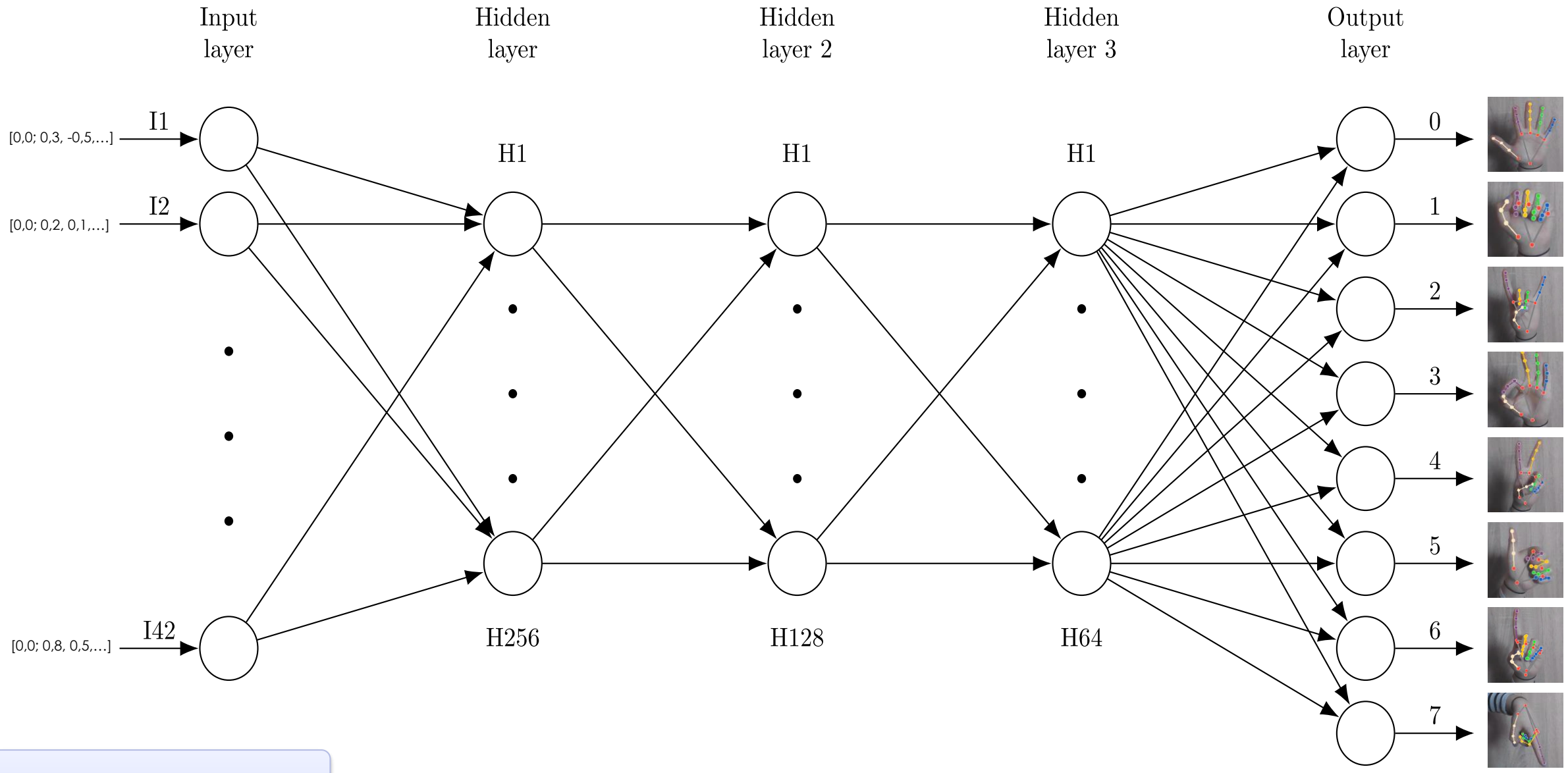


Data Collection





Data Normalization



Model for Gesture Recognition

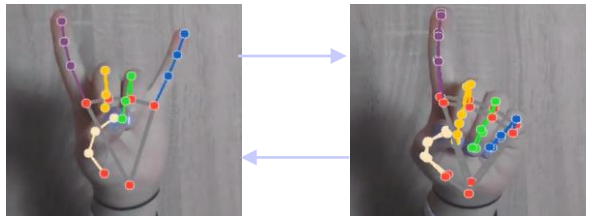
Training Results

→ Accuracy 99.62%

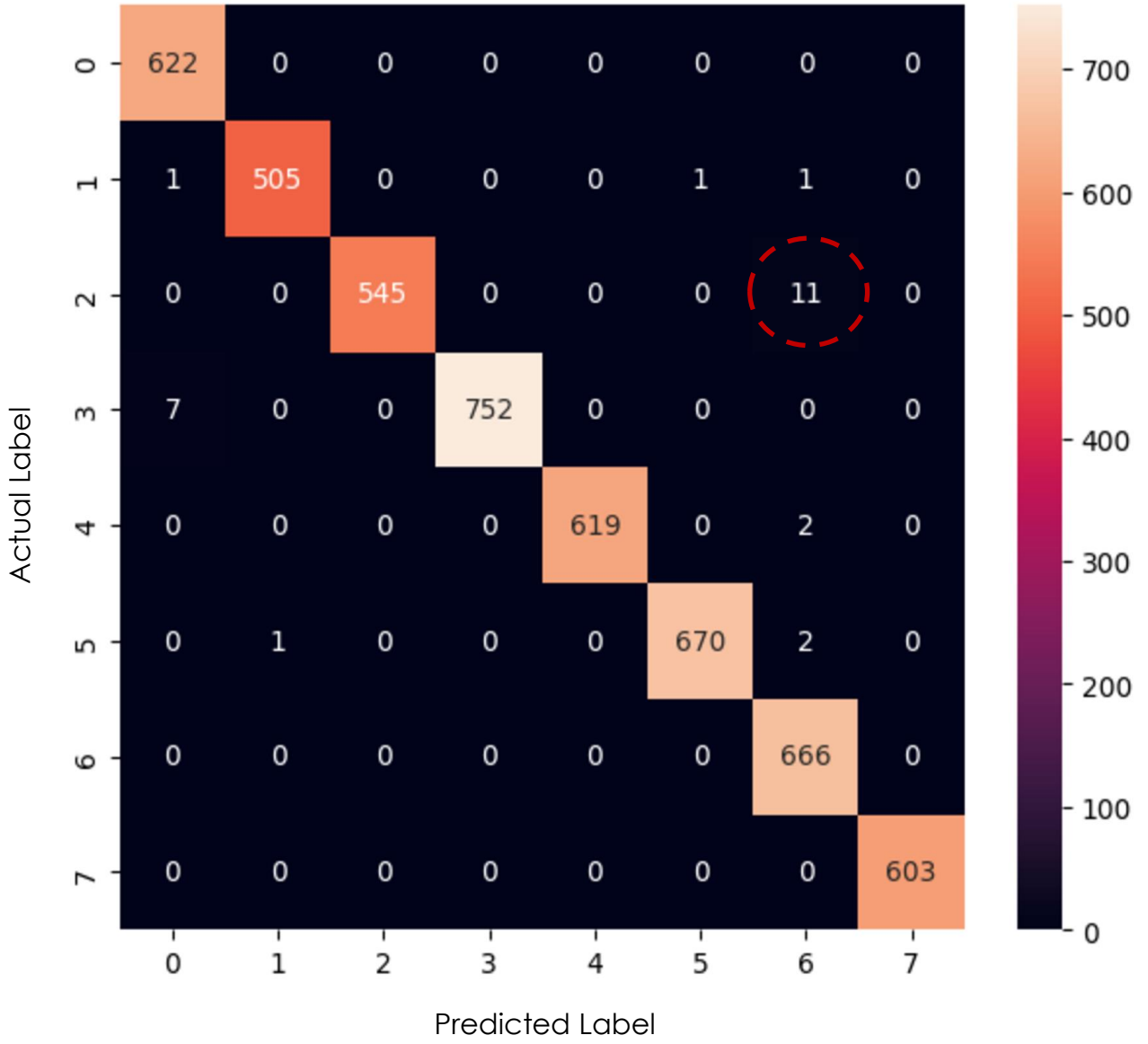
→ Loss 0.0309

→ 52 680 total parameters

Model for Gesture Recognition



2 6



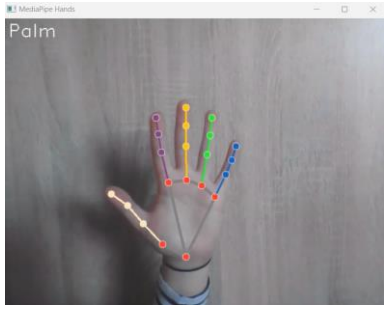
Training Results

→ RYZE Tech - Tello

