EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL INTELLIGENCE

DeepRehab: Real Time **Deep** Pose Refinement on the Edge

A **Rehab**ilitation Use Case

BRUNO CARLOS DOS SANTOS MELÍCIO
GÁBOR BARANYI
Computer Science Msc,
Eötvös Loránd University, Faculty of Informatics

SUPERVISOR | DR. ANDRÁS LŐRINCZ
Senior Researcher, PhD
Eötvös Loránd University, Faculty of Informatics

Budapest, 25 June 2021

Content

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL INTELLIGENCE

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL INTELLIGENCE

01. Introduction

Motivation; Problem; Objective.

Introduction - Motivation

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL INTELLIGENCE

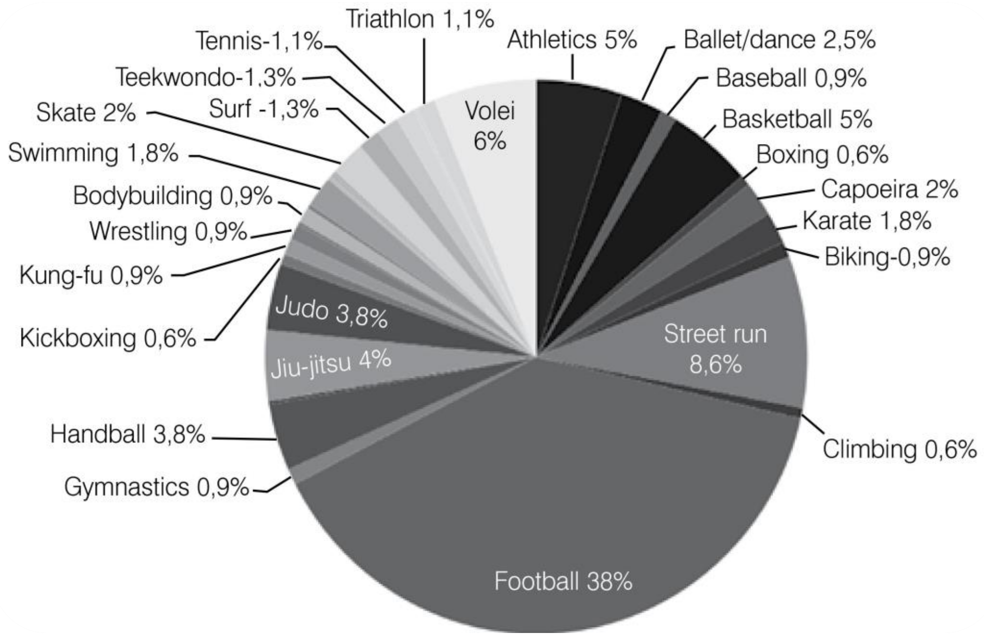


Fig. 1 - Division according to modalities practiced. [1]

Fig. 2 - Keypoint format for skeletons from COCO dataset.
Source: https://beyondminds.ai/blog/an-overview-of-human-pose-estimation-with-deep-learning/

Around **40%** of all sports injuries are related to the **knee joint**. [1]
500 000 total knee replacements are performed **annually** in the USA. [1]

Artificial Intelligence can aid patients during rehabilitation through **Human Body Pose estimation and tracking**.

[1] Nicolini, Alexandre Pedro et al. “Common injuries in athletes' knee: experience of a specialized center.” *Acta ortopedica brasileira* vol. 22,3 (2014): 127-31. doi:10.1590/1413-78522014220300475

01

Introduction – Problem

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL INTELLIGENCE

Most **Deep Learning** algorithms are large with **a lot of parameters**, which **restrict them to Cloud Computing** for **real time performance**.

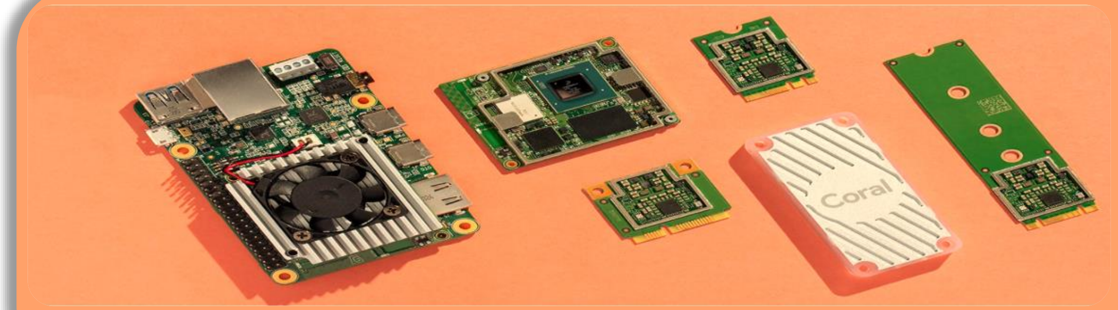


Fig. 3 – Edge TPU devices. Source: https://morioh.com/p/8cdc12e942f1

Google Coral made available **PoseNet** [4], a 2D Deep Learning model compatible with the Edge TPU devices, that estimates **17 keypoints**.
However, it is not so accurate, as it sacrifices accuracy for speed.

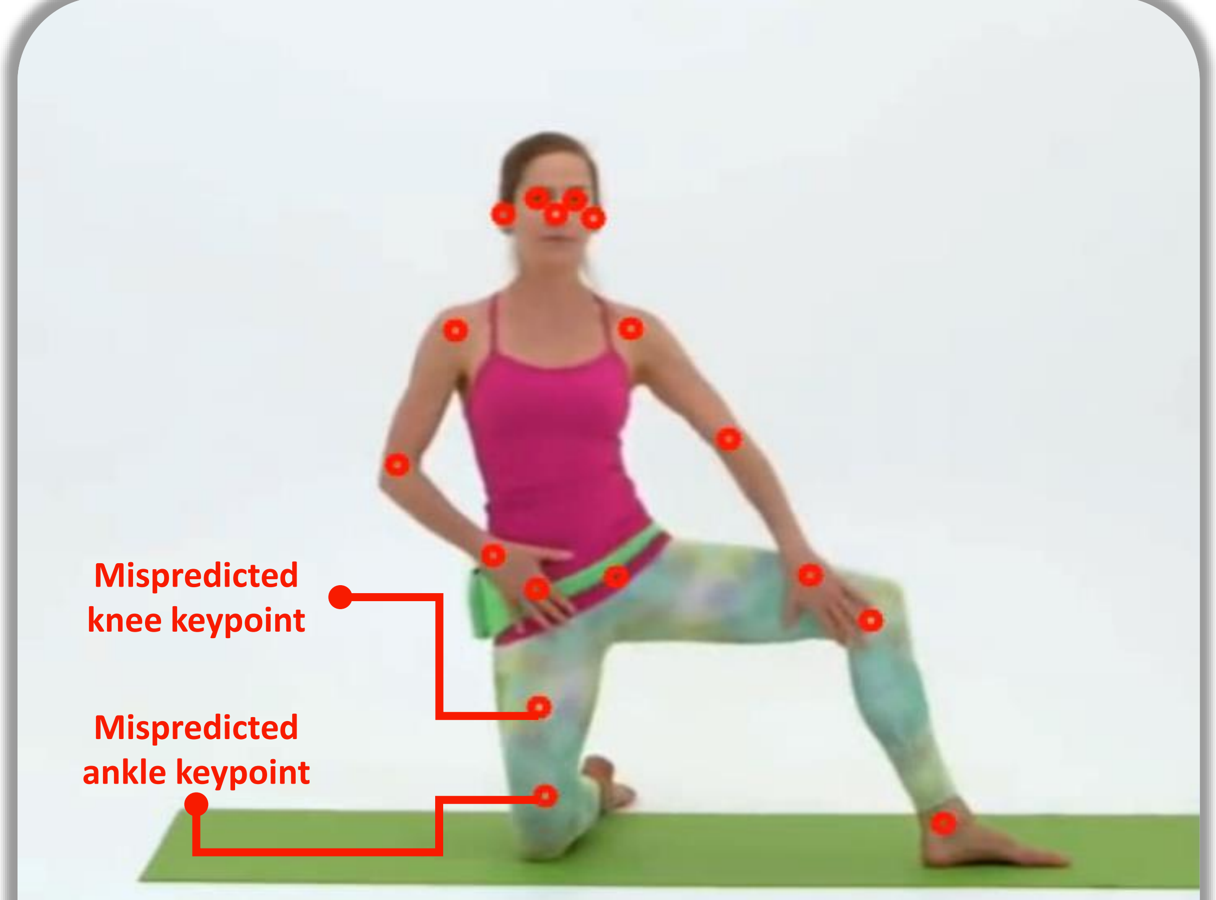
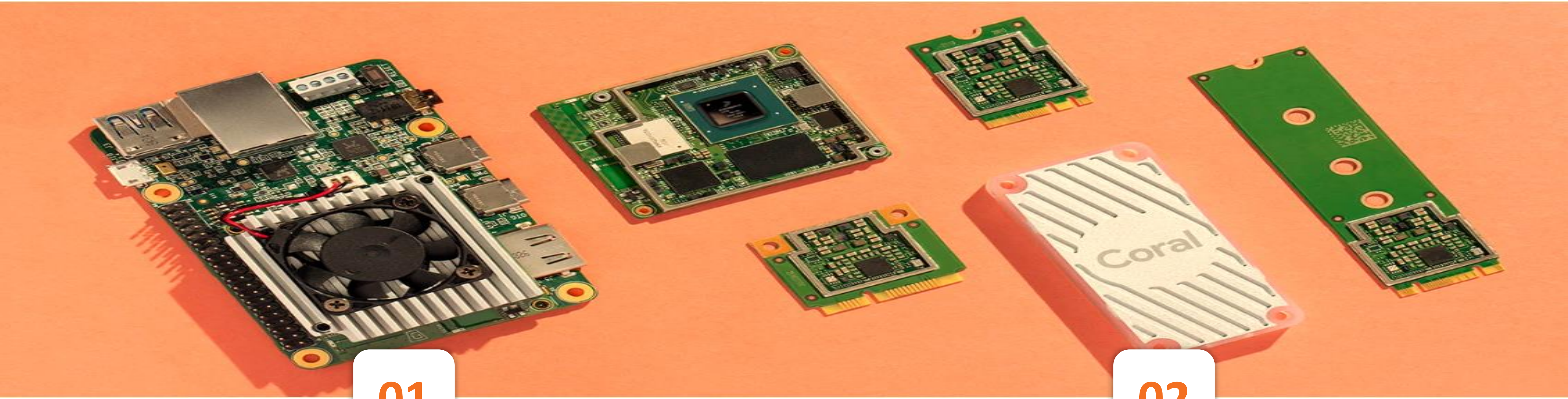


Fig. 4 – Google Coral PoseNet with MobileNet backbone results. [4]

[4] Google Coral PoseNet models: https://github.com/google-coral/project-posenet

Introduction – Objective

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL INTELLIGENCE



01

Reduce the error of PoseNet's pose estimations, focusing on the lower parts of the body.

