ETH zürich

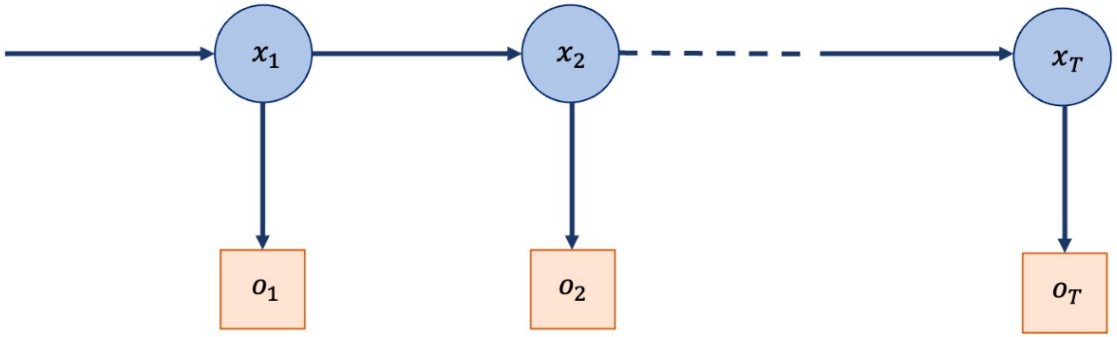
pd|z Product Development Group Zurich
Produktentwicklungsgruppe Zürich

Development of an Automated Task Sequencing Framework

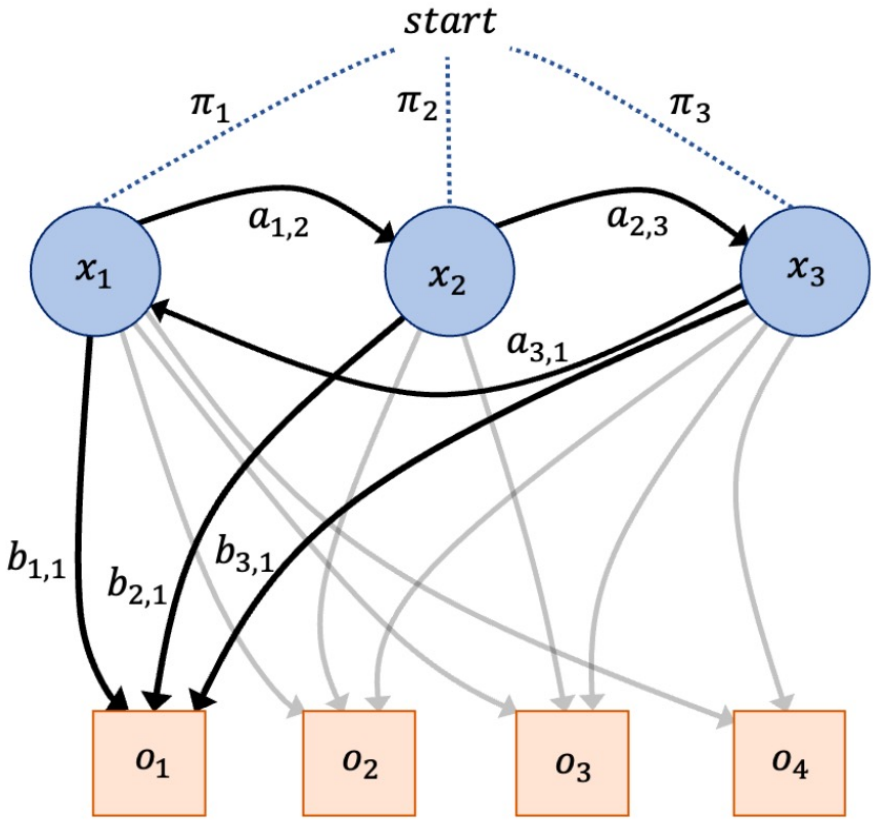
Cherif Ghazi
Supervised by Felix Wang

25. April 2022, Zurich

Introduction



- How the model works now.
- Prediction based on previous and current state and an ensemble of probabilities.



Example Data

Cap Off - Removal of the pen cap.
Significant Area of interest (AOI): Cap, Pen
Apply Tip - Applying the disposable needle to the pen.
Significant AOIs: Safety, Tip, Pen
Setting Units - Setting or checking the desired number of units using the dosage knob.
Significant AOIs: Gauge
Priming-Testing the function of the needle, removing trapped air by holding the device upright and releasing 1-2 units.
Significant AOIs: Tip, Pen, App
Injection - Placing the needle into injection area, injecting 5-10 units for 10 seconds.
Significant AOIs: Pad, Tip, App
Remove Tip - Removal of the disposable needle from the pen.
Significant AOIs: Safety, Tip, Pen
Cap On - Applying the Cap to the Pen.
Significant AOIs: Cap, Pen

Start_time	End_time	App	Cap	Gauge	Pen	Safety	Tip	Action
0	0.5	400	111	0	380	1600	0	Apply Tip
0.5	1	417	85	87	480	1600	1600	Apply Tip
1.5	2	300	84	1600	515	1600	500	Setting Units
2	2.5	200	1	1600	600	1600	200	Apply Tip

