THE FEASIBILITY OF AI CAMERAS AS A PLATFORM FOR AUTONOMOUS, REAL-TIME REHABILITATION WITH PERFORMANCE FEEDBACK

Charli Posner (le19806)

Supervisor: Alessandro Masullo

1. Introduction

- Over 90% of physical therapy sessions take place in a home setting, without professional supervision [1]
- There is a need for a system that can monitor patient adherence and prevent injuries caused by poor exercise performance
- AI cameras are small, low-cost devices capable of pose tracking without GPU dependence, but little research has been done on the feasibility of using them for exercise monitoring in a clinical setting

Project aim: to create an automated rehabilitation system for an AI camera that can capture a subject's exercises and provide performance feedback in real-time.

2. Objectives

- Pretrain an image classification neural network architecture on human activity recognition sequences
- Retrain the network to differentiate between rehabilitation exercises and assess general performance
- Implement the system on the OpenCV AI Kit with Depth (OAK-D) [2] AI camera by merging it with a pose estimation pipeline

3. Pretraining on human activity recognition

- Pretraining was done with MoVi [3], a large motion capture dataset comprised of 20 activity classes
- A sample of joint position estimates was taken from the centre of each sequence and converted into an image, as shown in **Figure 1**
- ResNet-50 [4], a pre-trained image recognition architecture, was used to predict the activity classes from the image inputs
- The network achieved over 90% accuracy on validation data