02

Develop a **fast** and **accurate 2D Deep Learning model compatible with the Edge TPU devices** and able to estimate **additional feet keypoints.**

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL INTELLIGENCE

02. Related Work

Related Work

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL INTELLIGENCE

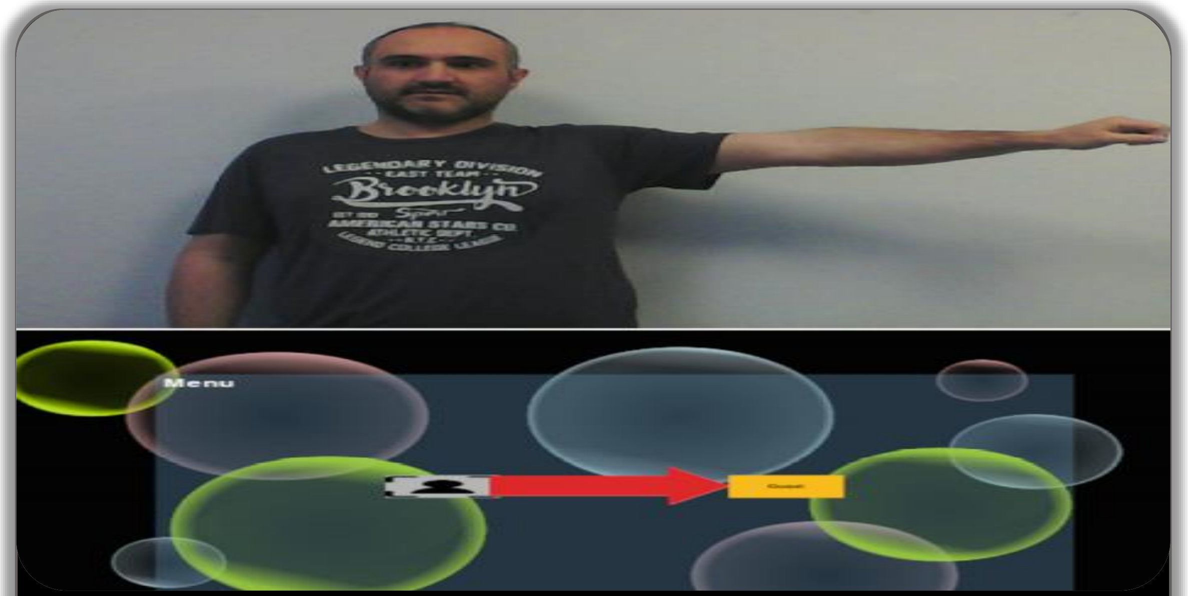


Fig. 5 - Exergame menu. Pardos et al. (2020) [5]

Fig. 6 - Golfer posture estimated by PoseNet. Kim et al. (2020) [6]

- Edge-native platform for exergames in a rehabilitation center;
- Real-time pose and gesture detection with Coral USB Accelerator;
- PoseNet with MobileNet backbone.

- Identify metrics of the golfer such as posture, swing tempo, and swing consistency in order to provide feedback. (Google Coral Dev board);
- PoseNet with MobileNet backbone
- Refine the outputs of the model using a Savitzky-Golay filter.

[5] Pardos et al. "Introducing an Edge-Native Deep Learning Platform for Exergames". In: Artificial Intelligence Applications and Innovations. Cham: Springer International Publishing, 2020, pp. 88–98. ISBN: 978-3-030-49186-4.
[6] Kim et al. "Applying Pose Estimation to Predict Amateur Golf Swing Performance using Edge Processing (February 2020)". In: IEEE Access PP (Aug. 2020), pp. 1–1. DOI: 10.1109/ACCESS.2020. 3014186.

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL INTELLIGENCE

03. Methods

PoseNet at the edge; Refined PoseNet with filtering methods; DeepRehab framework; Foot-PoseNet.

Methods – PoseNet at the Edge

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL INTELLIGENCE

We chose the version of **PoseNet with MobileNet** as backbone and HD image resolution of **1281x721x3** pixels as baseline.

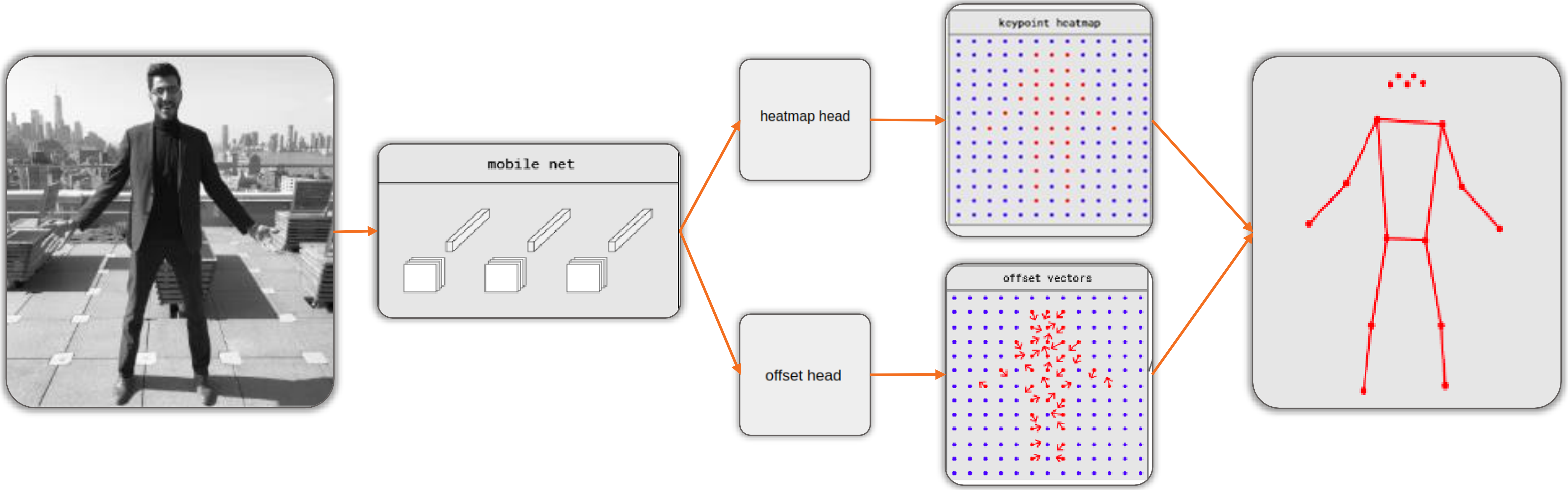


Fig. 7 – PoseNet with MobileNet backbone architecture [9]

[9] Dan Oved. Real-time Human Pose Estimation in the Browser with TensorFlow.js. 2020 (accessed November 27, 2020). URL: https://medium.com/tensorflow/real-time-human-pose-estimation-in-the-browser-with-tensorflowjs-7dd0bc881cd5.

03 Methods – PoseNet at the Edge

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL INTELLIGENCE

We chose the version of **PoseNet with MobileNet** as backbone and HD image resolution of **1281x721x3** pixels as baseline.

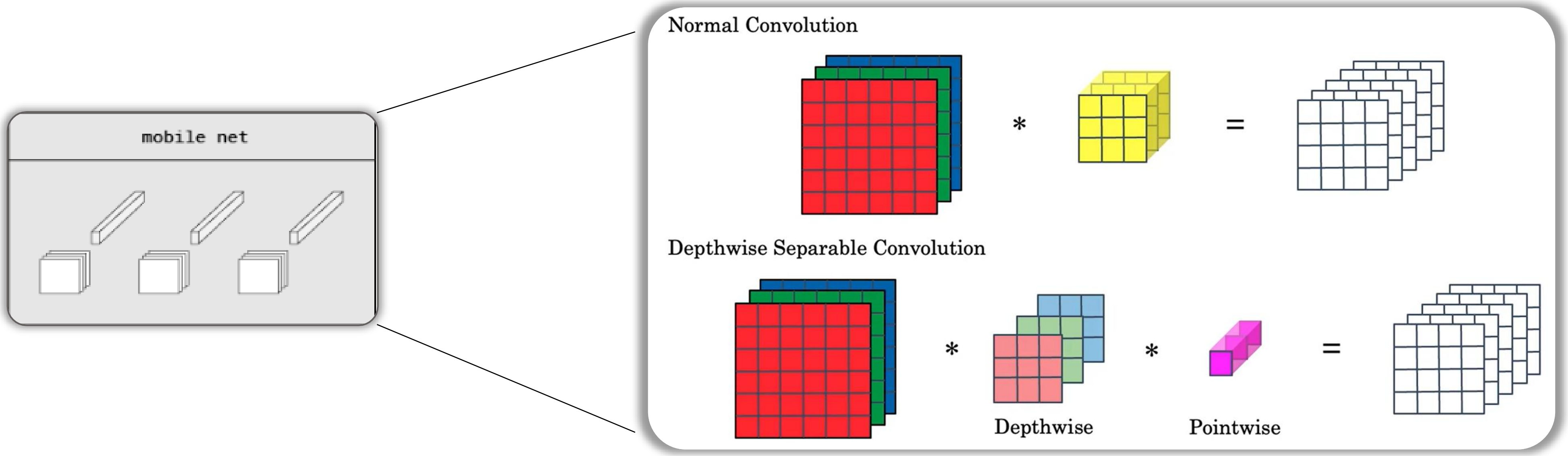


Fig. 8 – MobileNetV1 backbone architecture [10]

