University of Stavanger

Activity Recognition

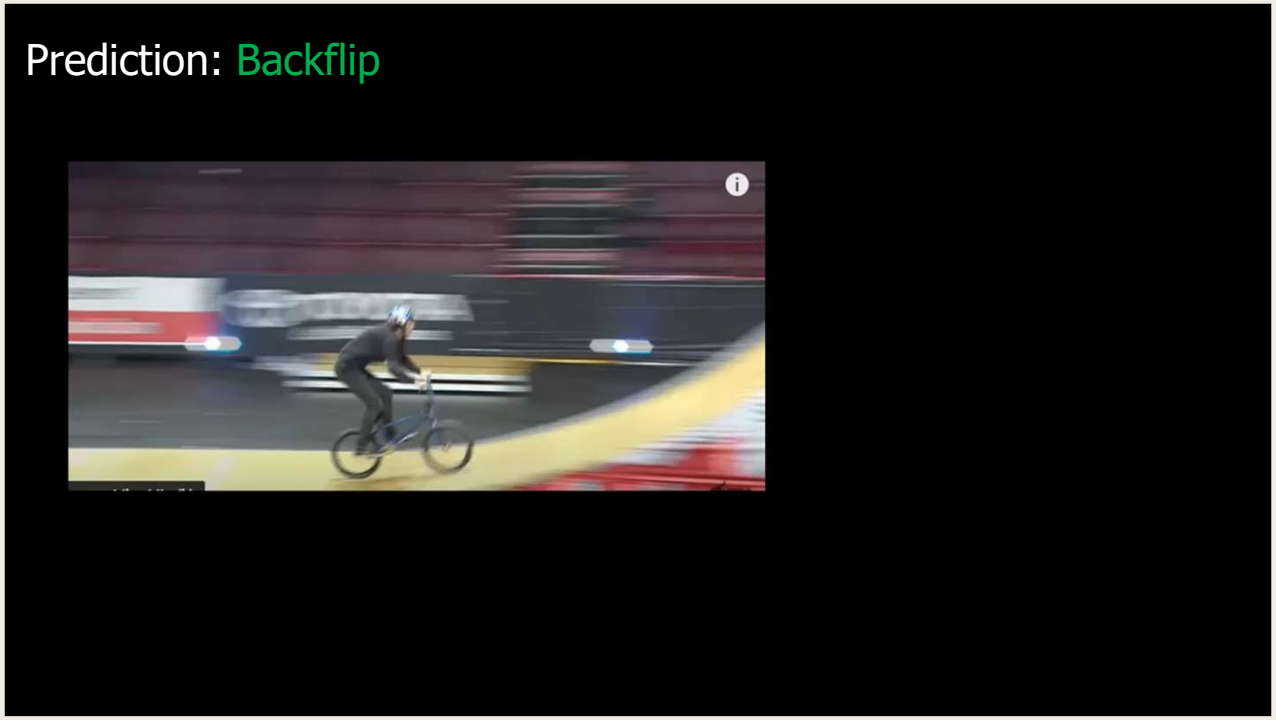
Øyvind Meinich-Bache
Associate Professor II

Activity Recognition – frame rate

Analyze a single frame from a video

Activity Recognition – frame rate

Analyze all frames in a video (high fps)



Activity Recognition – frame rate

Analyze few frames in a video

Prediction: Jumping

Activity Recognition

Activity classification/detection – each decision is based on analysing a single frame

Activity recognition - decision based on analysing sequential frames (video)

