Optimizing Physical Hand Training Effectiveness Using Deep Learning and Unsupervised Learning

Names: Peter Asbæk Skøt (S185185) and Jacob Asbæk Wolf (S236897)

Course: Applied Machine Learning and Big Data

Course ID: 62T22

Date: May 30, 2025

Agenda

- 1 min Motivation
- 1 min Problem and Use Case
- 2 min CNN Model
- 2 min K-Means Model
- 2 min Software Demo (Live Inference)
- 2 min Conclusion and Perspectives

Motivation

The situation

- Imagine being 19 with CP and cognitive disabilities.
- Hand therapy can improve your quality of life, but it is demotivating.
- Jacob’s son had CP after a serious meningitis as a baby.

Our reflection

- Could machine learning motivate and support training at home?

Our solution

- A simple software solution providing live motivational feedback.
- We will now show you what we built…

Problem and Use Case

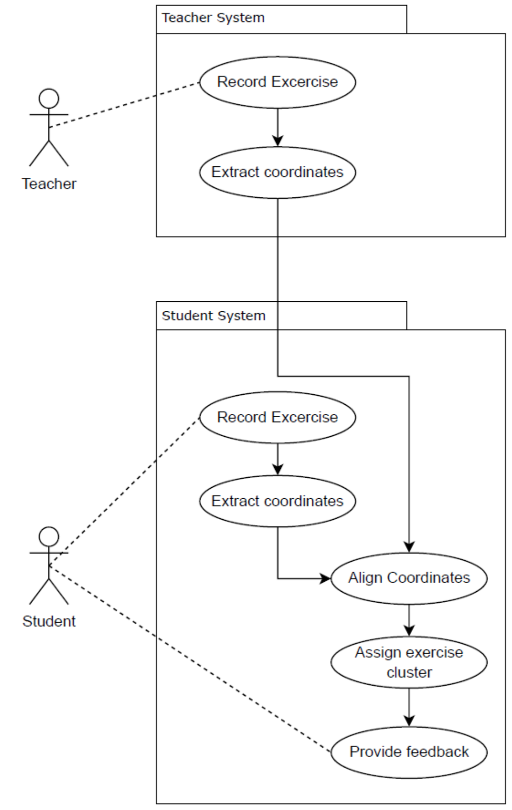
Problem

Hand exercise training targeted for patients with hand mobility issues with the goal of making the training more effective and the patient more motivated.

Hereunder:

1. Track hand positions.
2. Rate the students hand positions against the teacher's hand positions.
3. Apply supervised and unsupervised machine learning to a live stream

Use Case



Public

CNN Model

Purpose:

- Extract 3D landmarks from an image input.

Model data

Data source: HanCo

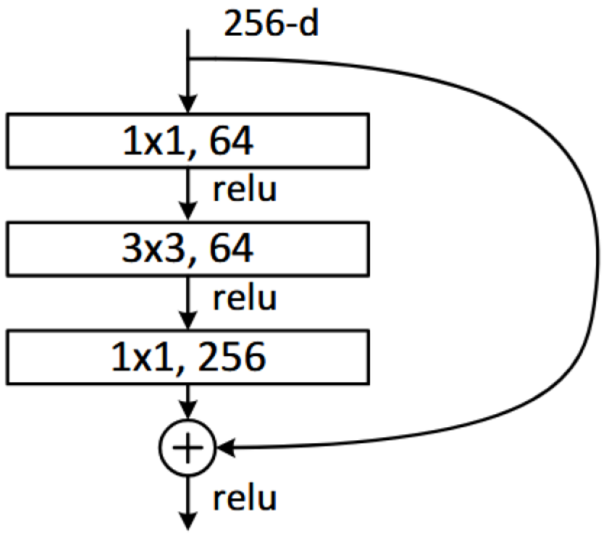
Input data: 224x224 RGB images of hands

Output data: 21 normalised 3D coordinates of landmarks.