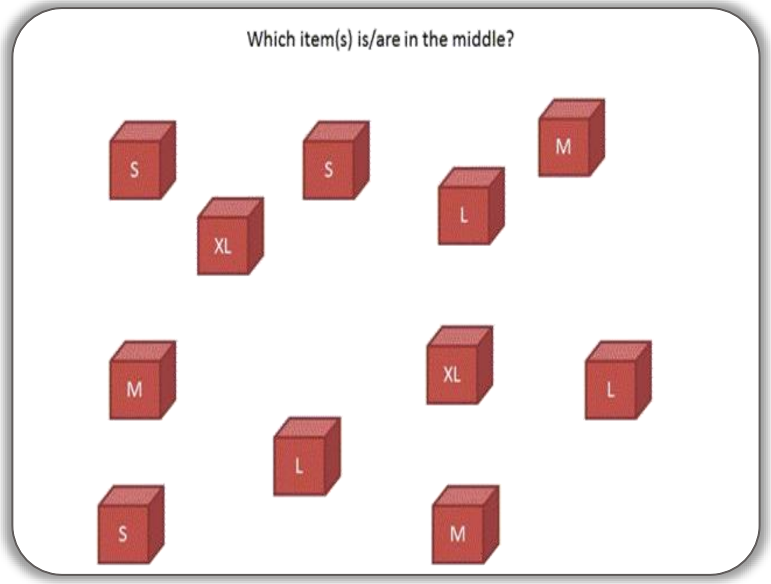
[10] Andrew G. Howard et al. “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications”. 2017. arXiv: 1704.04861 [cs.CV]

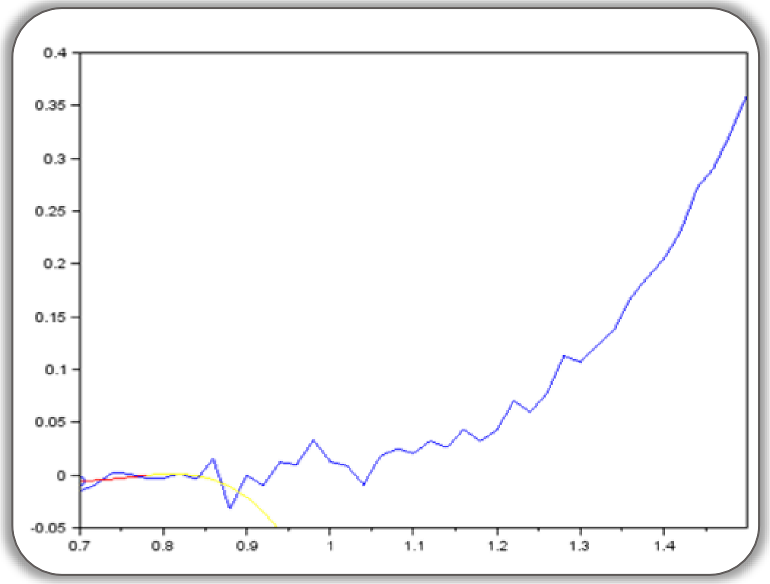
Methods – Refined PoseNet with filtering methods

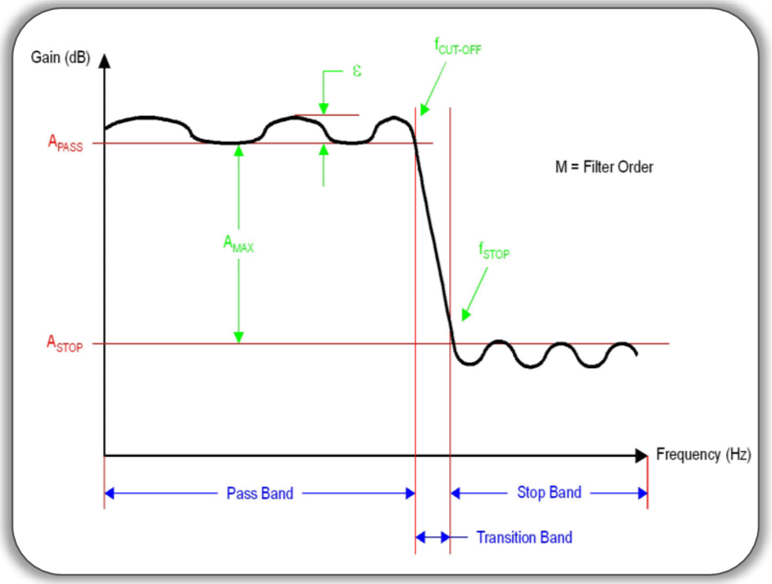
EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL INTELLIGENCE

PoseNet outputs contain errors, specifically when one or more keypoints are occluded, and it has **low precision in the bottom body parts**. In some situations, certain keypoints are not detected and that influences the estimation of others. These limitations can be addressed with **noise filtering**.

- **Median filter**
- **Savitkzy-Golay filter**
- **Chebyshev filter**







$$y[m] = median\{x[i], i \in \omega\}$$

$$(y_k)_s = \frac{\sum_{i=-M}^{M} A_i y_{k+i}}{\sum_{i=-M}^{M} A_i}, \qquad M = \frac{N-1}{2}$$

$$G_n(\omega) = \frac{1}{\sqrt{1+\varepsilon^2 T_n^2\left(\frac{\omega}{\omega_0}\right)}}$$

Methods – DeepRehab

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL INTELLIGENCE

By combining the outputs of PoseNet with the previously described filters, we developed a framework called **DeepRehab**.

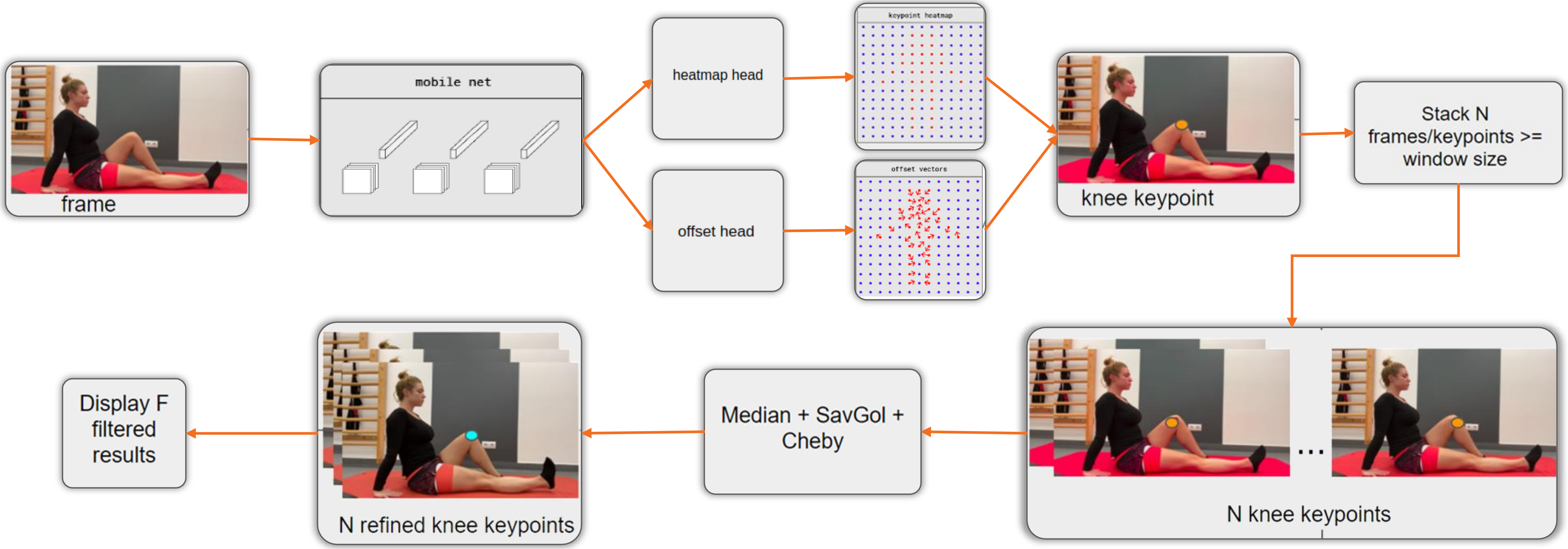


Fig. 9 – DeepRehab framework.

Methods – Foot-PoseNet

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL INTELLIGENCE

- **Filtering smooths the predictions over time, but is limited by the performance of PoseNet!**
- We developed **Foot-PoseNet with ResNet101** as backbone and image resolution of **224x224x3** pixels, a bottom-up approach trained on **23 keypoints** (17 body and 6 feet) from COCO-Whole Body dataset.