→ Programmable

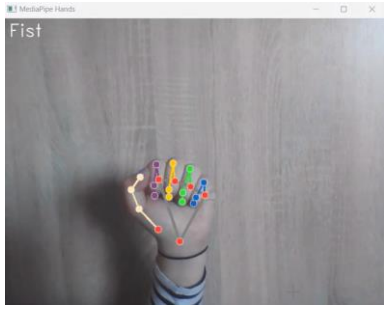
→ Onboard camera

Drone Control



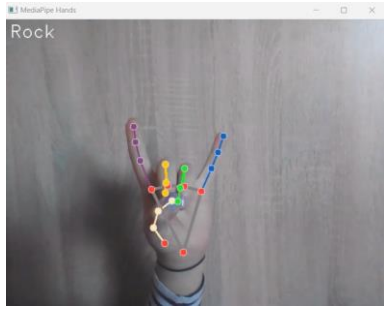
0 – Palm

move backward



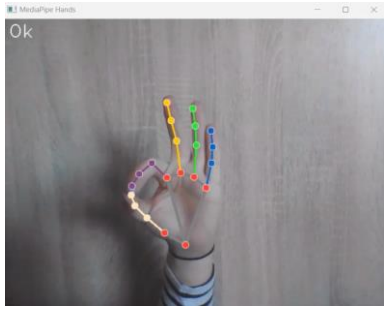
1 – Fist

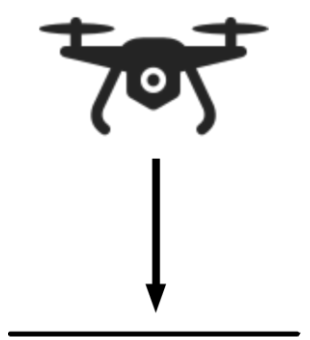
move forward



2 – Rock

perform a flip-forward

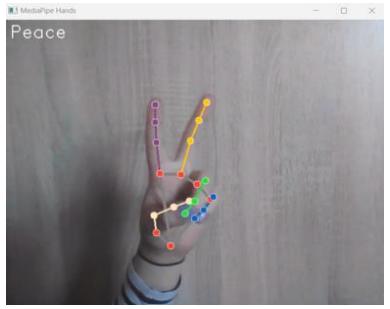




3 – Ok

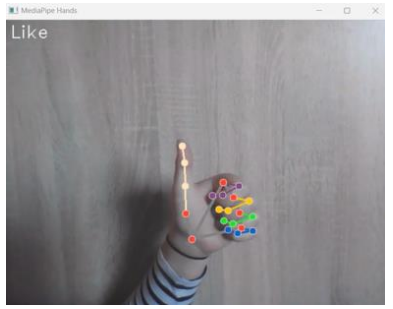