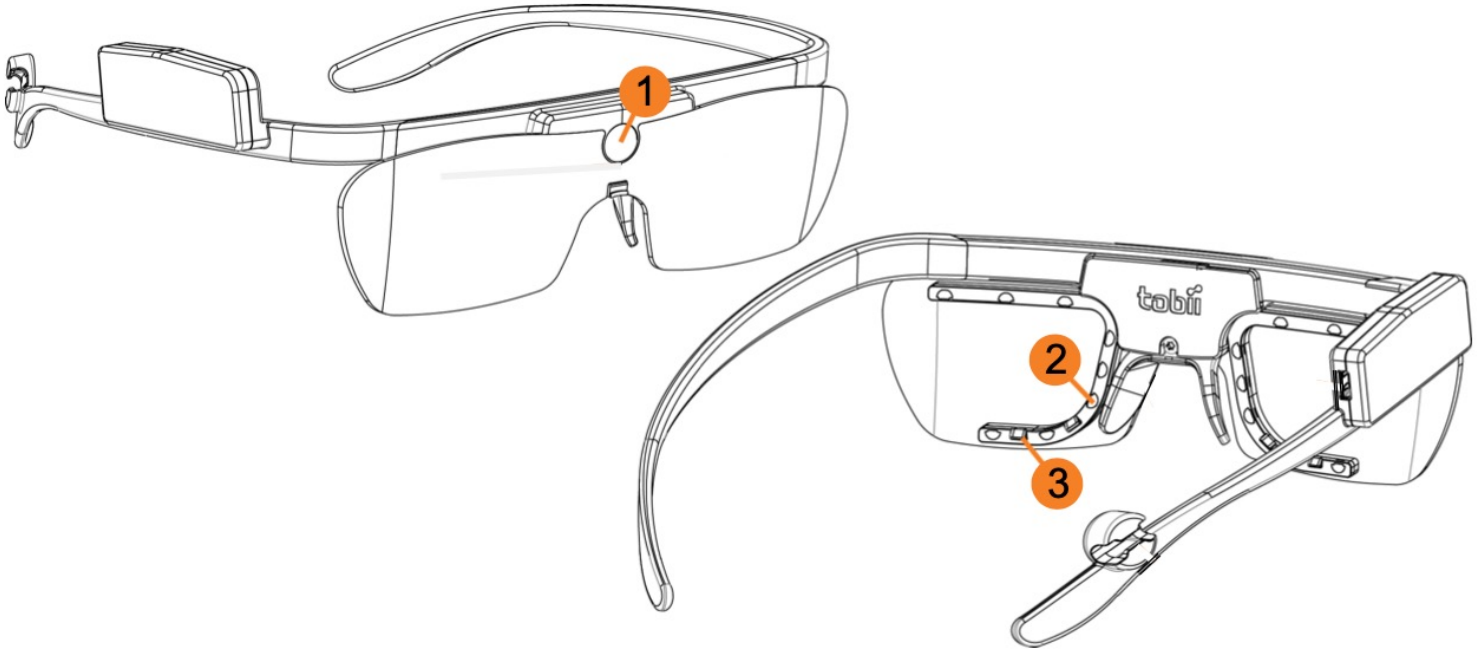
- Experiments have been carried out on several participants.
- They had to perform 7 different action.

Tobi Pro Glasses 2

pd|z Product Development Group Zurich
Produktentwicklungsgruppe Zürich

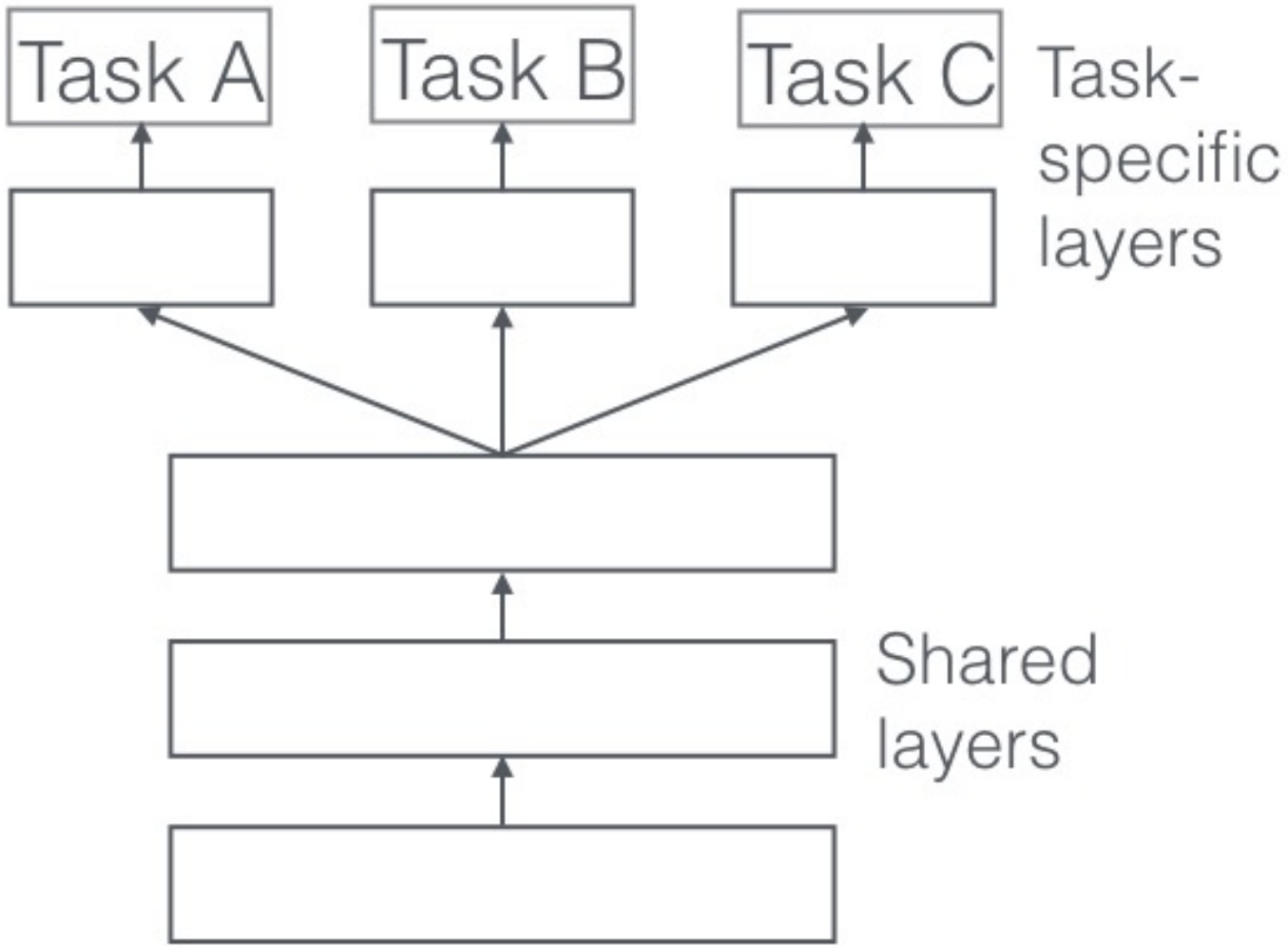
1. *High- definition scene camera - Captures a Full HD video of what is in front of the participant.*
2. *IR illuminators - Illuminates the eyes to support the eye tracking sensors.*
3. *Eye tracking sensors - Records eye orientation e.g. the direction of the eye gaze*

