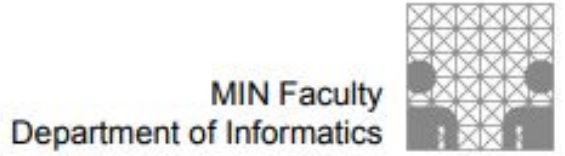


Gesture Recognition with CNN

University of Hamburg
Faculty of Mathematics, Informatics and Natural Sciences
Department of Informatics
Technical Aspects of Multimodal Systems

Ahmed Abdelghany

20 January 2020

Outline

- Motivation for Gesture Recognition
- Taxonomy of GR
- Sensors for Gesture Recognition
- GR for Human Robot Interaction
- Convolutional Neural Network
- Architectures of CNN for GR
  - CNN, Multi Channel CNN, CNN with LSTM
- Experiments & Results
- Conclusion & Future work

Motivation

- Gesture Recognition is one of the most interesting and challenging areas in Human-Robot-Interaction (HRI)
- Both in research and industry
- Obstacles?
  - Image Segmentation
  - Temporal and Spatial feature extraction
  - Real time recognition

Research Question

- Is Convolutional Neural Network able to successfully handle Gesture Recognition tasks?

- Can Convolutional Neural Network be tuned to handle both static and dynamic Gesture Recognition?

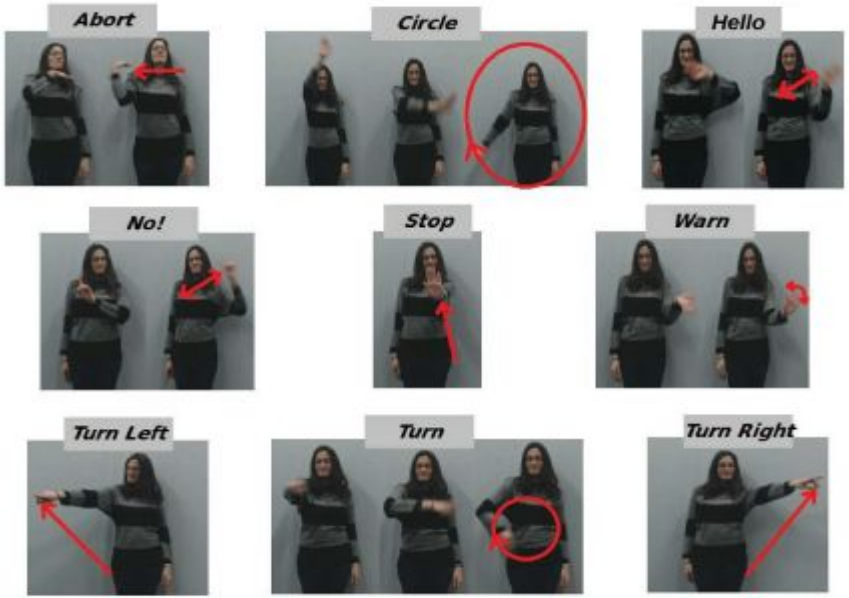
Taxonomy of Gestures

- Static: position does not change during the gesturing time, pose or configuration
- Dynamic: position changes continuously with time hands, arms, face, head, and/or body
- Both Static and Dynamic: Sign language

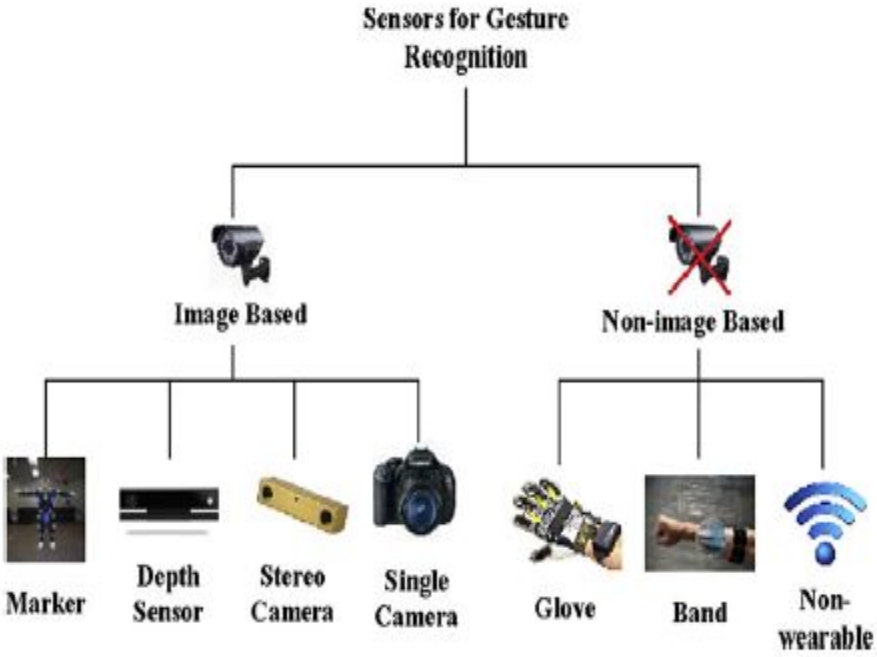
- The meaning of a gesture can be dependent on:
  - spatial information: where it occurs
  - pathic information: the path it takes

Gesture Recognition

Examples of Gestures:



