visuAAL

Privacy-Aware and Acceptable Video-Based Technologies and Services for Active and Assisted Living

Lifelogging and Behaviour Modelling

ESR 10 – 3rd PhD seminar

Vienna, Austria
1 December, 2023

Wiktor Mucha

Computer Vision Lab

TU Wien

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 861091".

Universitat d'Alacant
Universidad de Alicante
Project Coordinator

RWTH AACHEN UNIVERSITY

Stockholm University

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

TU WIEN

Presentation Agenda

1. Introduction
2. Contribution of this PhD
3. Progress to date
4. Dissemination overview
5. Future steps

Introduction

Aim and objectives of the PhD

Introduction

- Lifelogging → a **technology** that uses **wearable sensors** to gather and process **data** from the **daily lives** of an individual
- Using a wearable camera can **illustrate** in detail which **activities** the person wearing the camera has done during the **day**
- This **work** focuses on the possibilities of **egocentric** visual data processing for **health** and **lifestyle** improvement
- **Egocentric** → placing a camera on a **human body** giving a **view** from this **person's perspective**

User wearing a lifelogging device[1]

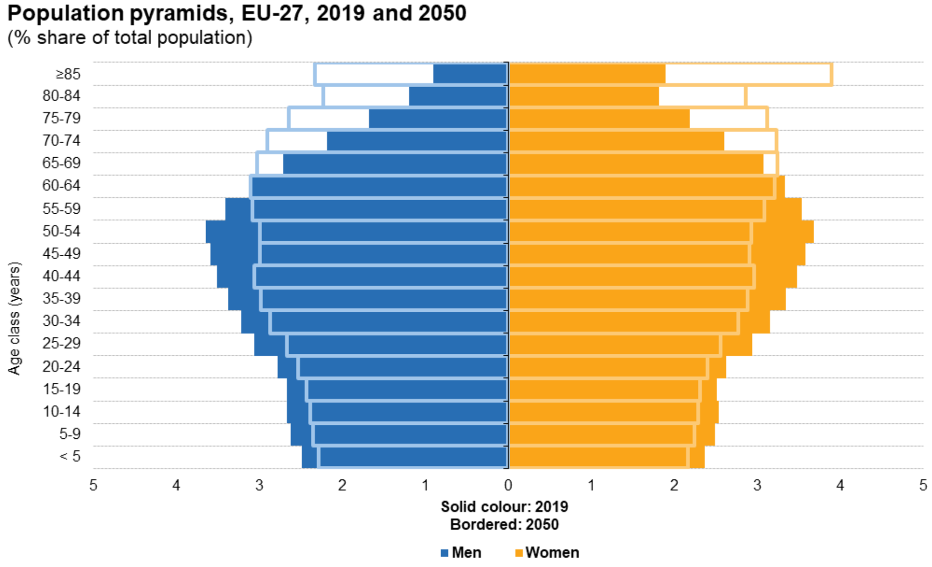
[1] https://newatlas.com/narrative-clip-2/35422/ visited on 20.01.2022

Motivation Behind the Thesis

- The **ageing population** demands new technological solutions to **reduce** the involved **medical personnel**
- The **technological progress** of electronic devices provides **new systems** for lifelogging and egocentric
- The **early stage** of **research** – no products on the market
- New devices on the market

RayBan Stories[3]

DJI Action 2 [4]



Population pyramid with future estimations [2]

[2] https://ec.europa.eu/eurostat/statistics-explained/index.php?title=File:Population_pyramids,_EU-27,_2019_and_2050_(%25_share_of_total_population)_AE2020.png visited on 24.09.2022

[3] https://www.ray-ban.com/canada/en/electronics/ray-ban%20stories%20%7C%20round-shiny%20black/8056597705035 visited on 22.09.2022

[4] https://www.gsmarena.com/the_dji_action_2_is_a_tiny_action_camera_made_big_by_its_multitude_of_accessories_and_mods-news-51608.php visited 14.04.2023

Application Examples

I. **Action recognition**

→ Task of discovering action in the image/clip

II. **Activity recognition**

→ **Activity** of Daily Living (ADL) **differs** from **action** detection in **length**

III. **Food scene monitoring**

→ Food scene understanding, food detection, environment analysis

III. **Social interaction monitoring**

→ Automated analysis for the social interaction pattern descriptions

→ Lack of social relations leads to a decrease in psychological well-being

Examples of actions in EPIC-KITCHEN dataset[5]

[5] https://epic-kitchens.github.io/2021 visited on: 24.09.2022

Contribution of this PhD

Research questions and planned progress

Research Questions

RQ1: What **actions** and behaviours that **affect health** and well-being can be **tracked** for health improvement using **egocentric** video-based lifelogging **systems**. Is it possible to process egocentric images to assist with health-related tasks such as **rehabilitation**, **taking medication.** Can we **recognise** struggles to provide assistance?

Research Questions

RQ2: How do **recent advances** in **egocentric hand pose** estimation compare with state-of-the-art techniques, including 3D pose-based methods for **action recognition**, in terms of the usability of 2D hand and object poses for egocentric action recognition tasks? Furthermore, how does performance vary when different types of pose input are used?

Research Questions

RQ3: What **strategies** and **methods** can be used to **improve performance** and **minimise** the differences **between datasets**, particularly in bridging the gap between laboratory conditions, with the aim of improving the generalisability and reliability of experimental results when using **hand pose based egocentric action recognition**?

Contribution to the State of the Art

I. **Improvement** of short term **video understanding** through **robust action recognition** for **fine-grained** scenarios

II. **Introduction** of **novel health-related applications** using **egocentric** imaging

III. **Bridging** the gap between **laboratory** experiments and **health-related** applications

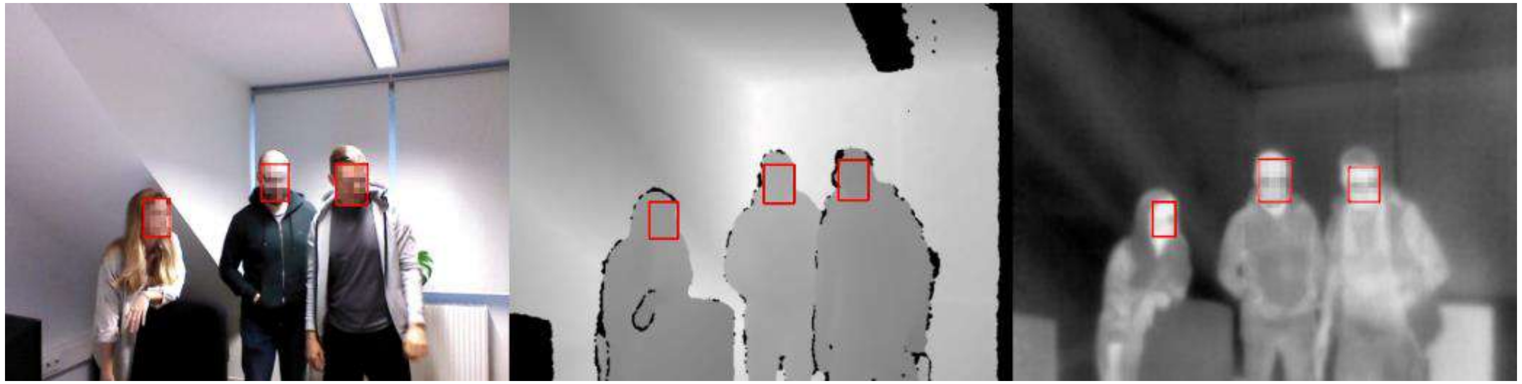
Progress to date

Summary of main findings

Image Modalities and Privacy

I. Research on various image modalities and their potential future applications for lifelogging

- Using additional modalities can lead to improvement in the results
- Comparison study of three image modalities (RGB, depth, thermal) in the example of Face Detection (FD) → FD is often a part of AAL systems, including egocentric lifelogging studies
- The RGB sensor is not always necessary, and it’s superior in the hardest FD scenario



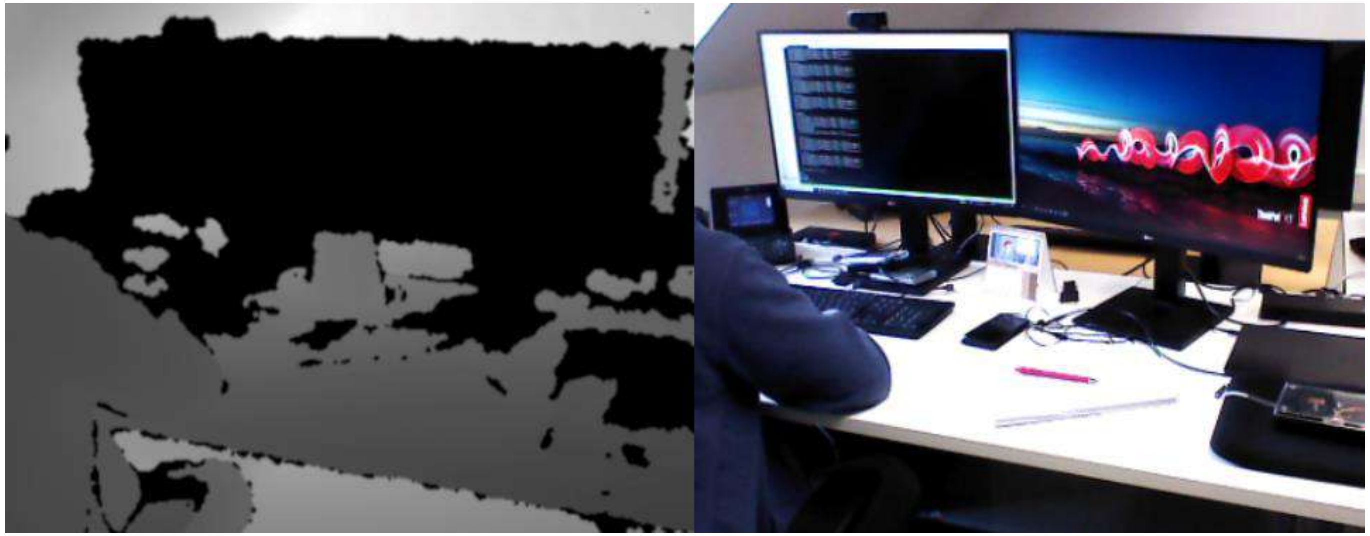
Example of a correct detection on each image modality, form left RGB, depth and thermal image

Mucha, W., & Kampel, M. (2022, February). **Depth and thermal images in face detection-a detailed comparison between image modalities**. In 2022 the 5th International Conference on Machine Vision and Applications (ICMVA) (pp. 16-21).

Image Modalities and Privacy

II. Privacy issues in lifelogging and AAL devices

- AAL vision-based life-logging systems raise privacy concerns due to monitoring indyviduals
- There are **contradictory** statements **about privacy** of depth cameras
- Contributions → **Factors** and scenarios **affecting privacy in depth** images
  → Face Recognition (FR) **performance study** to determine **possibility** of **identification**



Same scene visible in deptn and RGB image.

Image Modalities and Privacy

Conclusions

- **Depth** sensors **preserve more privacy** due to the lack of texture information. **FR** performs accurately **only** in **laboratory environments** with a small group of individuals and high-sensor resolution
- Lifelogging with depth sensors requires custom hardware, the market is evolving (e.g. mobile phones)
- **Enhancing** or **replacing** RGB images **with depth** modality is **beneficial** in **certain scenarios** (e.g., high privacy requirement, dietary monitoring), but at this moment it is **restricted** by **hardware** and data availability

Mucha, W., & Kampel, M. (2022, June). **Beyond Privacy of Depth Sensors in Active and Assisted Living Devices.** In Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments (pp. 425-429).

Mucha, W., & Kampel, M. (2022, July). **Addressing Privacy Concerns in Depth Sensors**. In Computers Helping People with Special Needs: 18th International Conference, ICCHP-AAATE 2022, Lecco, Italy, July 11–15, 2022, Proceedings, Part II (pp. 526-533). Cham: Springer International Publishing.

Egocentric Action Recognition with 2D Hand Pose

- Hand pose **simplifies** task of **action recognition**
- Action recognition finds application activities monitoring, e.g., for health reasons
- Release of **high quality**, comfortable **RGB** devices like RayBan Stories
- State-of-the-art methods for 3D egocentric hand pose result in **error** equal to 37 mm

  → (**20%** considering avg. hand)
- No wearable RGB-D devices on the market

DJI Action 2 [10]

Self-made wearable RGB-D camera and RayBan glasses[8]

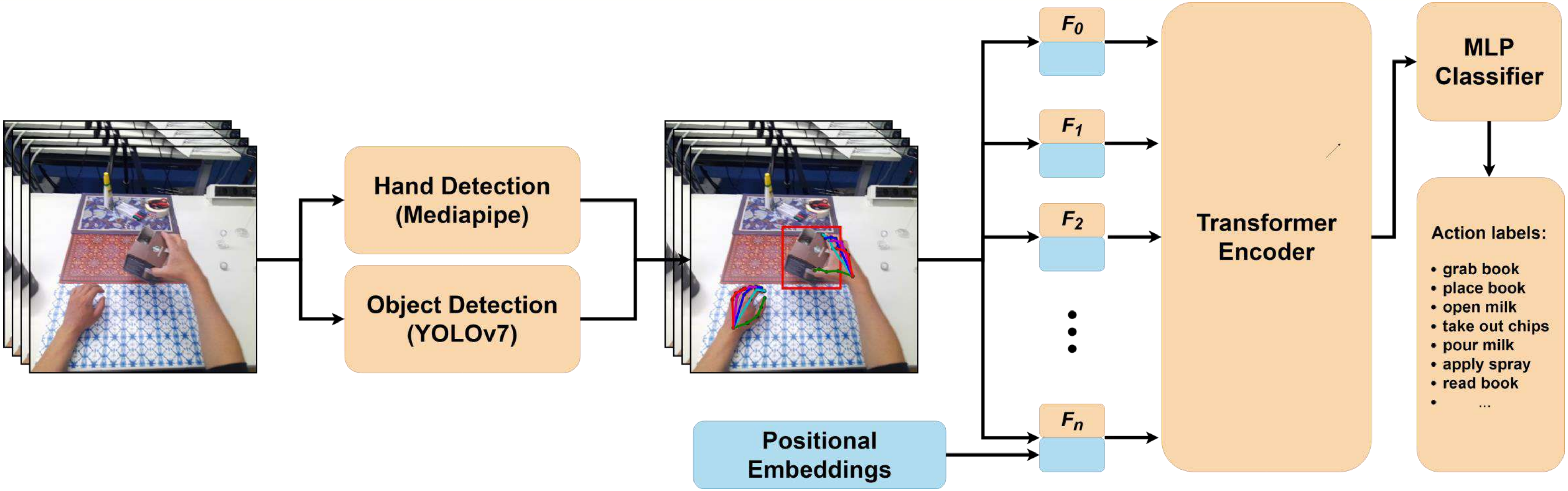
RayBan Stories[9]