Model Architecture

Backbone: ResNet50 + custom model head

Pretrained weights: ImageNet



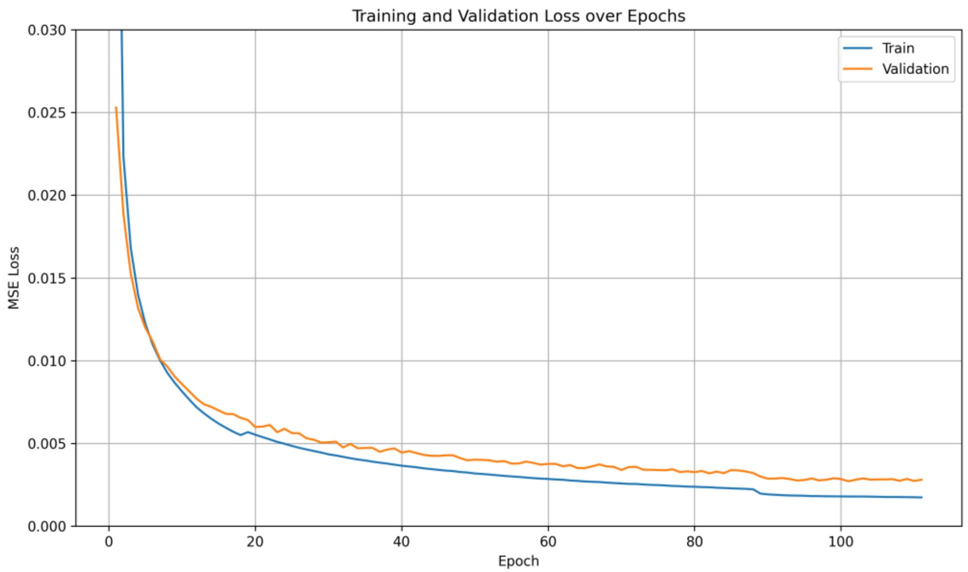
Layer	Dimensionality
FC + ReLU	2048 x 2048
FC + ReLU	2048 x 2048
FC	2048 x 63

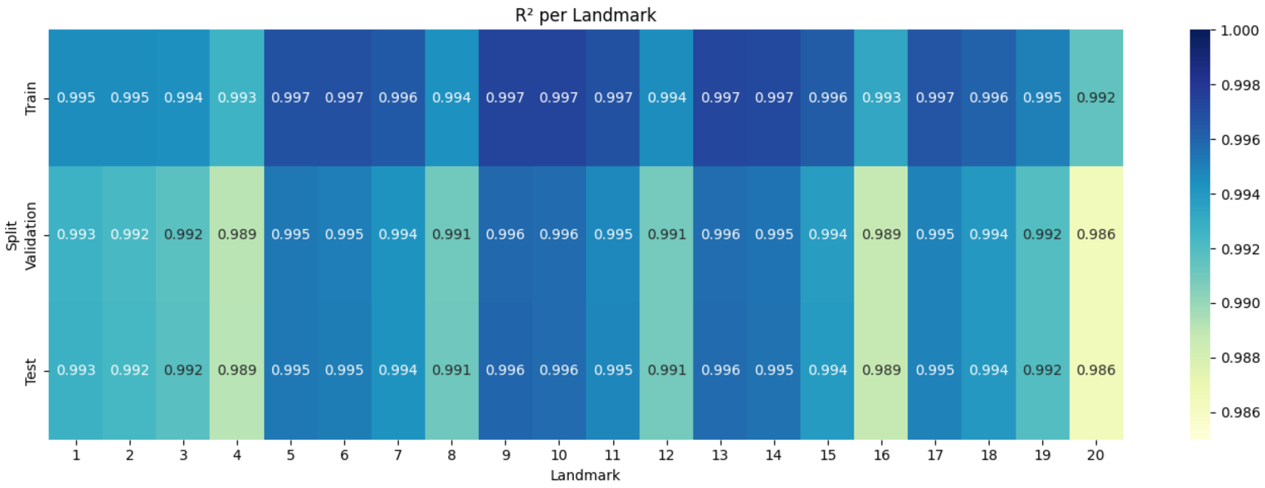
Table 1: Model head architecture with dimensionality of each layer.

CNN Results

- Strong within-dataset generalisation
  - R^2 of 0.982 on test set
  - Healthy loss curve with small train/val gap.
  - High variance explainability across all landmarks

Dataset	MSE Loss	R^2
Train	0.00172	0.985
Val	0.00272	0.982
Test	0.00271	0.982





K-Means Model

Purpose

Cluster landmarks into different exercise states

Model Data

Data Source: CNN prediction on self collected images (1174)

Input Data:

- 21 normalised 3D coordinates of landmarks
- 3 (1 good and 2 wrong) positions recorded (~400 each)
- Aligned to canonical coordinate system

Output Data: 63 dimensional centroid for each cluster.

