UniGe | MaLGa

Human in the image

Computational Vision, 09/05/2025

Matteo Moro

UniGe | MaLGa

Action classification – problem definition

Problem definition

Problem: Identify the action happening in a video clip

Challenges:

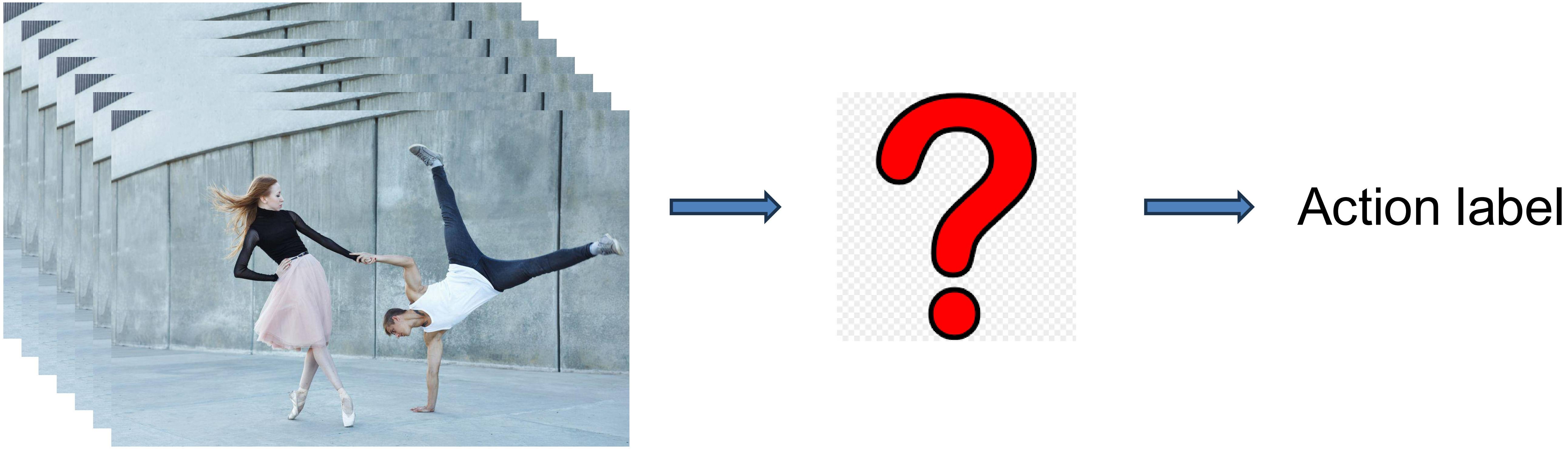
1) Number of people involved

2) Where the action is happening?