(Tobii AB 2018).



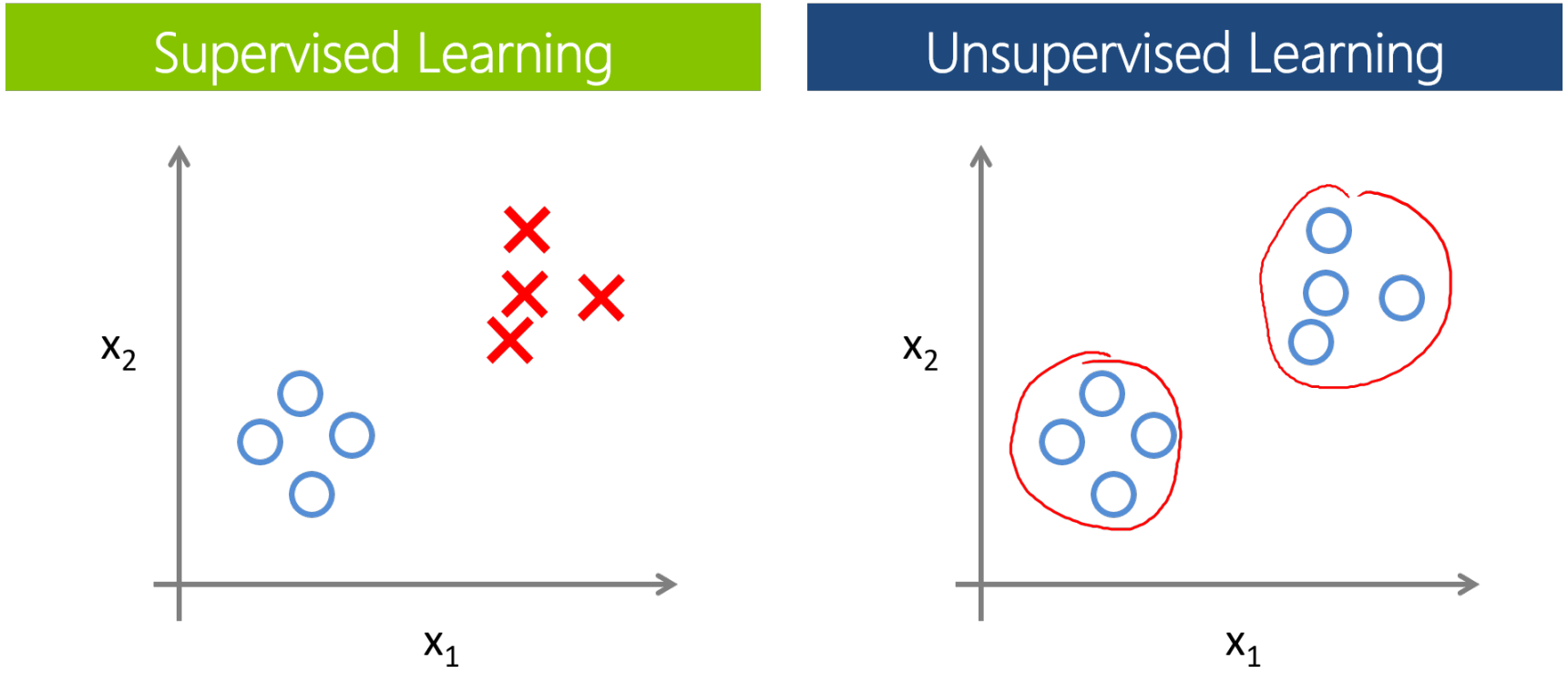
Multi Task Learning (MTL)

- Initial Idea.
- Sharing representations between related tasks.