Sensors for Gesture Recognition

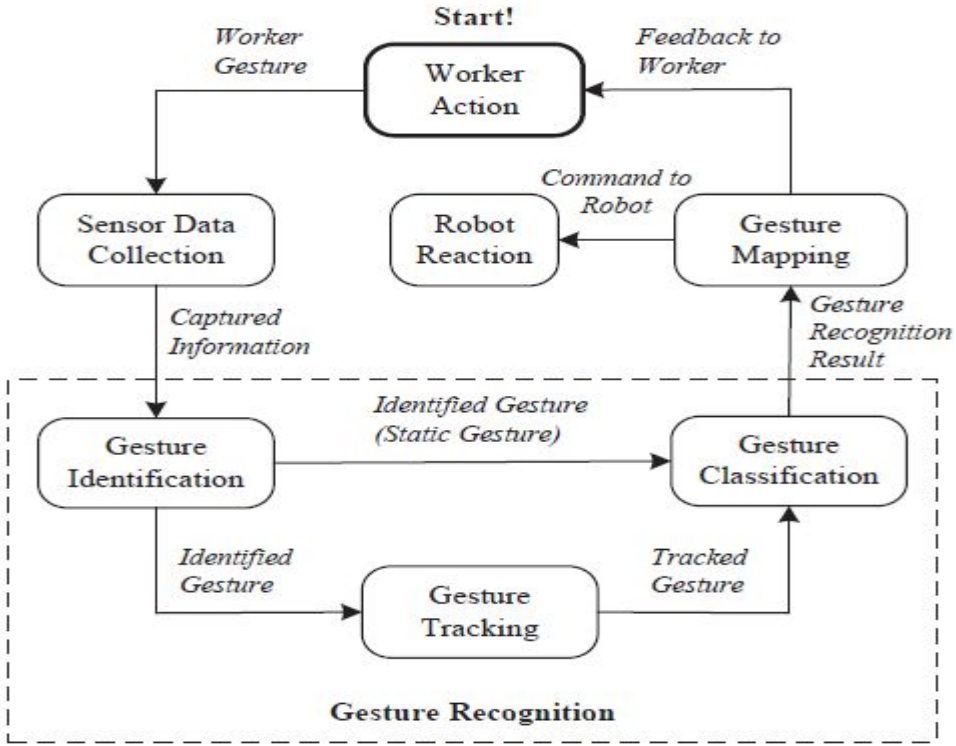


Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review [2]

Gesture Recognition in HRI

5 Steps:

- Sensor data collection
- Gesture identification
- Gesture tracking
- Gesture classification
- Gesture mapping



A review of vision based hand gestures recognition [3]

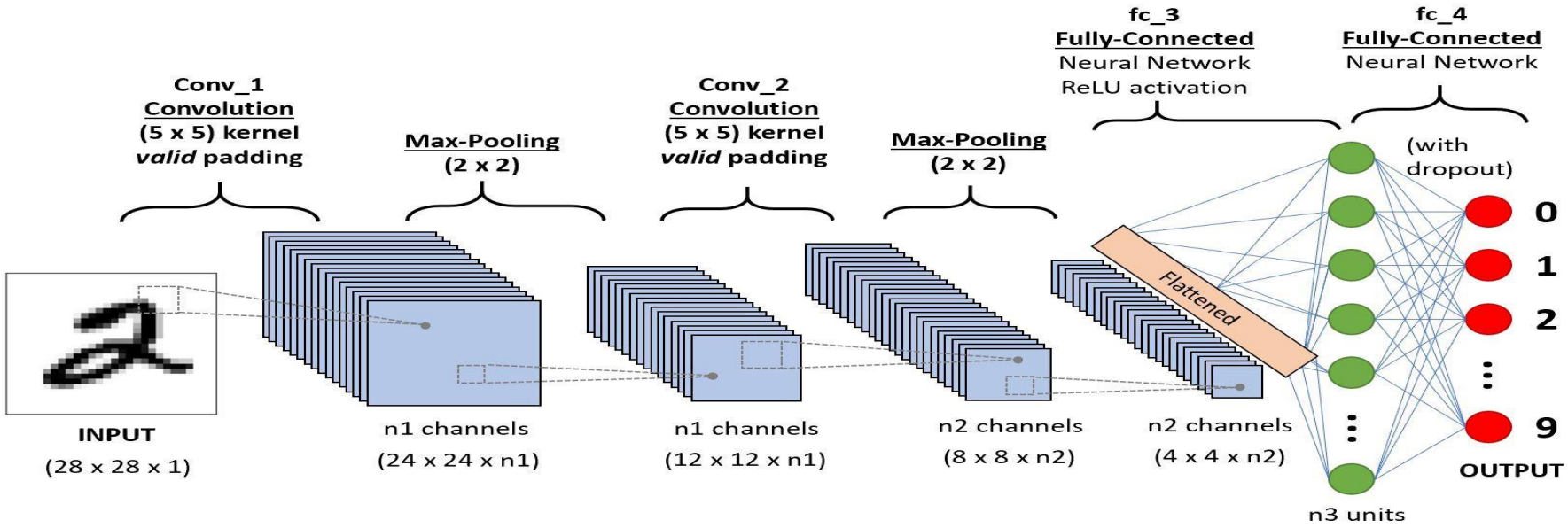
Gesture Recognition in HRI

https://www.youtube.com/watch?v=Vpr1cE44Lpw

Convolutional Neural Network: Why?