3) Background / context information

Problem definition

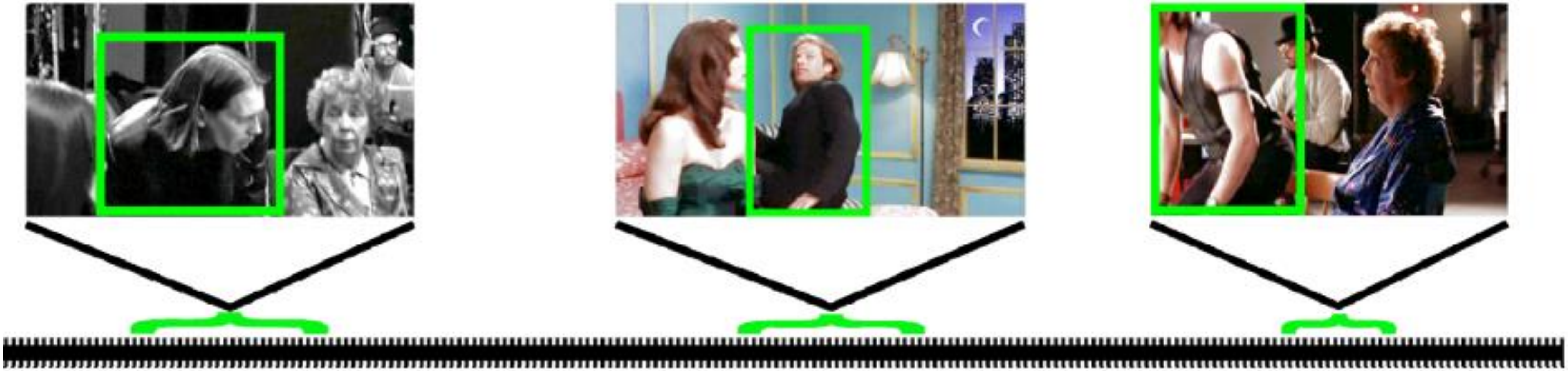
video



Action label

ZOTMAN
pizza pie

Action localization



Before the deep learning era

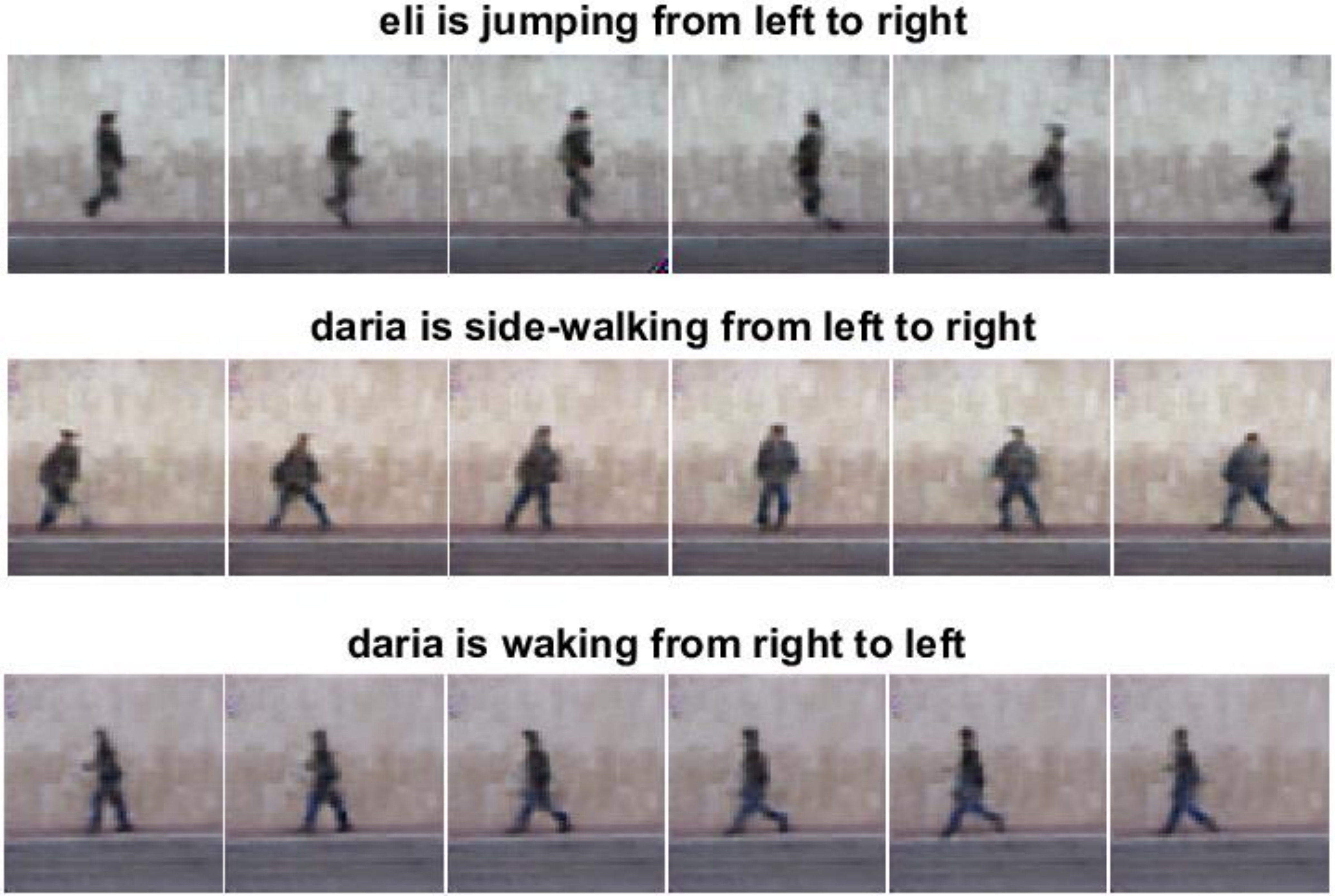
KTH Action dataset

hand waving

boxing

Schuldt, Christian, Ivan Laptev, and Barbara Caputo. "Recognizing human actions: a local SVM approach." Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.. Vol. 3. IEEE, 2004.