[8] Kwon, T et al. (2021). H2o: Two hands manipulating objects for first person interaction recognition. In Proceedings of the IEEE/CVF International Conference on Computer Vision

[9] https://www.ray-ban.com 14.09.2023

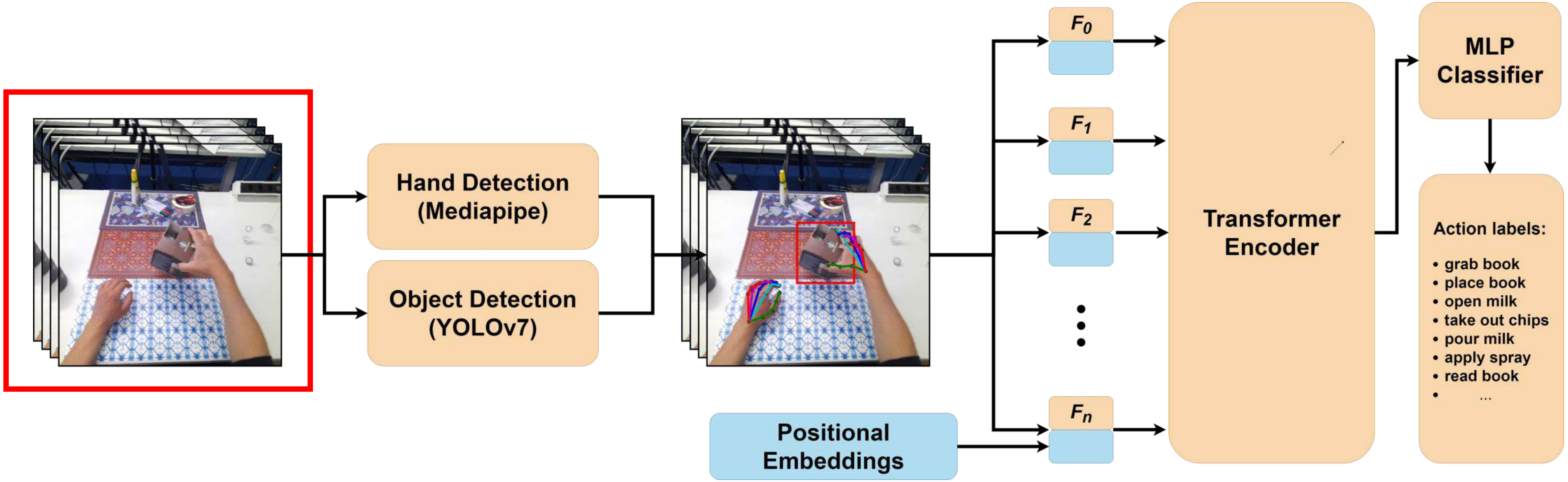
[10] https://www.gsmarena.com 14.09.2023

Hand-based Action Recognition – Overview



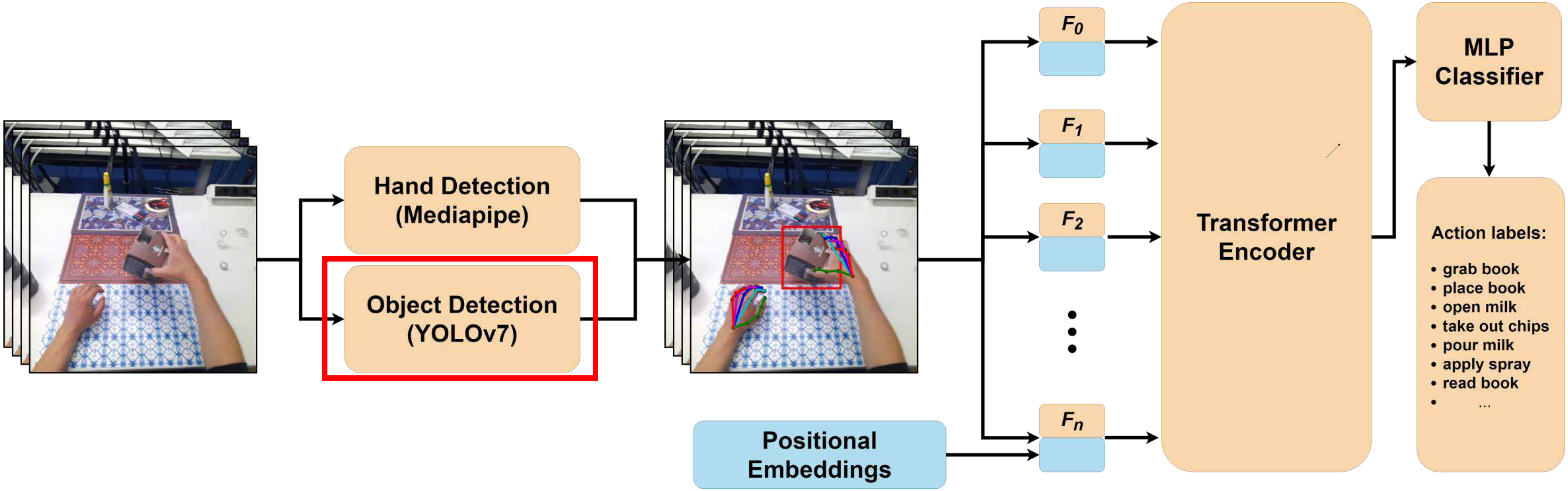
- Usage of **hands** and **objects** as **input** for supervised **sequence** model
- Allows to use of **pre-trained** models reducing the learning costs
- Allows adaptations for various **health-related tasks** which involve hands manipulation

Hand-based Action Recognition – Overview



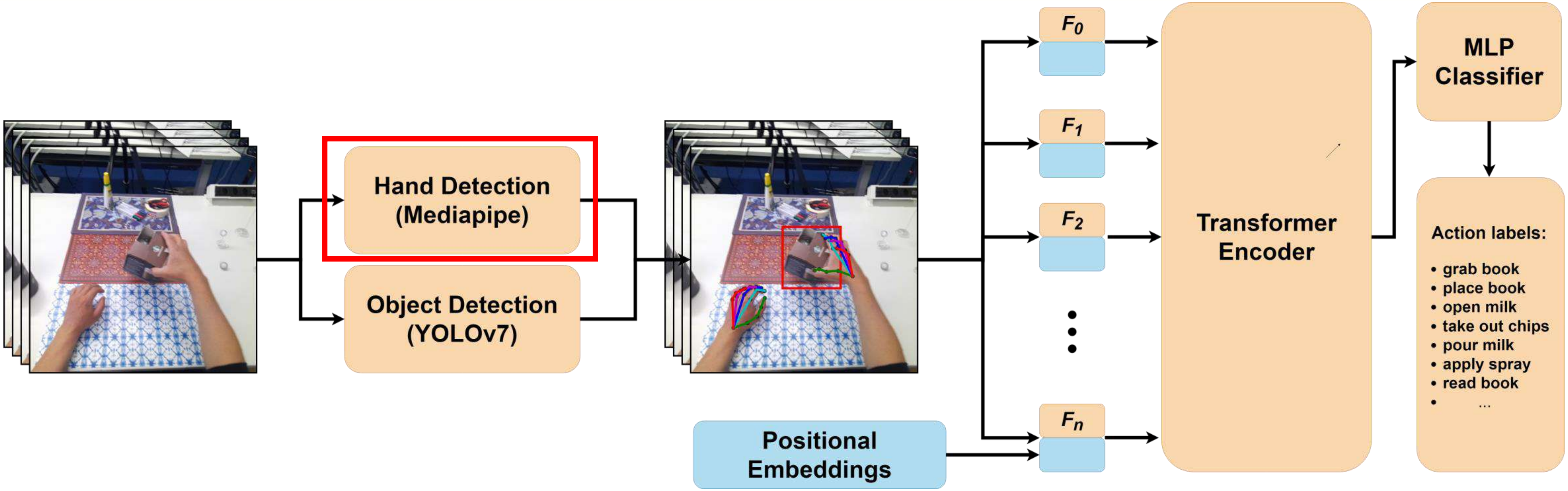
- Input sequence of frames $\boldsymbol{f}_1, \boldsymbol{f}_2 .. \boldsymbol{f}_n$ where $\boldsymbol{n} \in [\mathbf{1}, \mathbf{2} .. \boldsymbol{N}]$
- Actions **shorter** than *N* frames → **zero padding**
- Actions **longer** than *N* frames → **uniform subsampling**

Hand-based Action Recognition – Overview



- Implemented using state-of-the-art **YOLOv7** model
- Trained on ***H2O Dataset*** training subset
- Object described as $\boldsymbol{Po}_{bb}^{i}(\boldsymbol{x},\boldsymbol{y})$ where $\boldsymbol{i} \in [\mathbf{1..4}]$ corresponds to the **bounding box corners**
- $\boldsymbol{P}_{\boldsymbol{o}_{\boldsymbol{l}}}$ describes object label

Hand-based Action Recognition – Overview

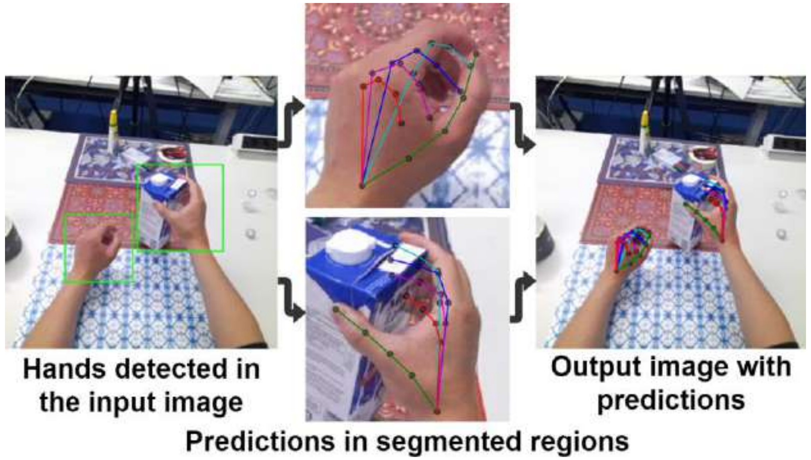


- Task of estimating the position of **21** hand **keypoints** in 2D space using **RGB** image

Egocentric 2D Hand Pose

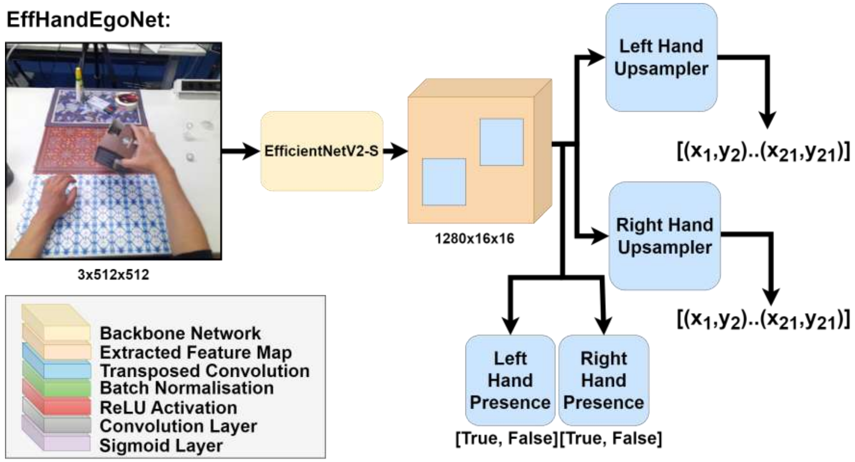
I. Single Hand Approach → *EffHandNet*:

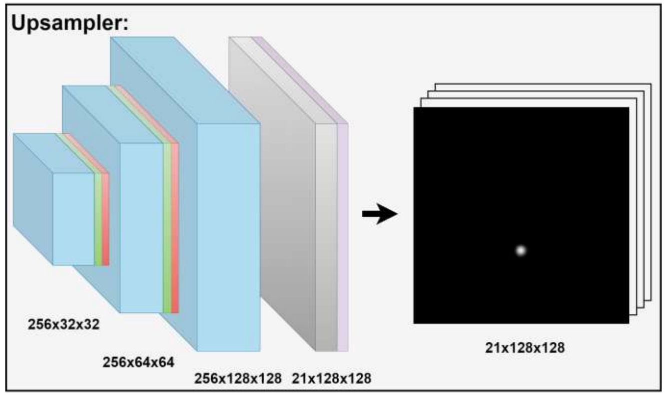
- Pre-trained **hand detector** in the egocentric input image
- Hand pose prediction in segmented regions $\boldsymbol{R_1, R_2}$
- Feature extractor:
  → *EfficientNetV2-S* [11]
- Prediction head:
  → Sequence of transposed convolution resulting in heatmaps


Hands detected in the input image
Output image with predictions
Predictions in segmented regions

II. Egocentric Approach → *EffHandEgoNet*

- Handness prediction module $\boldsymbol{H_L, H_R}$
- Two up-sampling heads
- Improves modelling of hand-object interaction
- Output hand pose: $\boldsymbol{Ph_l^i = (x, y)}$





[11] Mingxing Tan and Quoc Le, "Efficientnetv2: Smaller models and faster training," in International Conference on Machine Learning. PMLR, 2021, pp. 10096–10106.