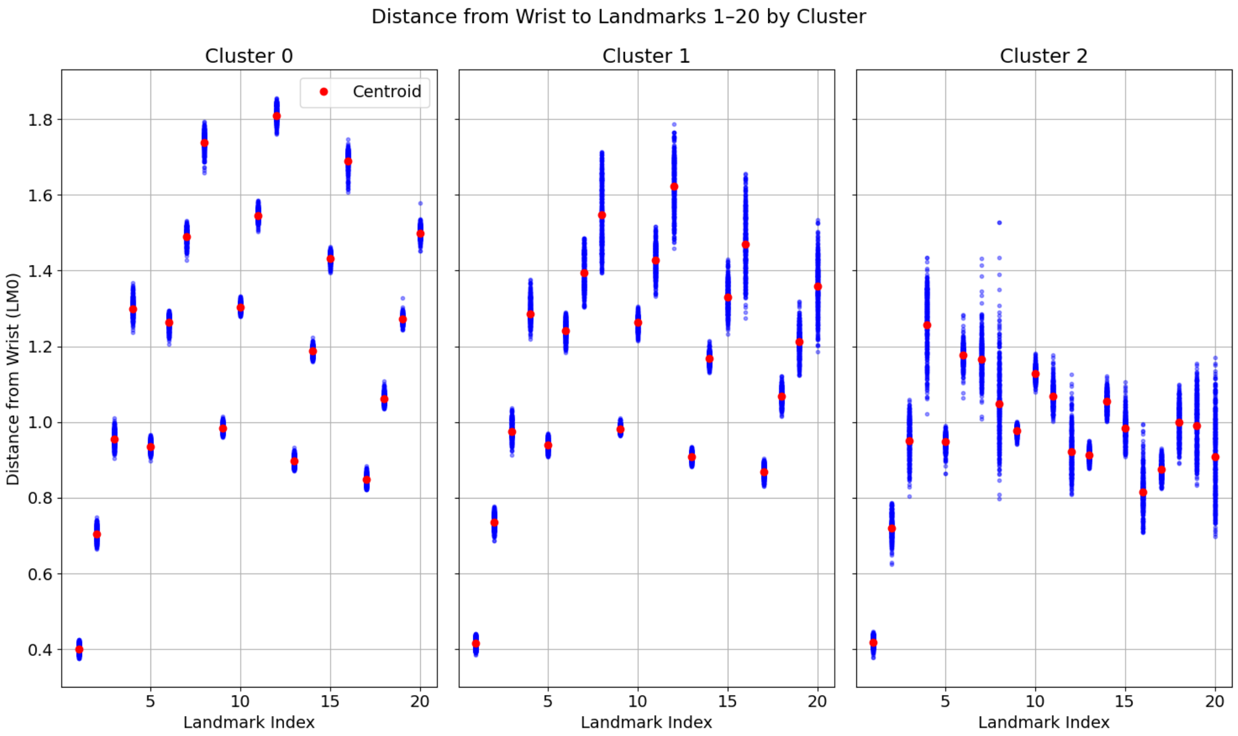
Cluster Selection

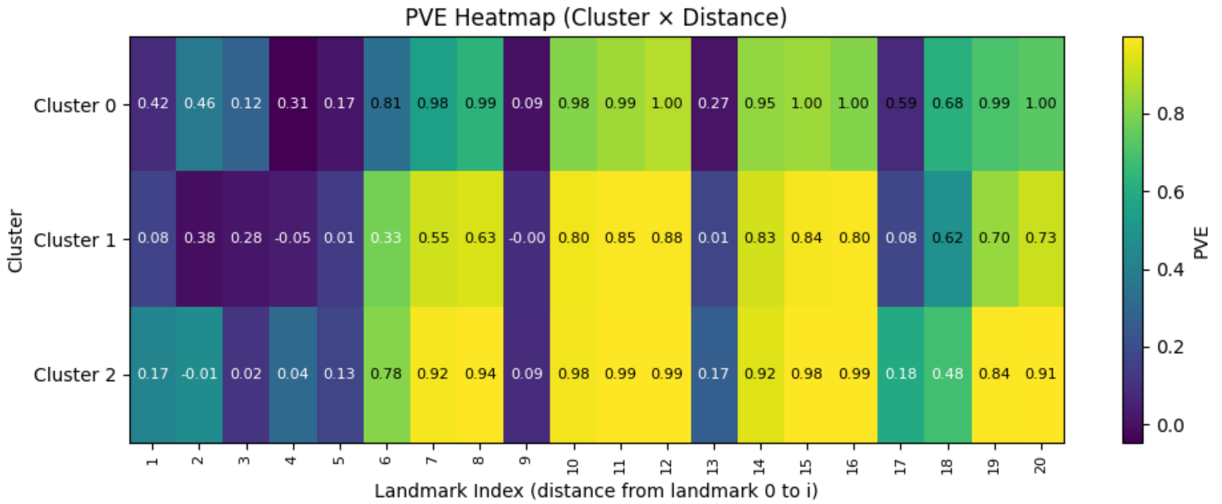
- Elbow method used
- Matches the expected 3 positions