Weizmann dataset



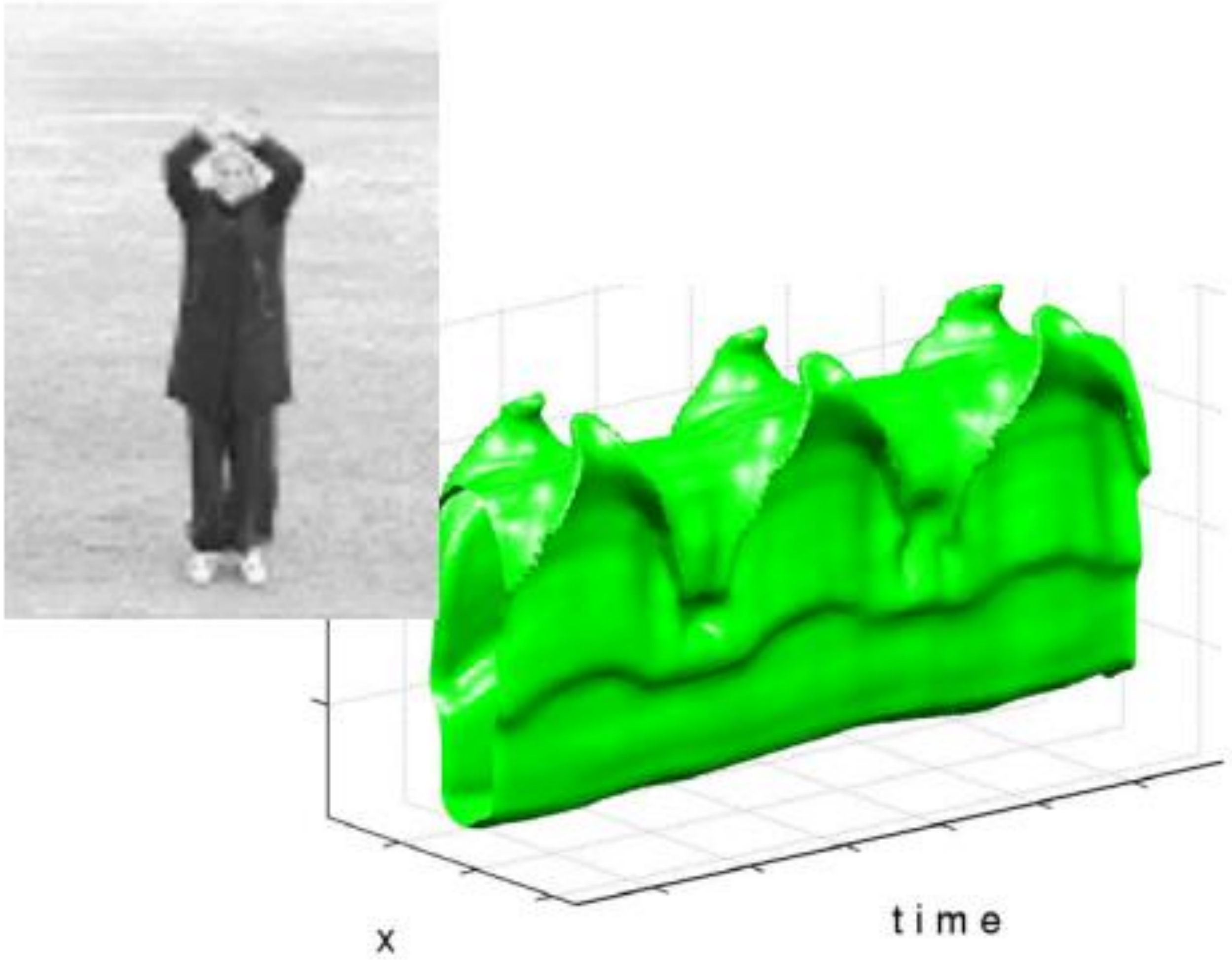
eli is jumping from left to right

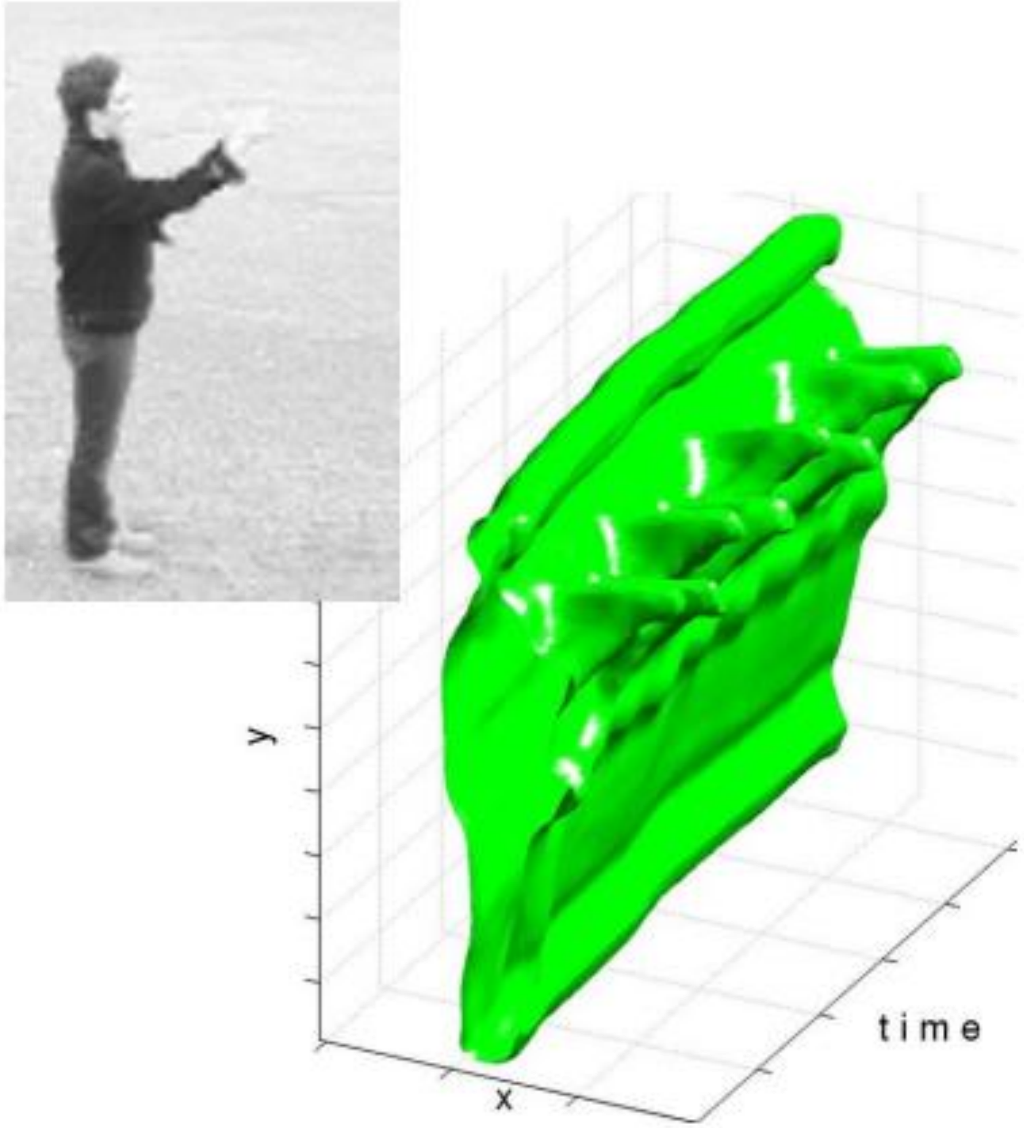
daria is side-walking from left to right