Egocentric 2D Hand Pose: Results

Single Hand Network:

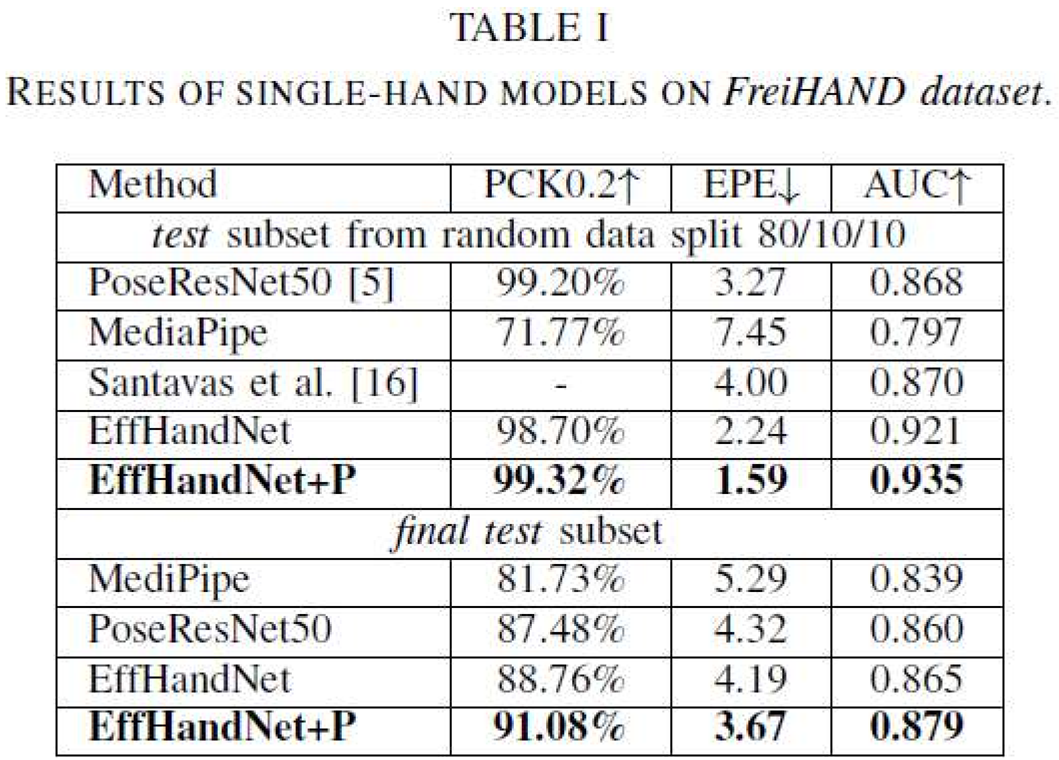


TABLE I

RESULTS OF SINGLE-HAND MODELS ON *FreiHAND dataset*.

Method	PCK0.2↑	EPE↓	AUC↑
test subset from random data split 80/10/10			
PoseResNet50 [5]	99.20%	3.27	0.868
MediaPipe	71.77%	7.45	0.797
Santavas et al. [16]	-	4.00	0.870
EffHandNet	98.70%	2.24	0.921
EffHandNet+P	**99.32%**	**1.59**	**0.935**
final test subset			
MediPipe	81.73%	5.29	0.839
PoseResNet50	87.48%	4.32	0.860
EffHandNet	88.76%	4.19	0.865
EffHandNet+P	**91.08%**	**3.67**	**0.879**

Egocentric Performance:

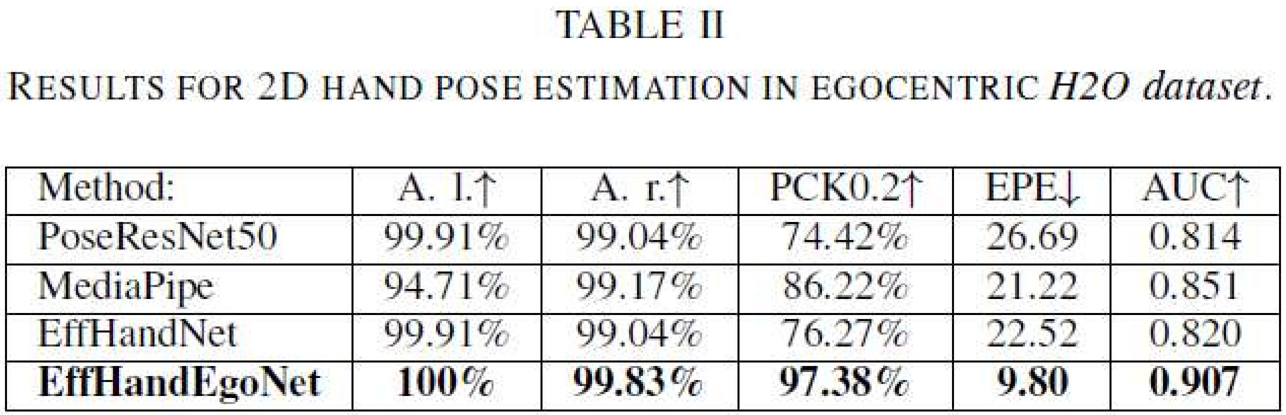


TABLE II

RESULTS FOR 2D HAND POSE ESTIMATION IN EGOCENTRIC *H2O dataset*.

Method:	A. l.↑	A. r.↑	PCK0.2↑	EPE↓	AUC↑
PoseResNet50	99.91%	99.04%	74.42%	26.69	0.814
MediaPipe	94.71%	99.17%	86.22%	21.22	0.851
EffHandNet	99.91%	99.04%	76.27%	22.52	0.820
EffHandEgoNet	**100%**	**99.83%**	**97.38%**	**9.80**	**0.907**

- Estimating **hand pose** in **egocentric** vision requires modelling **complex hand-object interactions**. For this scenario, the **performance** of the approach **based on hand detection** is **not sufficient**.

Egocentric 2D Hand Pose: Results

Single Hand Network:

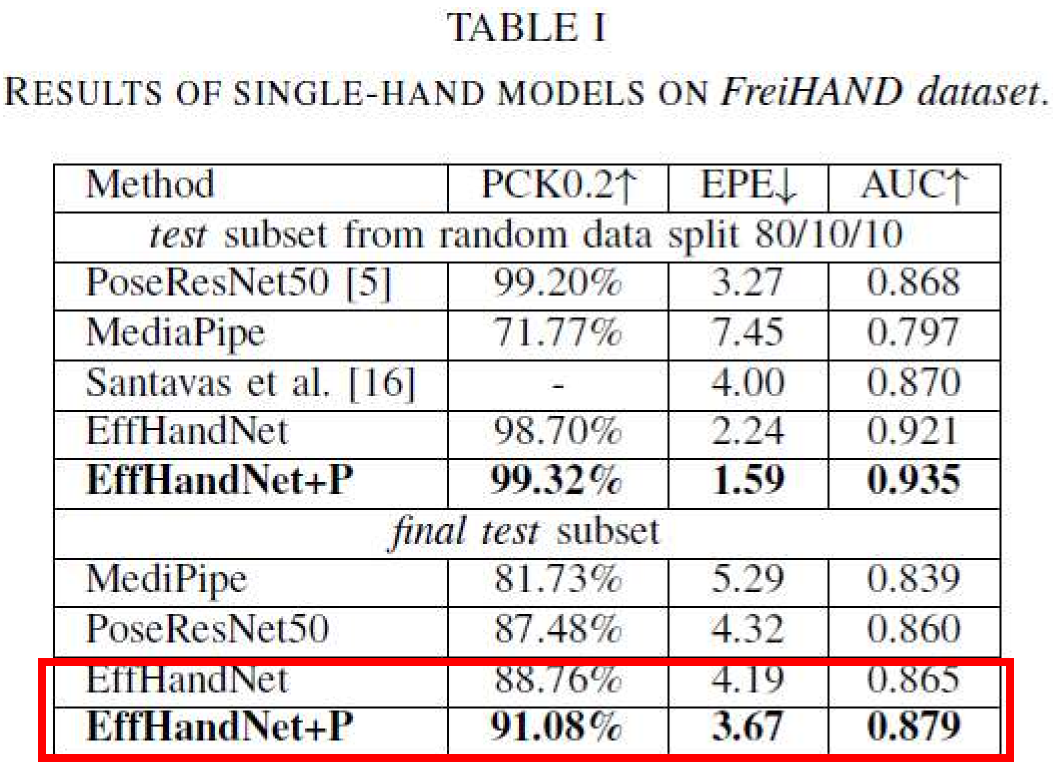


TABLE I

RESULTS OF SINGLE-HAND MODELS ON *FreiHAND dataset*.

Method	PCK0.2↑	EPE↓	AUC↑
test subset from random data split 80/10/10			
PoseResNet50 [5]	99.20%	3.27	0.868
MediaPipe	71.77%	7.45	0.797
Santavas et al. [16]	-	4.00	0.870
EffHandNet	98.70%	2.24	0.921
EffHandNet+P	**99.32%**	**1.59**	**0.935**
final test subset			
MediPipe	81.73%	5.29	0.839
PoseResNet50	87.48%	4.32	0.860
EffHandNet	88.76%	4.19	0.865
EffHandNet+P	**91.08%**	**3.67**	**0.879**

Egocentric Performance:

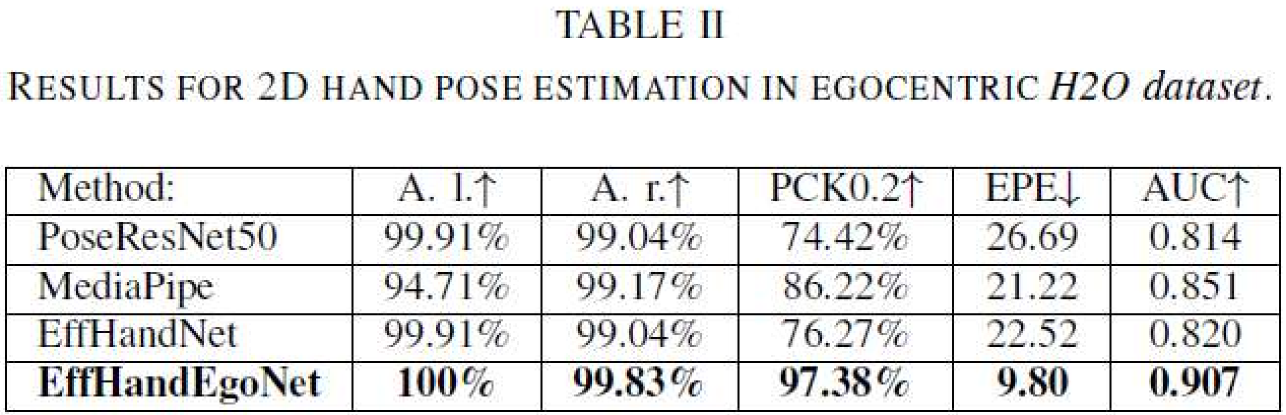


TABLE II

RESULTS FOR 2D HAND POSE ESTIMATION IN EGOCENTRIC *H2O dataset*.

Method:	A. l.↑	A. r.↑	PCK0.2↑	EPE↓	AUC↑
PoseResNet50	99.91%	99.04%	74.42%	26.69	0.814
MediaPipe	94.71%	99.17%	86.22%	21.22	0.851
EffHandNet	99.91%	99.04%	76.27%	22.52	0.820
EffHandEgoNet	**100%**	**99.83%**	**97.38%**	**9.80**	**0.907**

- Estimating **hand pose** in **egocentric** vision requires modelling **complex hand-object interactions**. For this scenario, the **performance** of the approach **based on hand detection** is **not sufficient**.

Egocentric 2D Hand Pose: Results

Single Hand Network:

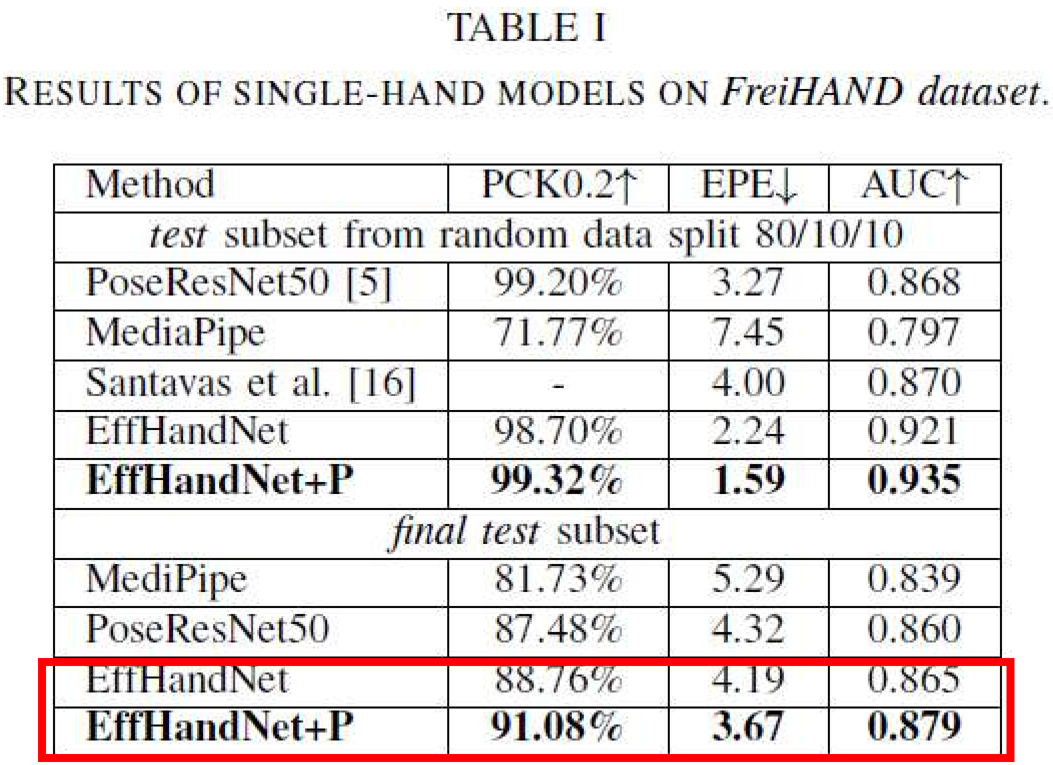


TABLE I

RESULTS OF SINGLE-HAND MODELS ON *FreiHAND dataset*.