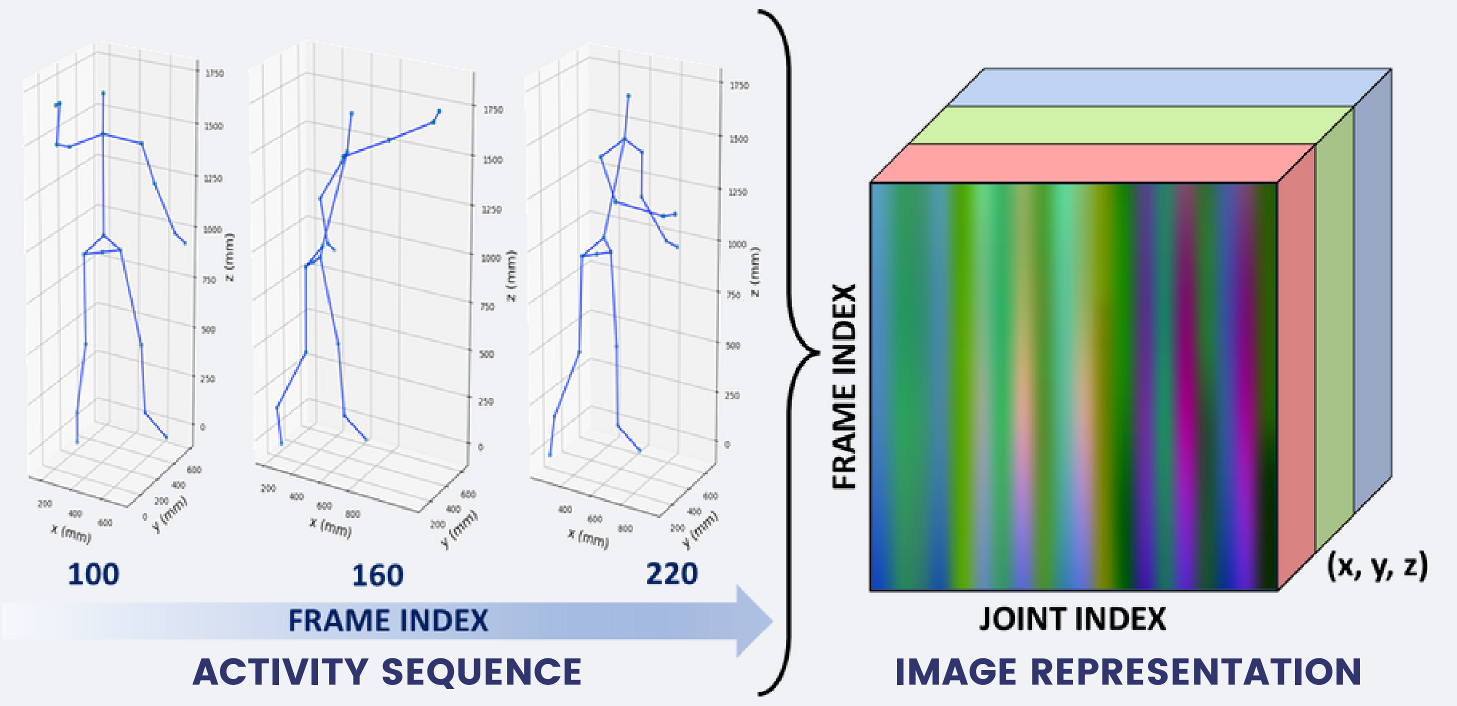

Figure 1: A visualisation of a "throw/catch" motion sample from MoVi (pre-normalisation), with dimensions [frames, joints, coordinates], and its ResNet-50 image representation with dimensions [224, 224, 3 (RGB)]

4. Retraining on rehabilitation sequences

- The KIMORE dataset [5], chosen for rehabilitation implementation, consists of five physical therapy exercises (**Figure 2**) performed by both healthy subjects and real patients
- Each sequence is paired with a performance assessment score out of 50, awarded by professional clinicians

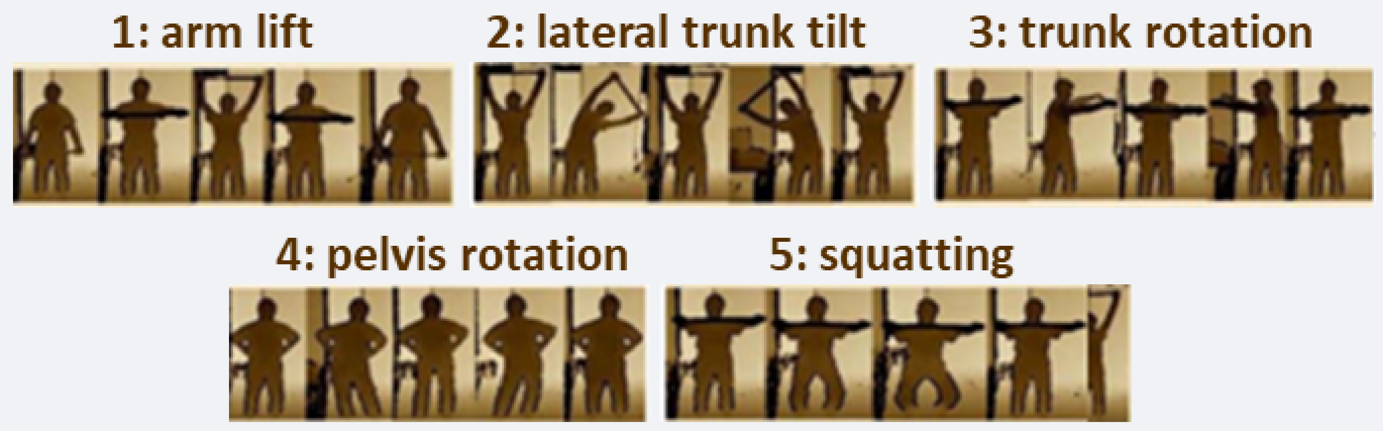

Figure 2: the five rehabilitation exercises in the KIMORE dataset, designed for patients suffering from lower back pain.