daria is waking from right to left

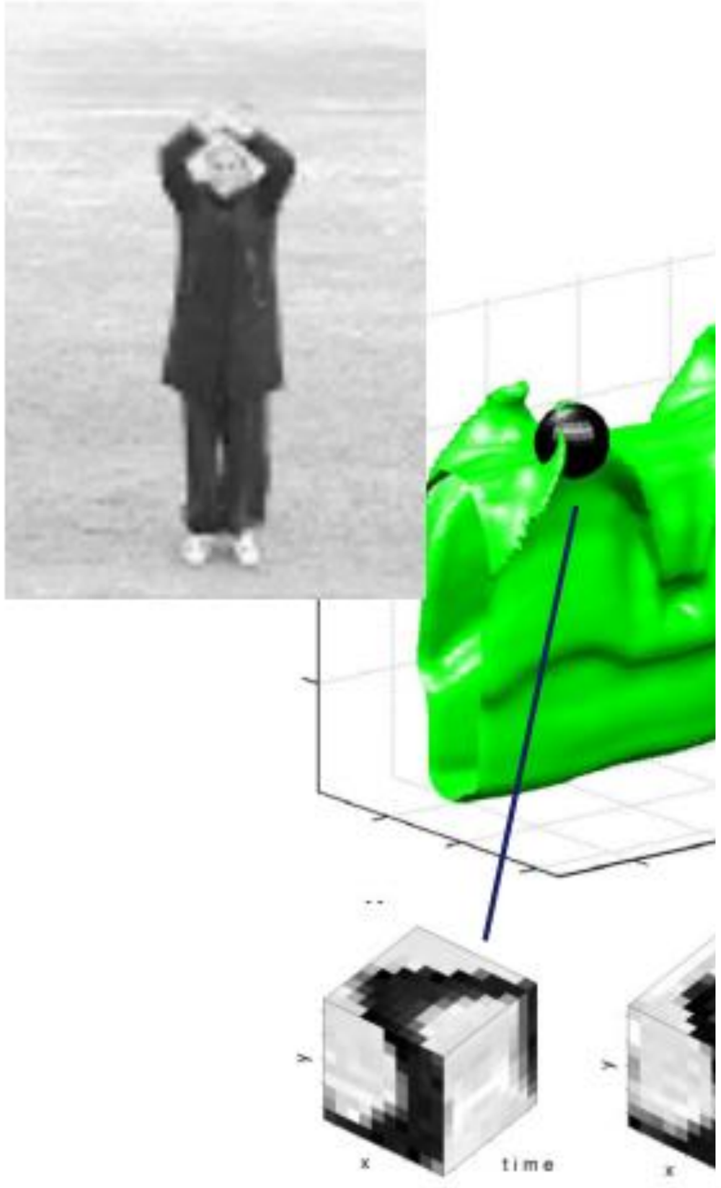
http://www.wisdom.weizmann.ac.il/~vision/SpaceTimeActions.html