Method	PCK0.2↑	EPE↓	AUC↑
test subset from random data split 80/10/10			
PoseResNet50 [5]	99.20%	3.27	0.868
MediaPipe	71.77%	7.45	0.797
Santavas et al. [16]	-	4.00	0.870
EffHandNet	98.70%	2.24	0.921
EffHandNet+P	**99.32%**	**1.59**	**0.935**
final test subset			
MediPipe	81.73%	5.29	0.839
PoseResNet50	87.48%	4.32	0.860
EffHandNet	88.76%	4.19	0.865
EffHandNet+P	**91.08%**	**3.67**	**0.879**

Egocentric Performance:

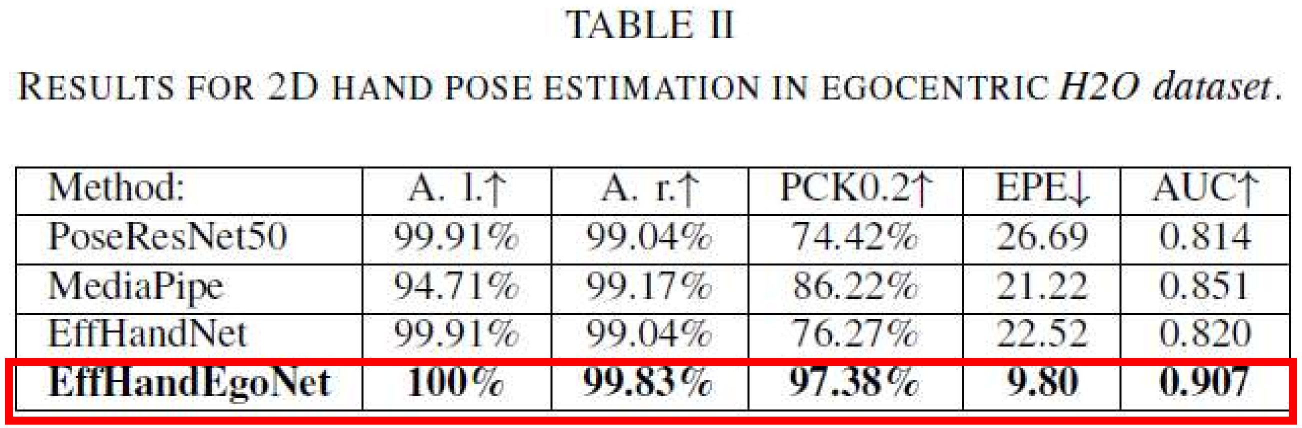


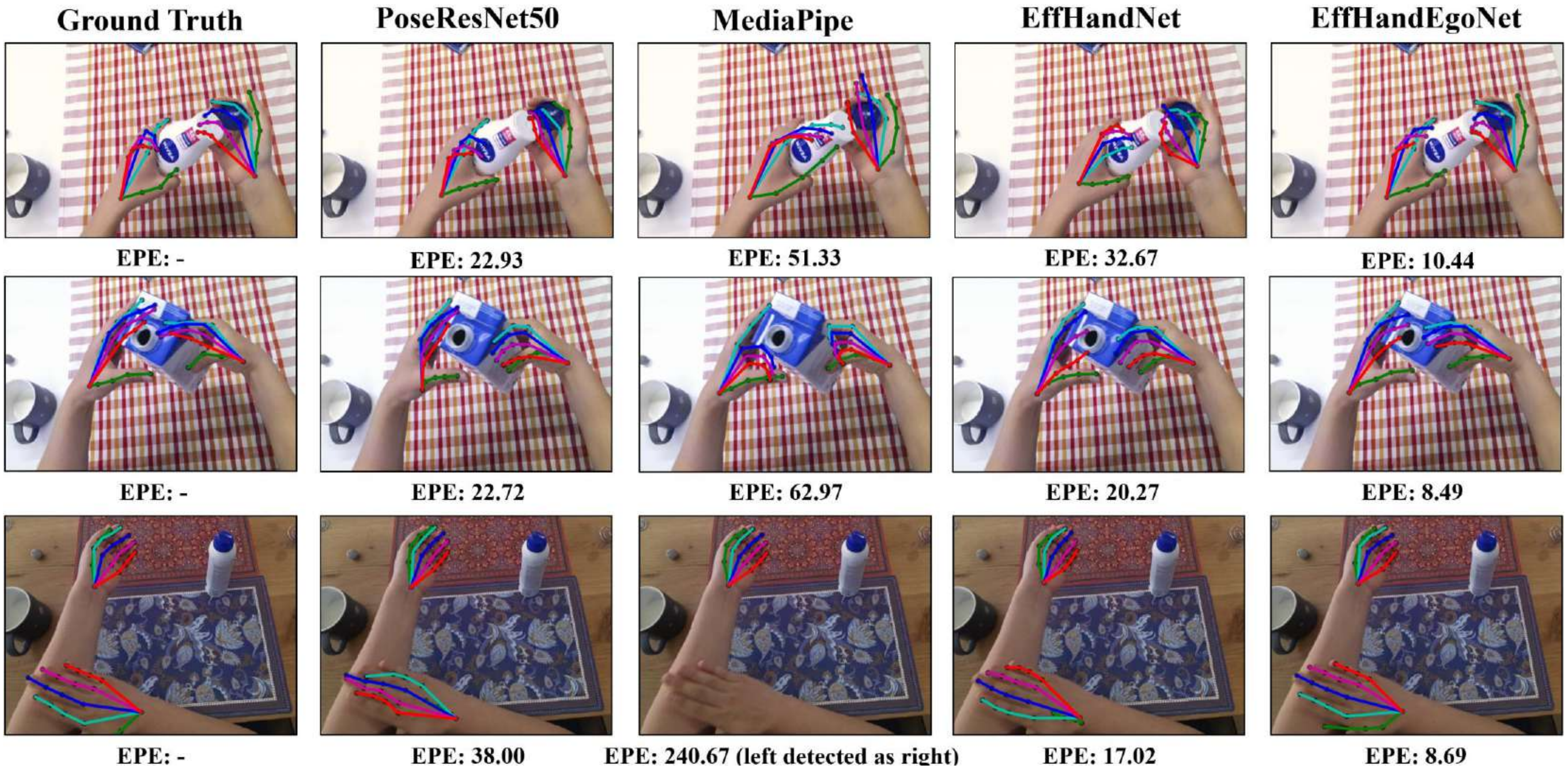
TABLE II

RESULTS FOR 2D HAND POSE ESTIMATION IN EGOCENTRIC *H2O dataset*.

Method:	A. l.↑	A. r.↑	PCK0.2↑	EPE↓	AUC↑
PoseResNet50	99.91%	99.04%	74.42%	26.69	0.814
MediaPipe	94.71%	99.17%	86.22%	21.22	0.851
EffHandNet	99.91%	99.04%	76.27%	22.52	0.820
EffHandEgoNet	**100%**	**99.83%**	**97.38%**	**9.80**	**0.907**

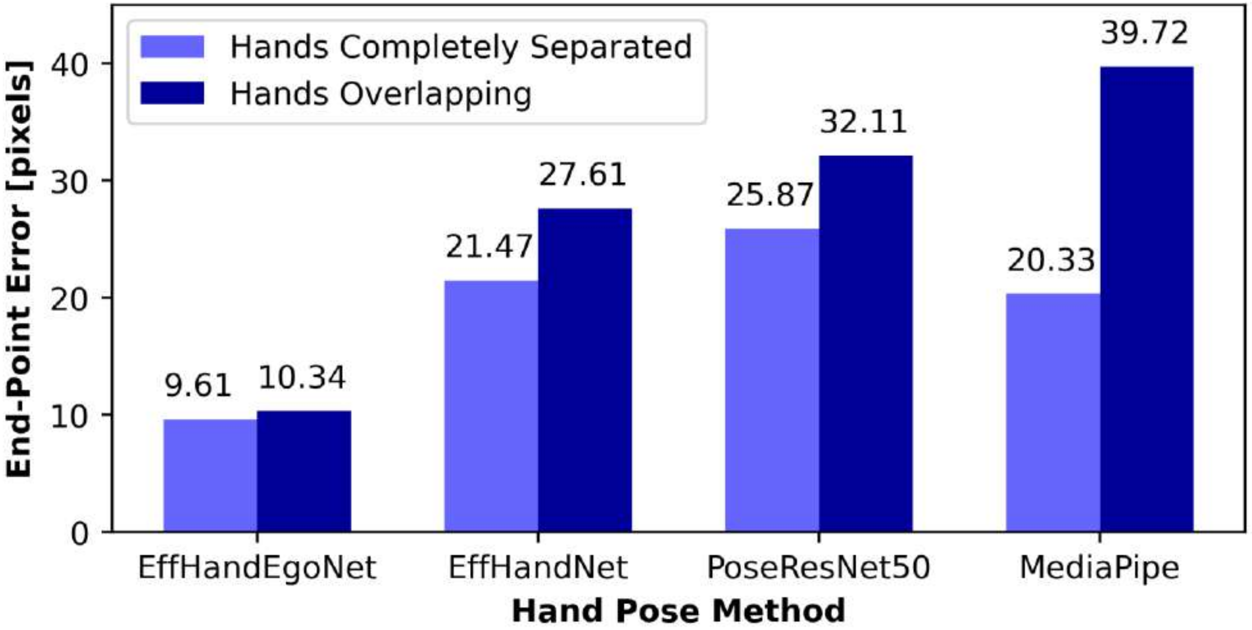
- Estimating **hand pose** in **egocentric** vision requires modelling **complex hand-object interactions**. For this scenario, the **performance** of the approach **based on hand detection** is **not sufficient**.

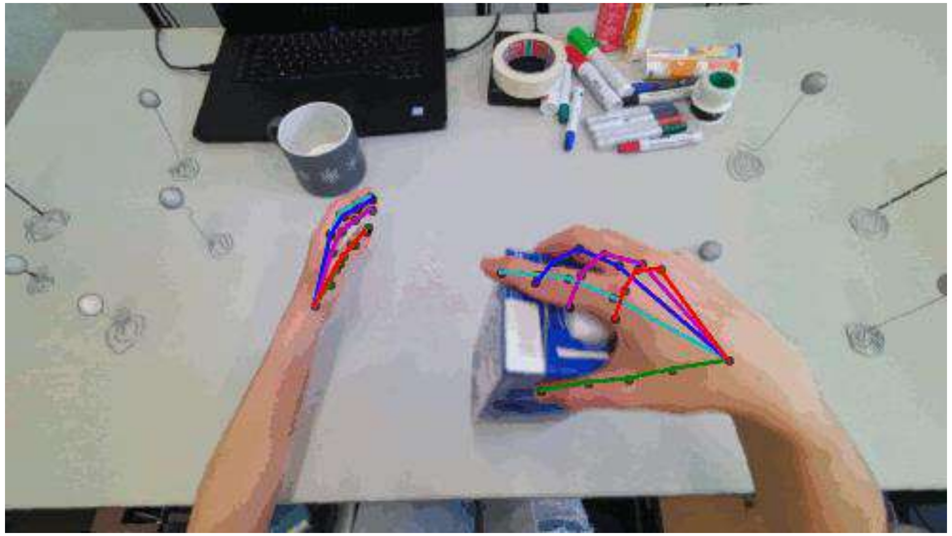
Egocentric 2D Hand Pose: Qualitative Results

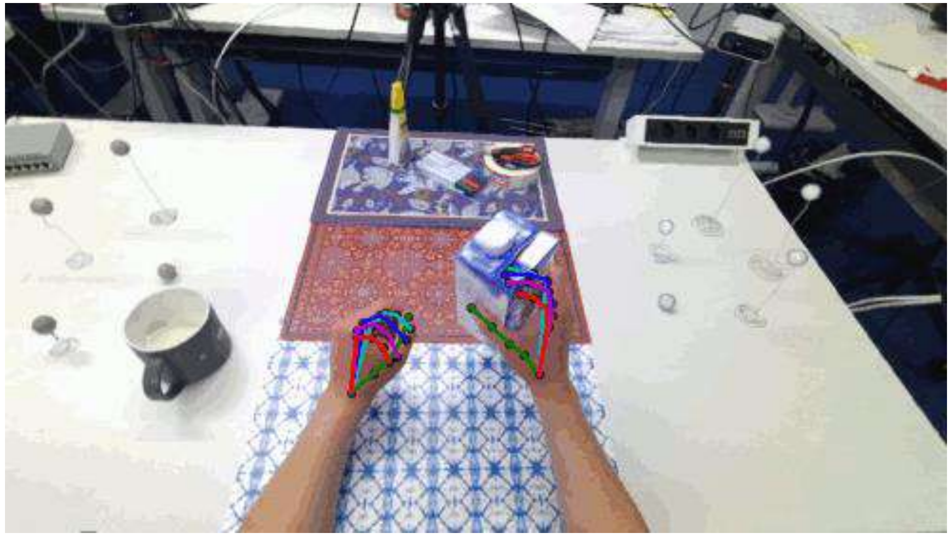


Egocentric 2D Hand Pose: Results

Performance in overlapping scenario:

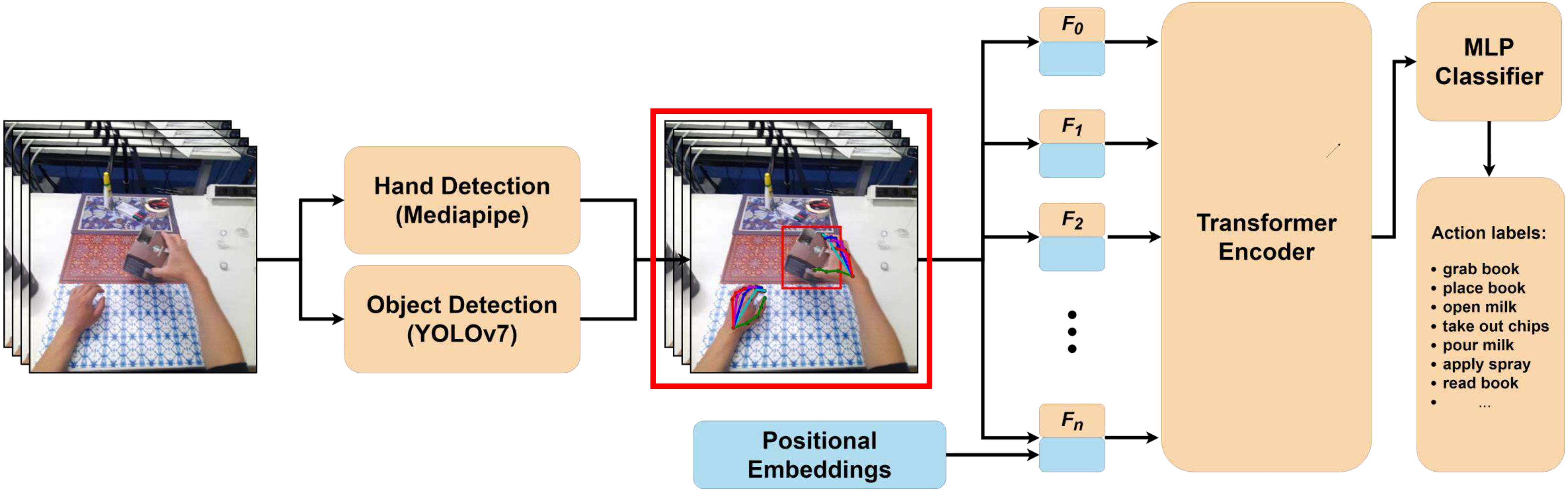






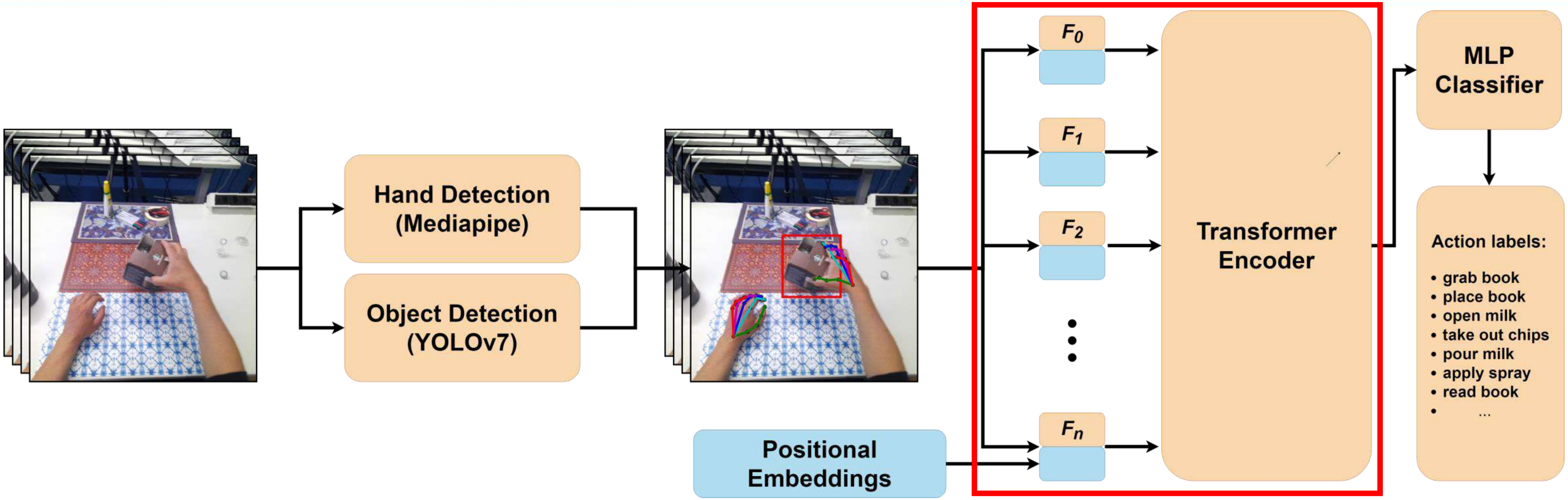
Submitted: Mucha W., Kampel M. (2023) “**In My Perspective, In My Hands: Accurate Egocentric 2D Hand Pose**”, The 18th IEEE International Conference on Automatic Face and Gesture Recognition - FG 2024