land or take off

Drone Control



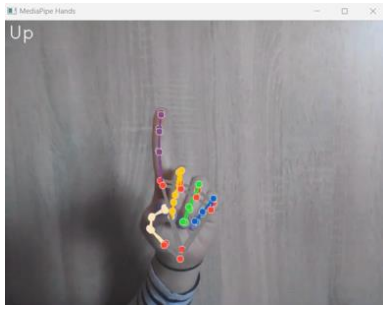
4 – Peace

take a picture



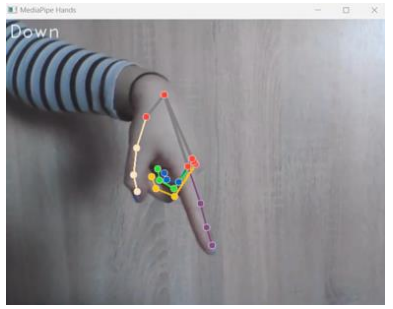
5 – Like

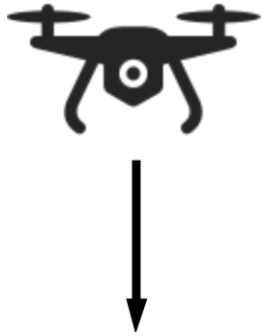
rotate 360°



6 – Up

move up

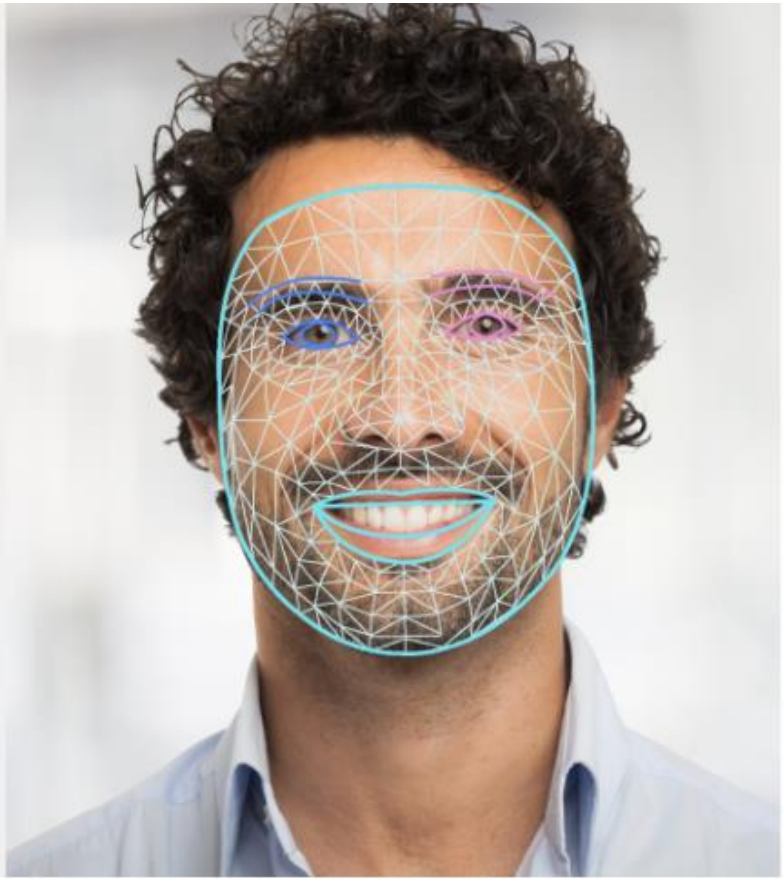




7 – Down

move down

Drone Control



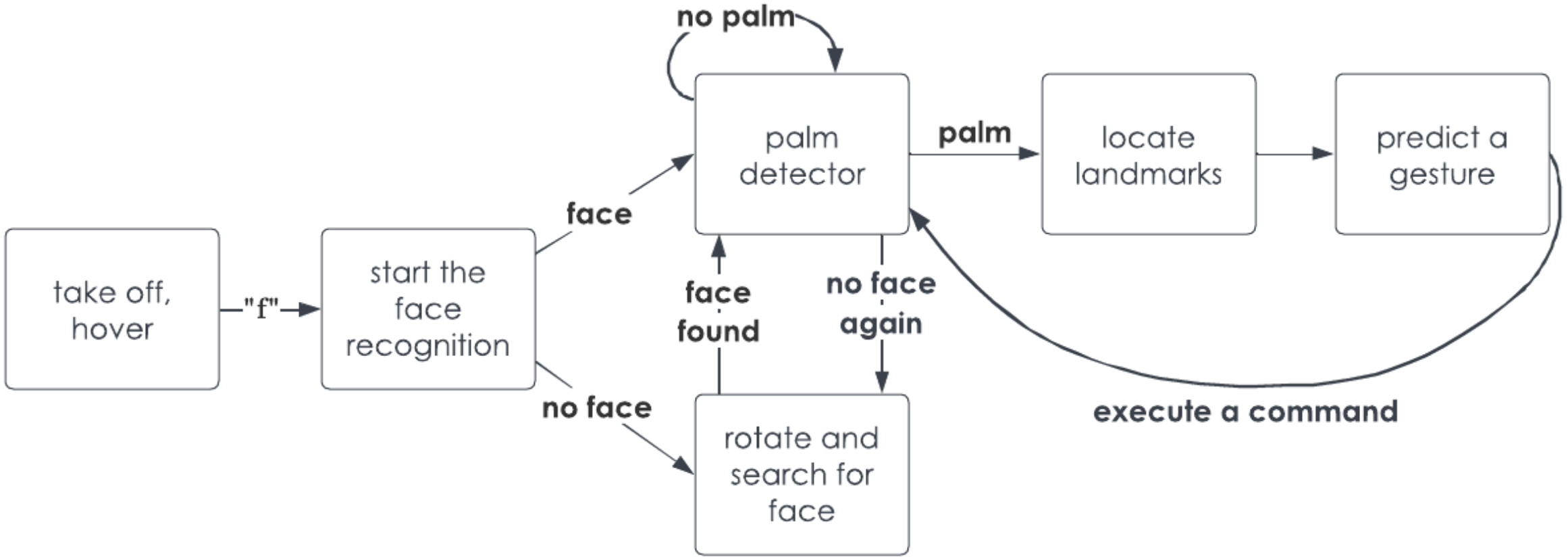
→ BlazeFace for detection

(6 landmarks)