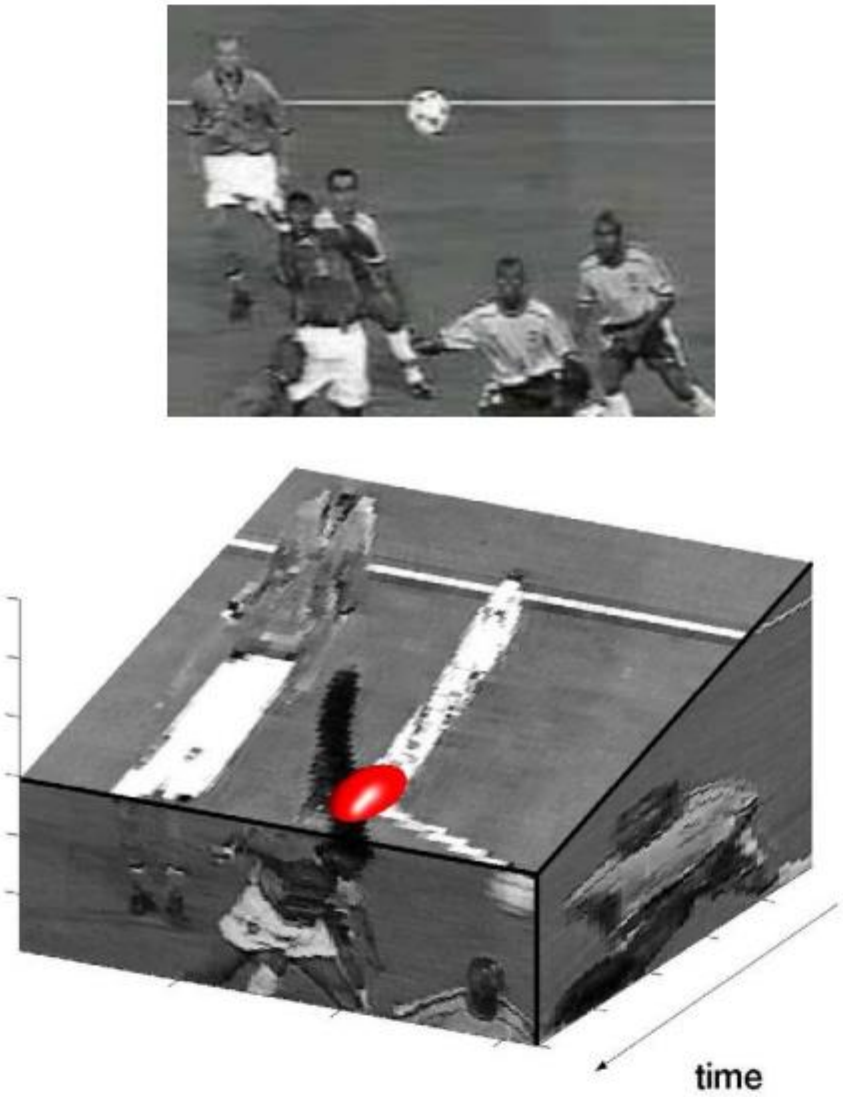
Actions == space-time objects

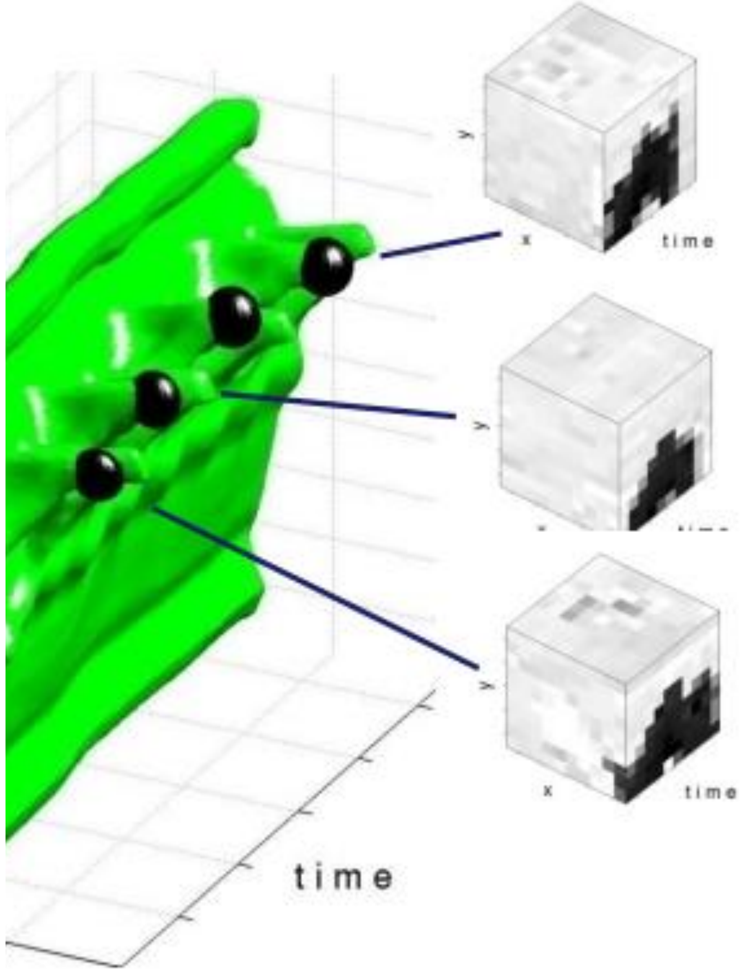




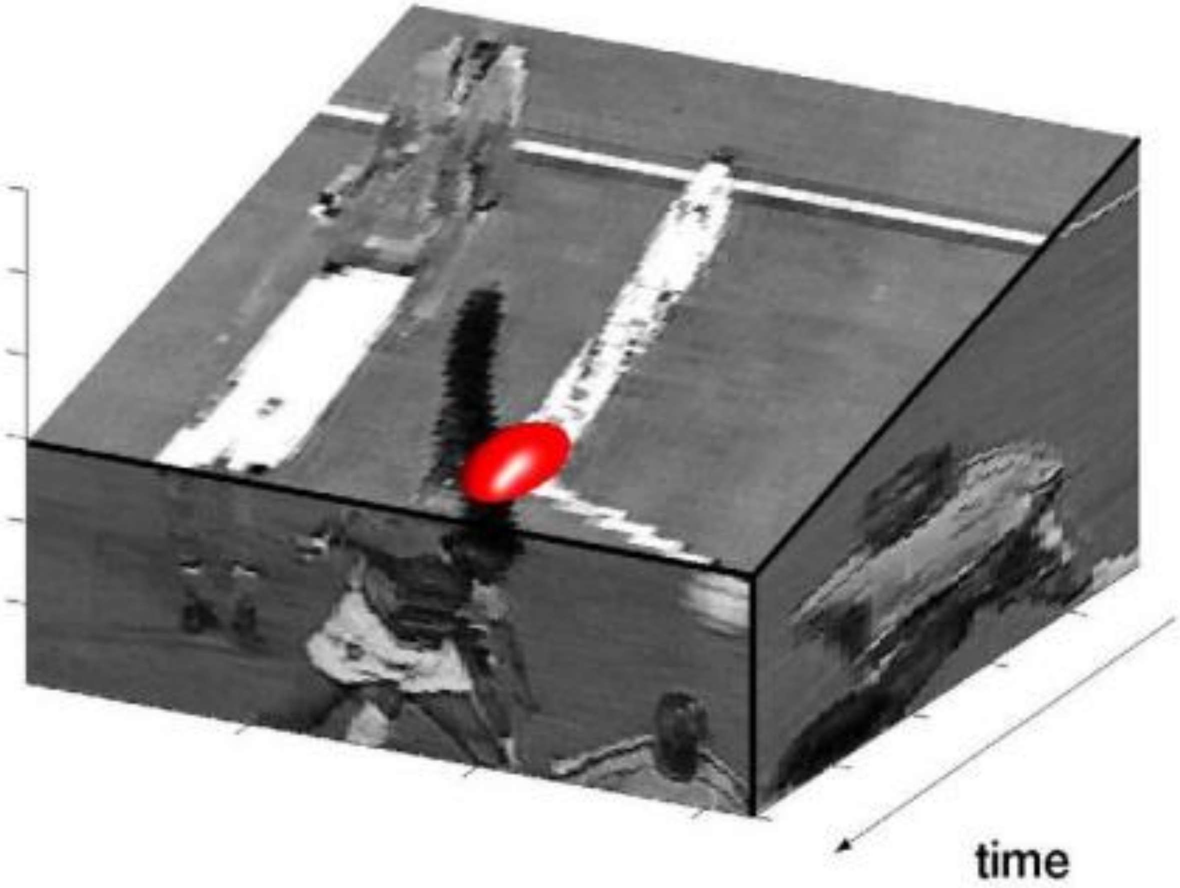
Local features



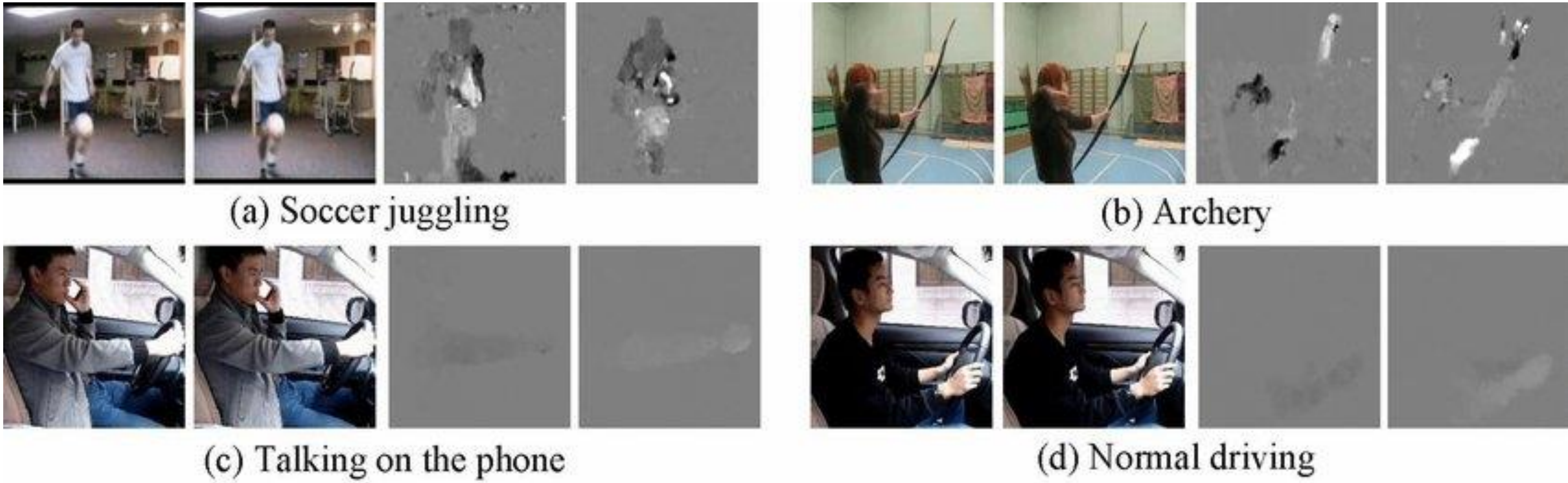




Local features



Sparse vs dense features