Supervised VS Unsupervised learning

- Different approaches.
- Clustering the data using Kmeans.
- Raw data as an input to a Neural Network (NN).

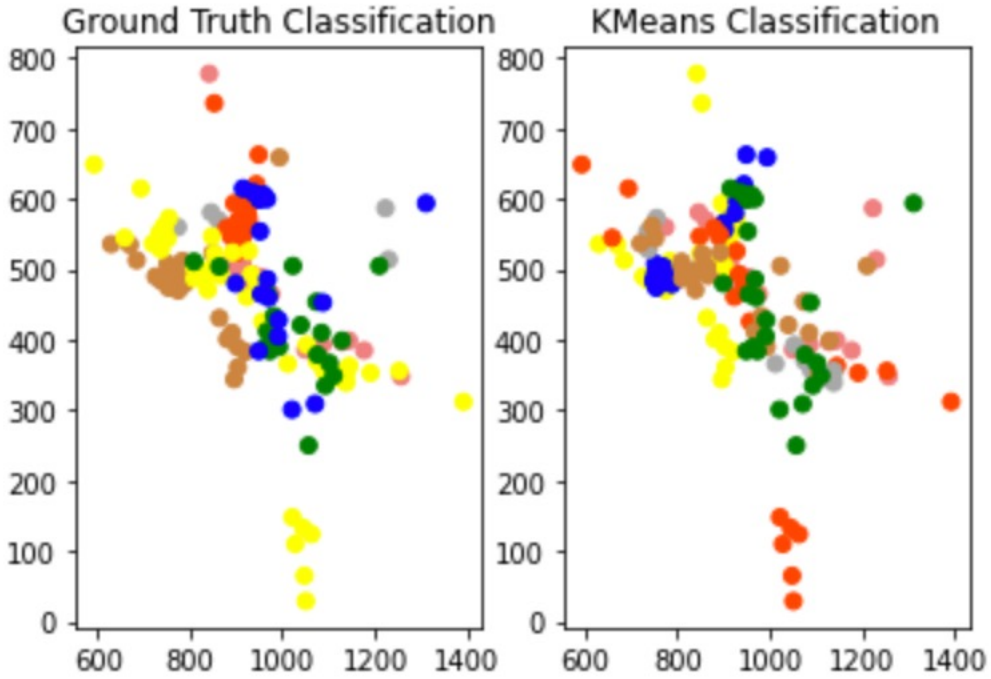


Advantages and Disadvantages

pd|z Product Development Group Zurich
Produktentwicklungsgruppe Zürich

	Advantages	Disadvantage
Supervised Learning	• Easy to find how many classes are there before giving data for training • Very efficient once it has been properly trained	• Needs a lot of computer resources • Risk of overfitting
Unsupervised Learning	• Does not need as much computer resources • No risk of overfitting	• Need to interpret the optimal number of clusters • It is not always certain that the obtained results will be useful since there is no label or output measure to confirm its usefulness.