- We abstracted the score into a three-star rating system, where over half of sequences were awarded 3* and only the lowest 15% given 1*
- Class weighting was used to prioritise the correct classification of 1* sequences, which are most likely to exacerbate injuries in practice
- The pre-trained ResNet-50 layers were utilised to predict the exercise and rating simultaneously
- Exercise classification accuracy on test data was over 95%
- The rating recall scores for the pre-trained network, as well as networks without pretraining, were evaluated, as shown in **Table 1**
- The pre-trained network showed an advantage for identifying 1* ratings, which is the main priority for this application

RATING CLASSIFICATION RECALL SCORES FOR DIFFERENT NETWORK CONFIGURATIONS

NETWORK WEIGHTS	1*	2*	3*	WEIGHTED AVERAGE
MoVi Pre-trained	**0.977**	0.875	0.972	0.943
ResNet-50 Imported	0.912	0.883	0.977	0.940
ResNet-50 Random	0.871	0.895	0.961	0.923

Table 1: the rating recall scores for the network with MoVi pre-training, compared to ResNet-50 with both imported and randomly-initialised weights, evaluated on a test set comprised of 25% of the subjects. The 2* class likely has the lowest recall due to the high frequency of scores close to the thresholds.

5. AI camera implementation

- BlazePose [6] was chosen for 3D pose estimation due to its fast run-time on mobile devices with little accuracy trade-off
- Since the network will never be trained on BlazePose data, the topologies of all skeletons were normalised according to the model shown in **Figure 3**

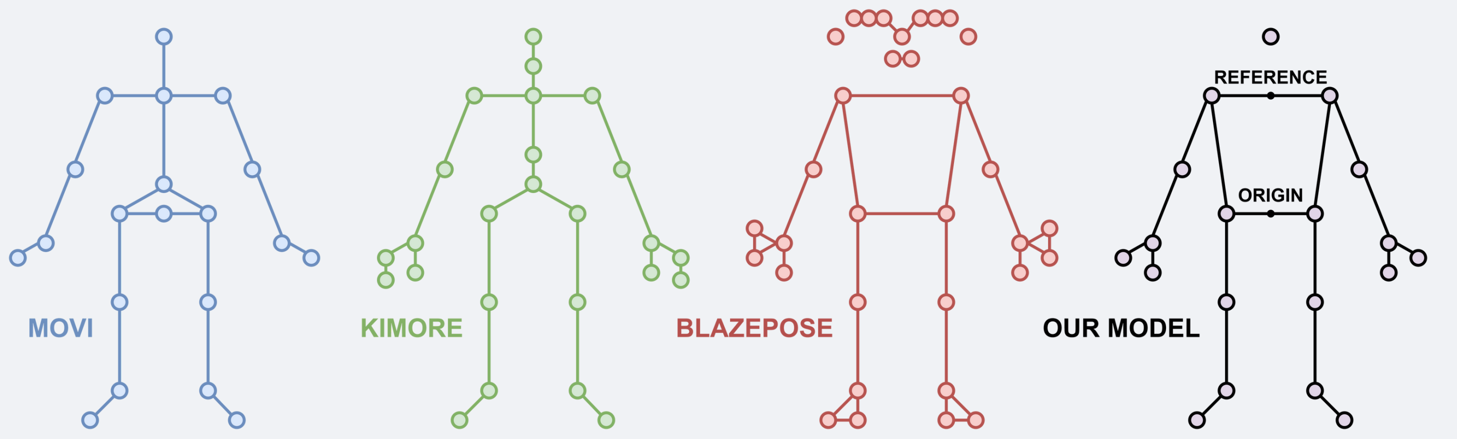

Figure 3: The skeleton topologies. They differ from one another with respect to the joints, the coordinate units and axes, and the origin location. We kept only the common joints, set the origin to the mid-hips, and scaled each skeleton according to the distance between the origin and mid-shoulders.

- Model conversion was applied to transform the rehabilitation network to a compatible format for the OAK-D hardware
- An existing BlazePose AI camera pipeline [7] was modified to produce normalised RGB images, which were fed as input to the network
- Preliminary tests have indicated that the network can reliably classify KIMORE exercises from normalised BlazePose inputs