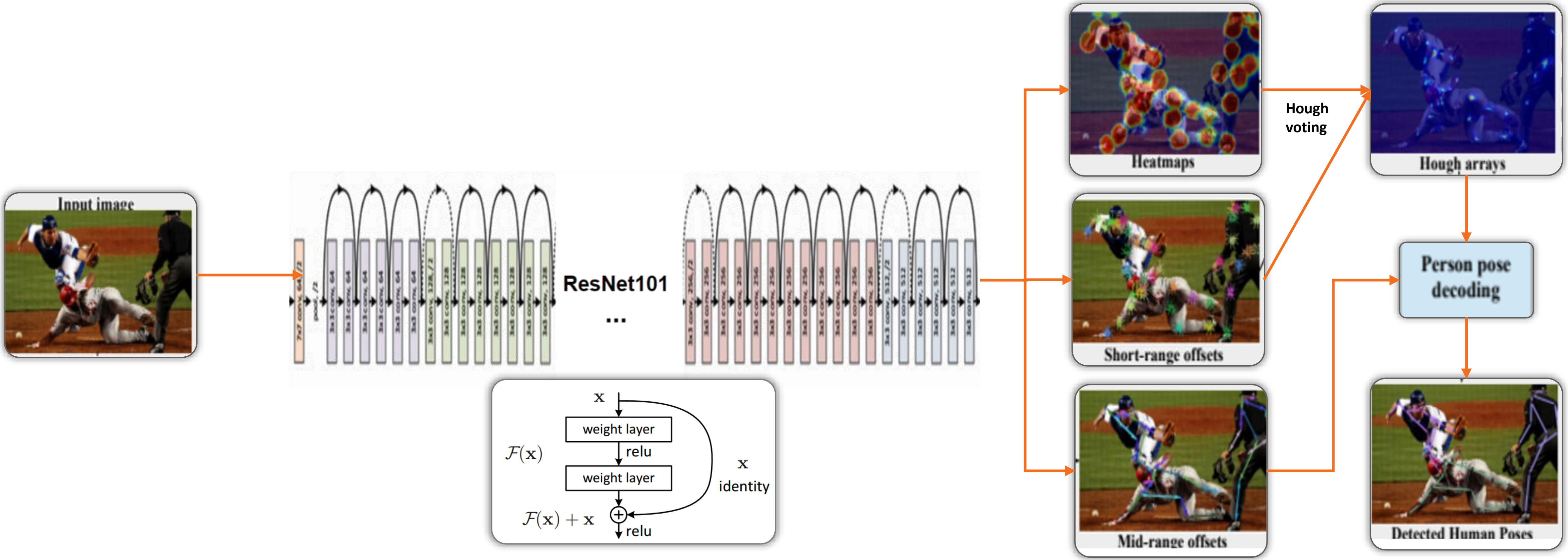


Fig. 10 – F-PoseNet with ResNet101 as backbone architecture

Architecture inspired on: George Papandreou et al. PersonLab: Person Pose Estimation and Instance Segmentation with a Bottom-Up, Part-Based, Geometric Embedding Model. 2018. arXiv: 1803.08225 [cs.CV].

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL INTELLIGENCE

04. Results

Datasets; PoseNet at the edge; Refined PoseNet with filtering methods; Foot-PoseNet.

04

Results - Datasets

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL
INTELLIGENCE

We collected a total of **30 videos** from YouTube resulting in approximate **50 000 frames**.

We chose a subset of videos from the testing set of **MPI-INF-3DHP** dataset and we use the **2D annotations**.

Fig. 11 – Rehab videos – collected dataset

Fig. 12 – Subset of MPI-INF-3DHP [11]

[12] D. Mehta et al. “Monocular 3D Human Pose Estimation in the Wild Using Improved CNN Supervision”. In: 2017 International Conference on 3D Vision (3DV). 2017, pp. 506–516. DOI: 10.1109/3DV.2017.00064

Results – PoseNet at the Edge

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL INTELLIGENCE

The models were tested on the **Coral USB accelerator** plugged into a laptop with an **Intel Core i7 6700HQ, 4 cores: max. 3.1 GHz**.

Fig 13. MobileNet 1281x721

Results – PoseNet at the Edge

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL INTELLIGENCE

The models were tested on the **Coral USB accelerator** plugged into a laptop with an **Intel Core i7 6700HQ, 4 cores: max. 3.1 GHz**.

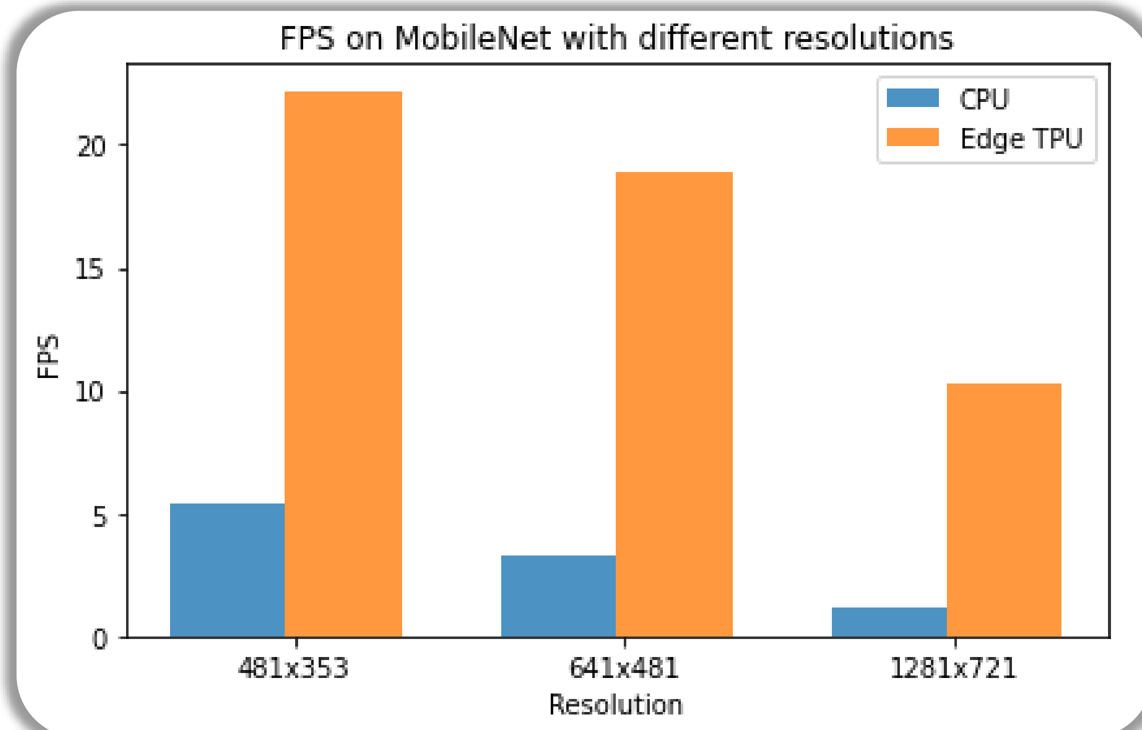


Fig 15. MobileNet FPS on CPU vs Edge TPU

Results – Refined PoseNet with filtering methods

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL INTELLIGENCE

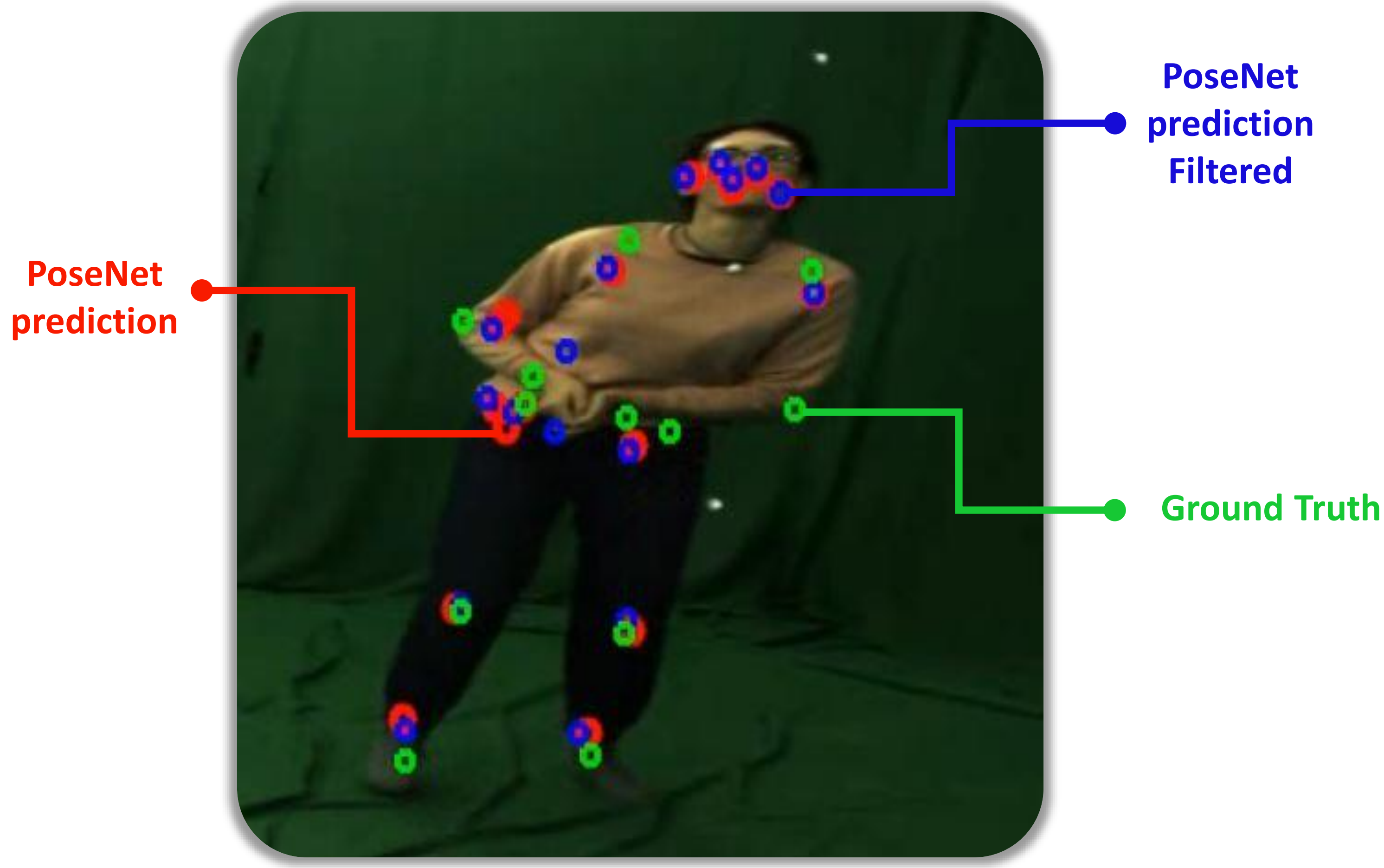


Fig 17. Results on video from MPI-INF-3DHP

Filter configurations that achieved the best results:

+ Median and SavGol window-size: 15
+ SavGol polynomial degree: 3
+ Chebyshev order: 3
+ Chebyshev ripple: 1
+ Chebyshev cutoff frequency: 0.2

04

Results – Refined PoseNet with filtering methods

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL INTELLIGENCE

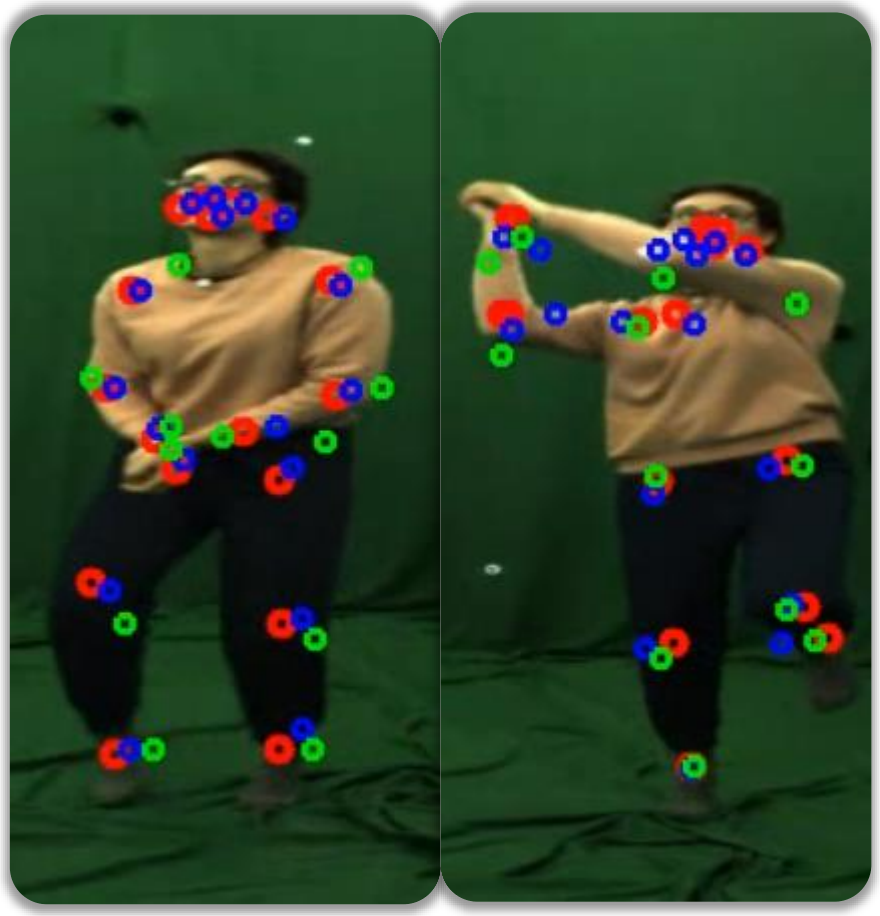
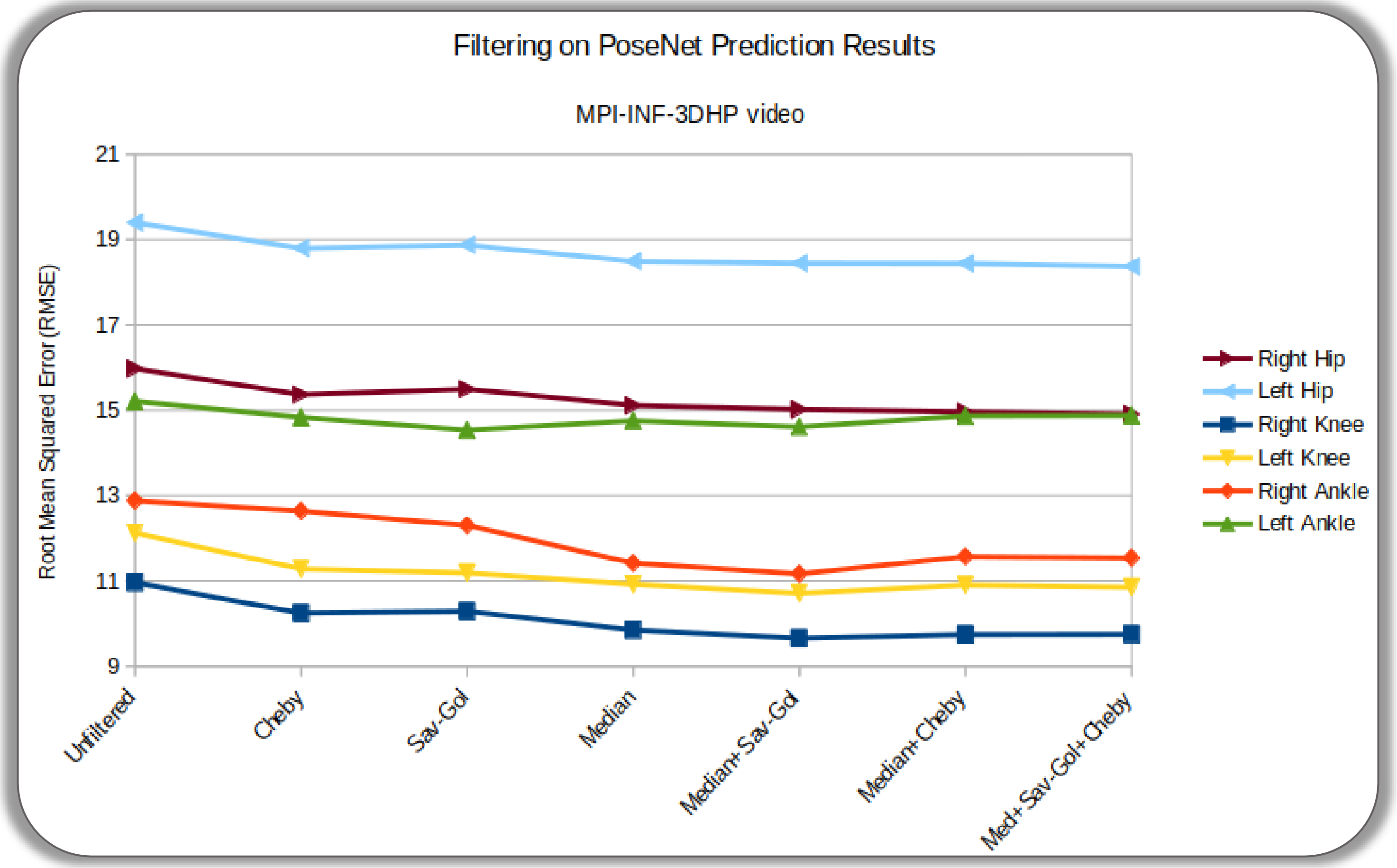
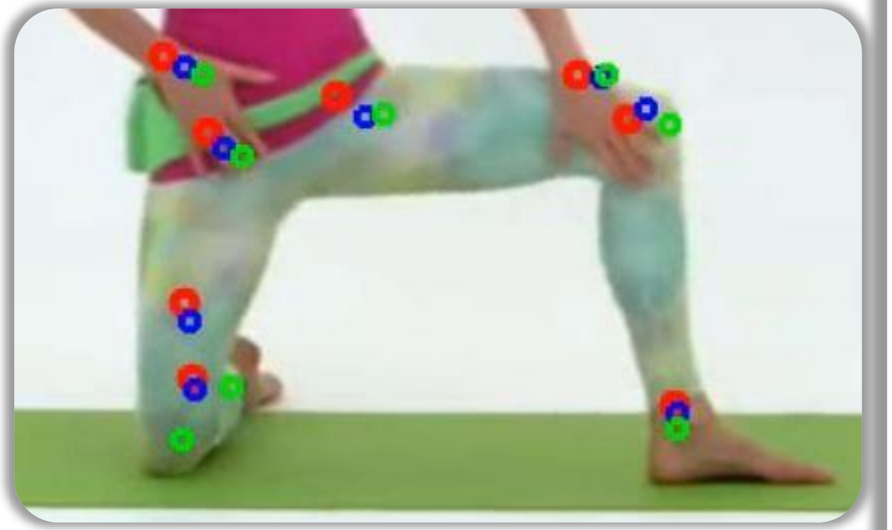