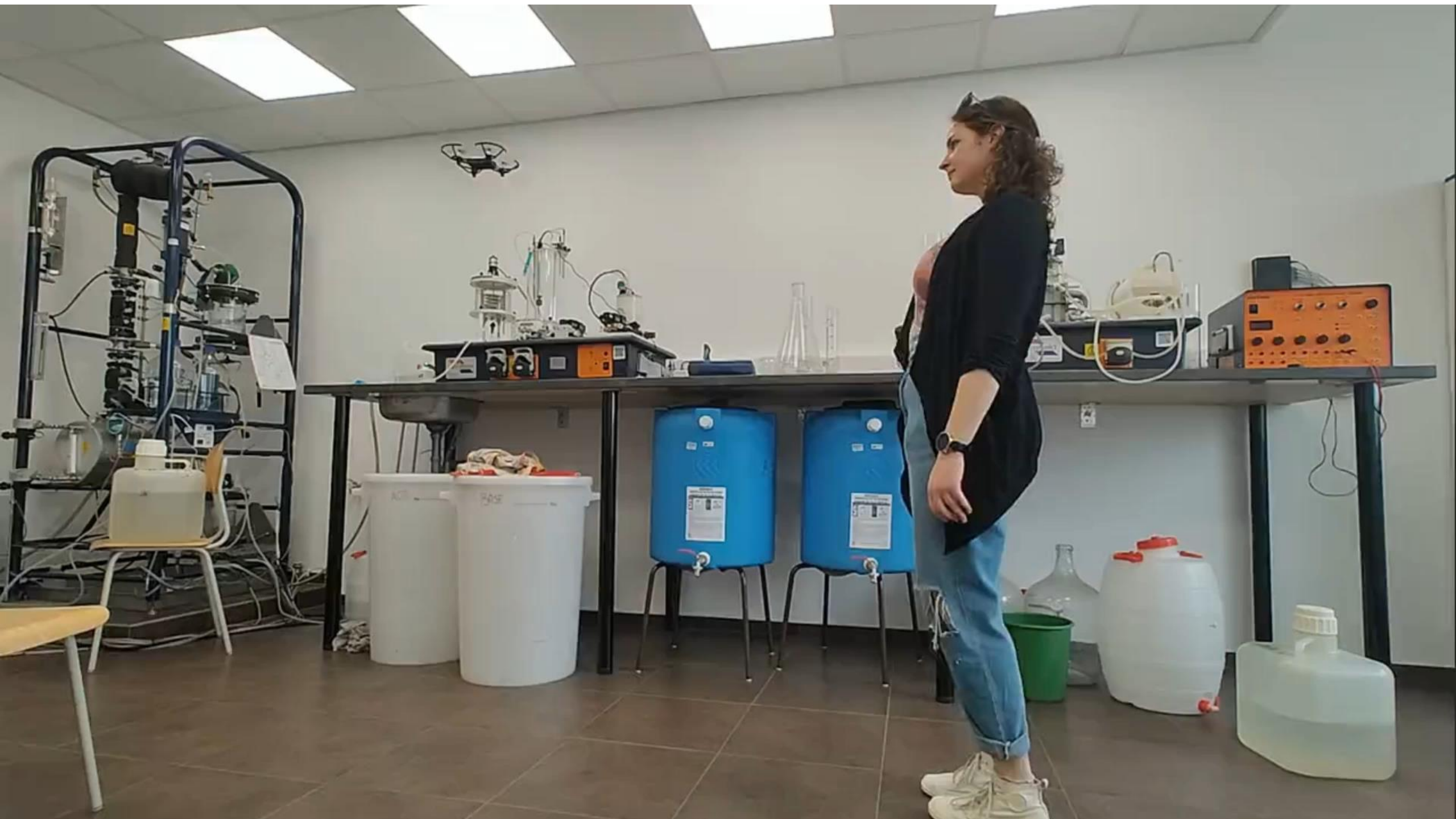
→ Face Mesh for recognition

(478 landmarks)

Face Detection



Face Detection




ACID
BASE

- → Data collection
- → NN for gesture prediction
- → 99.62% accuracy
- → Drone command execution
- → Implementing face detection
- → Possible expansions

Conclusions

Total samples: 20 030, 5008 is the validation set

- → Data collection
- → NN for gesture prediction
- → 99.62% accuracy
- → Drone command execution
- → Implementing face detection
- → Possible expansions

Conclusions

The presented dataset includes samples with the same settings. What steps must be taken to generalize the proposed classification model and ensure it works in different environments (e.g., various skin tones, bright settings, camera noise)?

→ MediaPipe bandle

Questions for the author

What are the benefits of controlling the drone using an integrated camera over the external one, which could increase the drone's range and mobility?

→ Intuition, purpose of use

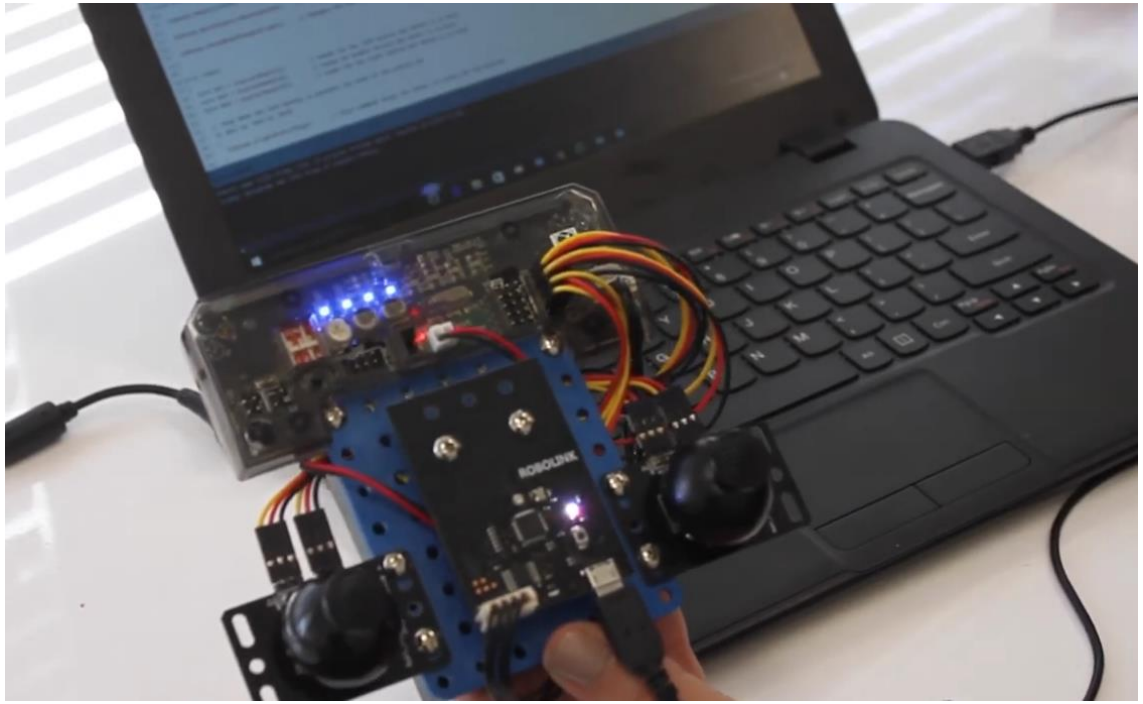
Questions for the author

Is it possible to implement the proposed navigation system directly on the drone without the need for an external computer? If not, what are the limiting factors?