(a) Soccer juggling

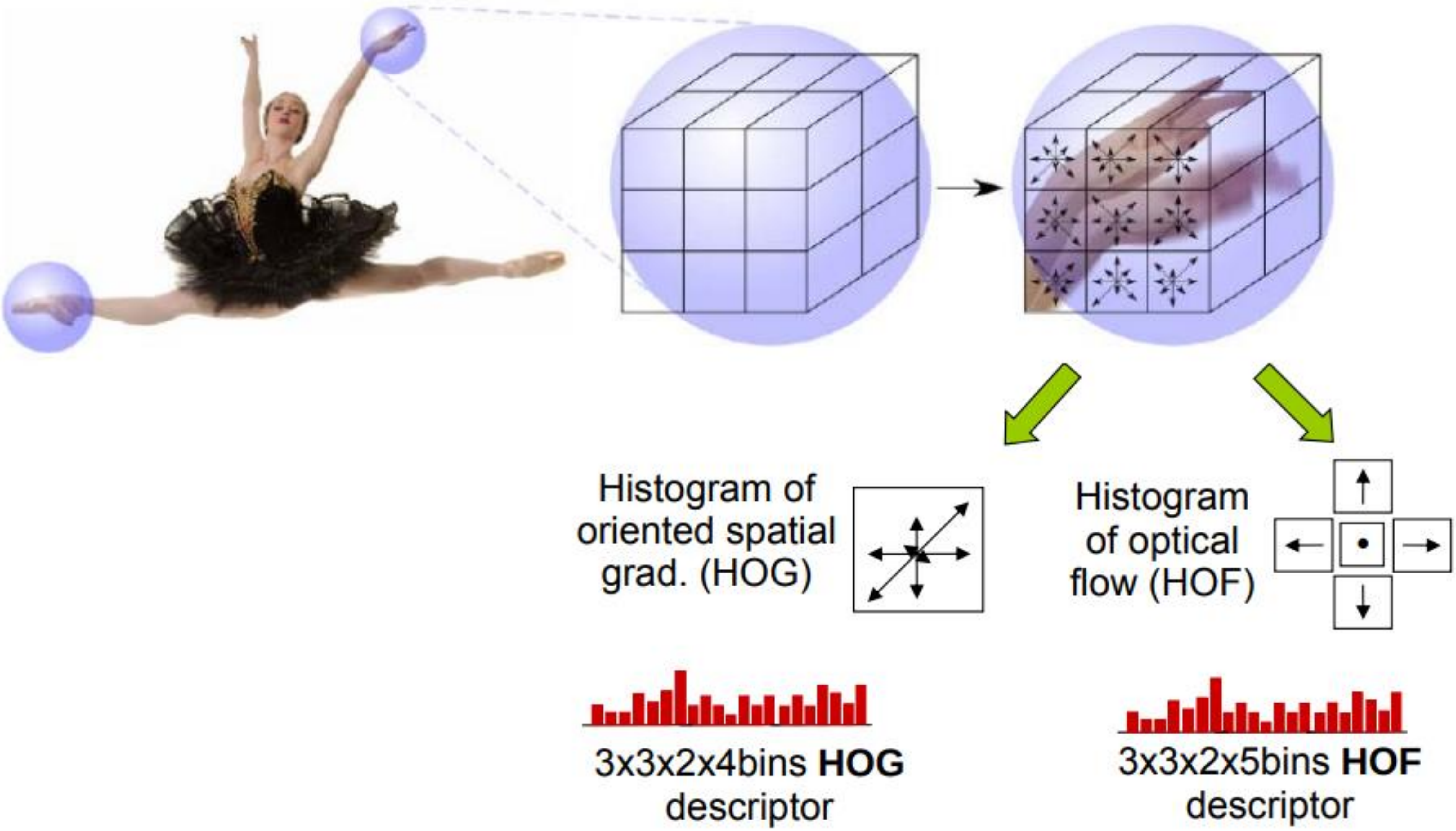
(b) Archery

(c) Talking on the phone

(d) Normal driving

Space-time descriptors

Multi-scale space-time patches



Histogram of oriented spatial grad. (HOG)

Histogram of optical flow (HOF)