- Ability to extract the temporal and spatial features of a gesture sequence
- The specification of gesture start and end points in the frames of movement is needed
- Temporal segmentation is required for the recognition of continuous gestures

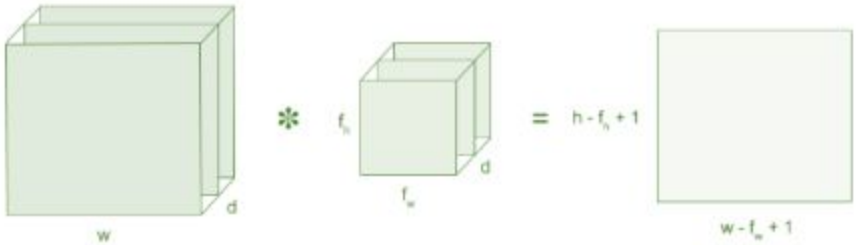
CNN Architecture



https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

CNN Architecture

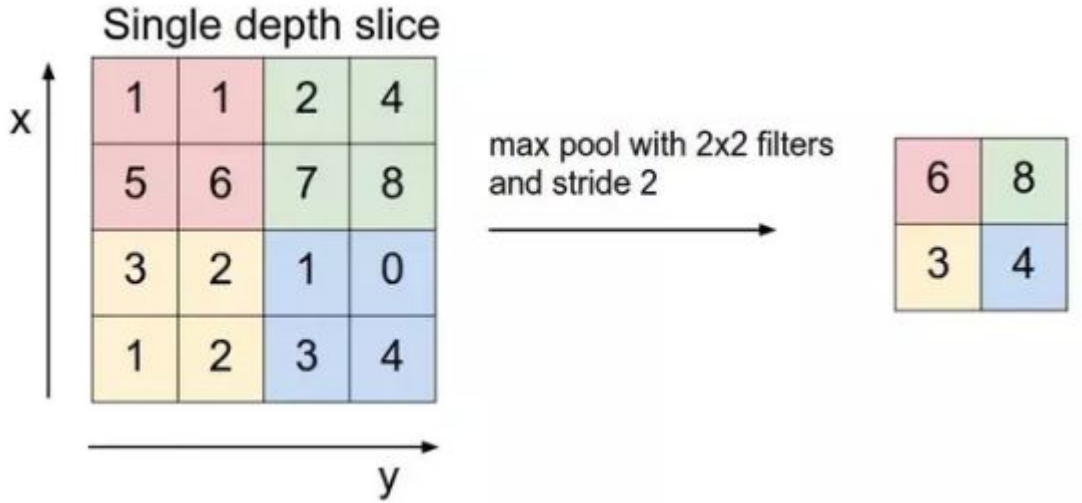
- Convolution Layer: image multiplies kernel or filter matrix, creates feature maps



https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks

CNN Architecture

- Pooling Layer:
  - Reduce the number of parameters
  - Can be max pooling, average pool or sum pooling

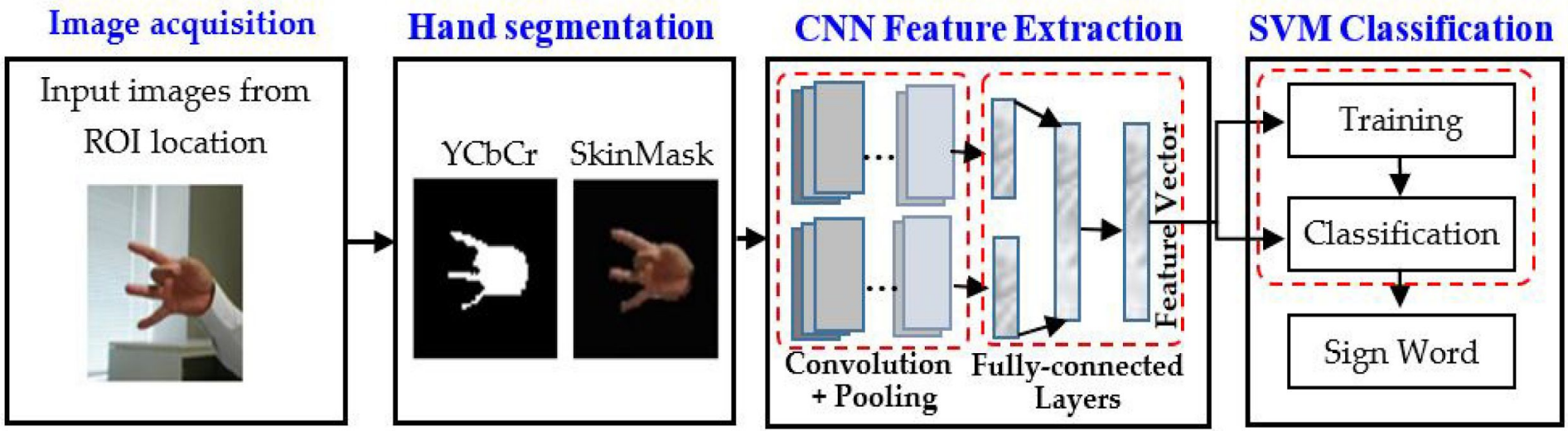


https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks

Drawback: Are CNN's flawless?