Fig 18. Results on video from MPI-INF-3DHP



Results – Refined PoseNet with filtering methods

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL INTELLIGENCE



Right Ankle X coordinate

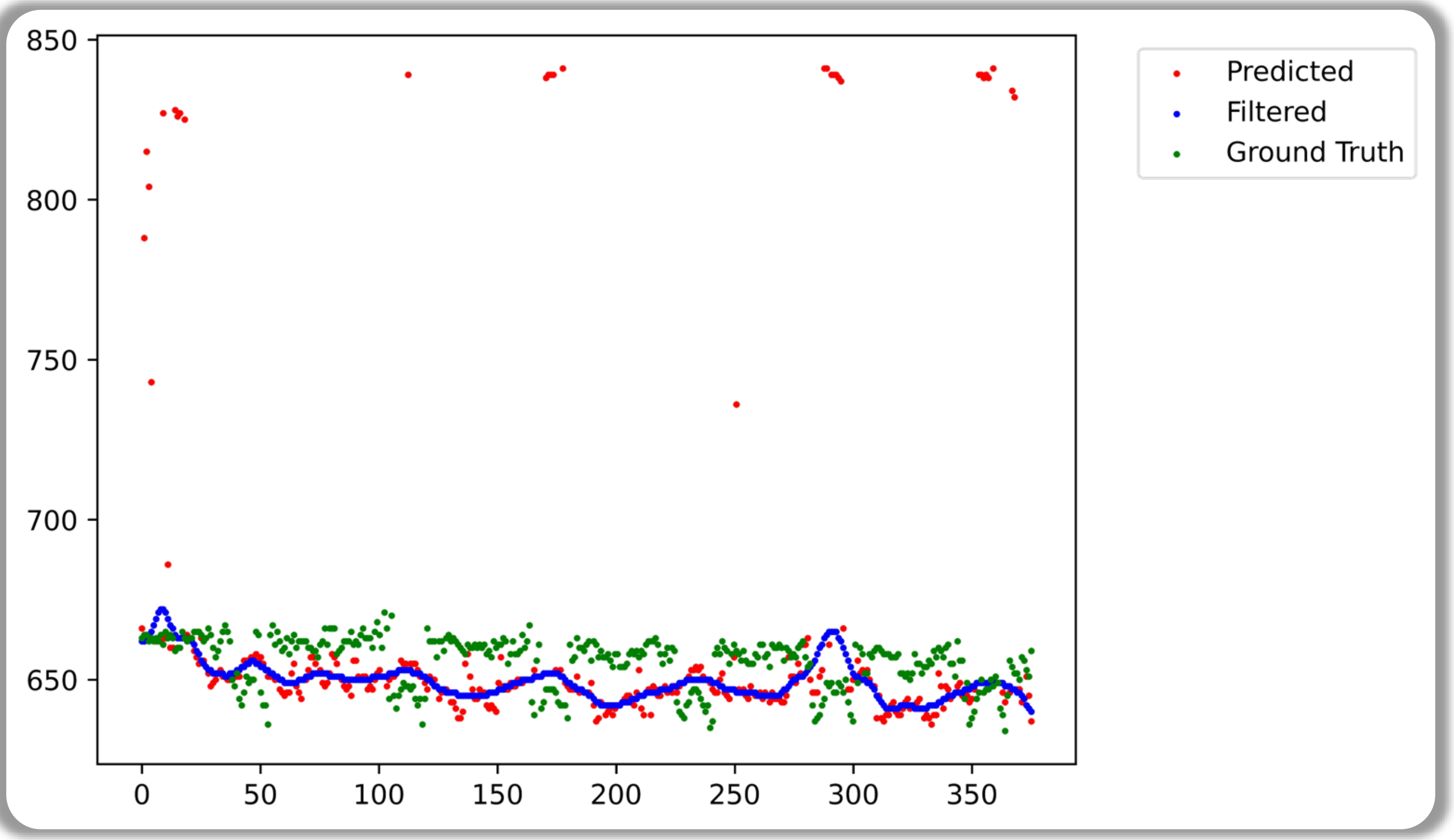
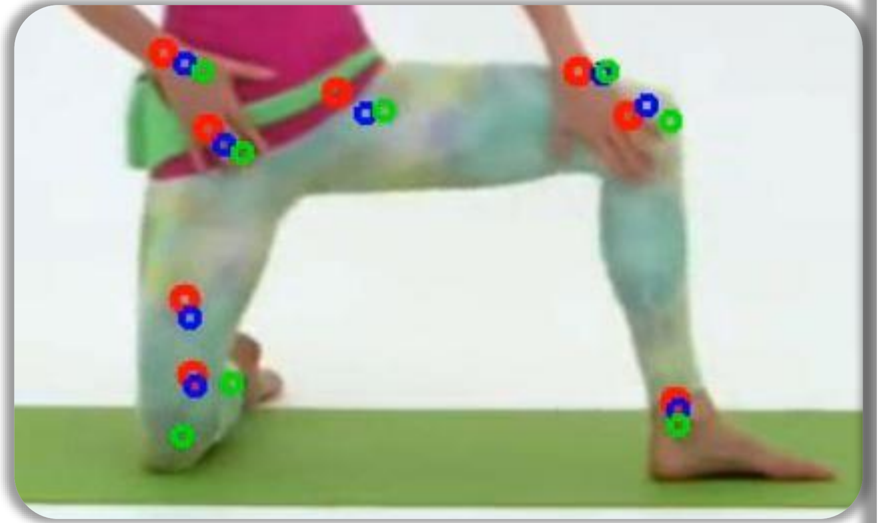


Fig 19. Results on video from Rehab videos

Results – Refined PoseNet with filtering methods

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL INTELLIGENCE



Left Ankle Y coordinate

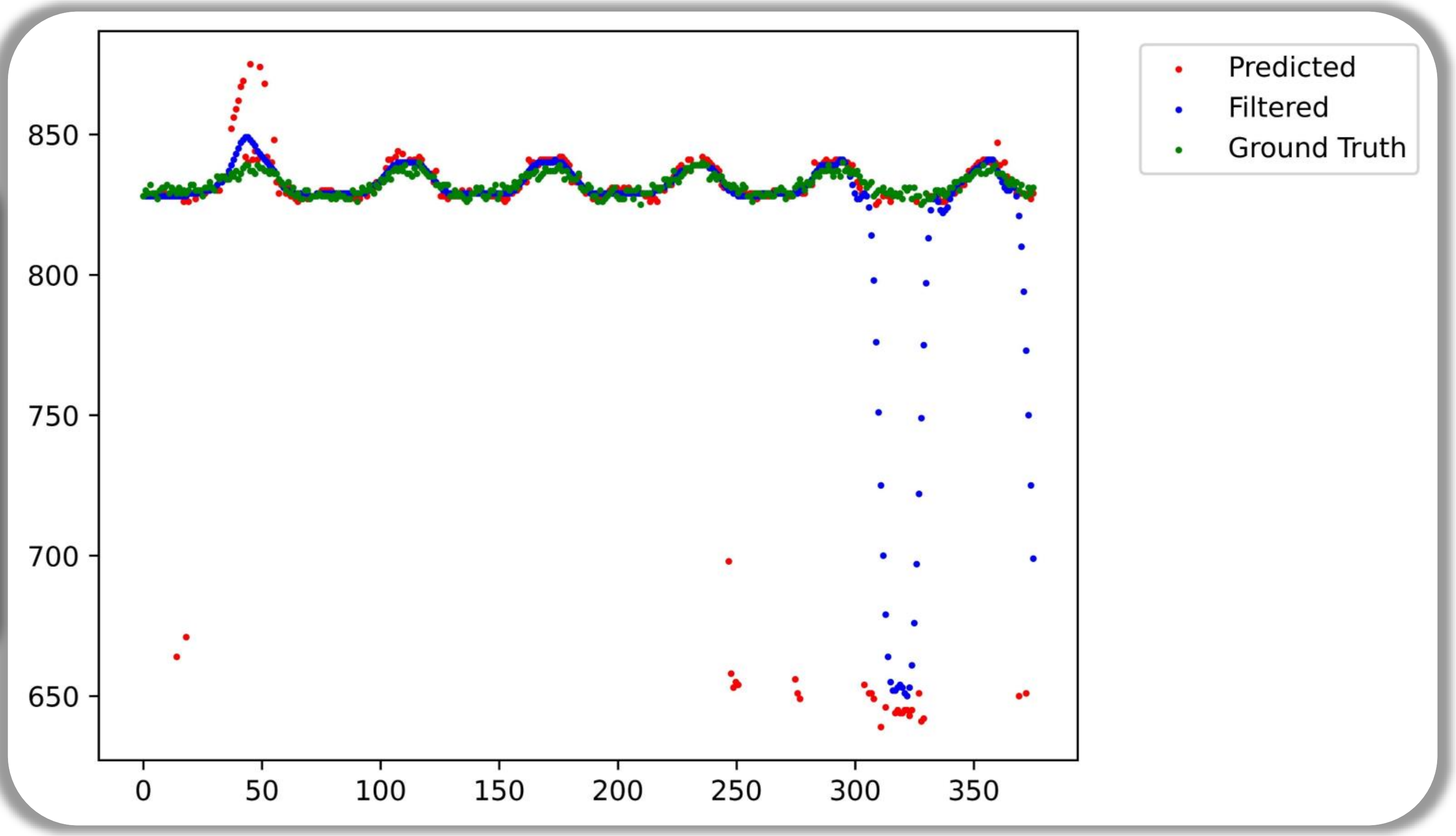


Fig 20. Results on video from Rehab videos

04

Results – Foot-PoseNet

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL INTELLIGENCE

Fig 21. Foot-PoseNet 23 keypoints estimation

Fig 22. Foot-PoseNet outputs

Results – Foot-PoseNet

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL INTELLIGENCE

Fig. 23. Foot-PoseNet 23 keypoints estimation vs PoseNet 17 keypoints estimation

Results – Foot-PoseNet

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL INTELLIGENCE

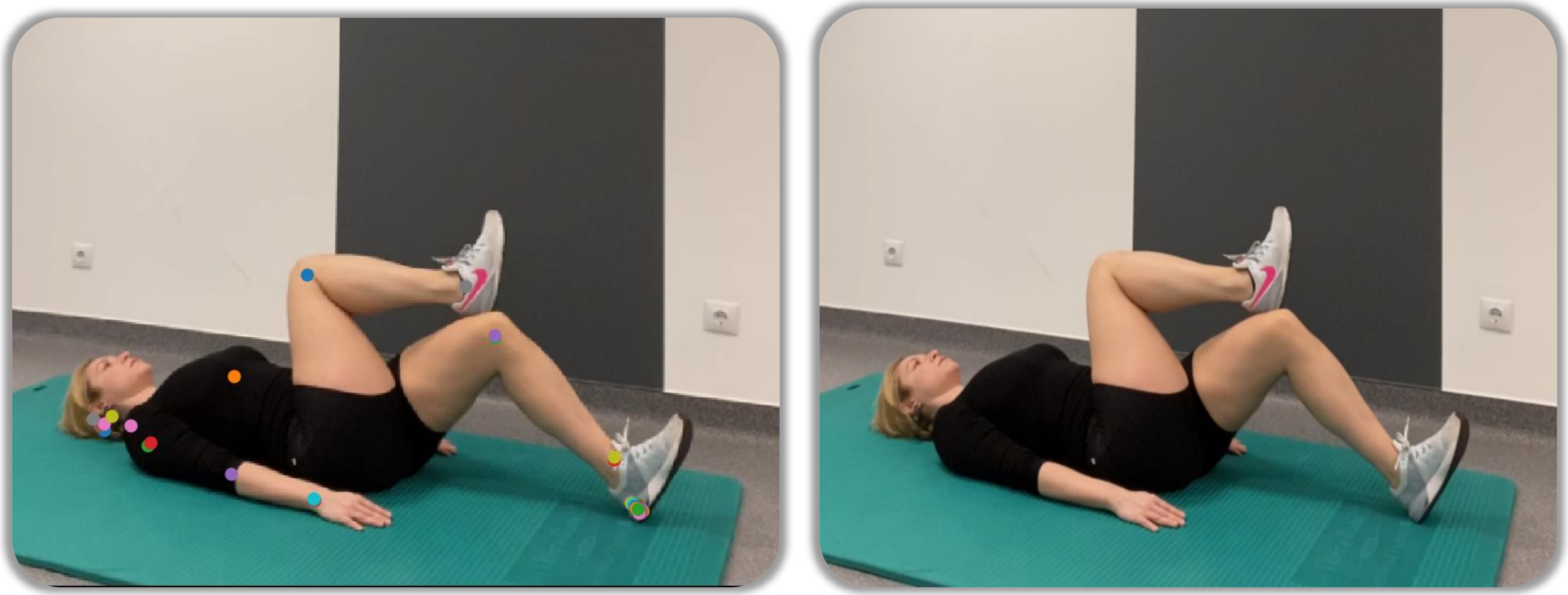


Fig. 24. F-PoseNet 23 keypoints estimation vs PoseNet 17 keypoints estimation

Results – Foot-PoseNet

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL INTELLIGENCE

Table 1. F-PoseNet results on COCO validation set compared to SOTA models with similar image resolution.