→ Onboard controller, weight and stability issues

Questions for the author

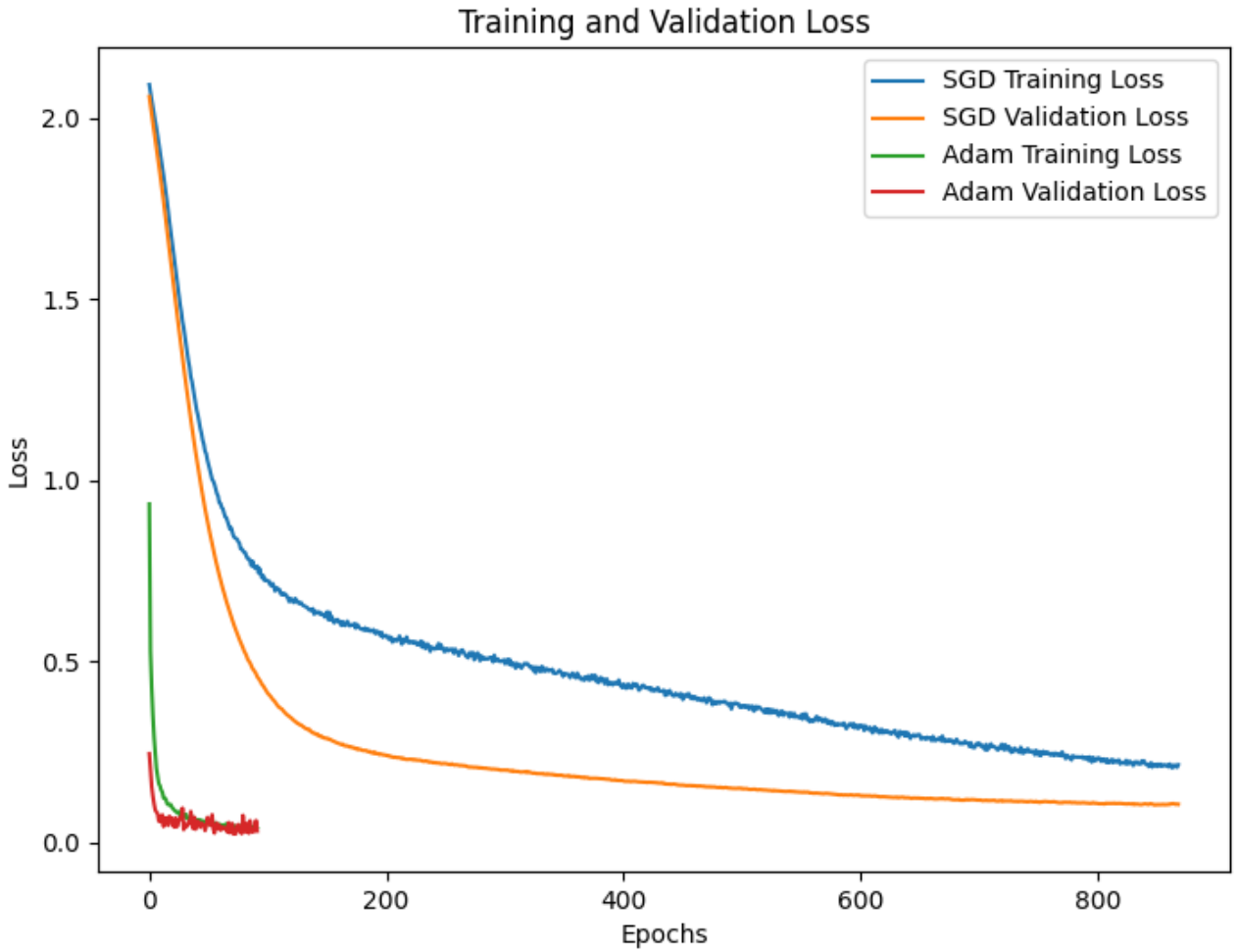
Questions for the author




extension module
Tello

Adjustments

- → More dense layers
- → Optimizer
- → Learning rate
- → Batch size
- → Early stopping