K-Means Results

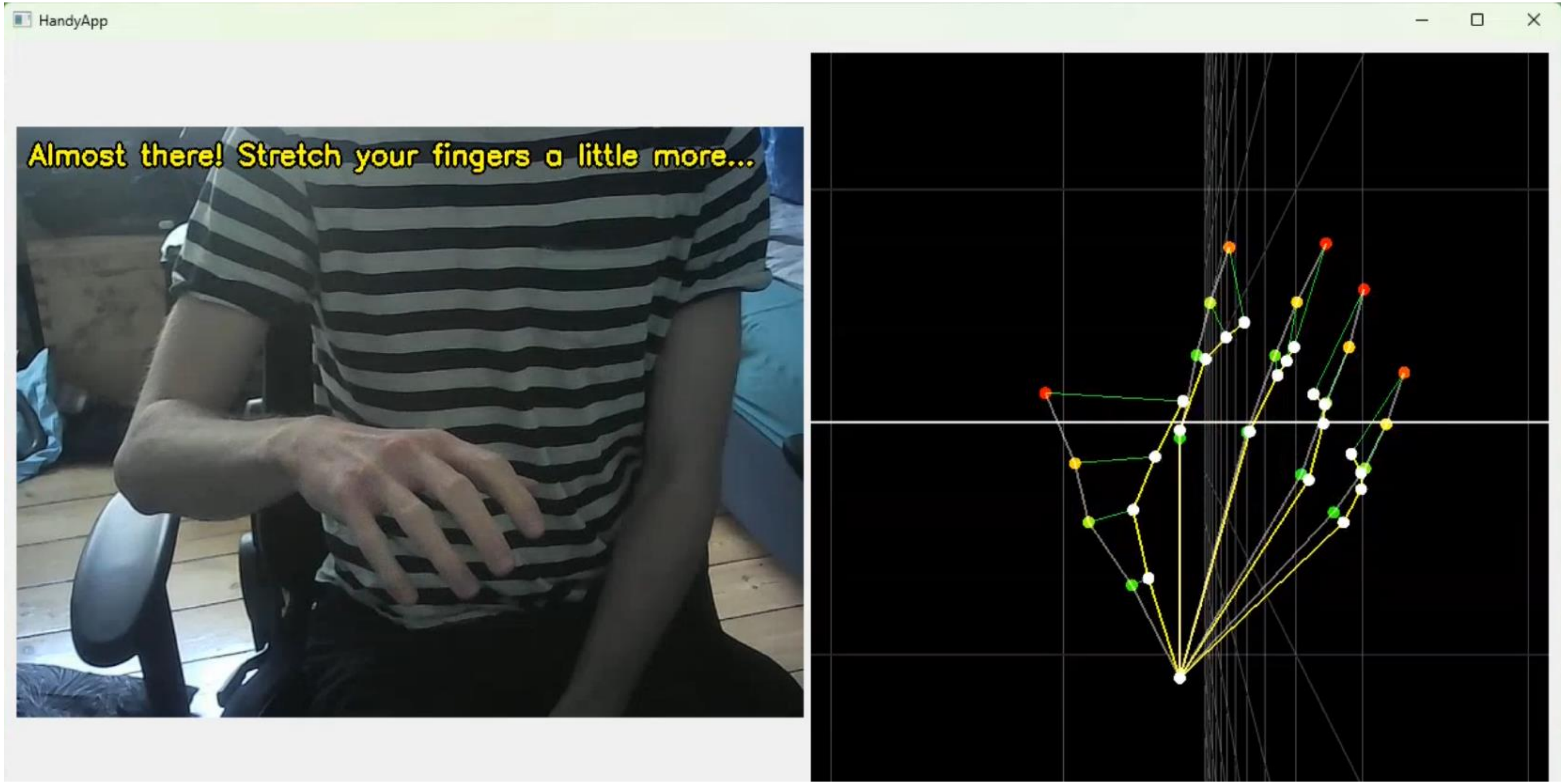
Centroids match the exercise’s three positions well.