Model	Resolution	Body AP	Foot AP	Edge TPU Compatible
HRNet w32 [8]	**256 x 192 x 3**	**0.70**	**0.57**	No
HRNet w48 [8]	256 x 192 x 3	**0.70**	**0.67**	No
DarkPose w32	256 x 192 x 3	0.69	0.57	No
PoseNet (MobileNet)	1280x720	0.59	N/A	Yes
F-PoseNet (ours)	224 x 224 x 3	0.65	0.61	Yes

[8] HrNet: Jingdong Wang et al. Deep High-Resolution Representation Learning for Visual Recognition. 2020. arXiv: 1908.07919 [cs.CV].
[13] Feng Zhang et al. "Distribution-Aware Coordinate Representation for Human Pose Estimation". In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020.

Results – Foot-PoseNet

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL INTELLIGENCE

PRIVATE HOSPITAL
EMINEO

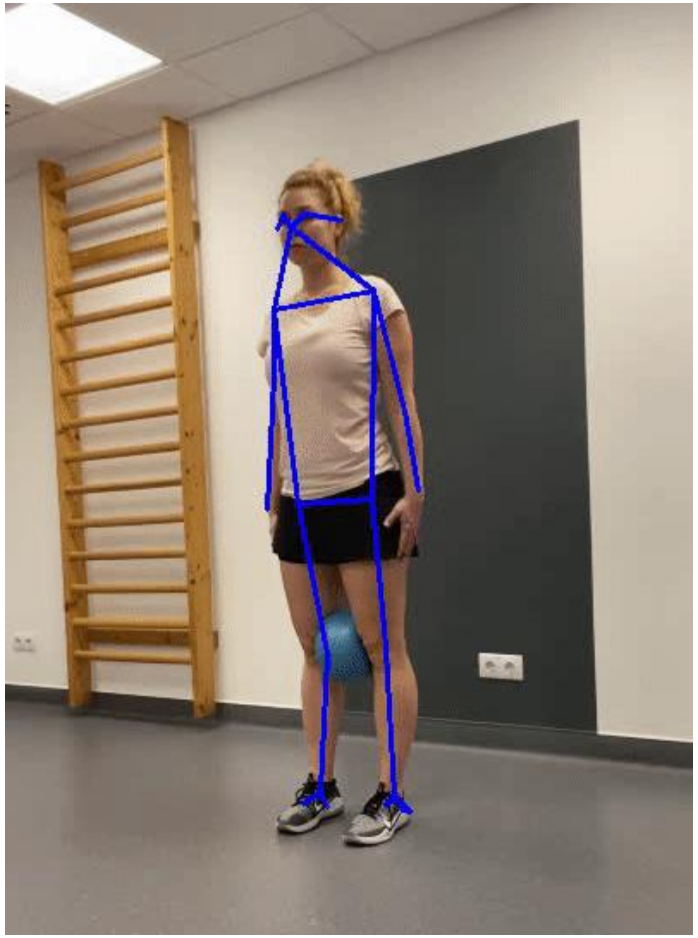


Fig. 26. F-PoseNet 23 keypoints estimation with Emineo clinic data.

Fig. 25 – Edge TPU (left) ; Coral USB accelerator (center) ; Coral Dev Board (right)

Image source: https://coral.ai/products/accelerator/

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL INTELLIGENCE

05. Conclusion

Summary; Future work.

Conclusion – Summary

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL INTELLIGENCE

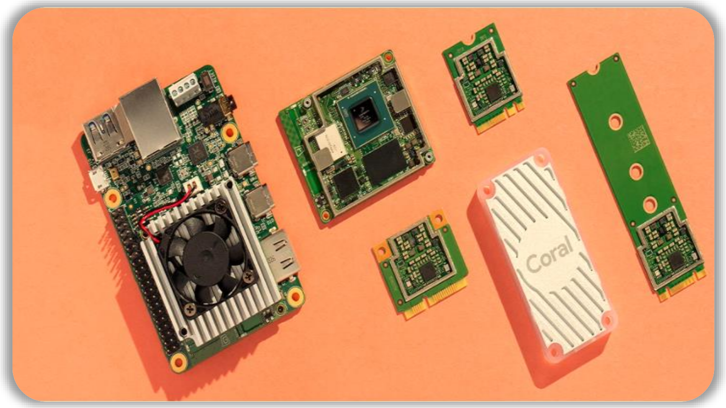
Our work targets a **real life application**, the **rehabilitation of knee injuries,** therefore high estimation precision is needed. **We present two main contributions:**

01

Using a combination of **Median, Savitzky-Golay and Chebyshev** low pass filters we were able to **reduce 10.7% of RMSE** around the **knee**

02

Developed **F-PoseNet, a 2D Deep Learning model optimized to run in real time on the edge** that achieved **0.65 Body AP and 0.61 Foot AP**



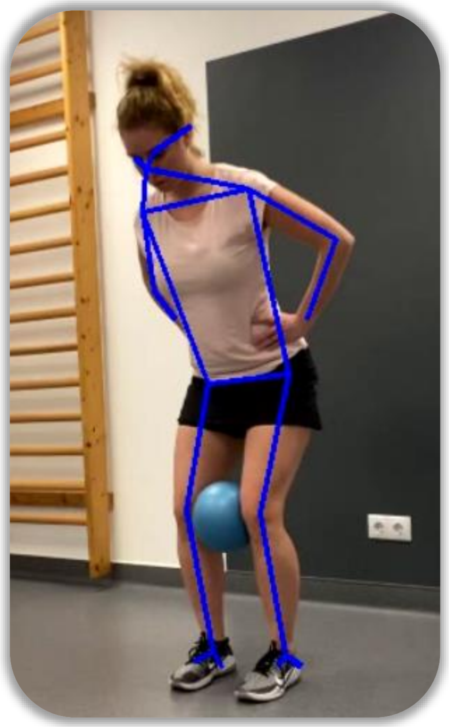
Conclusion – Future work

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL INTELLIGENCE

In the future:

- Use cropping and proper scaling
- Train F-PoseNet on a dataset more representative of our use case, in collaboration with Emineo Clinic.
- Extend F-PoseNet for 3D Pose Estimation

With this work, we hope to contribute for the advance in research of Pose estimation on the edge.



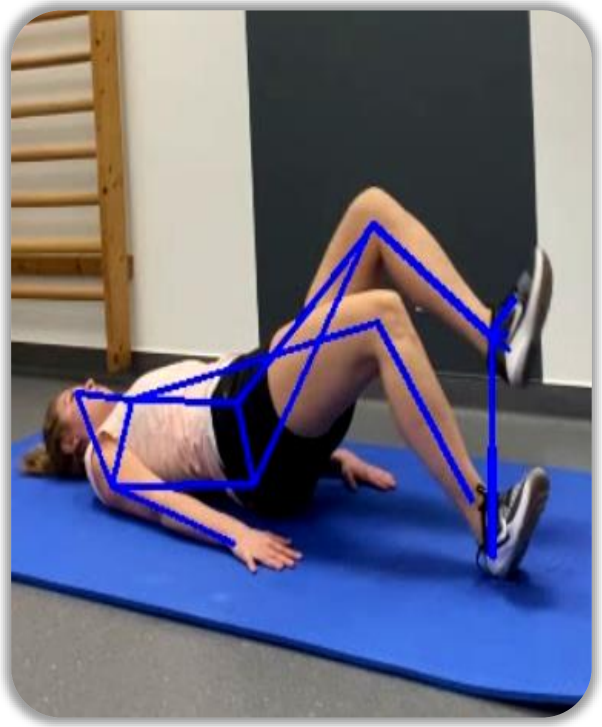


Fig. 26 – Foot-PoseNet results.

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL INTELLIGENCE

06. References

References; Acknowledgements

References - Acknowledgements

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL INTELLIGENCE

- This project has been supported by EIT Digital Grant No. 20670: "COVID-19 – DLUB - Don’t Leave Us Behind“
- Micro-project to the HumanE-AI EU ICT-48, European network of Human-Centered Artificial Intelligence

Partners:
1. Charles University
2. Emineo Clinic
3. Traumatology of the Markusovszky Hospital, Szombathely

- A special thank you for our supervisor Dr. András Lőrincz and our colleague Sohil Zidan.