Hand-based Action Recognition - Overview

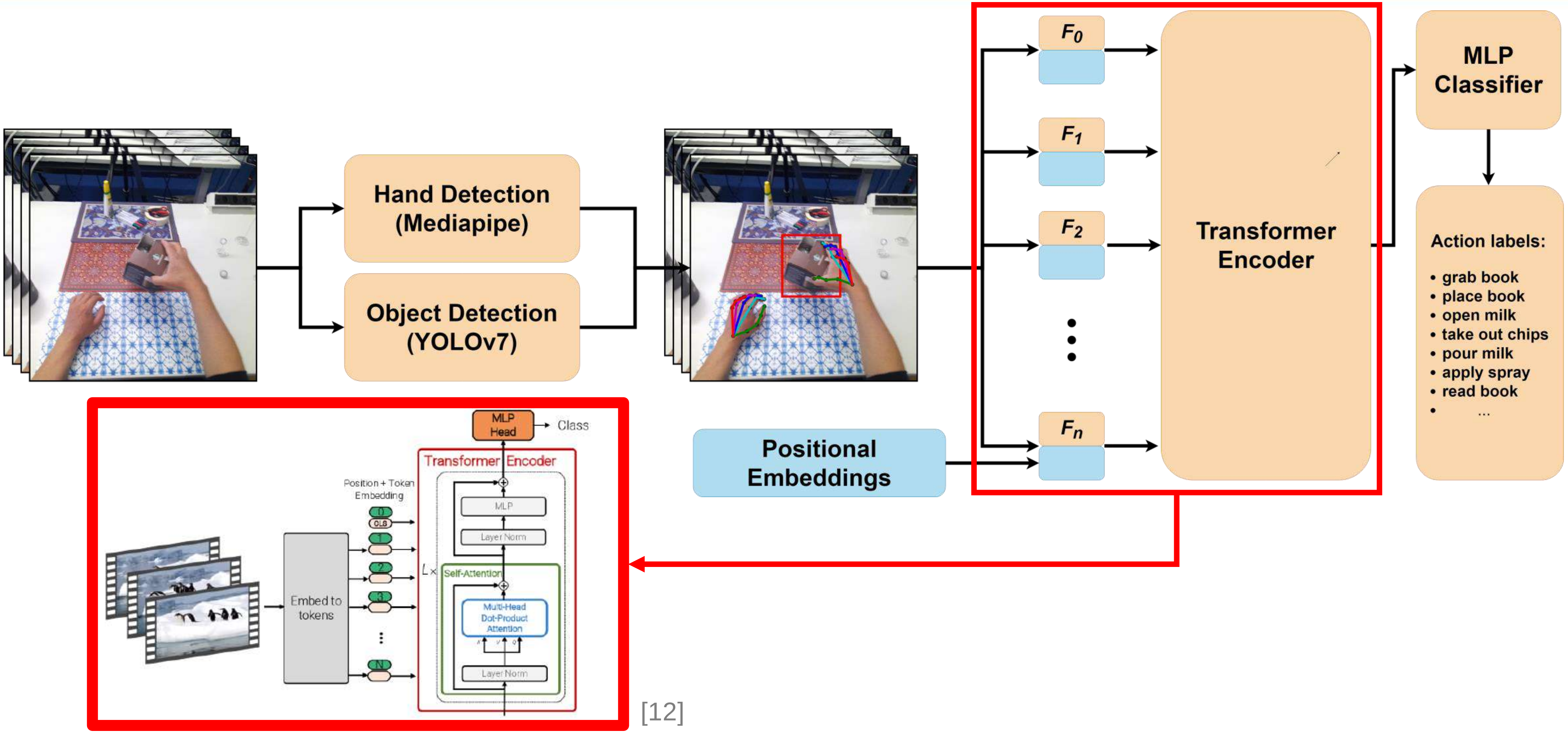


- Each frames describes: $\boldsymbol{f_n = Ph_l^i\,(x,y)\ Ph_r^i(x,y)Po_{bb}^i(x,y)P_{o_l}}$
- Sequence of frames: $\boldsymbol{V_{seq} = [f_1, f_2 .. f_n]}$

Hand-based Action Recognition - Overview



Hand-based Action Recognition - Overview



Hand Detection (Mediapipe)

Object Detection (YOLOv7)

F_0

F_1

F_2

F_n

Positional Embeddings

Transformer Encoder

MLP Classifier

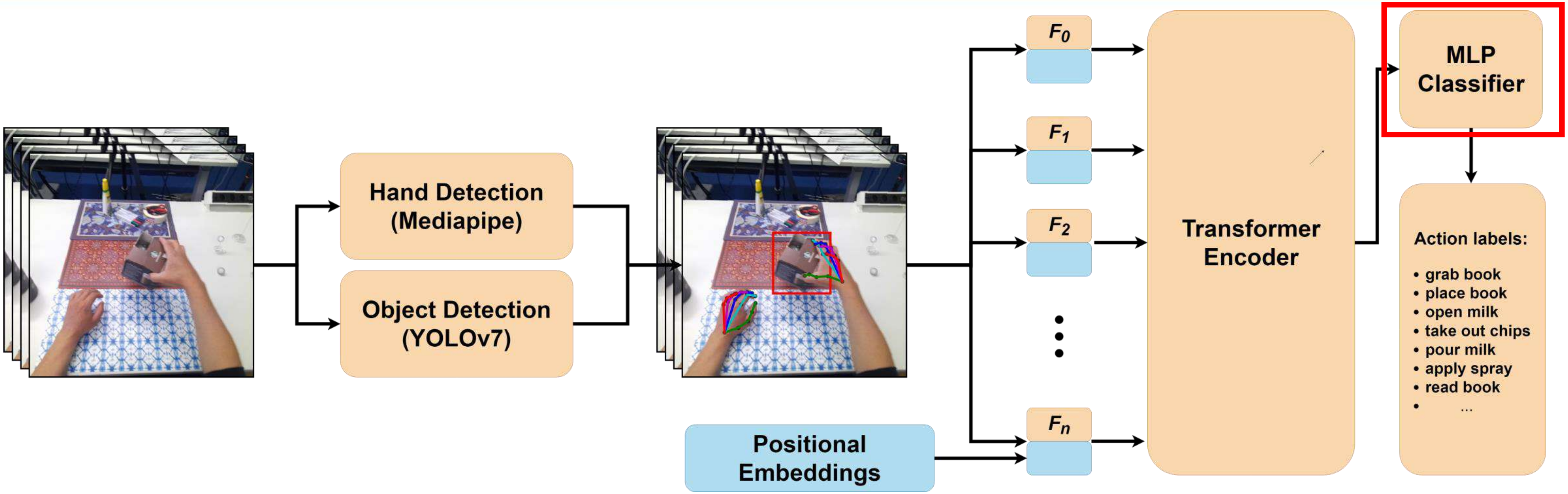
Action labels:

- grab book
- place book
- open milk
- take out chips
- pour milk
- apply spray
- read book
- ...

[12]

[12] Arnab et al., ViViT: A Video Vision Transformer. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*

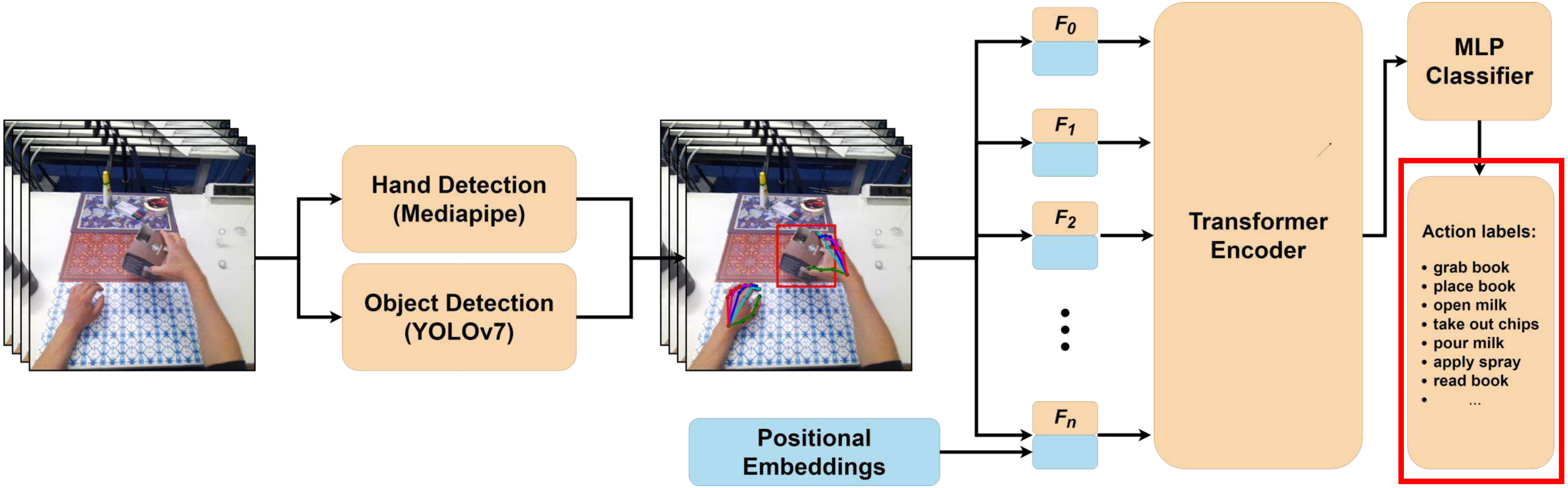
Hand-based Action Recognition - Overview



Action labels:

- grab book
- place book
- open milk
- take out chips
- pour milk
- apply spray
- read book
- ...

Hand-based Action Recognition - Overview



Evaluation

Action Recognition in *H2O Dataset*:

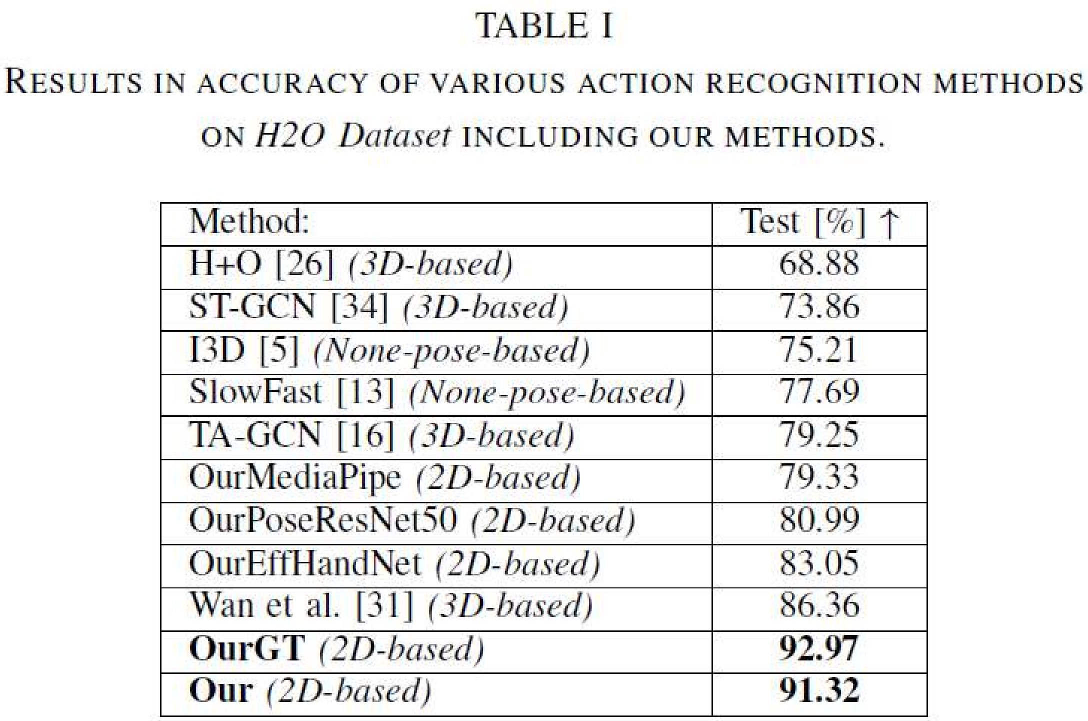


TABLE I

Results in accuracy of various action recognition methods on *H2O Dataset* including our methods.