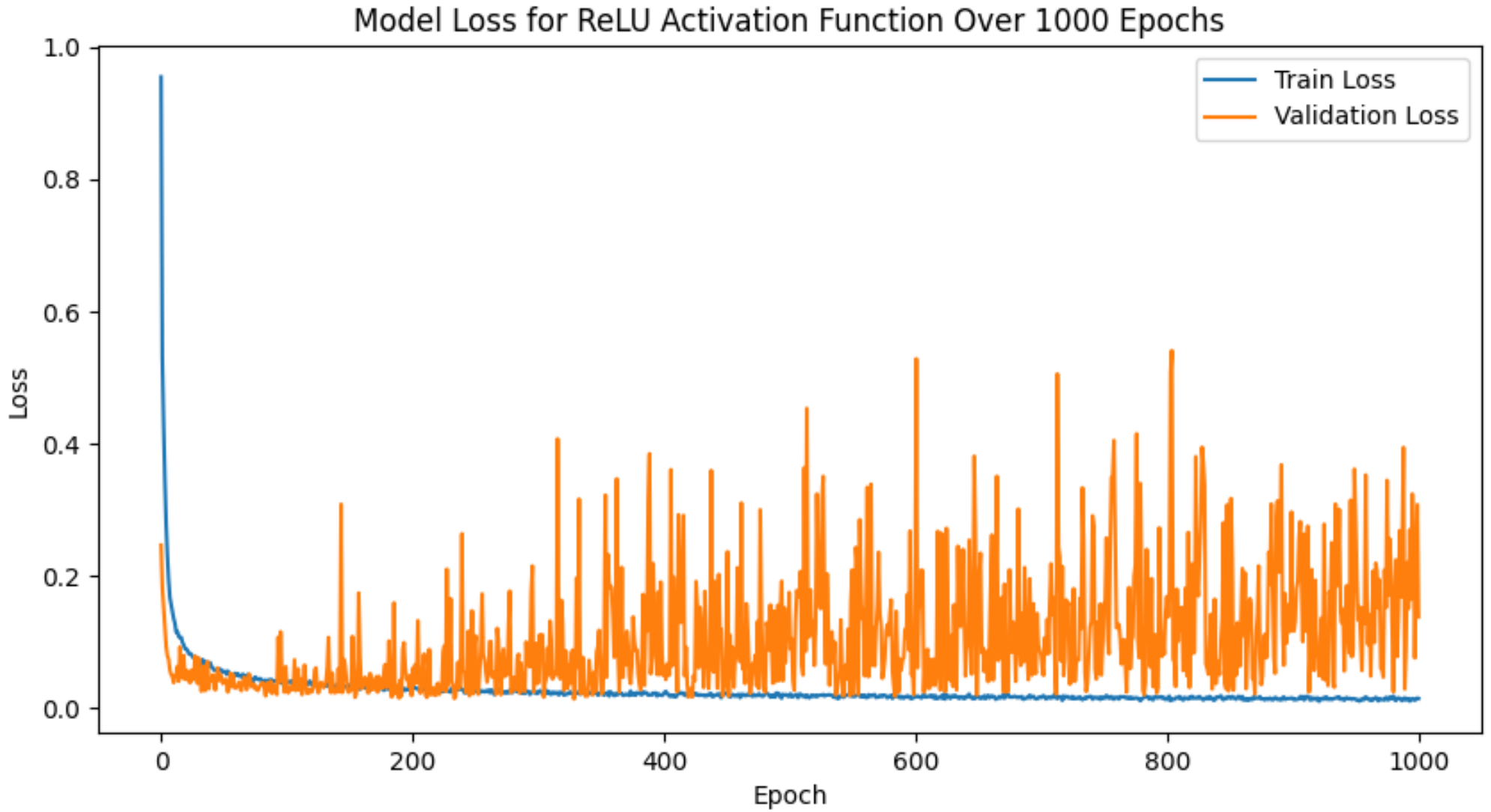


→ Adam optimizer is for our problem more efficient

→ Faster convergence

→ Models are generalizing well to the validation data

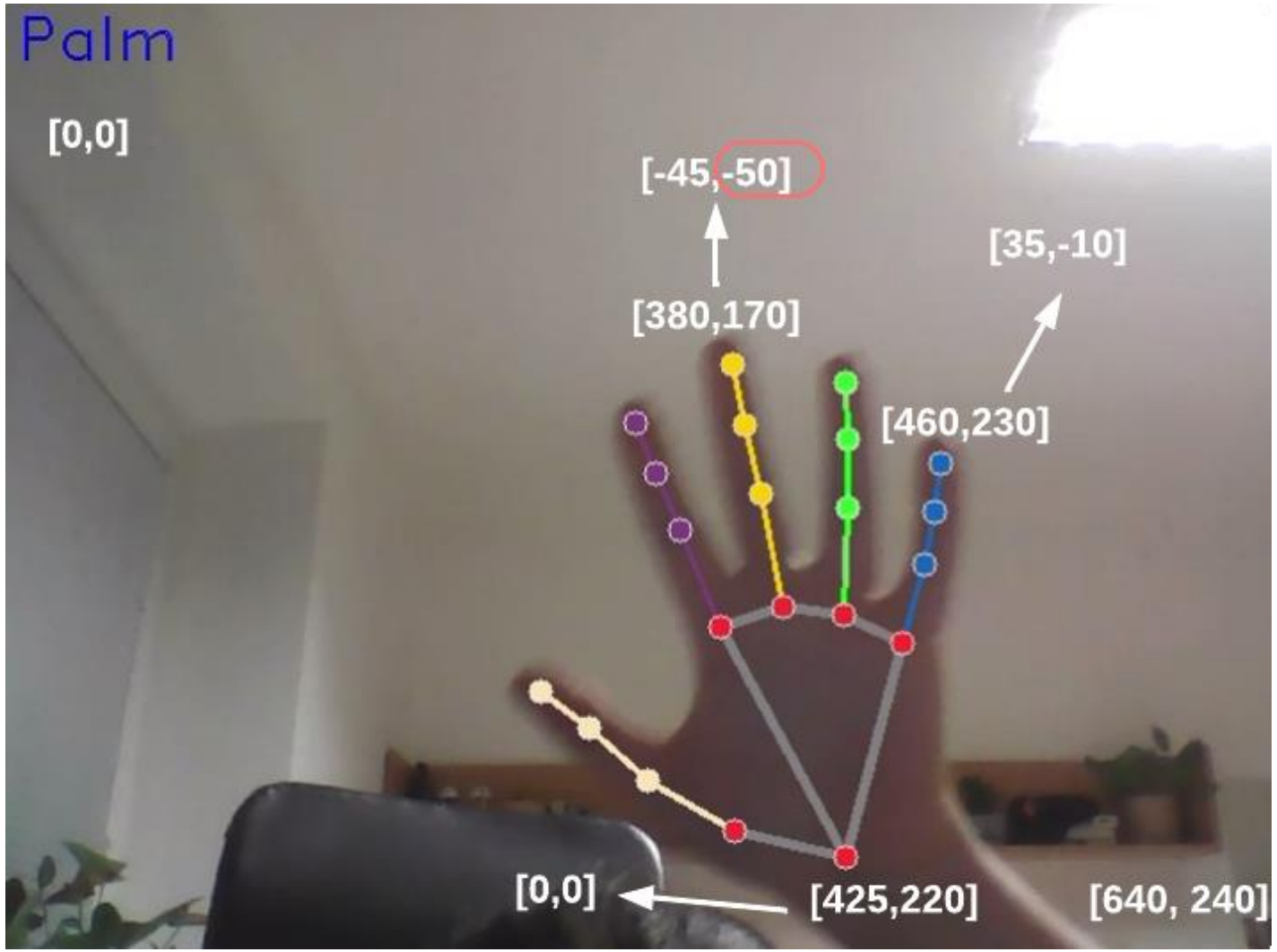
Model for Gesture Recognition



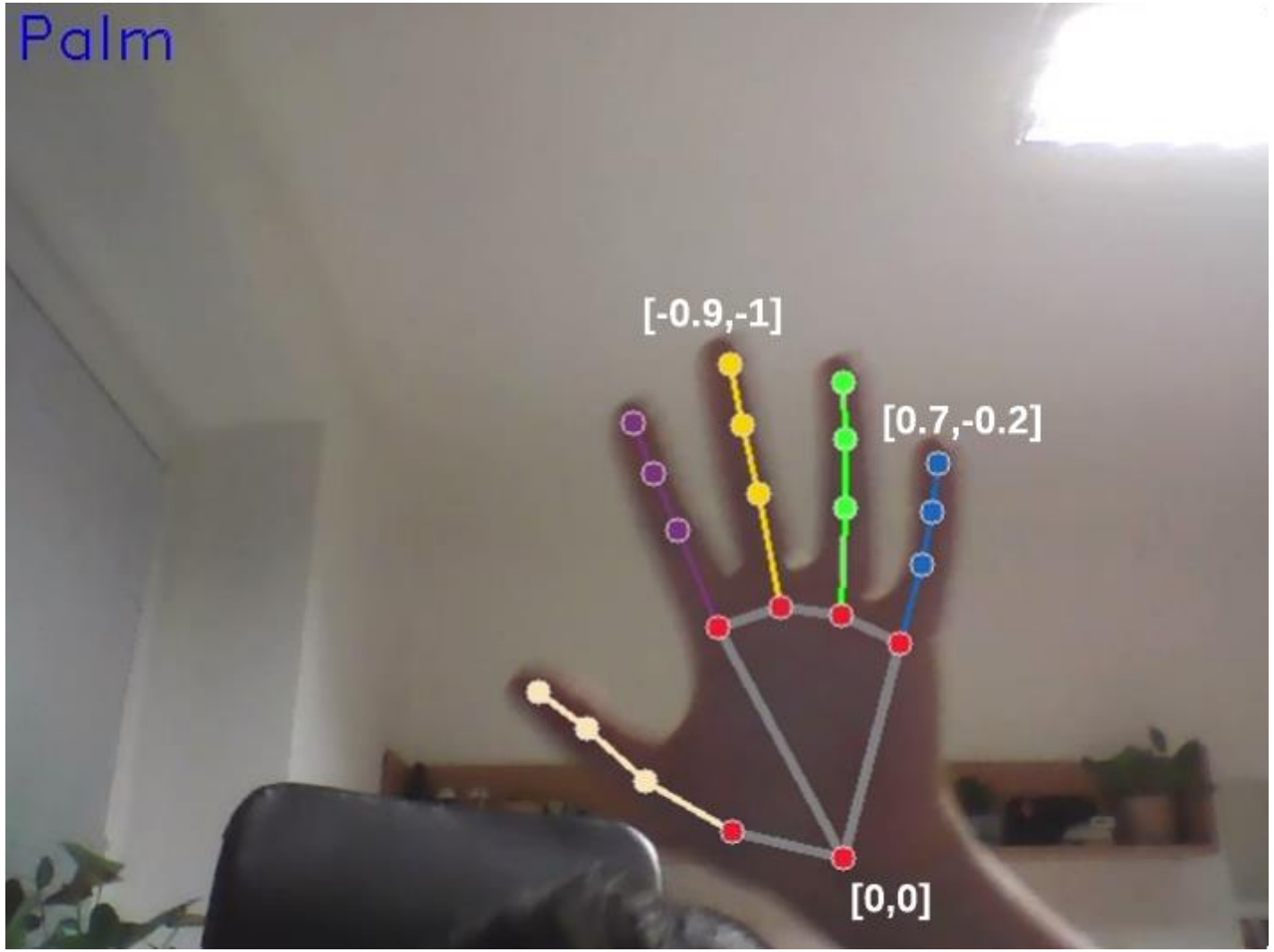
Model for Gesture Recognition

Parameter	Option	Validation Loss	Epochs
Optimizer	SGD	0.209	840
	Adam	0.063	98
Learning Rate	0.001	0.112	73
	0.0001	0.12	200
	$1 \cdot 10^{-5}$	0.18	1000
Batch Size	16	0.112	46