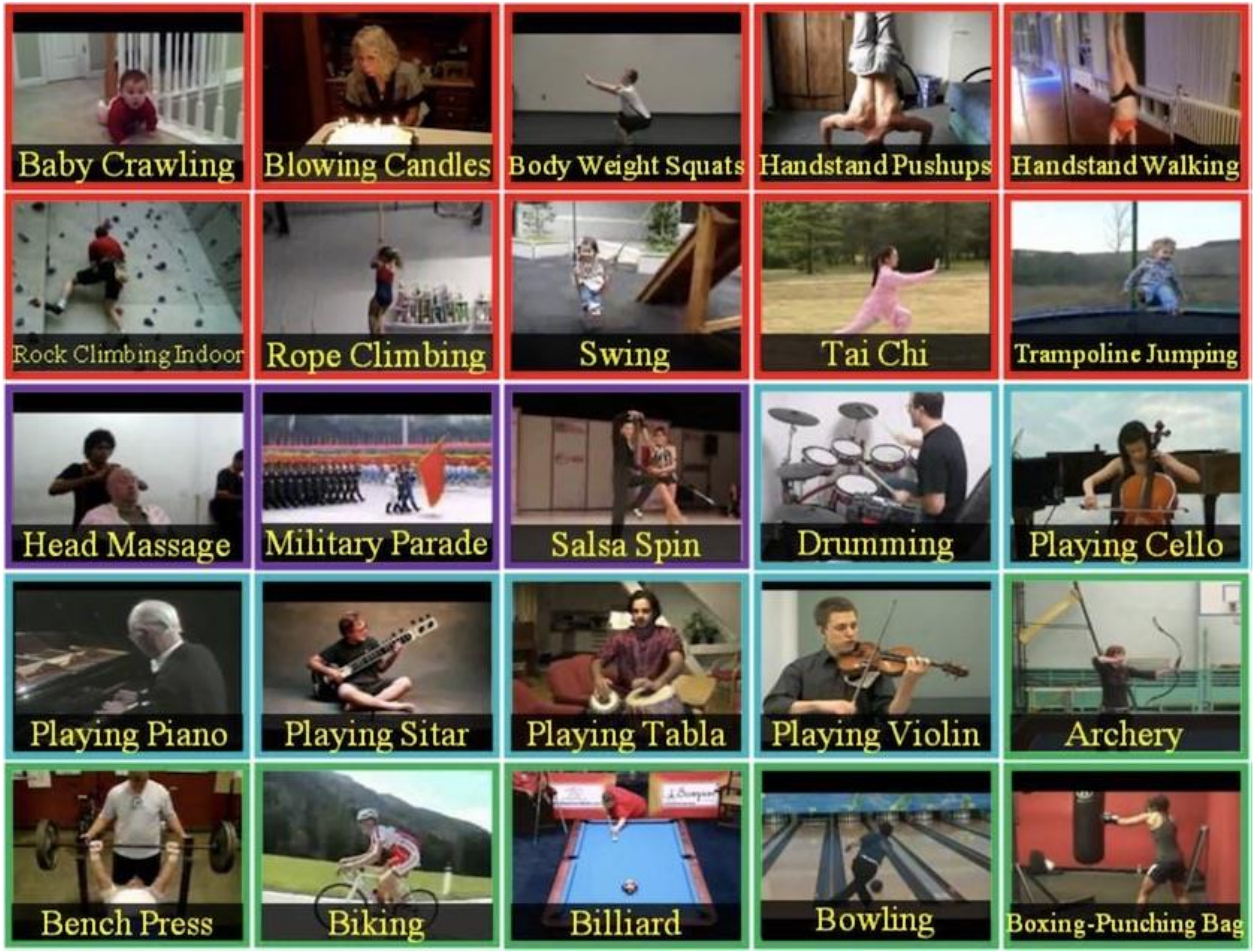
3x3x2x4bins **HOG** descriptor

3x3x2x5bins **HOF** descriptor

UniGe | MaLGa

Deep Learning era

UCF101



Soomro, Khurram, Amir Roshan Zamir, and Mubarak Shah. "UCF101: A dataset of 101 human actions classes from videos in the wild." arXiv preprint arXiv:1212.0402 (2012).