Unsupervised Learning Results

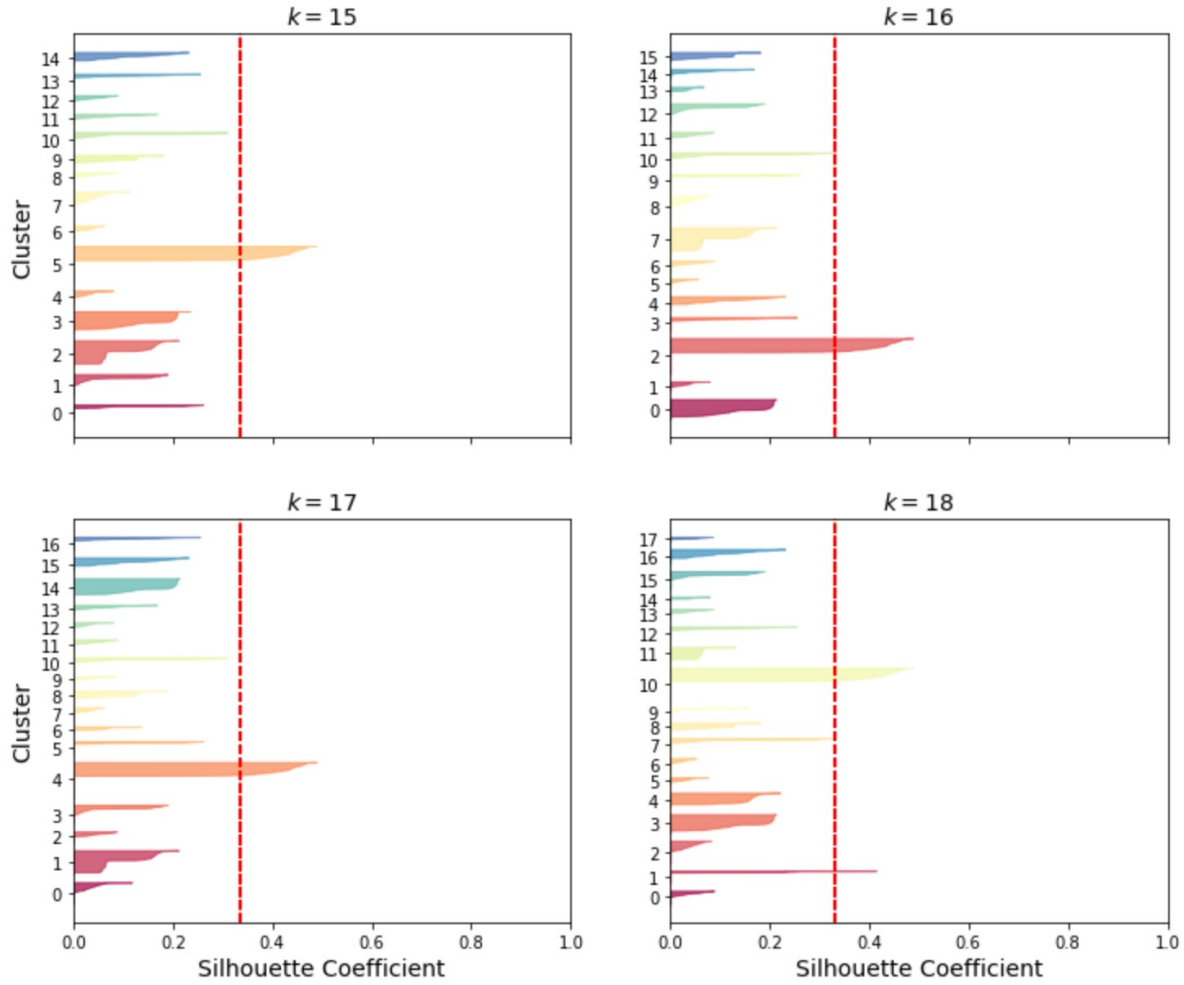


	precision	recall	f1-score	support
1	0.00	0.00	0.00	9
2	0.62	0.83	0.71	18
3	0.14	0.18	0.15	17
4	0.00	0.00	0.00	35
5	0.00	0.00	0.00	50
6	0.29	0.35	0.32	20
7	0.00	0.00	0.00	17
accuracy			0.15	166
macro avg	0.15	0.19	0.17	166
weighted avg	0.12	0.15	0.13	166

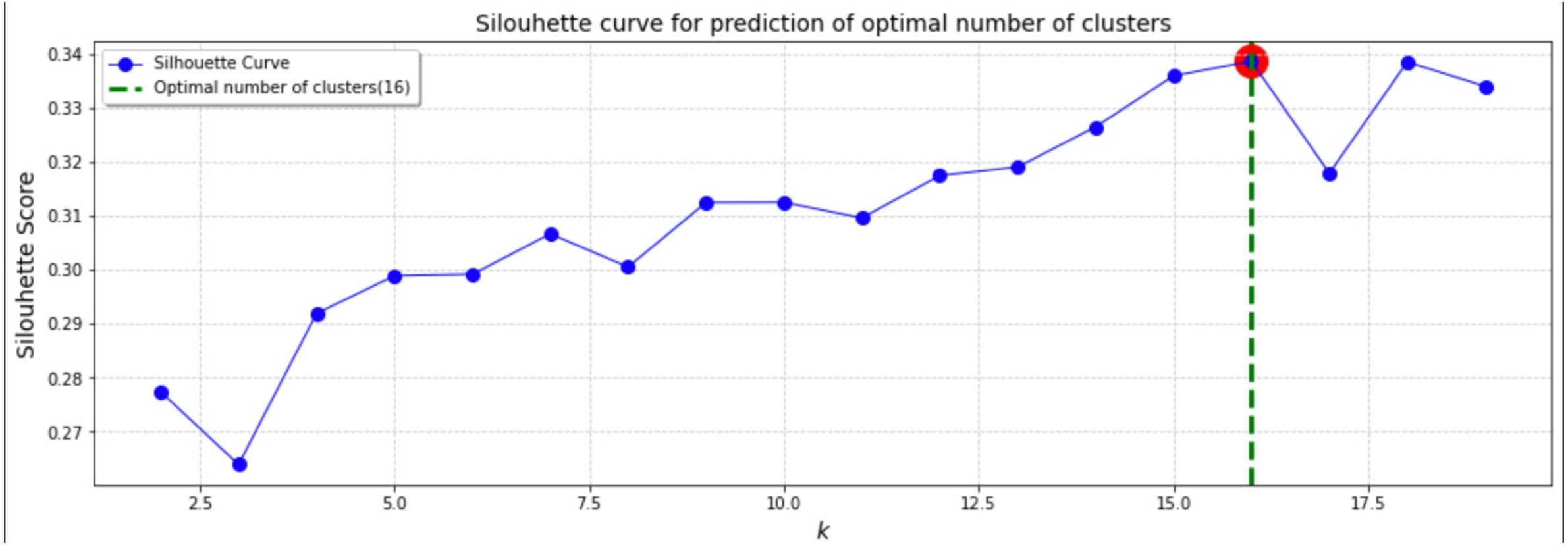
- Low accuracy (15%) probably due to high dimensionality (7 dimensional data).
- We had to find a new solution / approach.

Silhouette Coefficients

- Method of interpretation and validation of consistency within clusters of data
- How similar an object is to its own cluster compared to other clusters
- Ranges from -1 to +1
- High value indicates that the object is well matched to its own cluster
- Low value means clustering configuration may have too many or too few clusters



Silhouette Coefficients



- Method of interpretation and validation of consistency within clusters of data.
- Provides a succint graphical representation of how well each object has been classified.