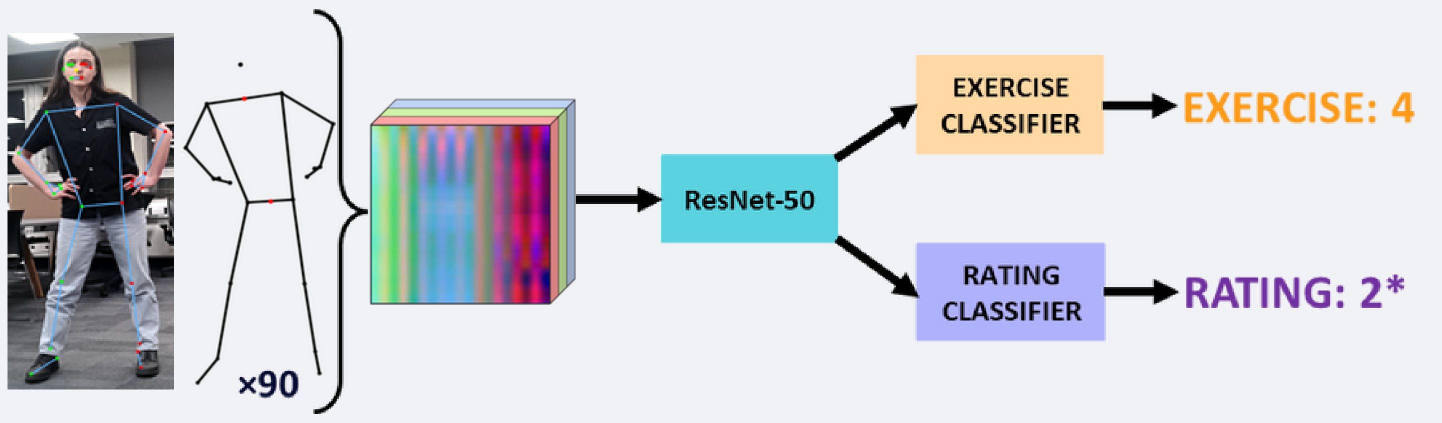

Figure 4: A diagram of the automated rehabilitation pipeline. 90 frames of BlazePose joint positions are normalised, as described in **Figure 3**, and processed into images, as shown in **Figure 1**. These are then fed as input to the neural network, which classifies the exercise and rating simultaneously.

6. Conclusion and future work

- The high recall values in **Table 1** indicate that the network can reliably identify exercise performance quality
- The network's ability to classify KIMORE exercises from BlazePose images suggests that normalisation succeeded in enabling system compatibility
- The final project tasks include finishing the automated rehabilitation pipeline on the OAK-D and assess its accuracy and frame-rate
- Further work may include providing joint-specific feedback for improving rating scores, or implementing the system on a lightweight alternative to ResNet-50 and assessing performance trade-offs
- The system's rating classification accuracy will also need to be verified in a clinical setting to ensure that the feedback is reliable

References

[1] R. Komatireddy, "Quality and quantity of rehabilitation exercises delivered by a 3-d motion controlled camera: A pilot study," International Journal of Physical Medicine Rehabilitation, vol. 02, 08 2014.
[2] "OpenCV AI Kit with Depth," OAK-D Documentation. [Online]. Available: https://docs.luxonis.com/projects/hardware/en/latest/pages/BW1098OAK.html
[3] S. Ghorbani, K. Mahdaviani, A. Thaler, K. Kording, D. J. Cook, G. Blohm, and N. F. Troje, "Movi: A large multi-purpose human motion and video dataset," PLOS ONE, vol. 16, no. 6, pp. 1–15, 06 2021. [Online]. Available: https://doi.org/10.1371/journal.pone.0253157
[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," CoRR, vol. abs/1512.03385, 2015. [Online]. Available: http://arxiv.org/abs/1512.03385
[5] M. Capecci, M. G. Ceravolo, F. Ferracuti, S. Iarlori, A. Monteriù, L. Romeo, and F. Verdini, "The KIMORE Dataset: KInematic Assessment of MOvement and Clinical Scores for Remote Monitoring of Physical REhabilitation," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 27, no. 7, pp. 1436–1448, 2019
[6] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, "BlazePose: On-device Real-time Body Pose tracking," CoRR, vol. abs/2006.10204, 2020. [Online]. Available: https://arxiv.org/abs/2006.10204
[7] geaxgx, "BlazePose Tracking with DepthAI." [Online]. Available: https://github.com/geaxgx/depthaiblazepose