Method:	Test [%] ↑
H+O [26] *(3D-based)*	68.88
ST-GCN [34] *(3D-based)*	73.86
I3D [5] *(None-pose-based)*	75.21
SlowFast [13] *(None-pose-based)*	77.69
TA-GCN [16] *(3D-based)*	79.25
OurMediaPipe *(2D-based)*	79.33
OurPoseResNet50 *(2D-based)*	80.99
OurEffHandNet *(2D-based)*	83.05
Wan et al. [31] *(3D-based)*	86.36
OurGT *(2D-based)*	**92.97**
Our *(2D-based)*	**91.32**

Action Recognition in *FPHA Dataset*:

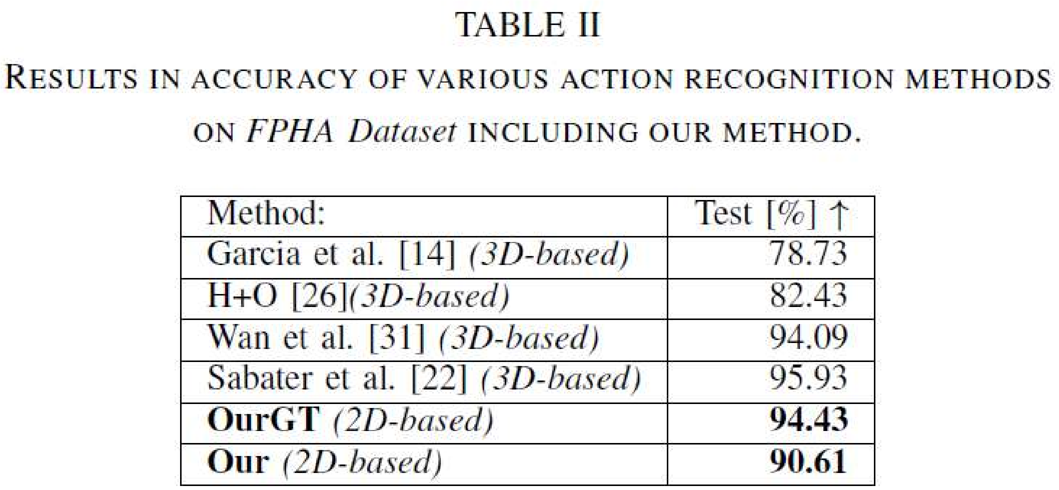


TABLE II

Results in accuracy of various action recognition methods on *FPHA Dataset* including our method.

Method:	Test [%] ↑
Garcia et al. [14] *(3D-based)*	78.73
H+O [26]*(3D-based)*	82.43
Wan et al. [31] *(3D-based)*	94.09
Sabater et al. [22] *(3D-based)*	95.93
OurGT *(2D-based)*	**94.43**
Our *(2D-based)*	**90.61**

- **2D Hand Post** is **robust** for egocentric action recognition
- **Accurate pose** estimation is **essential** for action recognition

Evaluation

Action Recognition in *H2O Dataset*:

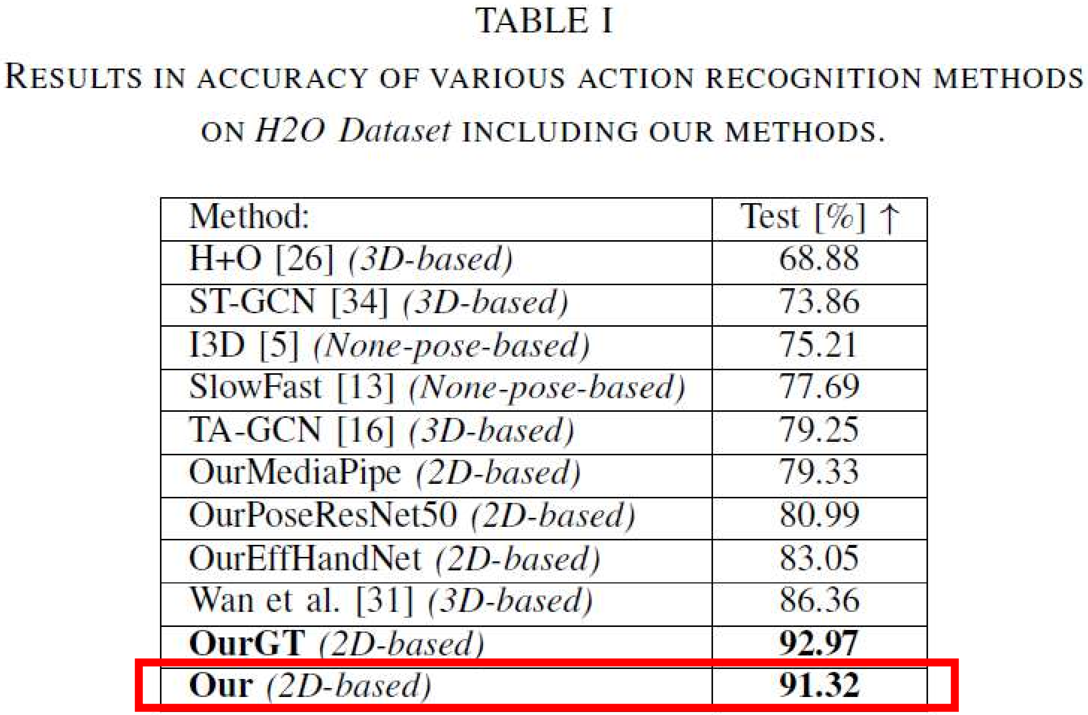


TABLE I

RESULTS IN ACCURACY OF VARIOUS ACTION RECOGNITION METHODS ON *H2O Dataset* INCLUDING OUR METHODS.

Method:	Test [%] ↑
H+O [26] *(3D-based)*	68.88
ST-GCN [34] *(3D-based)*	73.86
I3D [5] *(None-pose-based)*	75.21
SlowFast [13] *(None-pose-based)*	77.69
TA-GCN [16] *(3D-based)*	79.25
OurMediaPipe *(2D-based)*	79.33
OurPoseResNet50 *(2D-based)*	80.99
OurEffHandNet *(2D-based)*	83.05
Wan et al. [31] *(3D-based)*	86.36
OurGT *(2D-based)*	**92.97**
Our *(2D-based)*	**91.32**

Action Recognition in *FPHA Dataset*:

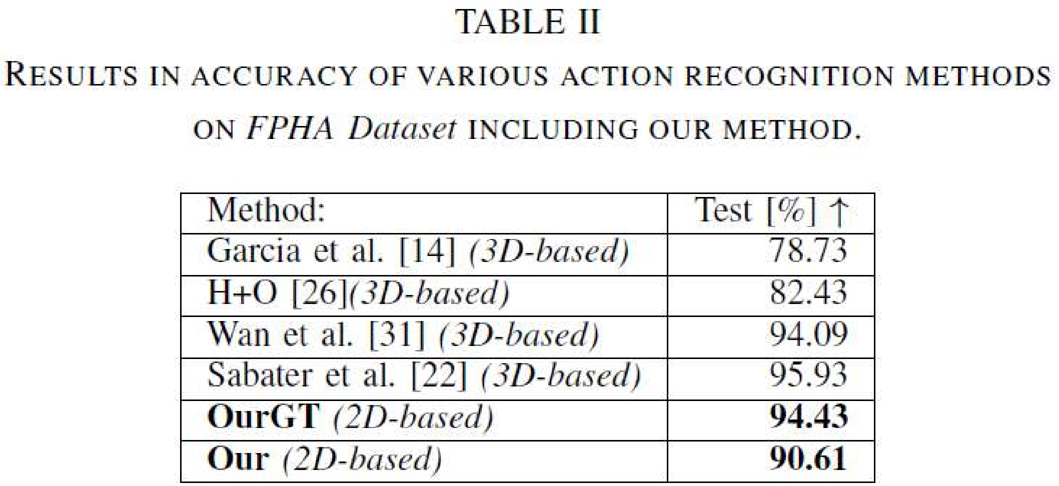


TABLE II

RESULTS IN ACCURACY OF VARIOUS ACTION RECOGNITION METHODS ON *FPHA Dataset* INCLUDING OUR METHOD.

Method:	Test [%] ↑
Garcia et al. [14] *(3D-based)*	78.73
H+O [26]*(3D-based)*	82.43
Wan et al. [31] *(3D-based)*	94.09
Sabater et al. [22] *(3D-based)*	95.93
OurGT *(2D-based)*	**94.43**
Our *(2D-based)*	**90.61**

- **2D Hand Post** is **robust** for egocentric action recognition
- **Accurate pose** estimation is **essential** for action recognition

Evaluation

Action Recognition in *H2O Dataset*:

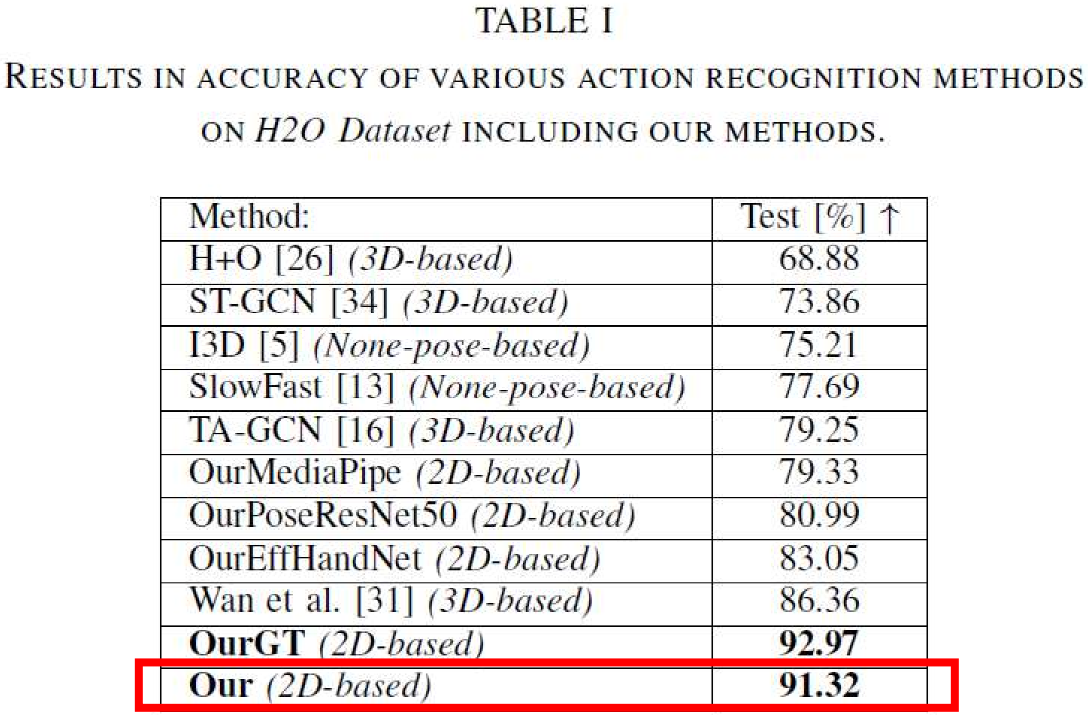


TABLE I

RESULTS IN ACCURACY OF VARIOUS ACTION RECOGNITION METHODS ON *H2O Dataset* INCLUDING OUR METHODS.

Method:	Test [%] ↑
H+O [26] *(3D-based)*	68.88
ST-GCN [34] *(3D-based)*	73.86
I3D [5] *(None-pose-based)*	75.21
SlowFast [13] *(None-pose-based)*	77.69
TA-GCN [16] *(3D-based)*	79.25
OurMediaPipe *(2D-based)*	79.33
OurPoseResNet50 *(2D-based)*	80.99
OurEffHandNet *(2D-based)*	83.05
Wan et al. [31] *(3D-based)*	86.36
OurGT *(2D-based)*	**92.97**
Our *(2D-based)*	**91.32**

Action Recognition in *FPHA Dataset*:

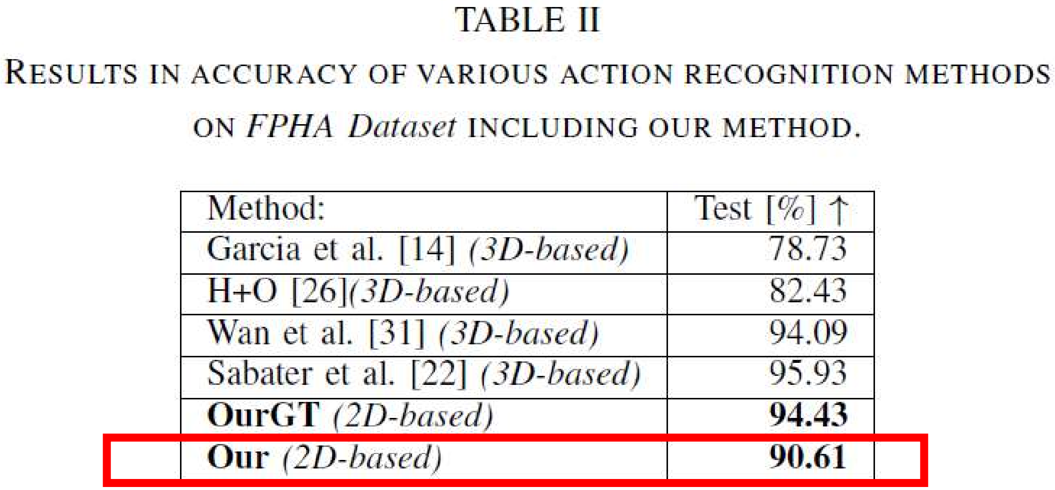


TABLE II

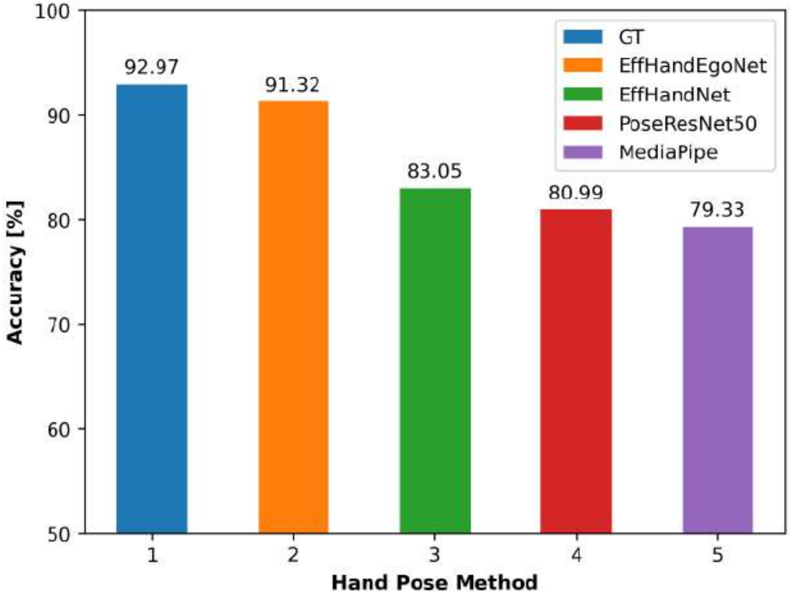
RESULTS IN ACCURACY OF VARIOUS ACTION RECOGNITION METHODS ON *FPHA Dataset* INCLUDING OUR METHOD.