- Backpropagation not always an efficient way of learning, because it needs huge dataset
- Convolution is a slow operation, therefore high computational cost
- CNNs do not encode the orientation of object
- Pooling layers loses a lot of valuable information

Gesture Recognition with CNN

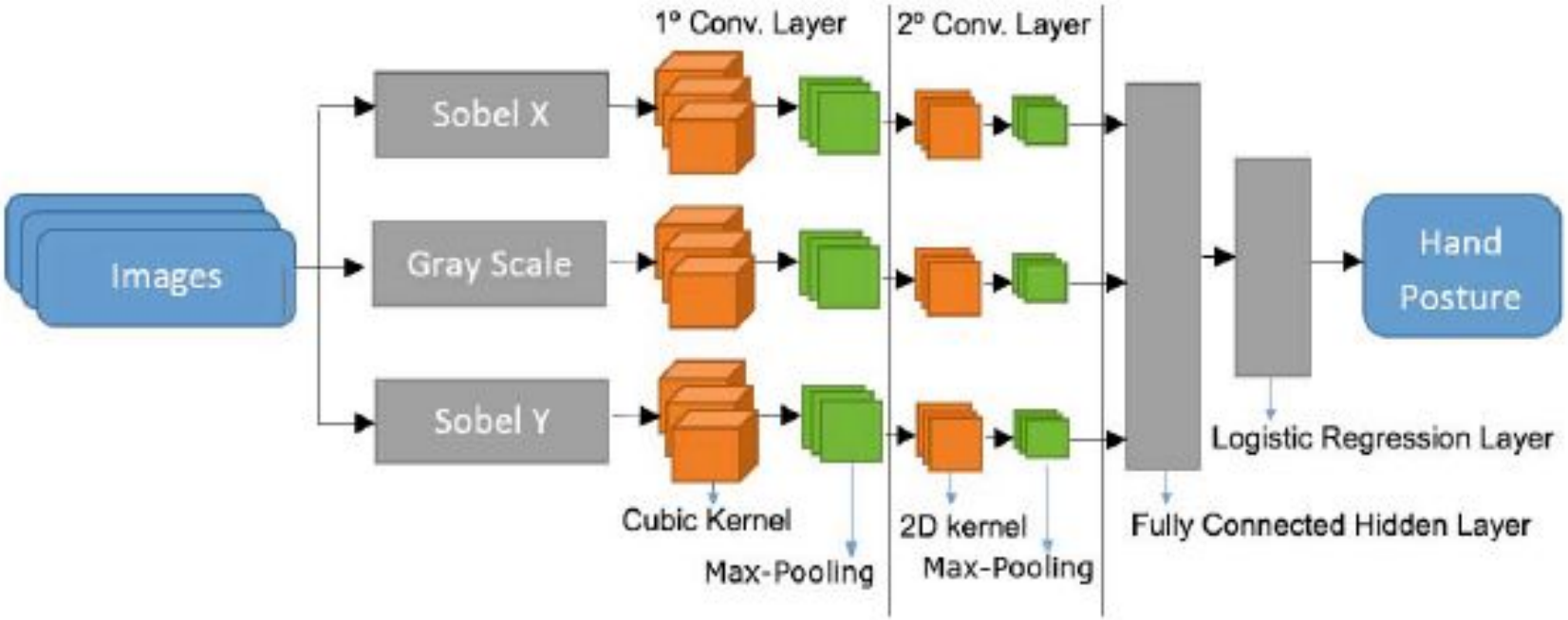


https://www.mdpi.com/2076-3417/9/18/3790/htm

Multi Channel CNN

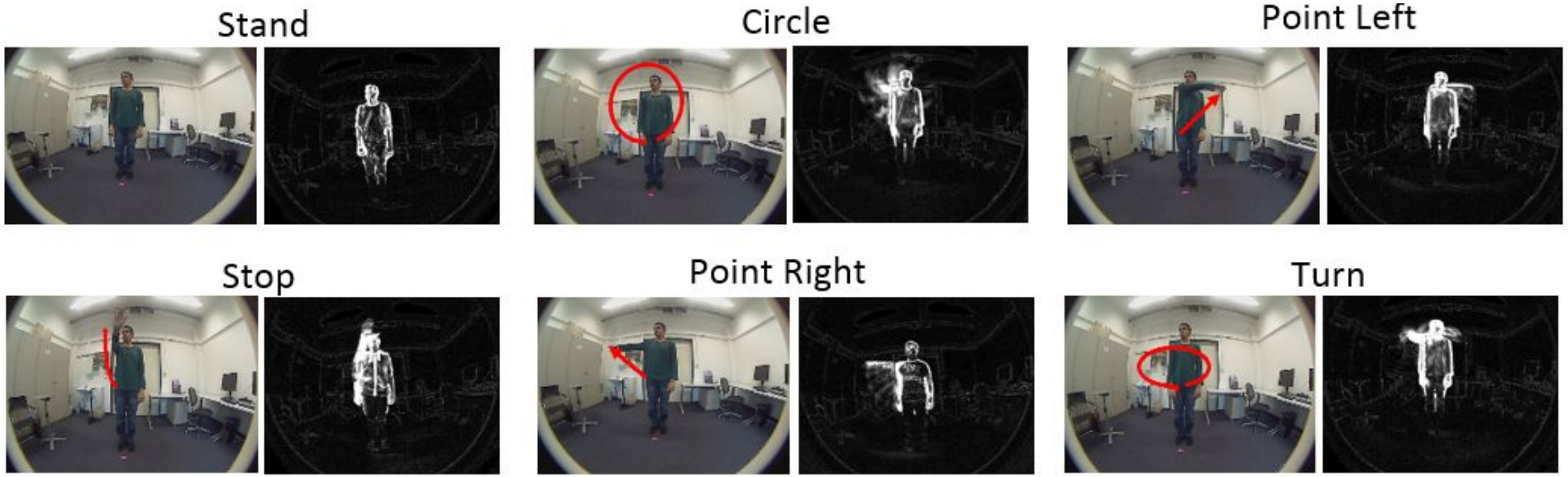
- Convolution with 3D kernels capturing motion information along the frames of an action stream, improves feature enhancement
- Uses multi channels to tune filters (Sobel operators)
  - The feature maps are created using different kernels to increase the diversity of features
- Instead of using single images for convolution, the whole computation is performed on a frame cube of predefined size (i.e. frames to consider in the video)