- Wrist-landmark distances in each cluster matches our expectations from the three recorded image scenarios.
- The datapoints distribute close to their centroids
- PVE shows good explainability for landmarks further from the wrist (exluding the thumb).

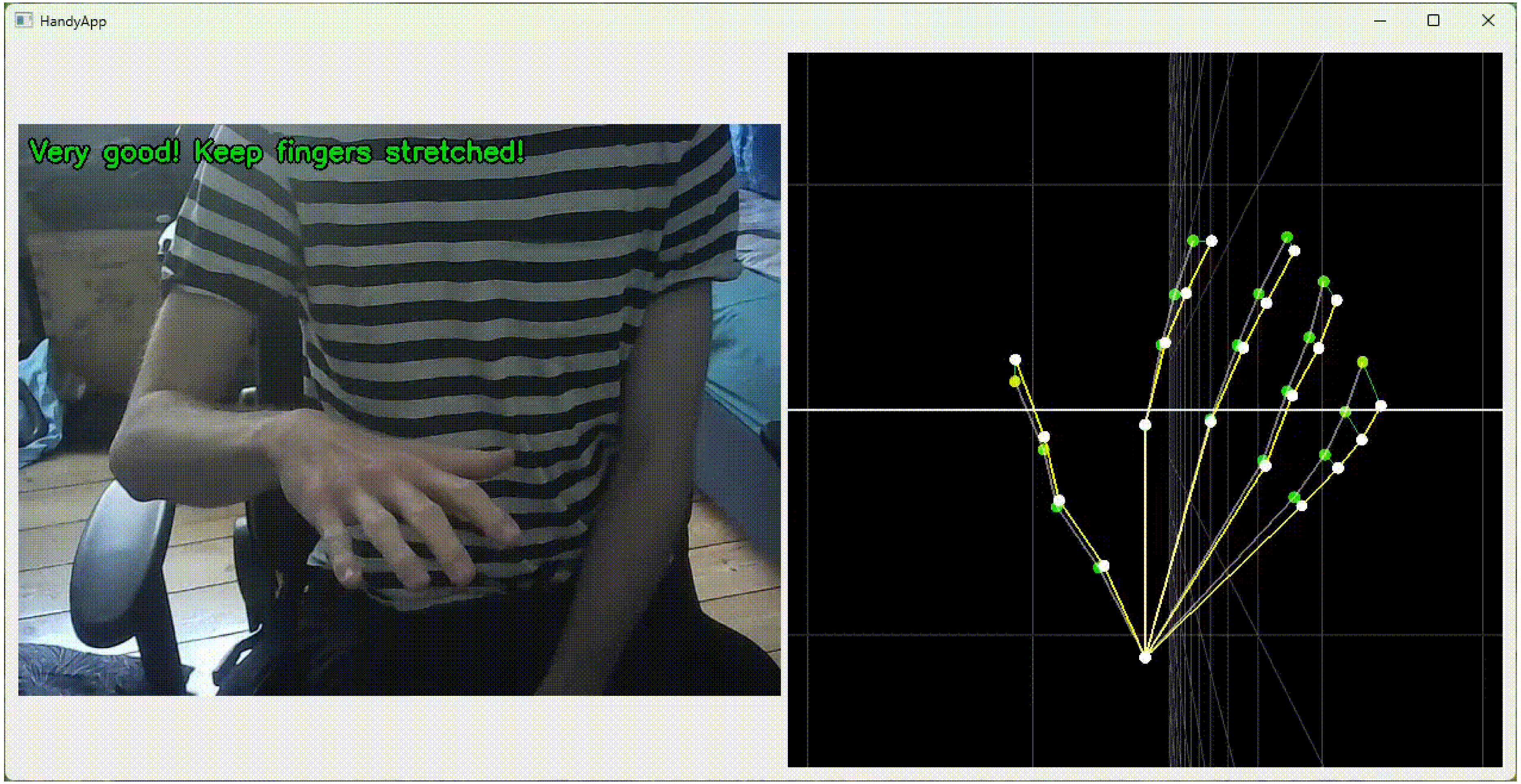




Software Demo (Live Inference)



Software Demo (Live Inference)



Conclusion

- Good CNN performance/generalisation with $R^2 = 0.982$
- Well performing K-means clustering matching our 3 scenarios
- Well perfoming live-inference software capable of guiding the patient

Perspectives

Dataset expansion/ Fine tuning

More exercises

Data collection

Regulatory compliance

Faster detection/ Mobile devices

Telemedicine

DTU