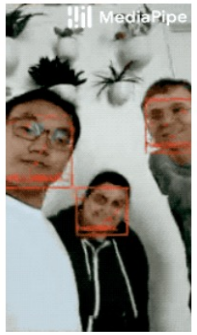
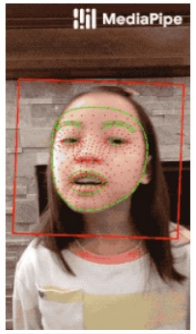
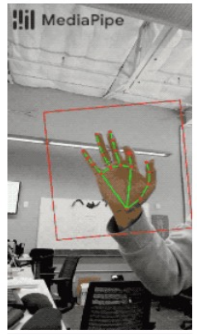
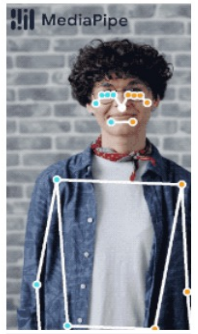
Supervised Learning Results

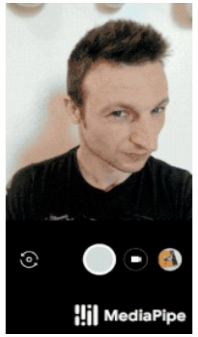
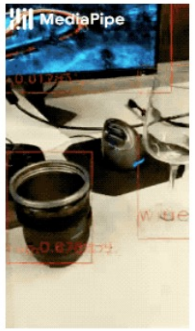
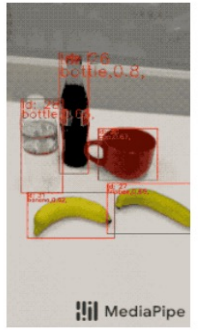
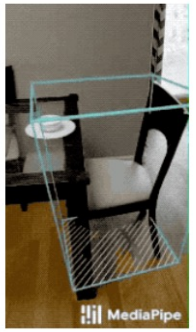

```
INFO:tensorflow:Calling model_fn.
INFO:tensorflow:Done calling model_fn.
INFO:tensorflow:Starting evaluation at 2022-04-11T13:47:24
INFO:tensorflow:Graph was finalized.
INFO:tensorflow:Restoring parameters from /tmp/tmp_jmczt57/model.ckpt-5000
INFO:tensorflow:Running local_init_op.
INFO:tensorflow:Done running local_init_op.
INFO:tensorflow:Inference Time : 0.72536s
INFO:tensorflow:Finished evaluation at 2022-04-11-13:47:25
INFO:tensorflow:Saving dict for global step 5000: accuracy = 0.2, average_loss = 1.98889, global_step = 5000, loss = 1.98889
INFO:tensorflow:Saving 'checkpoint_path' summary for global step 5000: /tmp/tmp_jmczt57/model.ckpt-5000

Test set accuracy: 0.200
```

- Low accuracy (20%) .
- Layers 100-80-50-20

Mediapipe

- Google Model.
- Trained over a million images.

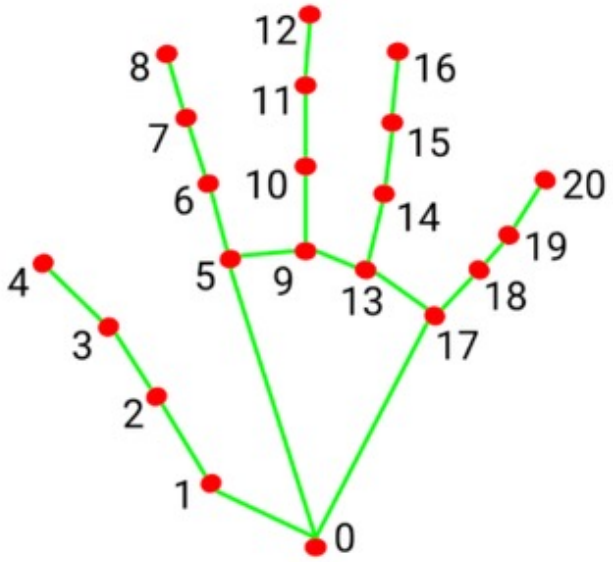
Face Detection	Face Mesh	Iris	Hands	Pose	Holistic
					

Hair Segmentation	Object Detection	Box Tracking	Instant Motion Tracking	Objectron	KNIFT
					