Multi Channel CNN



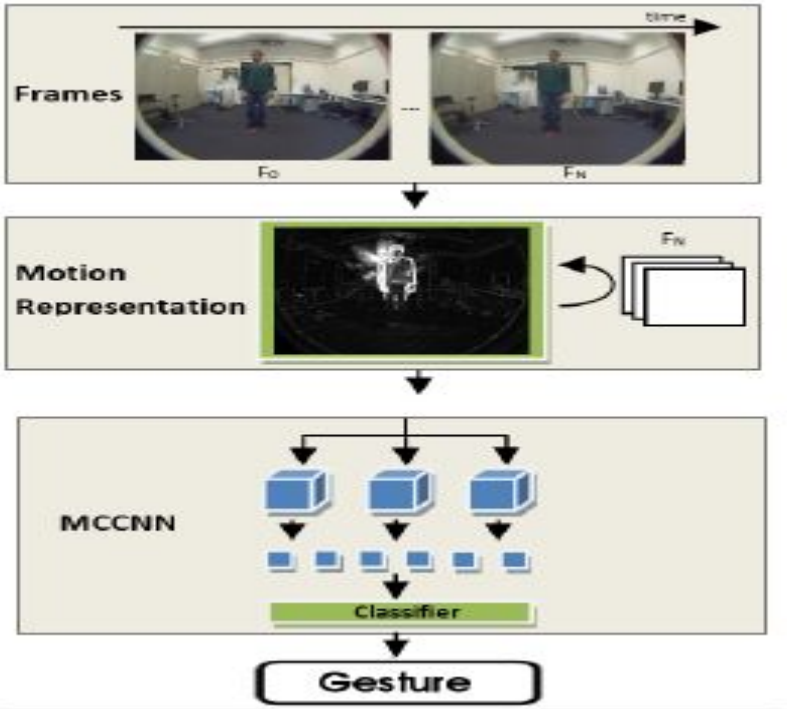
A Multichannel Convolutional Neural Network for Hand Posture Recognition [8]

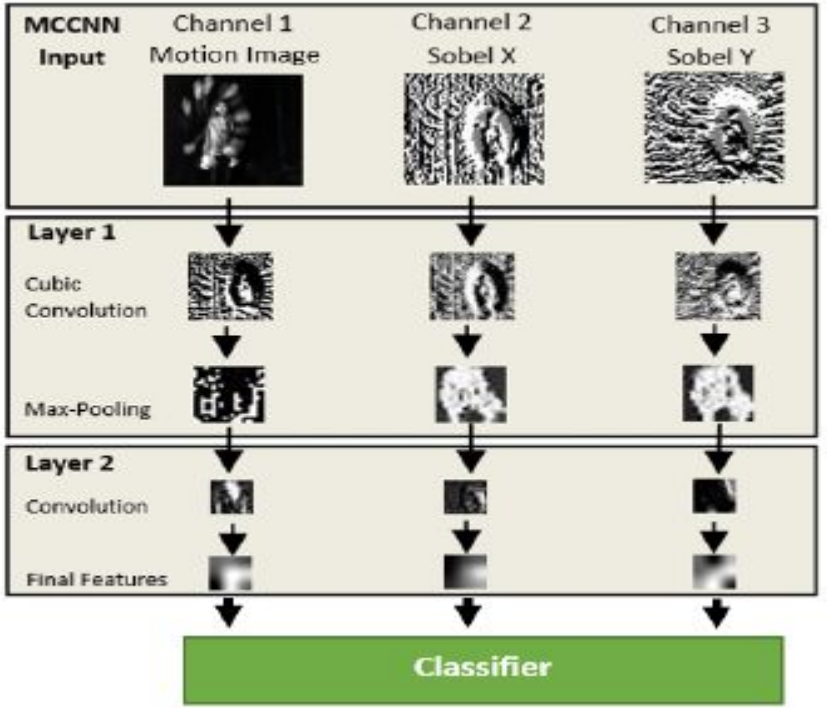
Experiment



A Multichannel Convolutional Neural Network for Hand Posture Recognition [8]

