Human Pose Recognition for Classifying Grappling Positions

Talhaa Hussain

University of Exeter

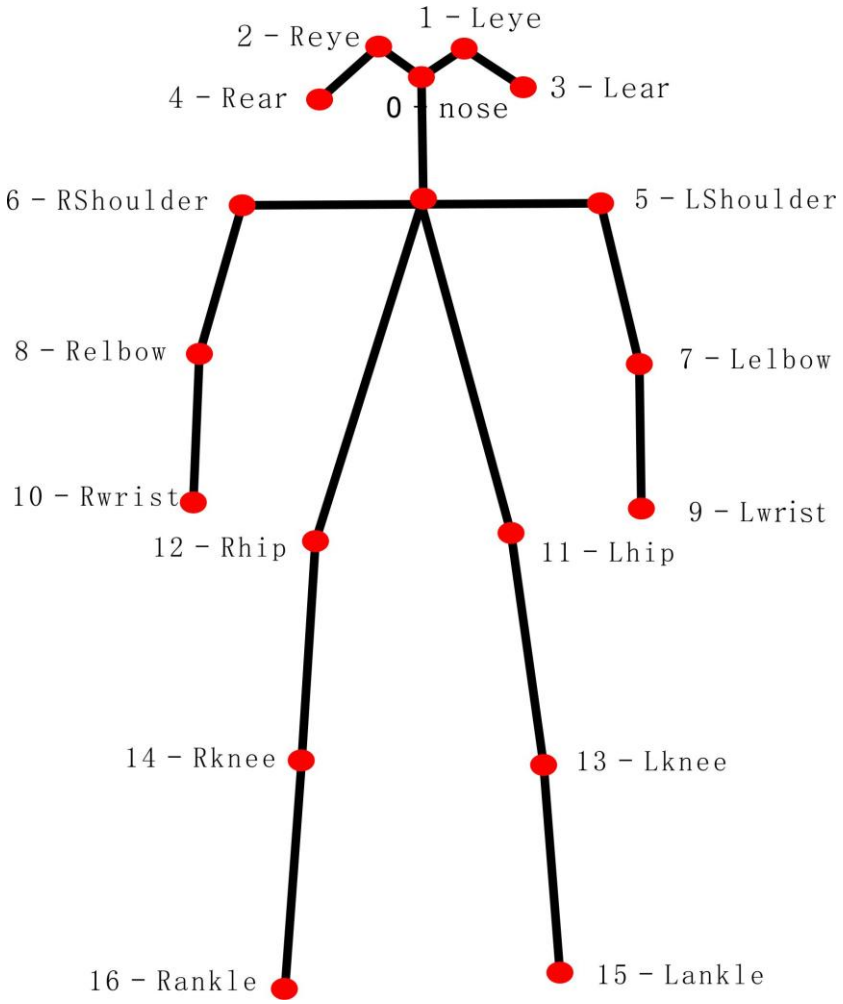
2024

Motivation

- Grappling is a form of unarmed combat that involves controlling an opponent, rather than striking them.
- Examples: Judo, Wrestling, Brazilian Jiu-Jitsu
- In all forms grappling, pinning an opponent to the ground is considered a dominant, winning position
- This project aims to leverage computer vision techniques to try to recognise if one athlete has another pinned to the ground

The Dataset

- The dataset is an adapted version of the dataset provided here: https://vicos.si/resources/jiujitsu/
- The original dataset has 18 different classes for Brazilian Jiu-Jitsu positions
- The adapted dataset simply has two classes – pin, or no-pin
- On the right we see an example of a 2-D kinematic model, in the MS-COCO format



Data Pre-processing

Techniques used:

- Random oversampling
- Feature scaling using min-max normalization

Changes made:

- Removed irrelevant data
- Renamed labels to be binary – representing pin or no-pin

76’343 DATAPOINTS:

- 62’792 NON-PINS
- 13’551 PINS

>

125’584 DATAPOINTS:

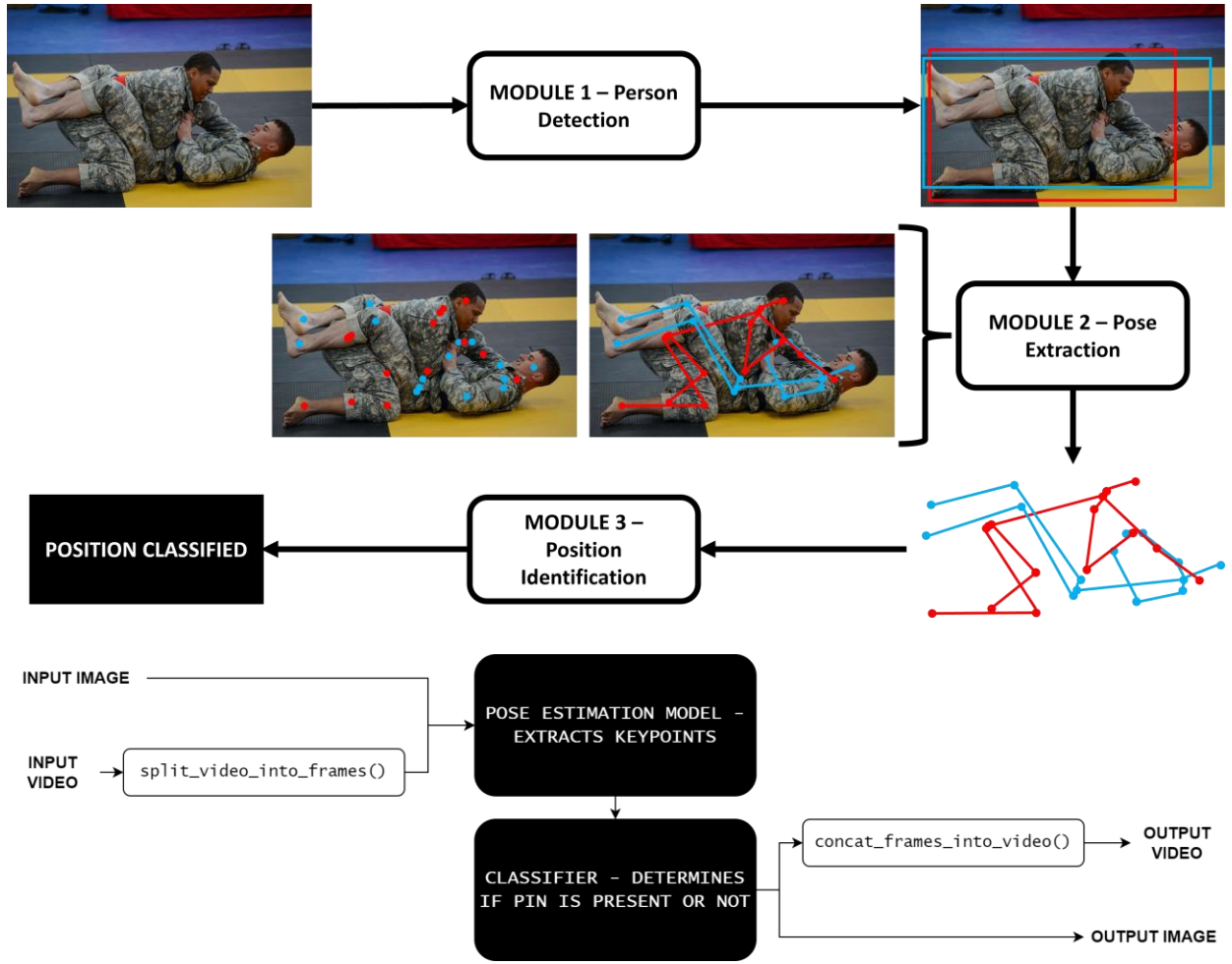
- 62’792 NON-PINS
- 62’792 PINS

Aims/Objectives

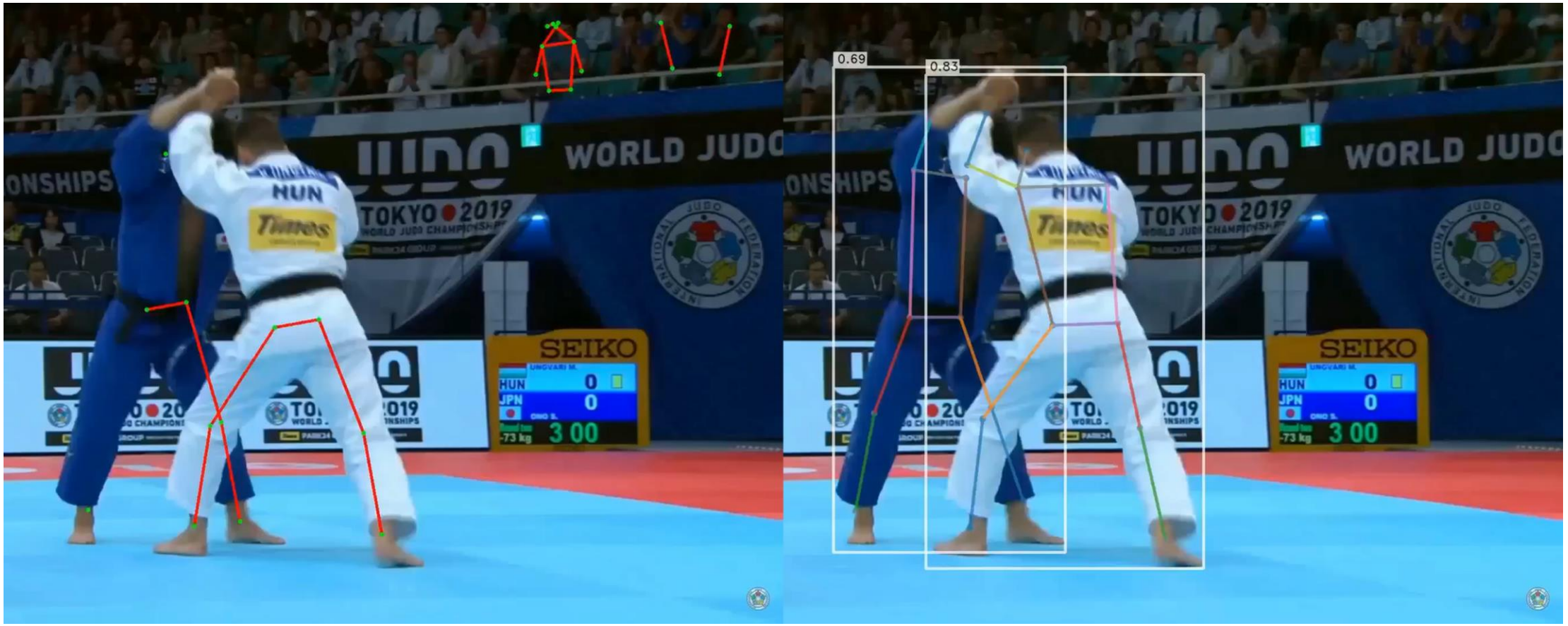
1. Make use of a HPE computer vision model to extract key-points for an athlete
   - Annotations must be green if pin is detected, red if not
2. Implement a binary classifier to decide if a given position is a pin or not
   - Binary classifier must achieve an F1-Score of 0.7 or higher
3. Implement a module for handling video, segmenting video into frames and rejoining them post-processing
   - Video frames must be processed exactly as isolated images are

Experimental Design

1. We take video/image input from user
2. We put video frames/image through a pose estimation model
3. We take key-points from pose estimation model and put them into machine learning model
4. We join frames up (if video) and output labelled video/image to user



Experimental Design (cont.)