Landmarks

pd|z Product Development Group Zurich
Produktentwicklungsgruppe Zürich

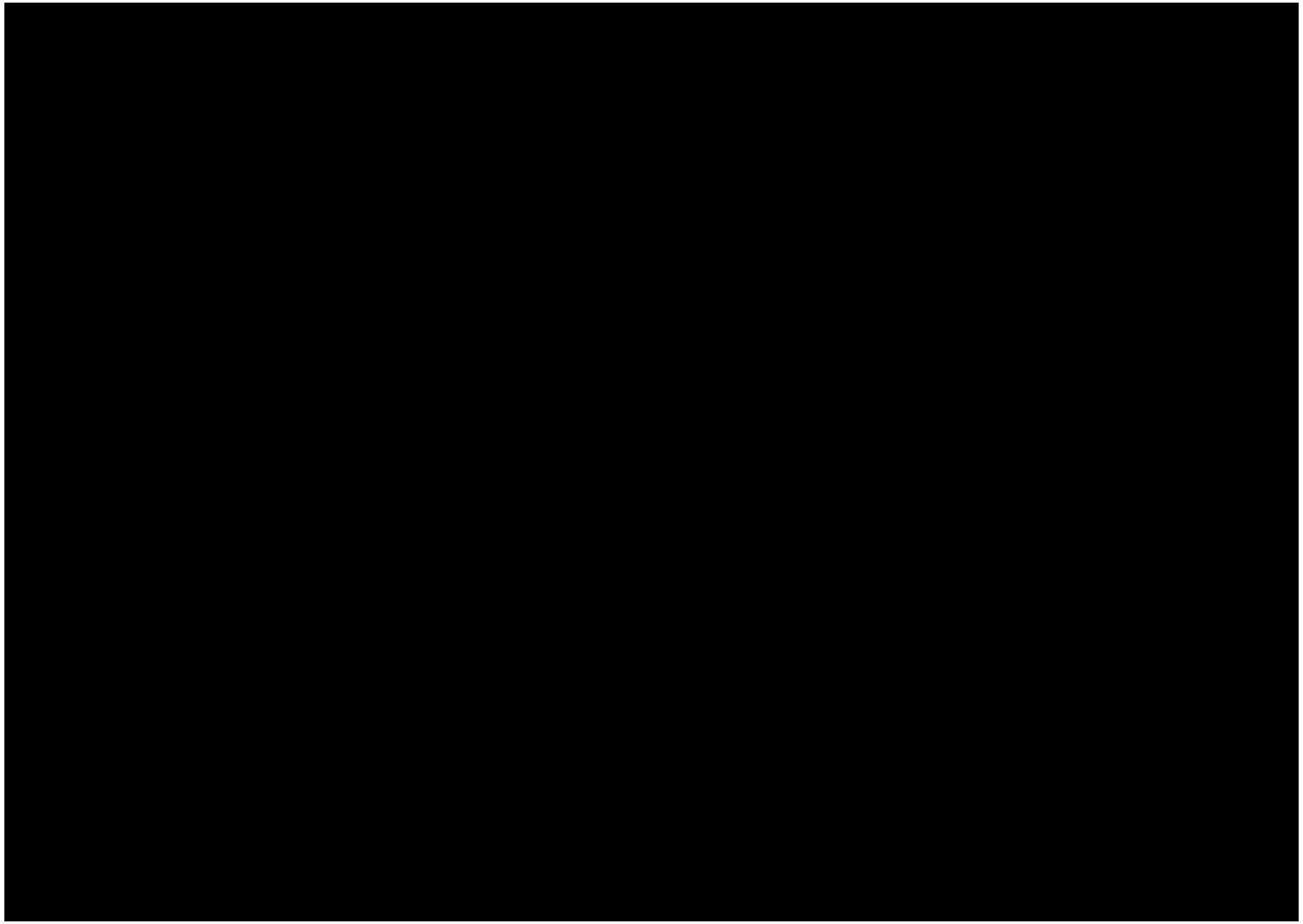
- Detecting a bunch of landmarks on the hands.
- Classified and identified.



0. WRIST
1. THUMB_CMC
2. THUMB_MCP
3. THUMB_IP
4. THUMB_TIP
5. INDEX_FINGER_MCP
6. INDEX_FINGER_PIP
7. INDEX_FINGER_DIP
8. INDEX_FINGER_TIP
9. MIDDLE_FINGER_MCP
10. MIDDLE_FINGER_PIP
11. MIDDLE_FINGER_DIP
12. MIDDLE_FINGER_TIP
13. RING_FINGER_MCP
14. RING_FINGER_PIP
15. RING_FINGER_DIP
16. RING_FINGER_TIP
17. PINKY_MCP
18. PINKY_PIP
19. PINKY_DIP
20. PINKY_TIP