TUDOMÁNYOS DIÁKKÖRI KONFERENCIA

OTDK | 35. ORSZÁGOS TUDOMÁNYOS DIÁKKÖRI KONFERENCIA 2021

ICANN21

eit Digital

HUMANE AI NET

PRIVATE HOSPITAL EMINEO

CHARLES UNIVERSITY

NIPG

NOSOCOMIUM SAVARIENSE MARKUSOVSZKY 1929

References

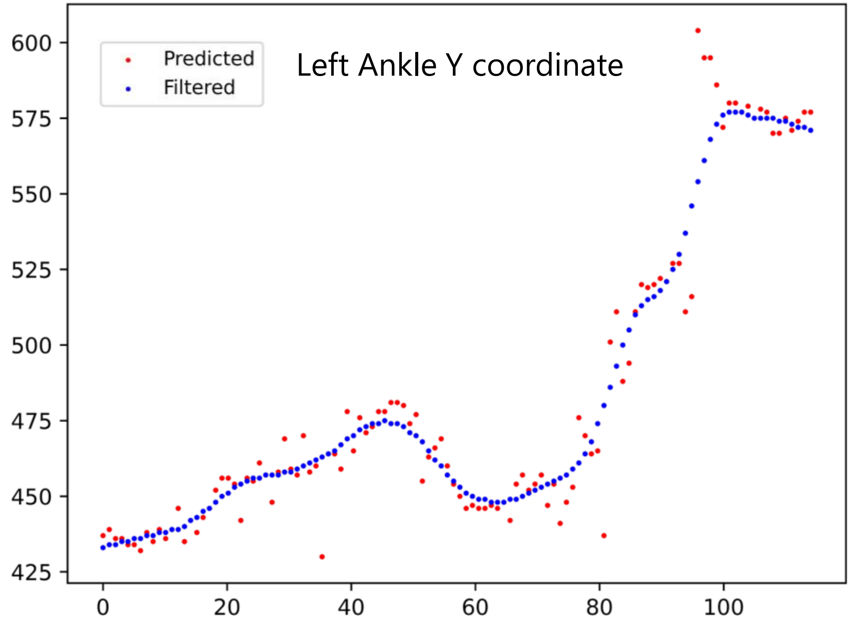
EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL
INTELLIGENCE

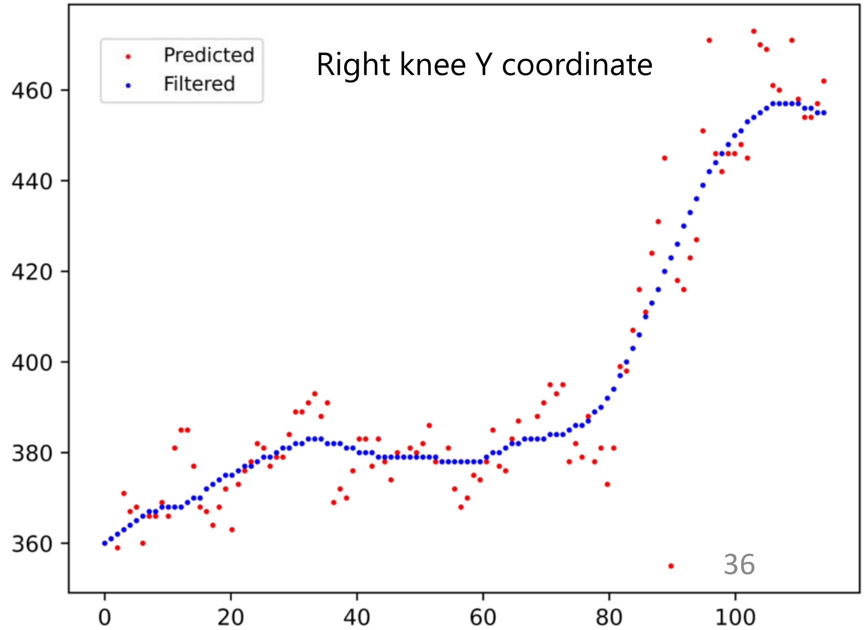
1. Nicolini, Alexandre Pedro et al. “Common injuries in athletes' knee: experience of a specialized center.” Acta ortopedica brasileira vol. 22,3 (2014): 127-31. doi:10.1590/1413-78522014220300475
2. Kurtz , Ong K, Lau E, Mowat F, Halpren M. “Projections of primary and revision hip and knee arthroplasty in the United States for 2005-2030.” J Bone Joint Surg Am. 2007;89:780-785
3. Tsung-Yi Lin et al. “Microsoft COCO: Common Objects in Context”. In: Computer Vision – ECCV 2014. Ed. by David Fleet et al. Cham: Springer International Publishing, 2014, pp. 740–755. ISBN: 978-3-319-10602-1.
4. https://github.com/google-coral/project-posenet
5. Antonis Pardos, Andreas Menychtas, and Ilias Maglogiannis. “Introducing an Edge-Native Deep Learning Platform for Exergames”. In: Artificial Intelligence Applications and Innovations. Ed. by Ilias Maglogiannis, Lazaros Iliadis, and Elias Pimenidis. Cham: Springer International Publishing, 2020, pp. 88–98. ISBN: 978-3-030-49186-4.
6. Theodore Kim, Mohamed Zohdy, and Michael Barker. “Applying Pose Estimation to Predict Amateur Golf Swing Performance using Edge Processing (February 2020)”. In: IEEE Access PP (Aug. 2020), pp. 1–1. DOI: 10.1109/ACCESS.2020. 3014186.
7. Alejandro Newell, Kaiyu Yang, and Jia Deng. Stacked Hourglass Networks for Human Pose Estimation. 2016. arXiv: 1603.06937 [cs.CV].
8. Jingdong Wang et al. Deep High-Resolution Representation Learning for Visual Recognition. 2020. arXiv: 1908.07919 [cs.CV].
9. Dan Oved. Real-time Human Pose Estimation in the Browser with TensorFlow.js. 2020 (accessed November 27, 2020). URL: https://medium.com/tensorflow/real-time-human-pose-estimation-in-the-browser-with-tensorflowjs-7dd0bc881cd5.
10. Andrew G. Howard et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications”. 2017. arXiv: 1704.04861 [cs.CV]
11. George Papandreou et al. PersonLab: Person Pose Estimation and Instance Segmentation with a Bottom-Up, Part-Based, Geometric Embedding Model. 2018. arXiv: 1803.08225 [cs.CV]
12. D. Mehta et al. “Monocular 3D Human Pose Estimation in the Wild Using Improved CNN Supervision”. In: 2017 International Conference on 3D Vision (3DV). 2017, pp. 506–516. DOI: 10.1109/3DV.2017.00064.
13. Feng Zhang et al. “Distribution-Aware Coordinate Representation for Human Pose Estimation”. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020.

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF INFORMATICS
DEPARTMENT OF ARTIFICIAL
INTELLIGENCE

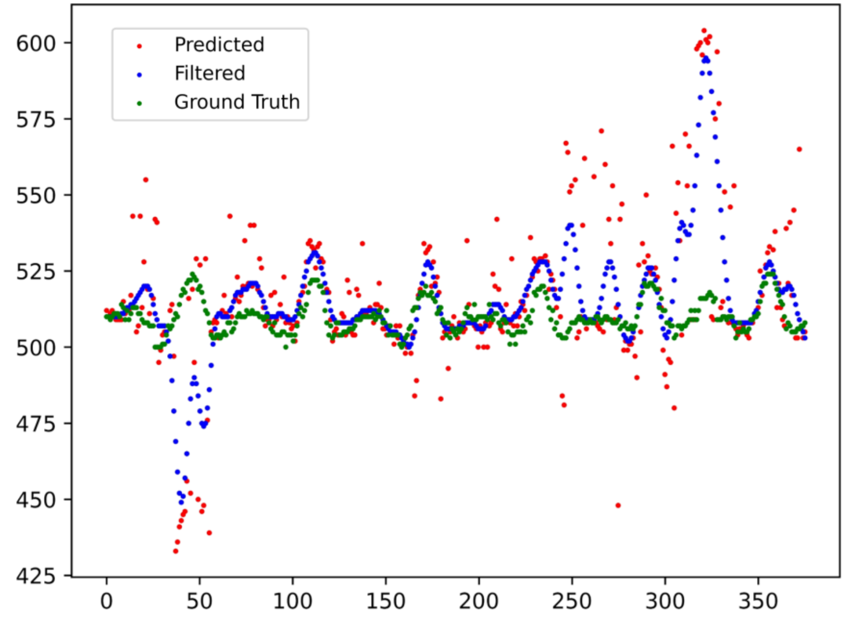
THANK YOU

4. **Results** - Refined PoseNet outputs with filtering methods

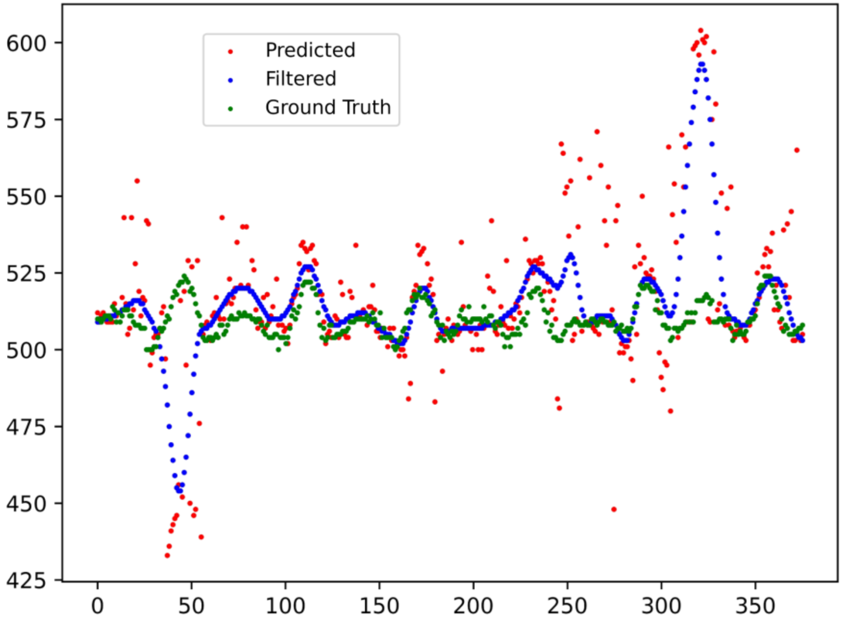




Window size = 5



Window size = 15



Window size = 49