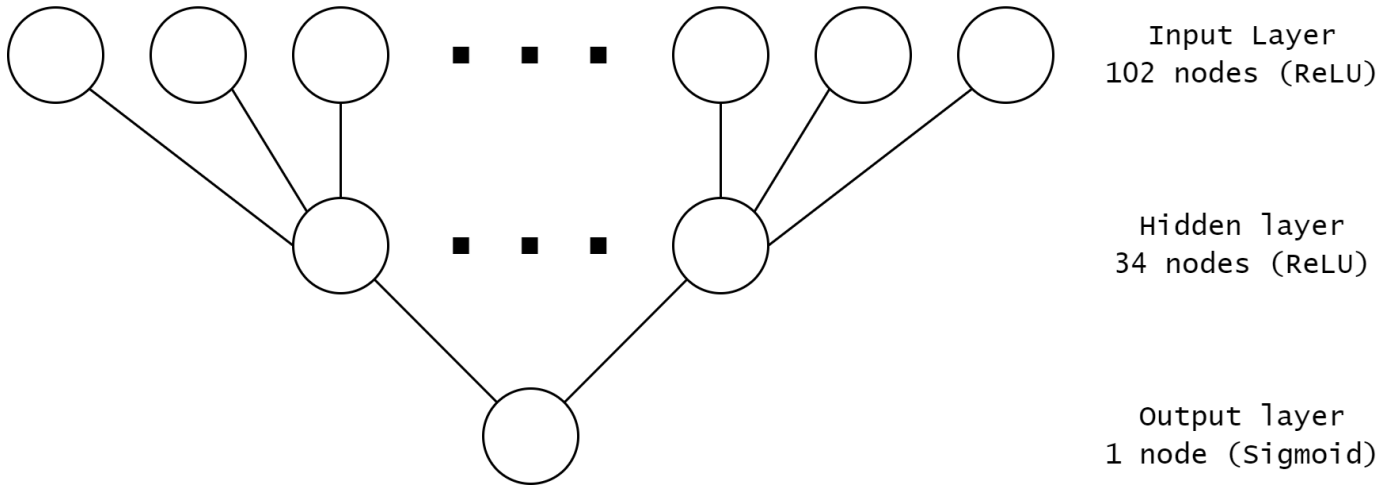


MoveNet Lightning by TensorFlow

YOLO-NAS-POSE-L by Deci AI

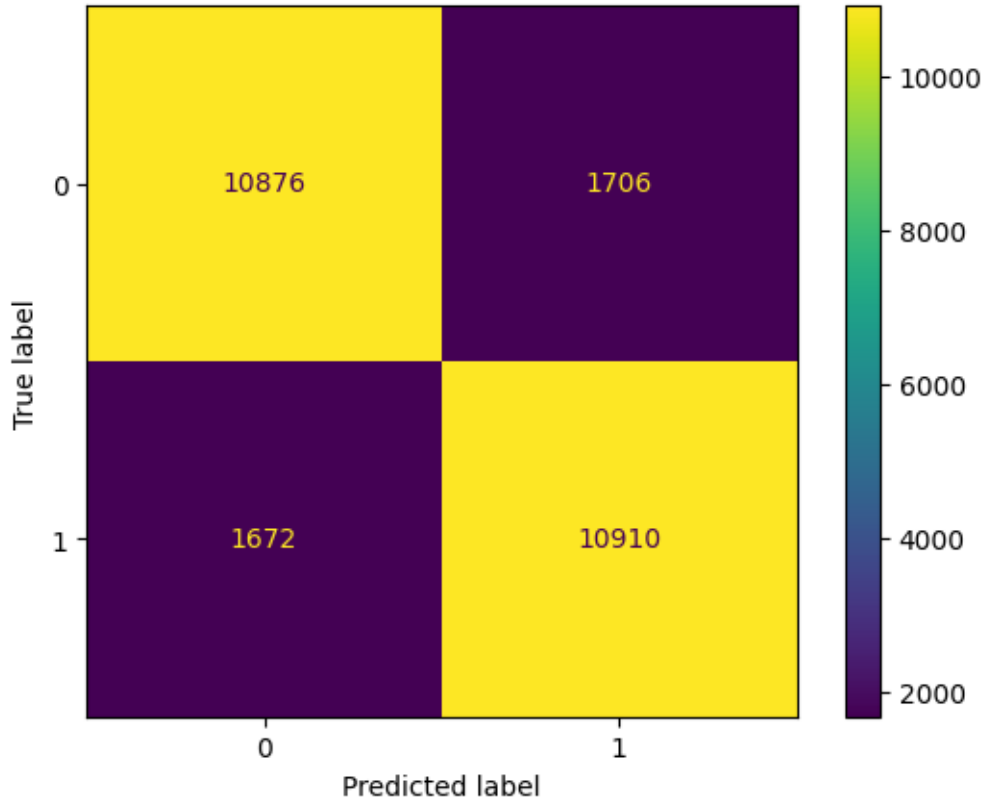
Experimental Design (cont.)