	64	0.114	72
	128	0.11	77
	256	0.12	94
	512	0.113	80
	1024	0.119	167

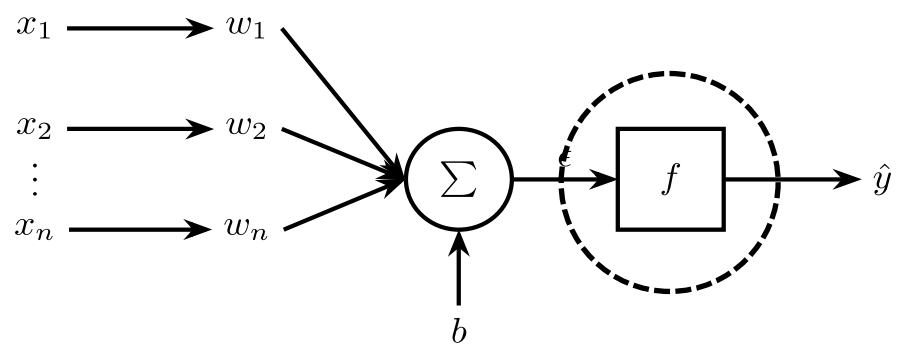
Model for Gesture Recognition



MediaPipe



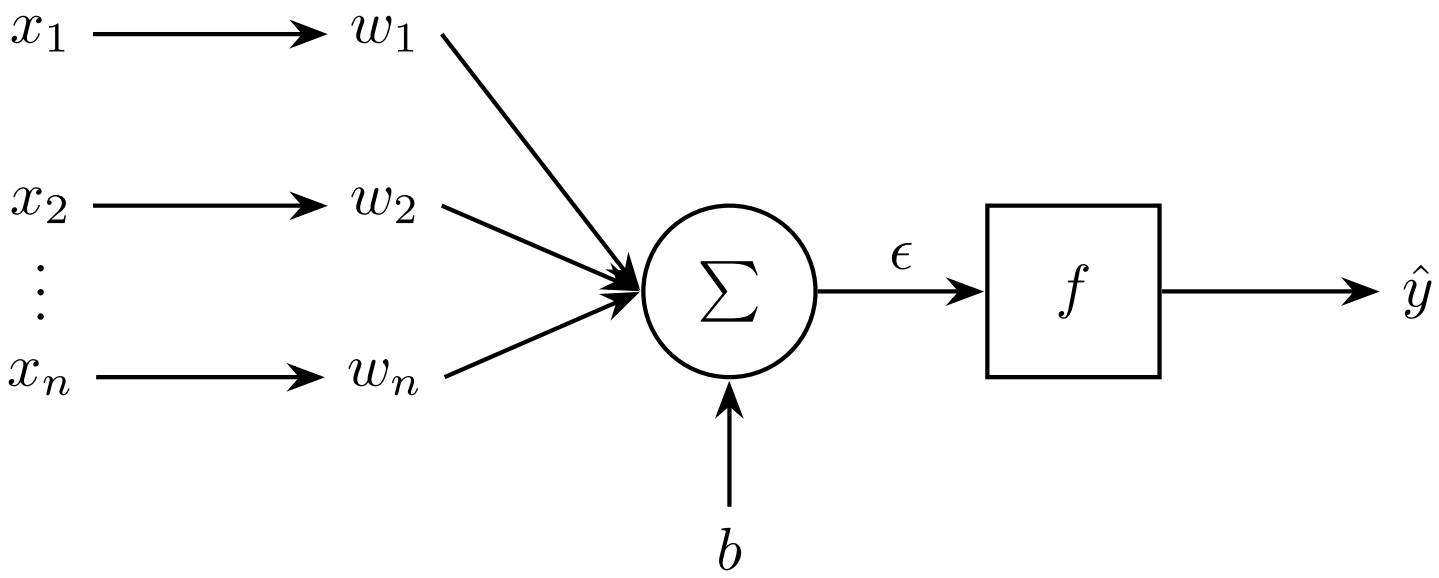
MediaPipe



Model Settings

- → Dropout drops a unit during training, randomly sets a fraction of inputs to 0 at each update
- → Early stopping: when there's increase in loss
- → Epoch: 1 iteration in training cycle – how many times the network sees the whole dataset
- → Sparse Categorical Crossentropy