Gesture Recognition with MC-CNN



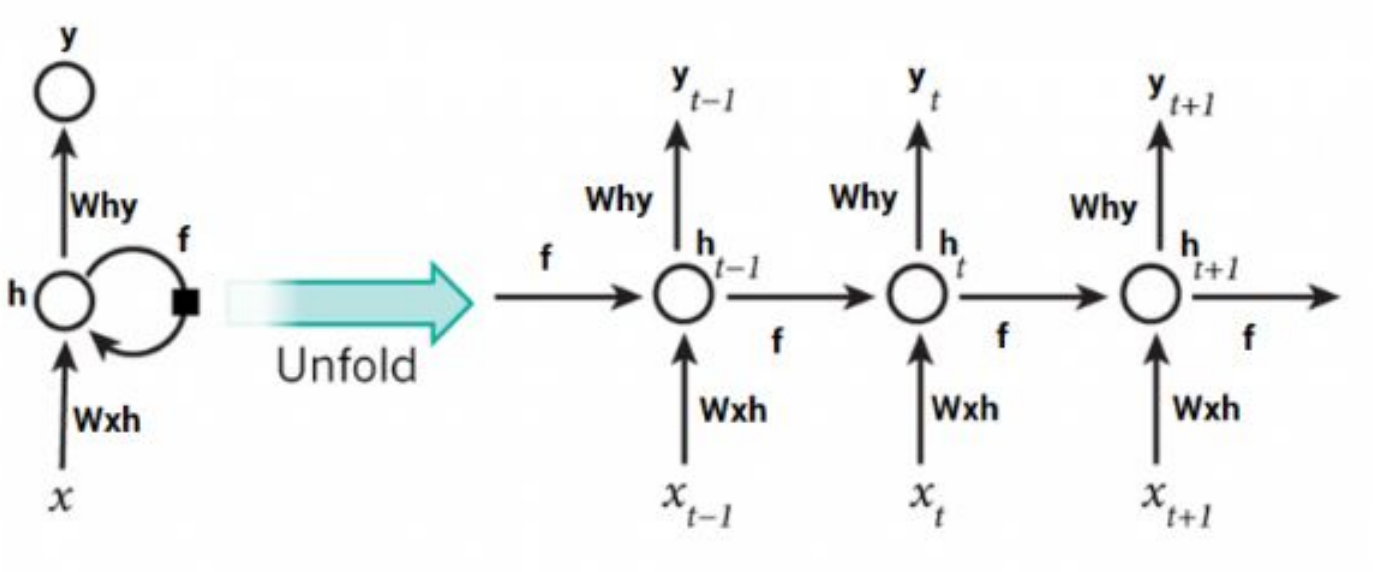


A Multichannel Convolutional Neural Network for Hand Posture Recognition [8]

CNN LSTM

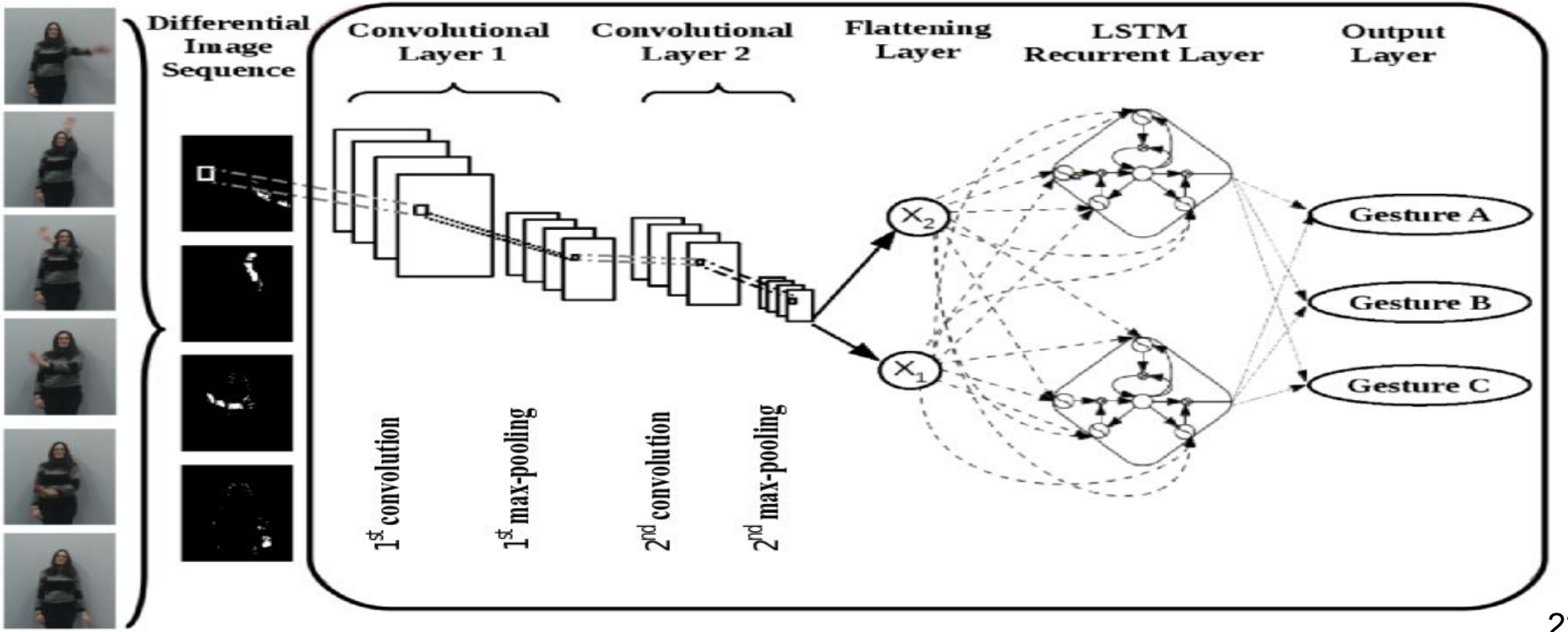
- CNN with Recurrent Neural Network (aka R CNN)
- Problem? lack of flexibility in learning sequences of different sizes
- Useful for dealing with long-range temporal dependencies
- Accordingly able to learn gestures varying in duration
- How? by the usage of Back Propagation Through Time (BPTT)