Method:	Test [%] ↑
Garcia et al. [14] *(3D-based)*	78.73
H+O [26]*(3D-based)*	82.43
Wan et al. [31] *(3D-based)*	94.09
Sabater et al. [22] *(3D-based)*	95.93
OurGT *(2D-based)*	**94.43**
Our *(2D-based)*	**90.61**

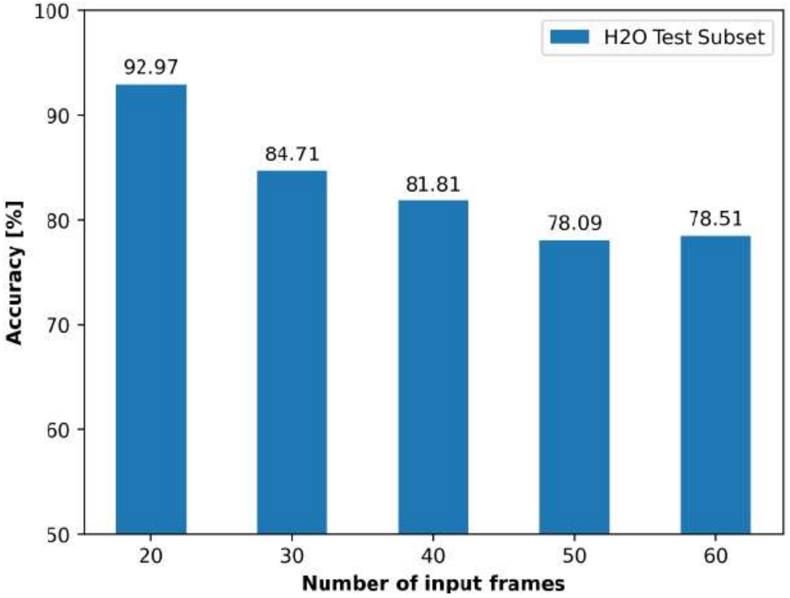
- **2D Hand Post** is **robust** for egocentric action recognition
- **Accurate pose** estimation is **essential** for action recognition

Evaluation

Impact of hand pose module:



Impact of numer of input frames:



Presented: Mucha W., Kampel M. (2023) "**Human Action Recognition in Egocentric Perspective Using 2D Object and Hands Pose**", EPIC Workshop, The Conference on Computer Vision and Pattern Recognition (CVPR2023) , June 2023, Vancouver, Canada

Mucha W., Kampel M. "**Hands, Objects, Action! Egocentric 2D Hand-based Action Recognition**", 14th International Conference on Computer Vision Systems (ICVS), September 2023, Vienna, Austria

Submitted: Mucha W., Kampel M. (2023) "**Towards Assistive Technology with Egocentric Action Recognition using 2D Hand and Object Pose**", The 18th IEEE International Conference on Automatic Face and Gesture Recognition - FG 2024

Struggle Determination

2nd secondment – University of Bristol

- Determination of struggle level in **three** different task
- Binary and 4-way determination

Tower of Hanoi

Tent Assembly

Pipes Assembly

Struggle Determination

Motivation:

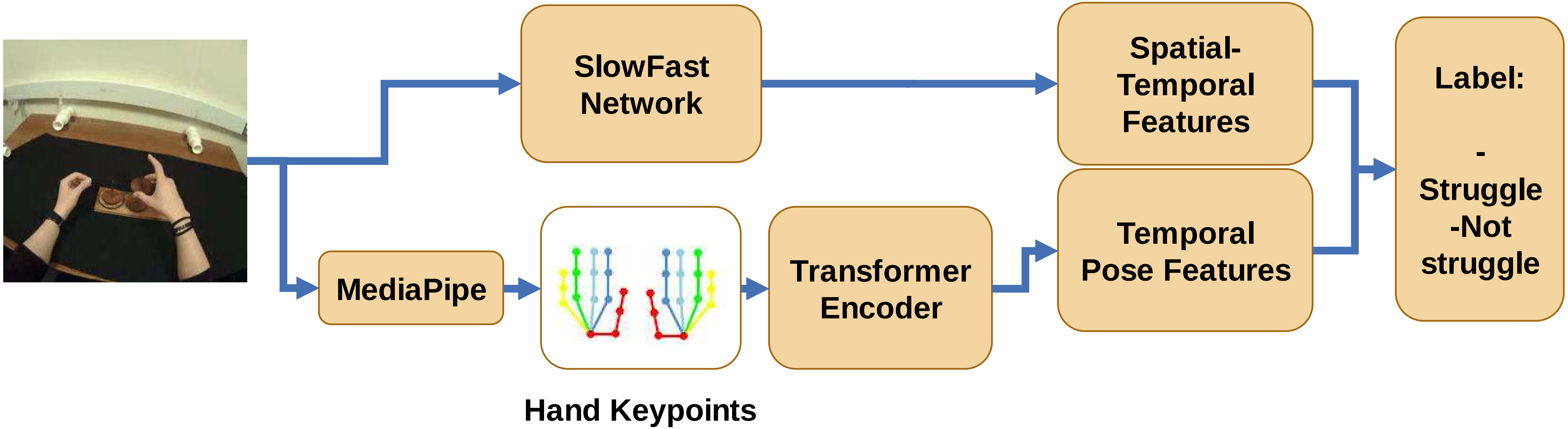
→ Correct struggle recognition leads to robust assistance for individuals

Current results and outcomes:

- **Binary** determination with **89%** of accuracy
- Best working with an approach **merging hand pose** information and **semantic** features from image
- State-of-the-art hand pose methods **do not work correctly** in these environments

Struggle Determination

Proposed approach:



SlowFast Network

Spatial-Temporal Features

MediaPipe

Hand Keypoints

Transformer Encoder

Temporal Pose Features

Label:

-

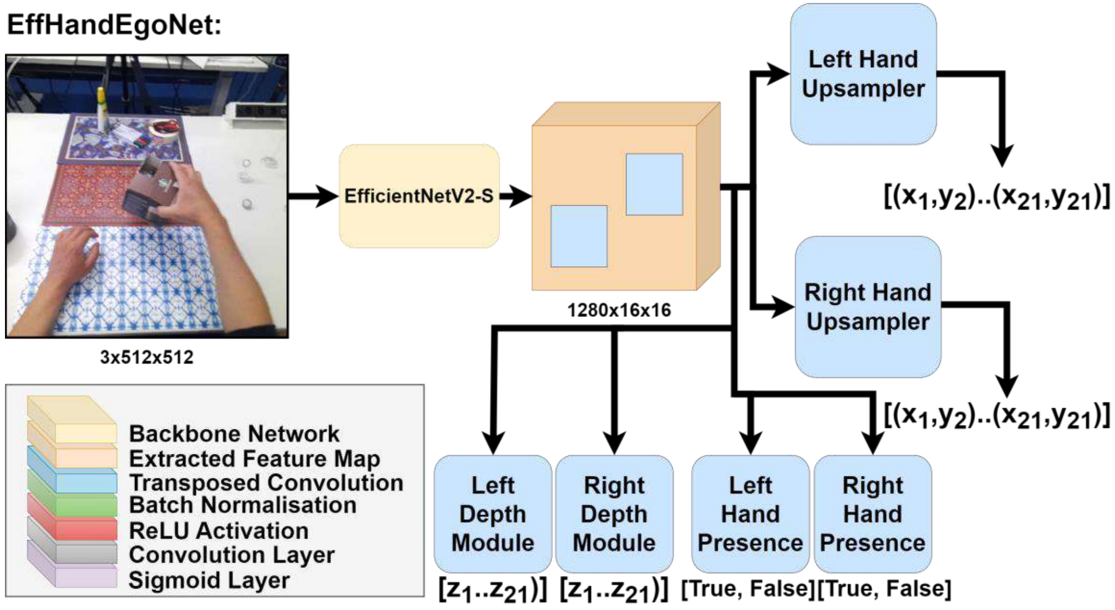
Struggle

-Not struggle

Egocentric 3D Hand Pose

I. Extending *EffHandEgoNet* to 3D

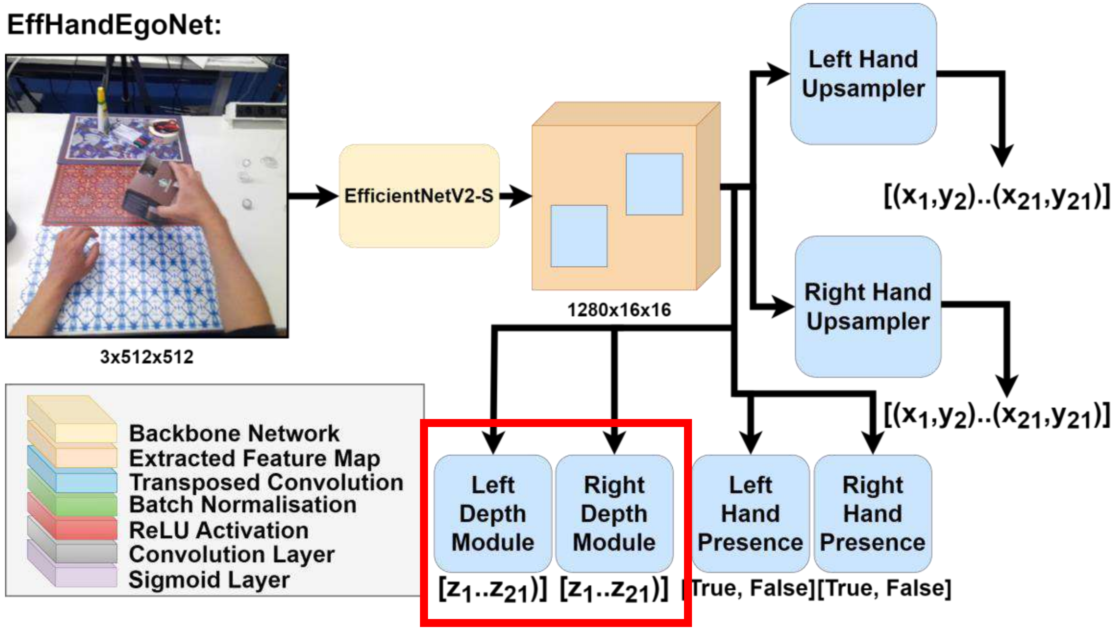
- Architecture with regressions module for estimation of **z** coordinate representing depth
- Regression head + upsampler = 2.5D coordinates (image space)
- Pinhole camera model transformation to 3D



Egocentric 3D Hand Pose

I. Extending *EffHandEgoNet* to 3D

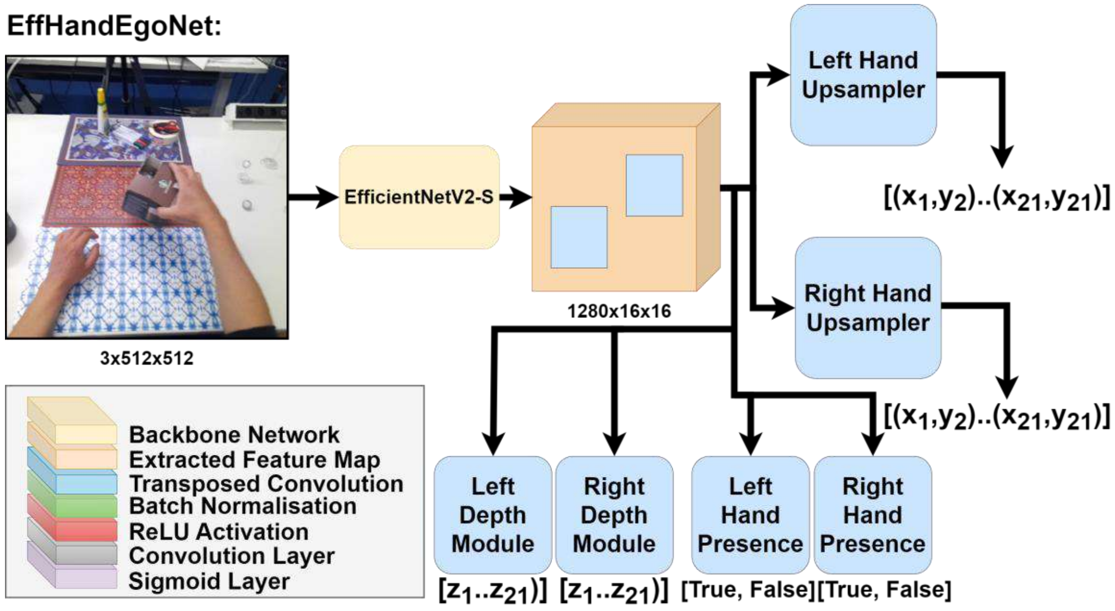
- Architecture with regressions module for estimation of **z** coordinate representing depth
- Regression head + upsampler = 2.5D coordinates (image space)
- Pinhole camera model transformation to 3D



Egocentric 3D Hand Pose

I. Extending *EffHandEgoNet* to 3D

- Architecture with regressions module for estimation of **z** coordinate representing depth
- Regression head + upsampler = 2.5D coordinates (image space)
- Pinhole camera model transformation to 3D



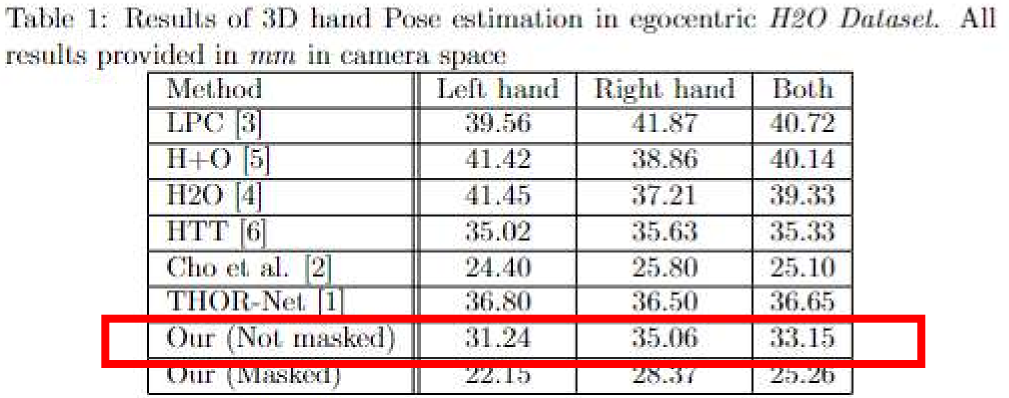


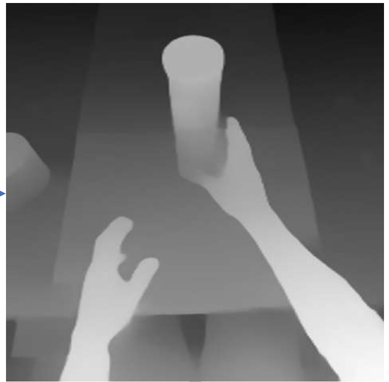
Table 1: Results of 3D hand Pose estimation in egocentric *H2O Dataset*. All results provided in *mm* in camera space