Activity classification

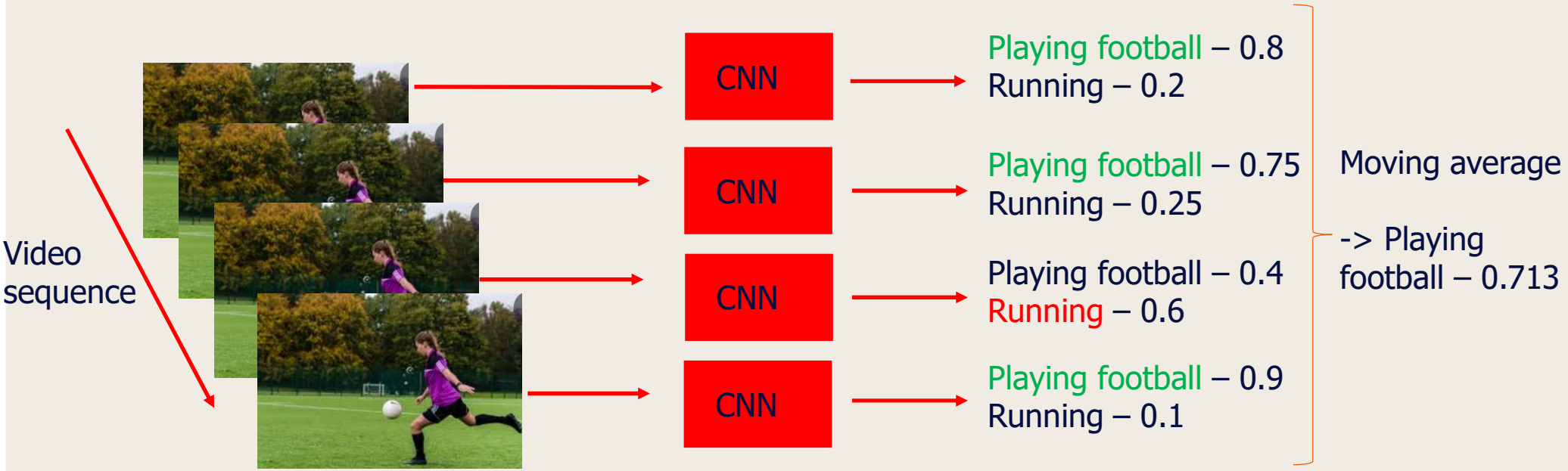
Prediction: Playing football

Prediction: Running

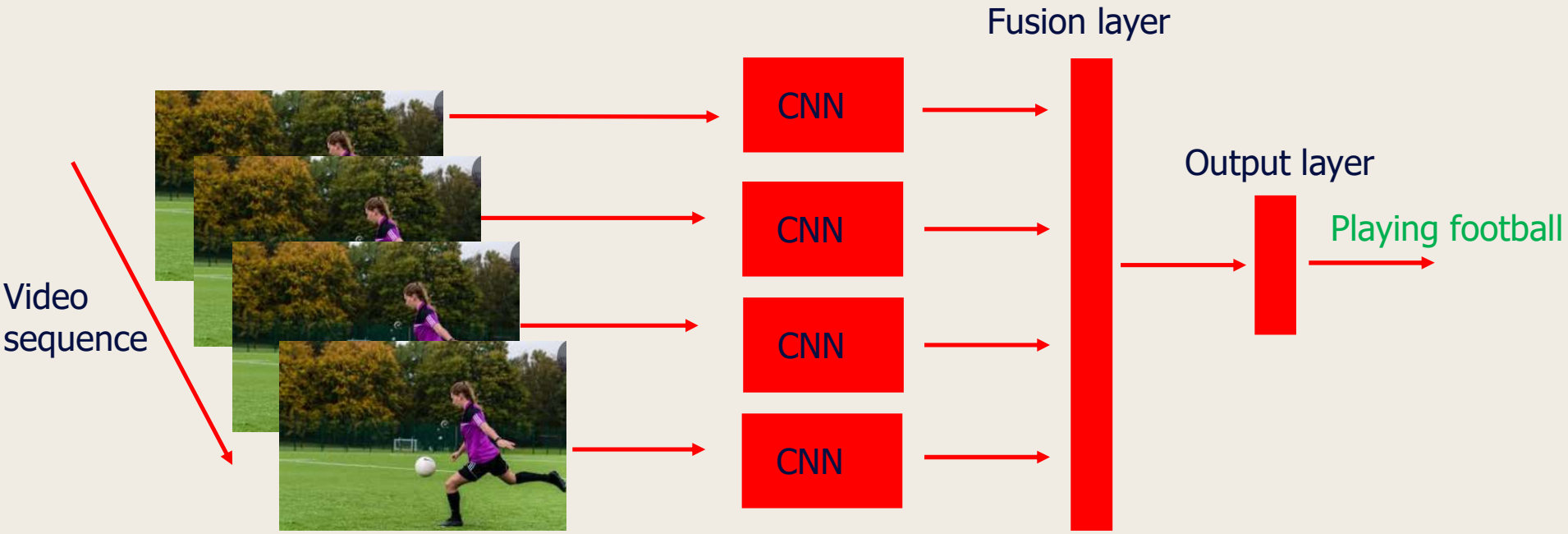
Enough image examples during training: Model can learn do distinguish between similar activities by learning the environmental context

