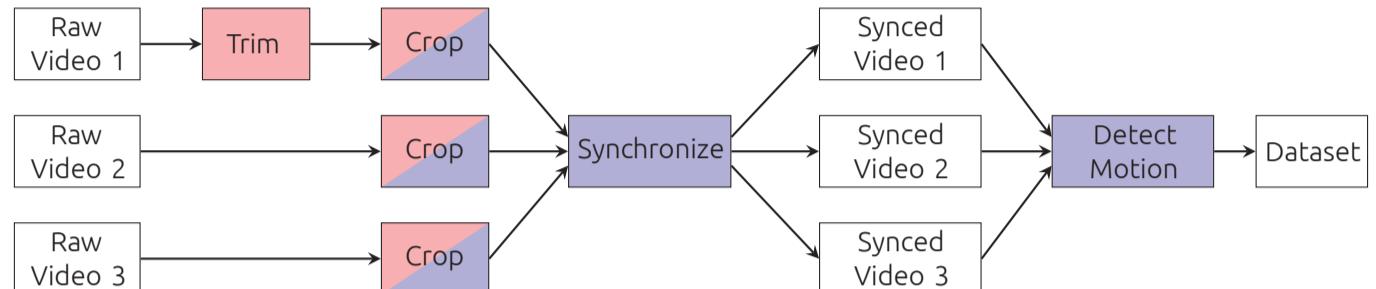


# Self-Supervised Learning for Recognition of Sports Poses in Image

Author: Bc. Daniel Konečný  
Supervisor: prof. Ing. Adam Herout, PhD.

## Dataset Preparation Tools

- Scene filmed from different viewpoints.
- Manual and Automatic operations.
- Video is automatically cropped and scaled to match the input size of the Encoder after a bounding box is manually chosen.
- Dense Optical Flow for synchronization.
- Sparse Optical Flow for motion detection.



## Datasets

- Hand Dataset:
  - Movement of wrist and fingers.
  - For development of Dataset Preparation Tools and Encoder.



- Upper Body Dataset:
  - Movement of arms in shoulders and elbows.
  - For encoding and classifying of sports poses.
  - 2 sets of labels:
    - Directions (4 classes),
    - Bent (16 classes).



## Encoder & Classifier

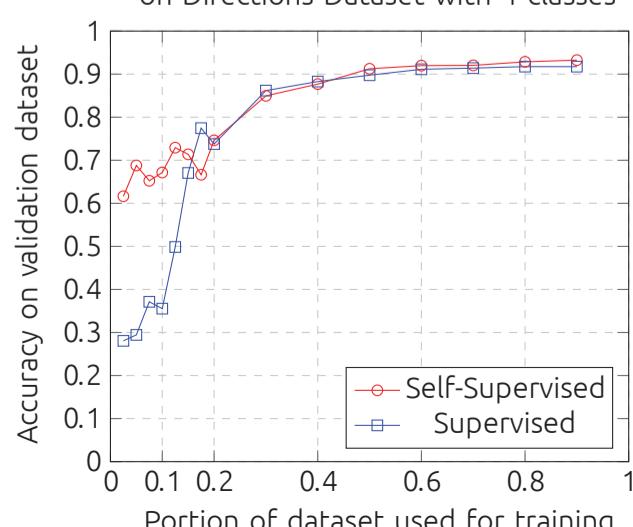
- Architecture of the Encoder and the Classifier in comparison to a normal supervised model.
- Number of parameters that need to be trained for different model architectures.

Model	Self-Supervised Encoder (fit)	Self-Supervised Encoder (fine-tune)	Self-Supervised Classifier	Supervised
Trained Parameters	131,136	23,665,728	4,420	23,542,788

Self-Supervised Model		Supervised Model	
Description	Layer – Shape	Layer – Shape	Description
Image	Input – (224, 224, 3)	Input – (224, 224, 3)	Image
Backbone ResNet50	Padding – (230, 230, 3) Pooling – (2048) Dense – (64)	Padding – (230, 230, 3) pooling – (2048) Dense – (4)	Backbone ResNet50
Embedding	L2 Normalize – (64) Dense – (64)	Dense – (4)	Label
Label	Dense – (4)		

## Self-Supervised vs. Supervised

Comparison of self-supervised and supervised models' accuracy on Directions Dataset with 4 classes



Comparison of self-supervised and supervised models' accuracy on Bent Dataset with 16 classes