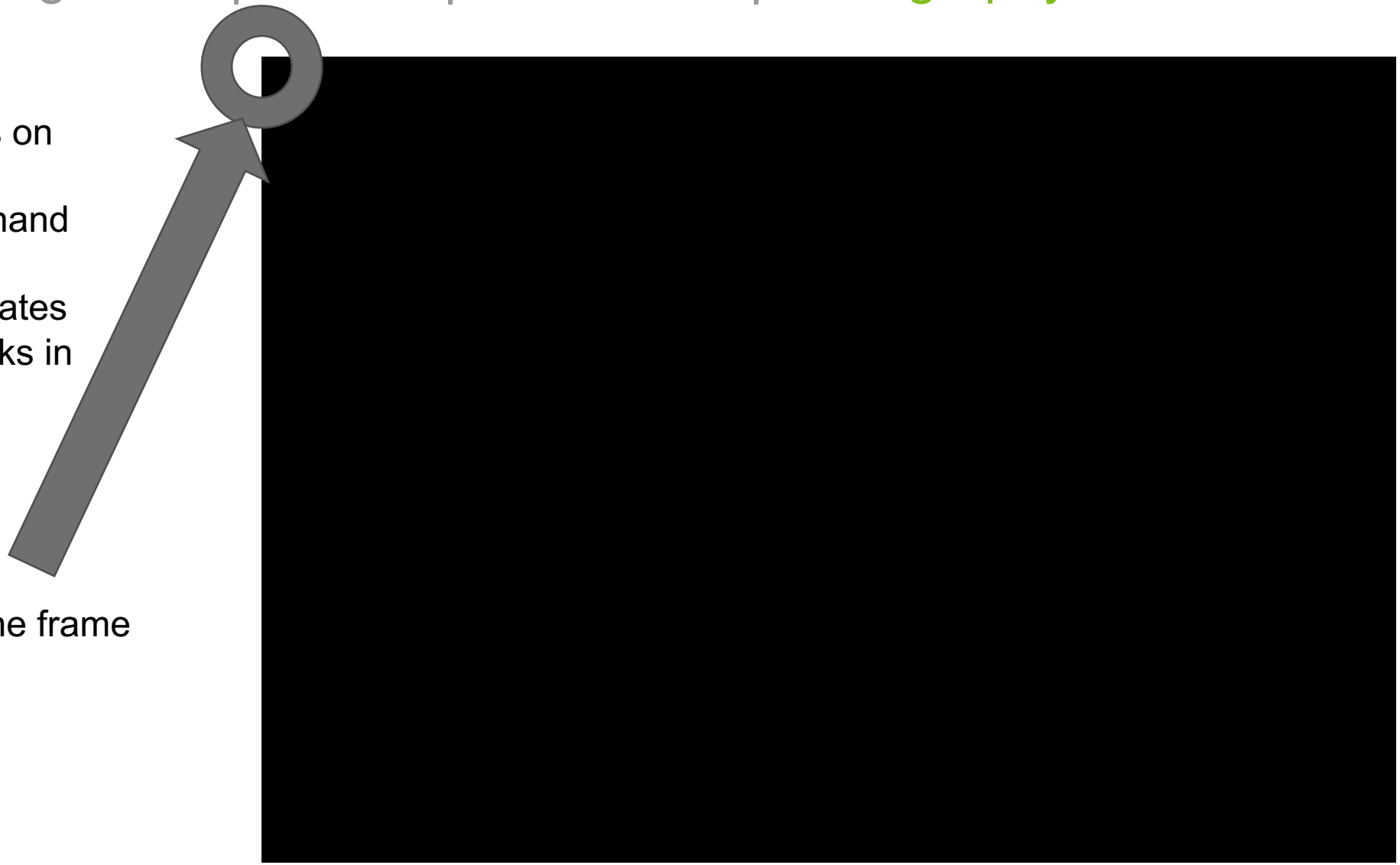
Unsupervised Learning Results

- Prints all the landmarks on the video.
- Detecting whether the hand is left or right.
- Ouputs x and y coordinates of the different landmarks in the image frame.



What we can achieve

- Prints all the landmarks on the video.
- Detecting whether the hand is left or right.
- Ouputs x and y coordinates of the different landmarks in the image frame.



Origin of the frame

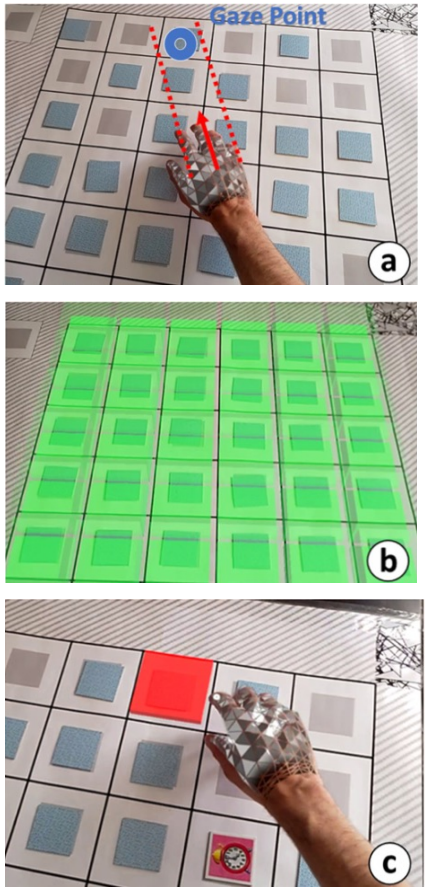
Metrics

pd|z Product Development Group Zurich
Produktentwicklungsgruppe Zürich

Hand Object

Gaze Object

Gaze Hand



Conclusion

Future

- Continue to focus on two approaches.
- Dig unsupervised learning deeper.
- Improve former algorithm based on Hidden Markov models or come up with a new neural Network if we do not succeed in improving it.

Possible issues

- Synchronization between algorithm (15 Hz) and glasses (25 Hz)
- Understanding clusters
- Combine the data and hand tracking

Bibliography

1. https://ruder.io/multi-task/index.html#fn3
2. https://lakshaysuri.wordpress.com/2017/03/19/machine-learning-supervised-vs-unsupervised-learning/
3. https://fr.wikipedia.org/wiki/Silhouette_(clustering)
4. https://google.github.io/mediapipe/
5. https://imotions.com/blog/10-terms-metrics-eye-tracking/
6. Gaze Comes in Handy: Predicting and Preventing Erroneous Hand Actions in AR-Supported Manual Tasks , Julian Wolf
7. https://www.dreamstime.com/object-form-geometric-part-angle-degrees-expert-hand-professional-to-take-measurements-calculate-right-image168287546
8. Development of a Peripheral Vision Based Approach to Human Task Classification, Thomas Kreiner

Felix Wang
Phd Student
wangfe@ethz.ch
ETH Zürich

Product Development Group
LEE O10
Leonhardstrasse 21
8092 Zürich