-> Image classifier could be enough for many cases
- Not activity **recogniton** – but is very efficient

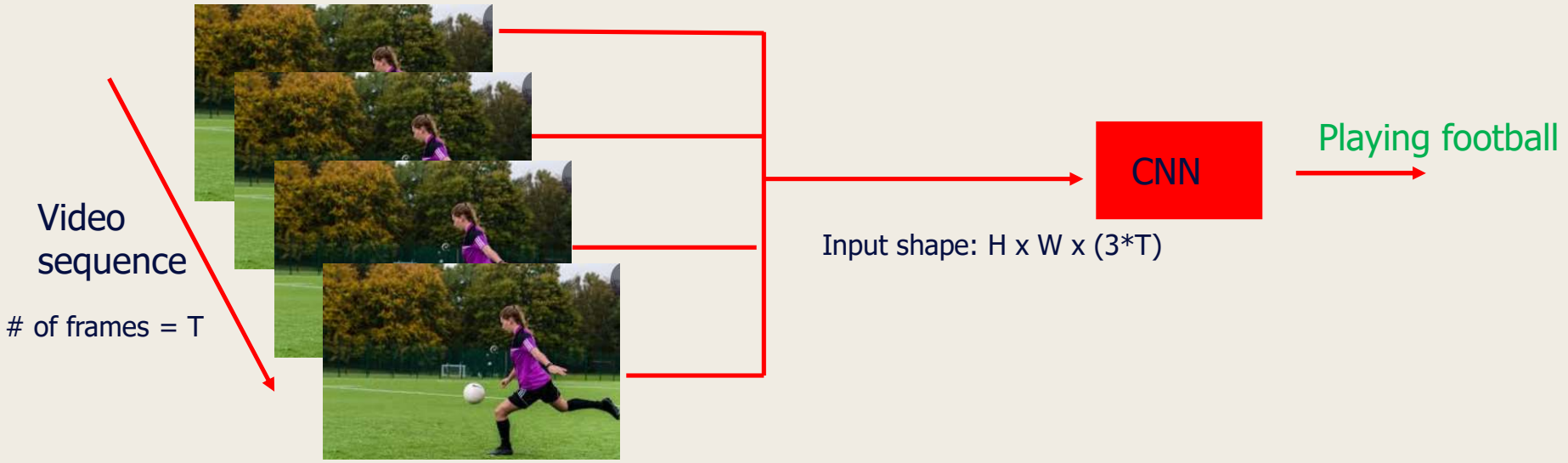
Method 1 – Single frame CNN



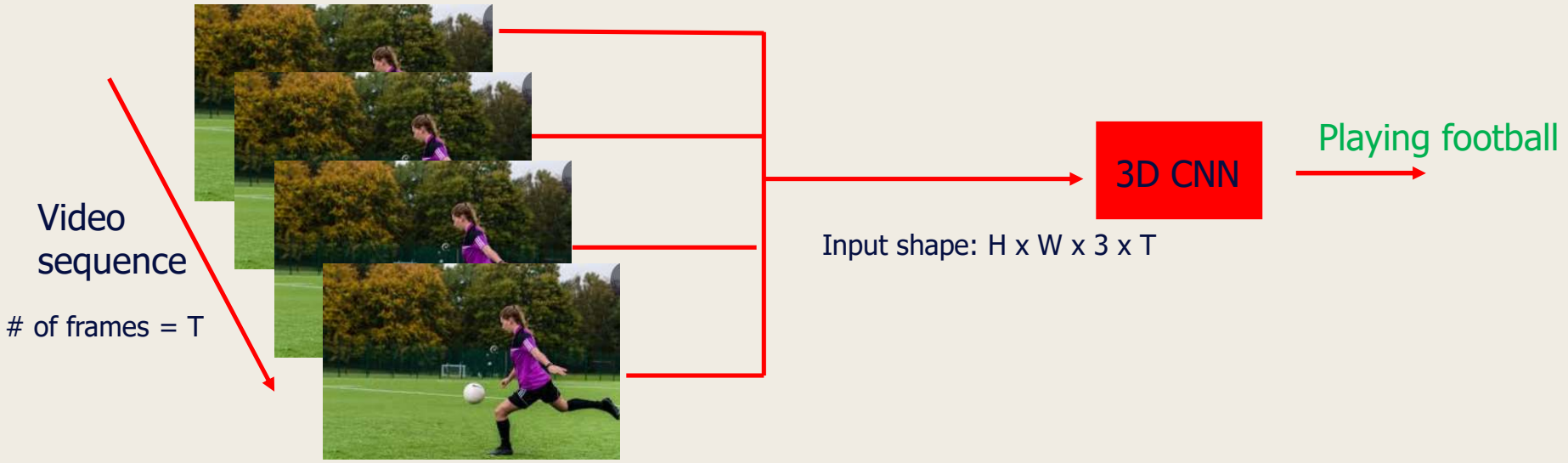
Method 2: Late fusion



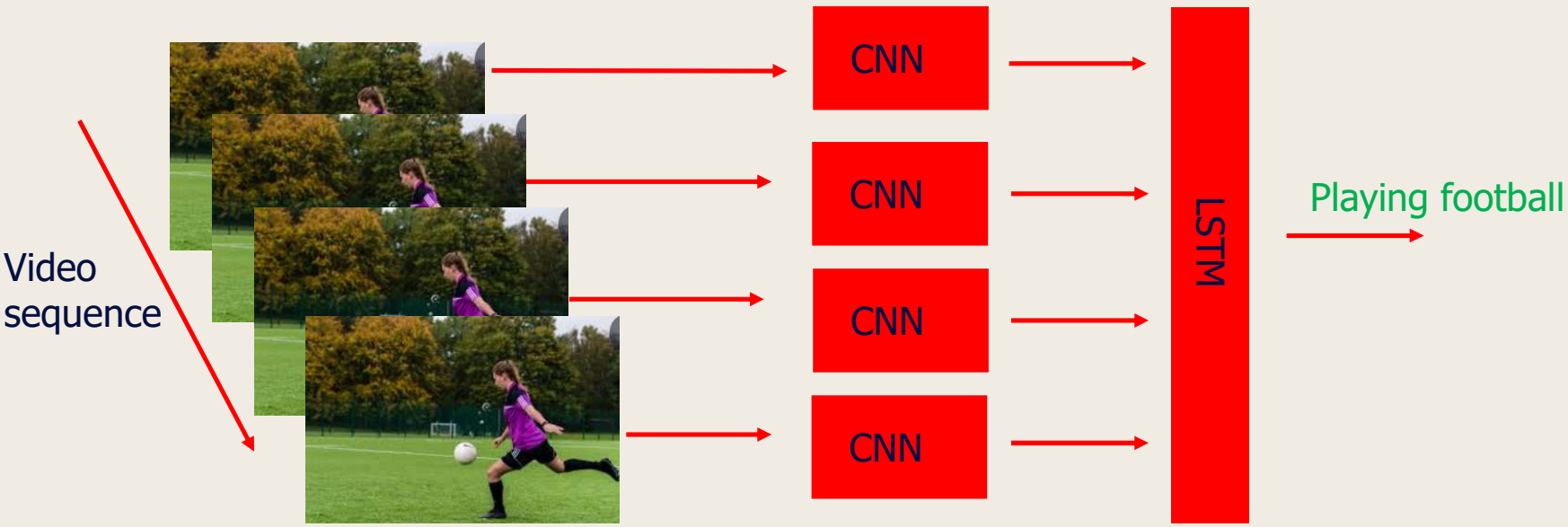
Method 3: Early fusion



Method 4 – Slow fusion



Method 5: CNN + LSTM



Method 6 – RGB + Optical flow – Two stream CNN

3D CNN

Input shape: H x W x 3 x T

Fusion layer: Combine feature vectors

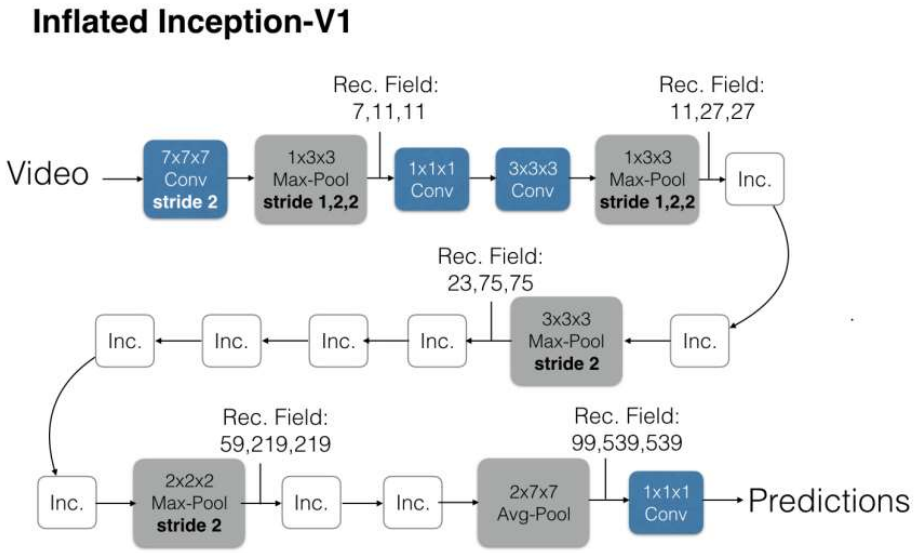
Output layer

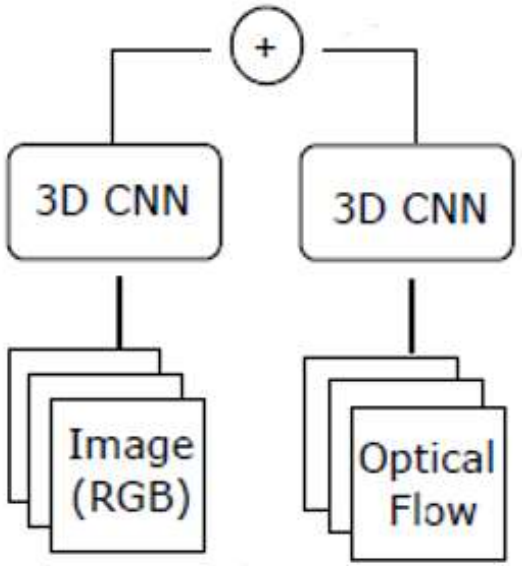
3D CNN

Input shape: H x W x 3 x T

Optical flow: Motion representation.
Generated by differences between sequential image frames