LSTM



https://www.analyticsvidhya.com/blog/2017/12/fundamentals-of-deep-learning-introduction-to-lstm/

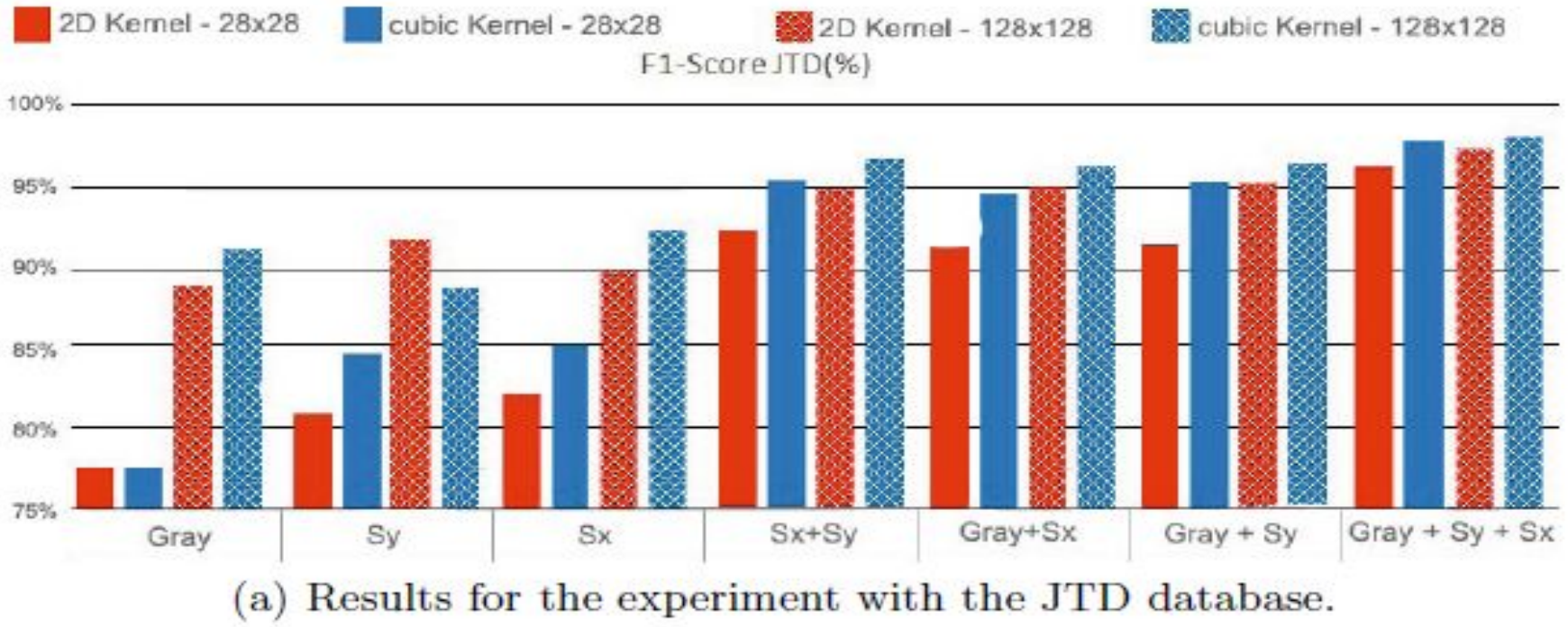
CNN with LSTM



MC-CNN Experiment & Results

- 2 datasets: JTD & NCD for hand postures
- 3 channels are used: raw image, horizontal and vertical Sobel filters
- Results for 1000 epochs were calculated
- F-1 score of 92% for JTD and 94% for NCD