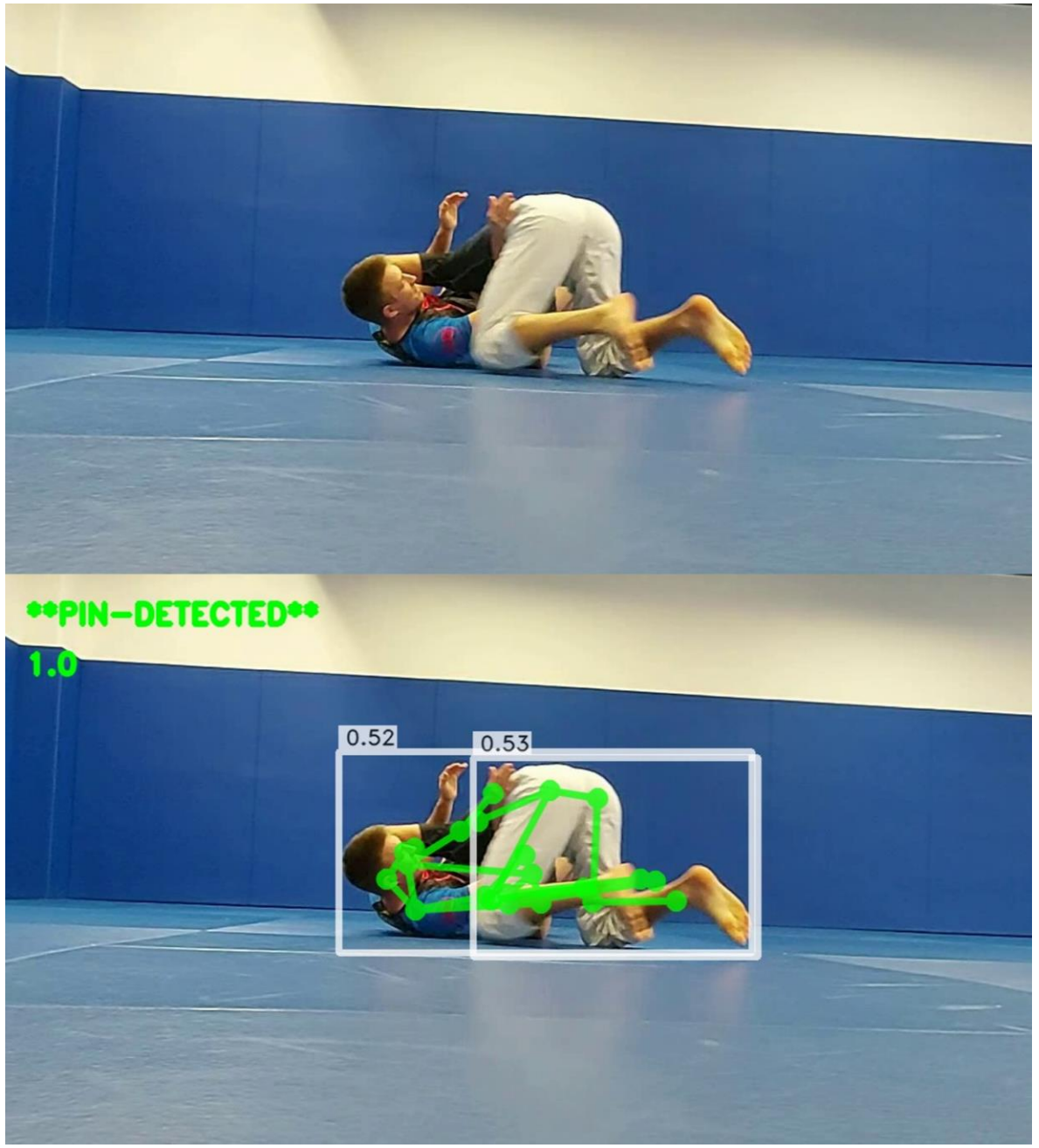
- Used a multi-layer perceptron as shown below
- Only one (sigmoid) output node needed, thanks to binary classification; no need for convolution!
- Architecture based off previous solutions to similar problems