Two stream 3D CNN – Inception 3D (I3D)





- 3D inflated version of the 2D CNN Inception v1
  - Weight in 2D filters are repeated
  - Filters are normalized
- Inception v1 weights are pretrained on imagenet before inflation
- I3D further trained (pretrained) on Kinetics dataset (video) after inflation
  - Flow model + RGB model
  - Pretrained models made available

I3D paper: https://arxiv.org/pdf/1705.07750.pdf

Code by creators: https://github.com/deepmind/kinetics-i3d

Datasets

UCF101: https://www.crcv.ucf.edu/data/UCF101.php
- 101 classes
- The activity categories can be divided into five types: 1)Human-Object Interaction 2) Body-Motion Only 3) Human-Human Interaction 4) Playing Musical Instruments 5) Sports.
- Download dataset:
  - https://www.crcv.ucf.edu/data/UCF101/UCF101.rar

Kinetics 400/600/700:
- A collection of large-scale, high-quality datasets of URL links of up to 650,000 video clips that cover 400/600/700 human action classes, depending on the dataset version
- The videos include human-object interactions such as playing instruments, as well as human-human interactions such as shaking hands and hugging. Each action class has at least 400/600/700 video clips. Each clip is human annotated with a single action class and lasts around 10 seconds.
- Download dataset and papers:
  - https://deepmind.com/research/open-source/kinetics

Activity recognition in newborn resuscitation videos