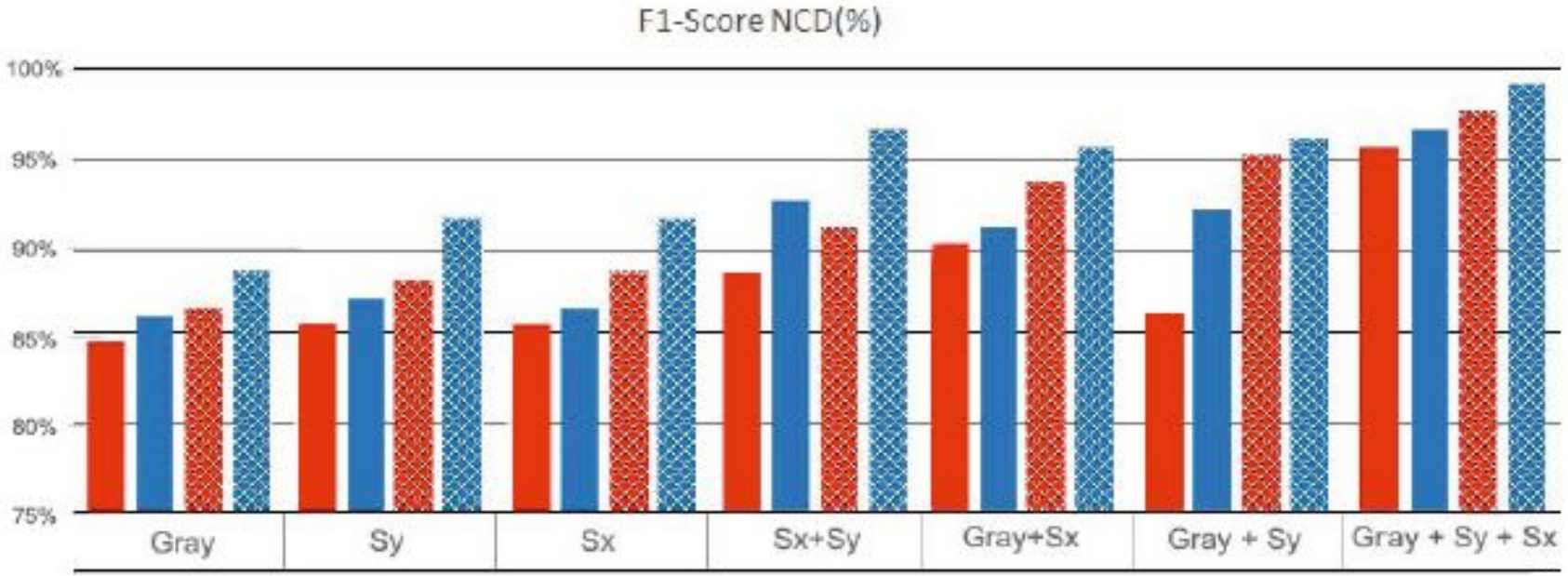
MC-CNN Experiment & Results



(a) Results for the experiment with the JTD database.

Gesture Recognition with a Convolutional Long Short-Term Memory Recurrent Neural Network [1]

MC-CNN Experiment & Results



(b) Results for the experiment with the NCD database.

Gesture Recognition with a Convolutional Long Short-Term Memory Recurrent Neural Network [1]

CNN-LSTM Experiment & Results

- TsironiGR-dataset, consists of 543 gesture sequences in total
- 9 different Human-Robot Interaction commands:
  - “abort”, “circle”, “hello”, “no”, “stop”,
  - “warn”, “turn left”, “turn” and “turn right”
- Each experiment was repeated five times

Model	Accuracy	Precision	Recall	F1-measure
CNN	77.78%±3.75%	79.87%±3.64%	77.78%±4.19%	76.56% ±4.27%
CNNLSTM	91.67%±1.13%	92.25% ±1.02%	91.67%±1.13%	91.63% ±1.15%

Gesture Recognition with a Convolutional Long Short-Term Memory Recurrent Neural Network [1]

Conclusion & Future

- CNN can be quite effective in Gesture Recognition tasks
- Research further CNN architectures for Gesture Recognition
  - Ex: Gated shape CNN, Max Pooling CNN
- Experiment mentioned architectures on facial expression datasets?
- Try Spatial Transformer Networks?
- What to teach robots using machine learning?

Thank you for your attention!

Questions?

References

1. Eleni Tsironi, Pablo Barros and Stefan Wermter, ”Gesture Recognition with a Convolutional Long Short-Term Memory Recurrent Neural Network”, Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN), pp. 213-218,Bruges, Belgium (2016)
2. Waseem Rawat, Zenghui Wang, Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review, Neural Computation 29, 2352–2449 (2017)
3. G. R. S. Murthy & R. S. Jadon, A review of vision based hand gestures recognition, International Journal of Information Technology and Knowledge Management, July-December 2009, Volume 2, No. 2, pp. 405-410
4. Pablo Barros, German I. Parisi, Doreen Jirak and Stefan Wermter, Real-time Gesture Recognition Using a Humanoid Robot with a Deep Neural Architecture, 2014 14th IEEE-RAS International Conference on Humanoid Robots (Humanoids) November 18-20, 2014. Madrid, Spain
5. Pramod Pisharady, Martin Saerbeck, Recent methods and databases in vision-based hand gesture recognition: A review, ElSevier 2015
6. Albert Clapes, Marco Bellantonio, Hugo Jair Escalante, Vıctor Ponce-Lopez, Xavier Baro, Isabelle Guyon, Shohreh Kasaei, Sergio Escalera, A survey on deep learning based approaches for action and gesture recognition in image sequences, 2017 IEEE 12th International Conference on Automatic Face & Gesture Recognition
7. Hongyi Liu, Lihui Wang, Gesture recognition for human-robot collaboration: A review, ElSevier 2017
8. Barros P., Magg S., Weber C., Wermter S. (2014) A Multichannel Convolutional Neural Network for Hand Posture Recognition. In: Wermter S. et al. (eds) Artificial Neural Networks and Machine Learning – ICANN 2014. ICANN 2014. Lecture Notes in Computer Science, vol 8681. Springer, Cham