Method	Left hand	Right hand	Both
LPC [3]	39.56	41.87	40.72
H+O [5]	41.42	38.86	40.14
H2O [4]	41.45	37.21	39.33
HTT [6]	35.02	35.63	35.33
Cho et al. [2]	24.40	25.80	25.10
THOR-Net [1]	36.80	36.50	36.65
Our (Not masked)	31.24	35.06	33.15
Our (Masked)	22.15	28.37	25.26

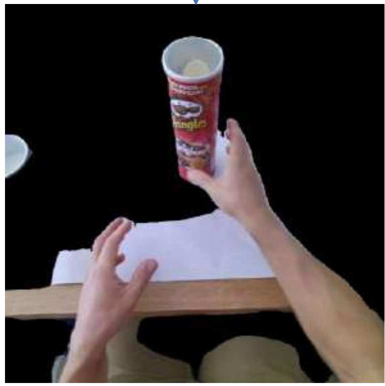
Egocentric 3D Hand Pose

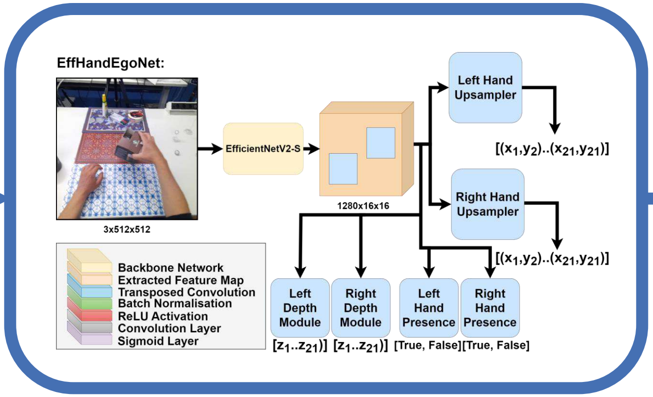
I. Masking using estimated depth information

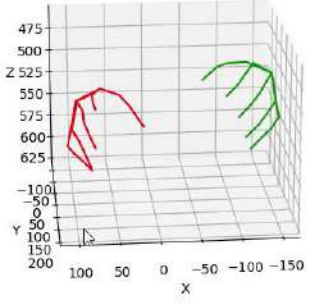
Depth Estimation
(DPT-Hybrid)



Masking







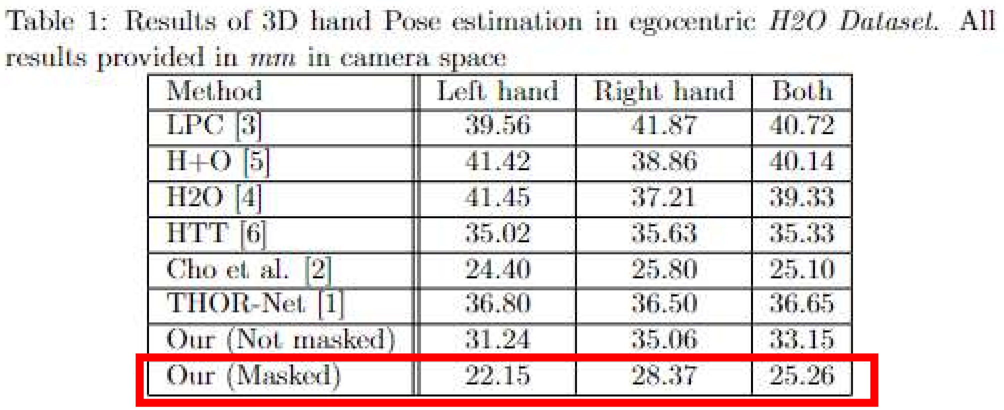
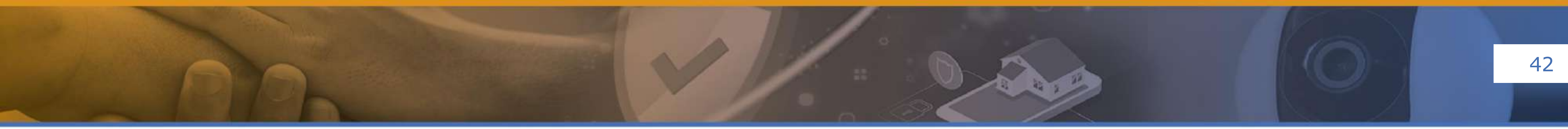


Table 1: Results of 3D hand Pose estimation in egocentric *H2O Dataset*. All results provided in *mm* in camera space

Method	Left hand	Right hand	Both
LPC [3]	39.56	41.87	40.72
H+O [5]	41.42	38.86	40.14
H2O [4]	41.45	37.21	39.33
HTT [6]	35.02	35.63	35.33
Cho et. al. [2]	24.40	25.80	25.10
THOR-Net [1]	36.80	36.50	36.65
Our (Not masked)	31.24	35.06	33.15
Our (Masked)	22.15	28.37	25.26



Dissemination overview

Finished and planned publications

Published Research:

Mucha W., Kampel M. (2022) "**Depth and Thermal Images in Face Detection – A Detailed Comparison Between Image Modalities**", The 5th International Conference on Machine Vision and Applications (ICMVA 2022), February 18-20, 2022, Singapore

Mucha W., Kampel M. (2022) "**Beyond Privacy of Depth Sensors in Active and Assisted Living Devices**", The 15th PErvasive Technologies Related to Assistive Environments Conference – PrivAw Workshop, June 29 – July 1, 2022, Corfu, Greece

Mucha W., Kampel M. (2022) "**Addressing Privacy Concerns in Depth Sensors**", Joint International Conference on Digital Inclusion, Assistive Technology & Accessibility – ICCHP-AAATE 2022, July 11-15, 2022, Lecco, Italy

Mucha W., Kampel M. "**Hands, Objects, Action! Egocentric 2D Hand-based Action Recognition**", Accepted in the 14th International Conference on Computer Vision Systems (ICVS), September 2023, Vienna, Austria

Planned Publications:

- **Papers under the review:**

Mucha W., Kampel M. "**In My Perspective, In My Hands: Accurate Egocentric 2D Hand Pose**", The 18th IEEE International Conference on Automatic Face and Gesture Recognition - FG 2024

Mucha W., Kampel M. "**Towards Assistive Technology with Egocentric Action Recognition using 2D Hand and Object Pose**", The 18th IEEE International Conference on Automatic Face and Gesture Recognition - FG 2024

- **Planned papers:**
  - Methodology:
    → 3D Hand Pose in egocentric vision
  - Applications:
    → Hand rehabilitation with egocentric vision
    → Medication intake monitoring

Next steps of my PhD

Work to be completed to finalise PhD

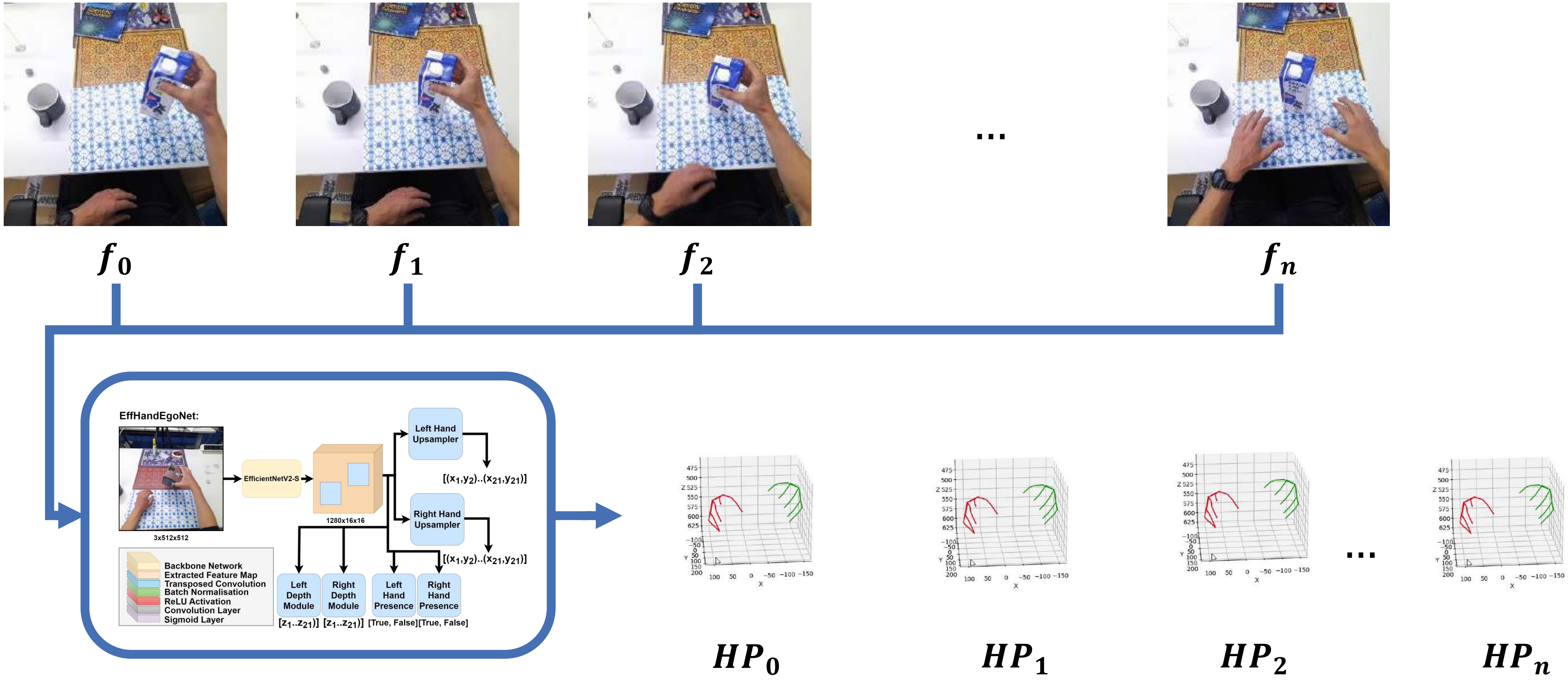
Struggle Determination

I. Further work for struggle determination project:

- Data collection in collaboration with the University of Bristol
  - → Introduction of more tasks
  - → Increasing data in current tasks
- Incorporate new methods such as self-supervision
- Transferring hand pose models to struggle dataset
- Struggle detection for future action support

Egocentric 3D Hand Pose

II. Embedding temporal information for improvement of 3D pose



Egocentric 3D Hand Pose

III. Domain Adaptation Improvement

Four datasets for egocentric hand pose estimation:

- FPHA
- H2O
- AssemblyHands
- HoloAssist

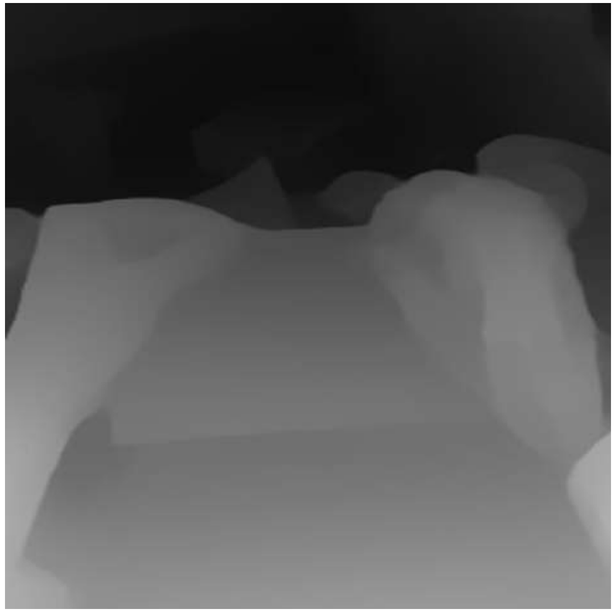
Difficult to annotate

Variance in image quality, distortion, type

Potential solution:

- Self-supervised learning through **depth estimation** task:

FPHA



H2O

Health Related Tasks

IV. Upper-limb rehabilitation with egocentric vision for stroke

Motivation:

- Stroke remains the **third** leading cause of **mortality** and
- **disability** worldwide, [1].
- Approximately **85% of stroke patients** worldwide experience **hand dysfunction** [2].
- No egocentric studies available

Challenges:

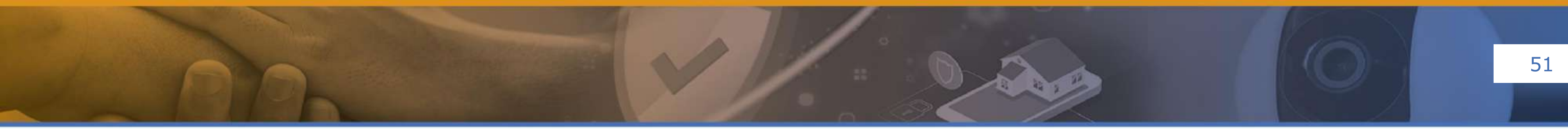
- Exercise recognition
- Repetition counting
- Exercise detection
- Form evaluation

[13] V. L. Feigin et al., “Global, regional, and national burden of stroke and its risk factors, 1990–2019: a systematic analysis for the global burden of disease study 2019,” The Lancet Neurology, vol. 20, no. 10, pp. 795–820, 2021

[14] D. Cao et al., “Efficacy and safety of manual acupuncture for the treatment of upper limb motor dysfunction after stroke: Protocol for a systematic review and meta-analysis,” Plos one, vol. 16, no. 11, p. e0258921, 2021.

Timeline

Task	2021			2022				2023					2024				2025
	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4		Q1	Q2	Q3	Q4	Q1
Initial research on life logging and AAL devices	🟩											Last visuAAL PhD seminar in Vienna, December 2023					
Modality comparison study		🟩	P														
Privacy with depth			🟩	P P													
Proficiency evaluation					R.P.	Pres.											
Hand based action recognition									P	🟩	P						
Egocentric hand pose paper							🟩	🟩			P						
3D hand pose for AR											🟩		P	🟩			
Struggle determination											🟩		🟩	🟩	P		
Hand rehabilitation with egocentric view											🟩		🟩	🟩	P		
Secondments					🟧						🟧						
Thesis Writing																Thesis	



Thank you!

Wiktor Mucha

TU Wien

wiktor.mucha@tuwien.ac.at