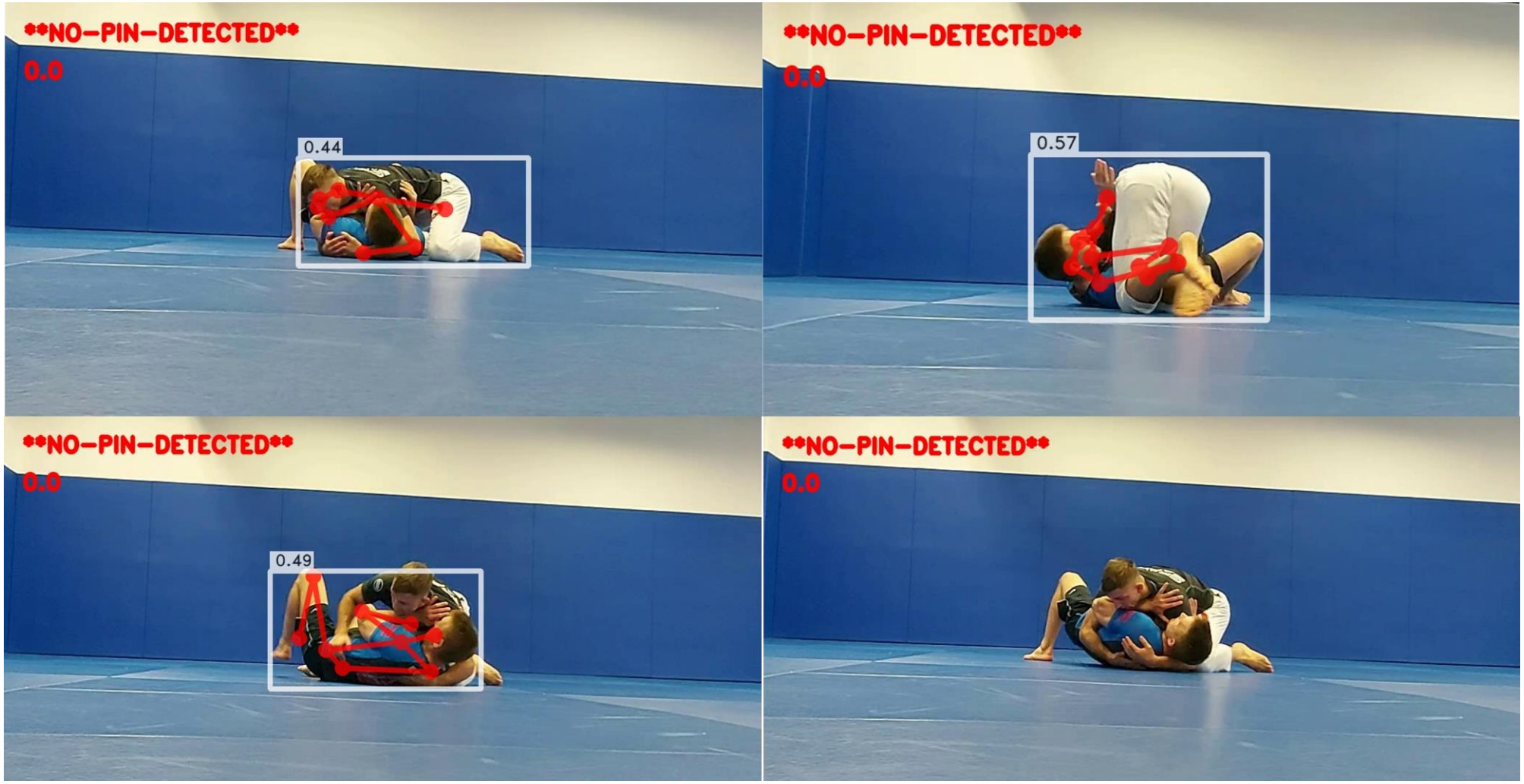
Results

Classifier achieved an F1-Score of 0.87 on unseen data after training for 1’000 epochs!





Issues


NO-PIN-DETECTED
0.0
0.44
NO-PIN-DETECTED
0.0
0.57
NO-PIN-DETECTED
0.0
0.49
NO-PIN-DETECTED
0.0

Discussion

Successes:

- Can process both video and images
- F1-Score of 0.87 on the classifier
- Predictions on images are made very fast and efficiently
- Minimal lag or overhead

Limitations:

- Human pose estimation model occasionally fails to recognise people
- More data is needed to effectively generalize
- Processing video frame by frame is less efficient than batch processing

The Future?

- Supporting more than one pair of combatants
- Using a more accurate and powerful HPE algorithm
- Leveraging deep learning and convolutional neural networks
- Working with multiple camera angle inputs at once

Important Links!

Link to GitHub repo: https://github.com/talhaahussain/grappling-pose-identification