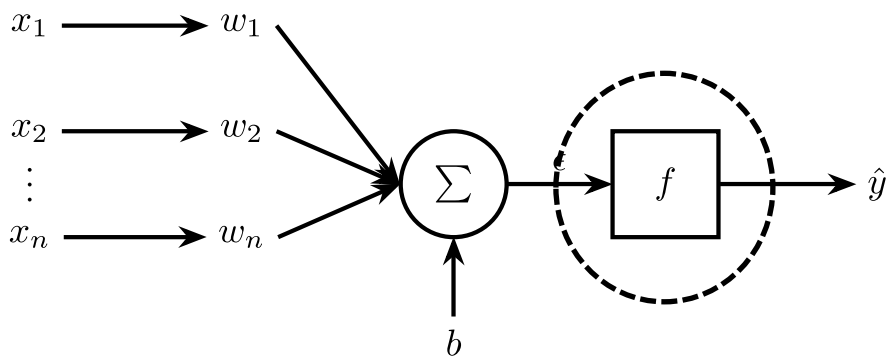
Model for Gesture Recognition



$$\hat{y} = f(\epsilon) = f(x_i w_i + b)$$

Model for Gesture Recognition

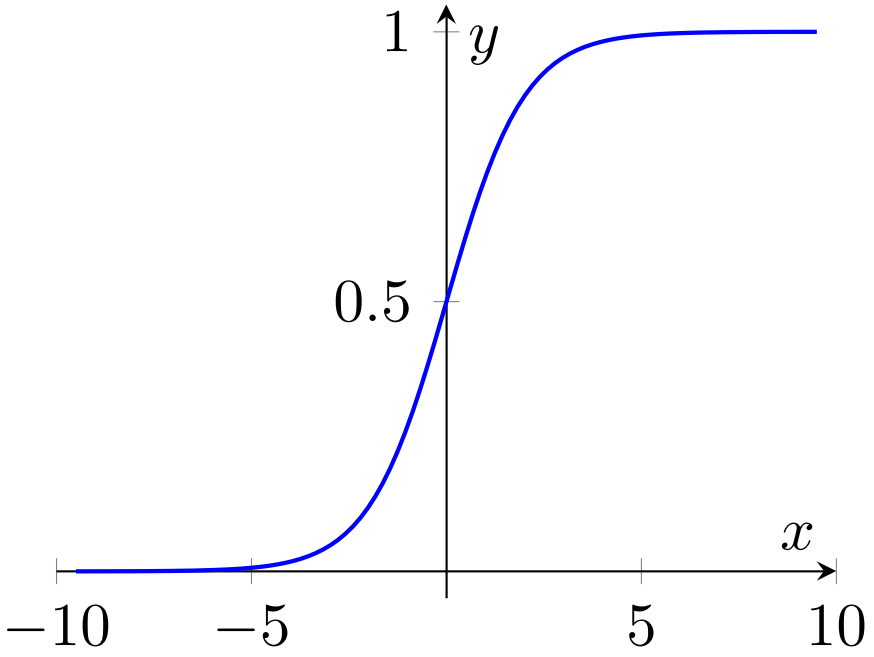
Activation Functions



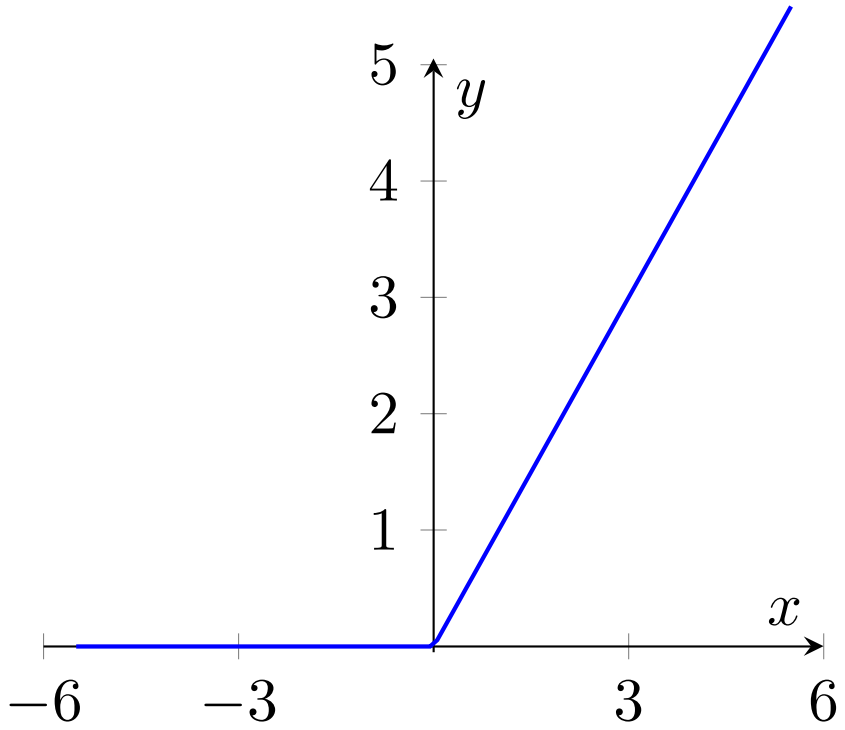
→ ReLU

→ SoftMax

Model for Gesture Recognition



SoftMax



ReLU

Model for Gesture Recognition

Learning

$$C = -\sum_{c=1}^{M} y_{o,c} \log(p_{o,c})$$

$$w_1 = w_1 - \alpha \frac{\partial C}{\partial w_1},$$

Model for Gesture Recognition

Resources

https://mediapipe-studio.webapps.google.com

https://www.ryzerobotics.com/tello

https://www.zero-x.com.au/product-force

https://lucid.app