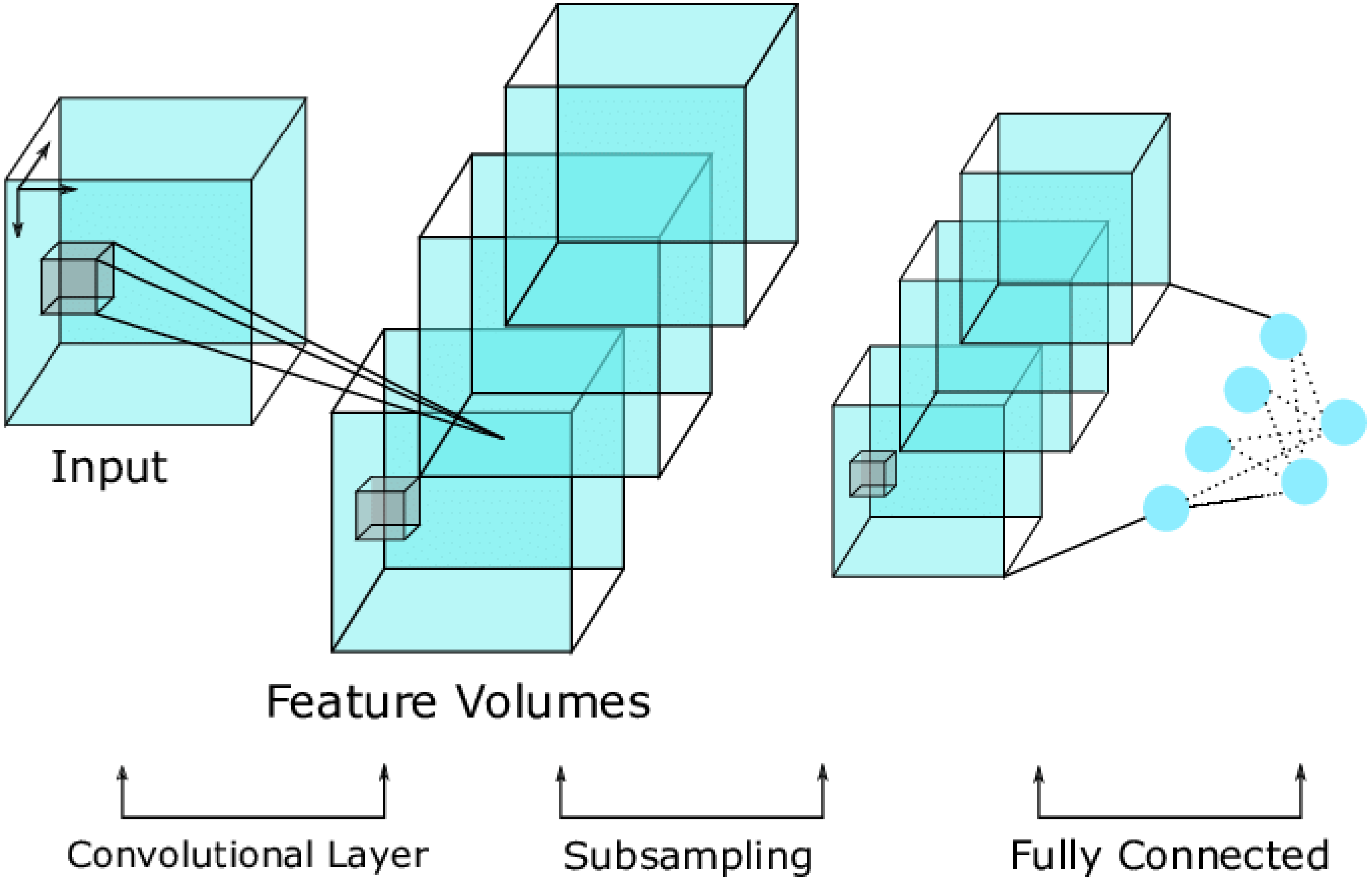
HMDB51

Brush air

Kick

Kuehne, Hildegard, et al. "HMDB: a large video database for human motion recognition." 2011 International conference on computer vision. IEEE, 2011.

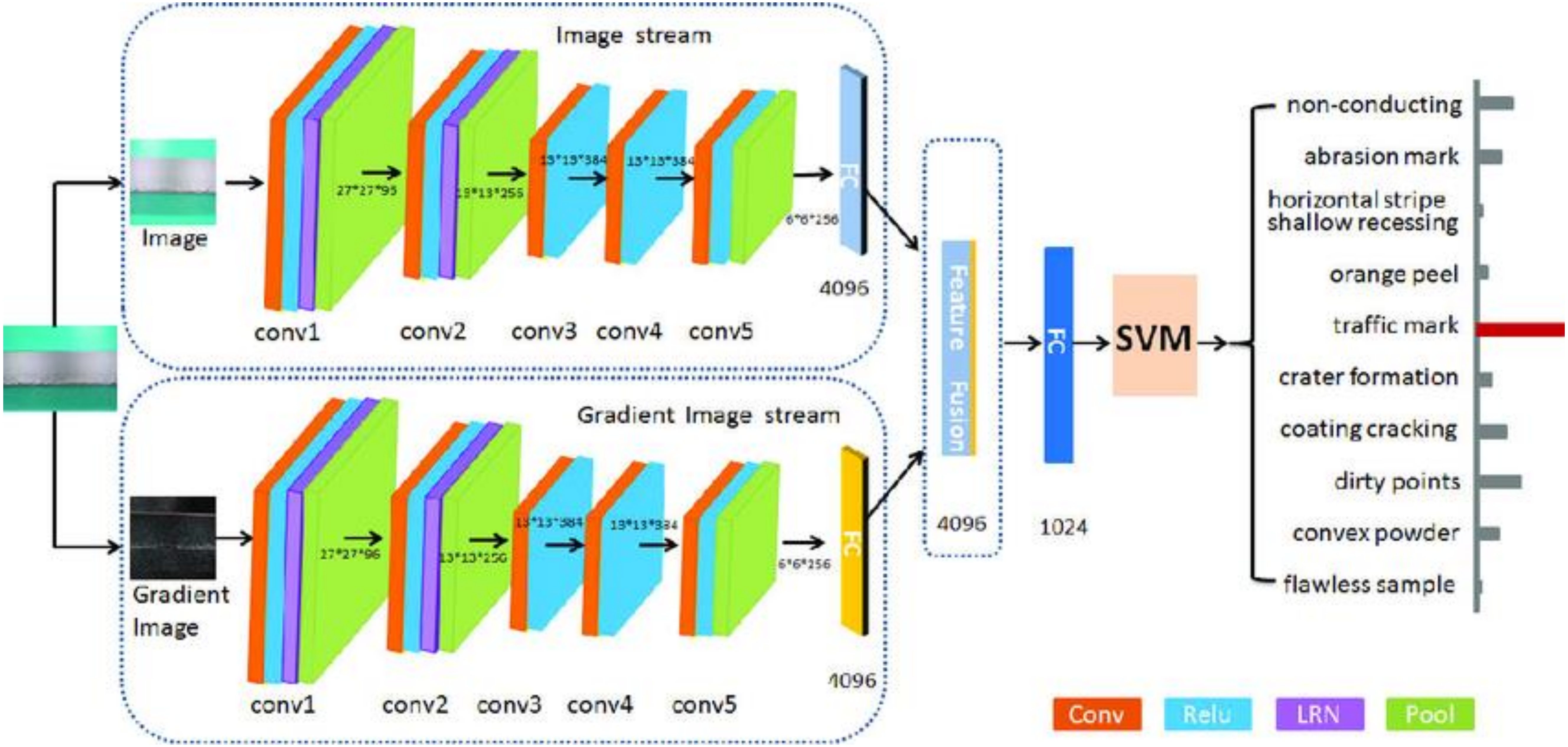
Spatio-temporal (3D) Convolutional Neural Networks



Carreira, Joao, and Andrew Zisserman. "Quo vadis, action recognition? a new model and the kinetics dataset." proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017.

Two-stream Networks

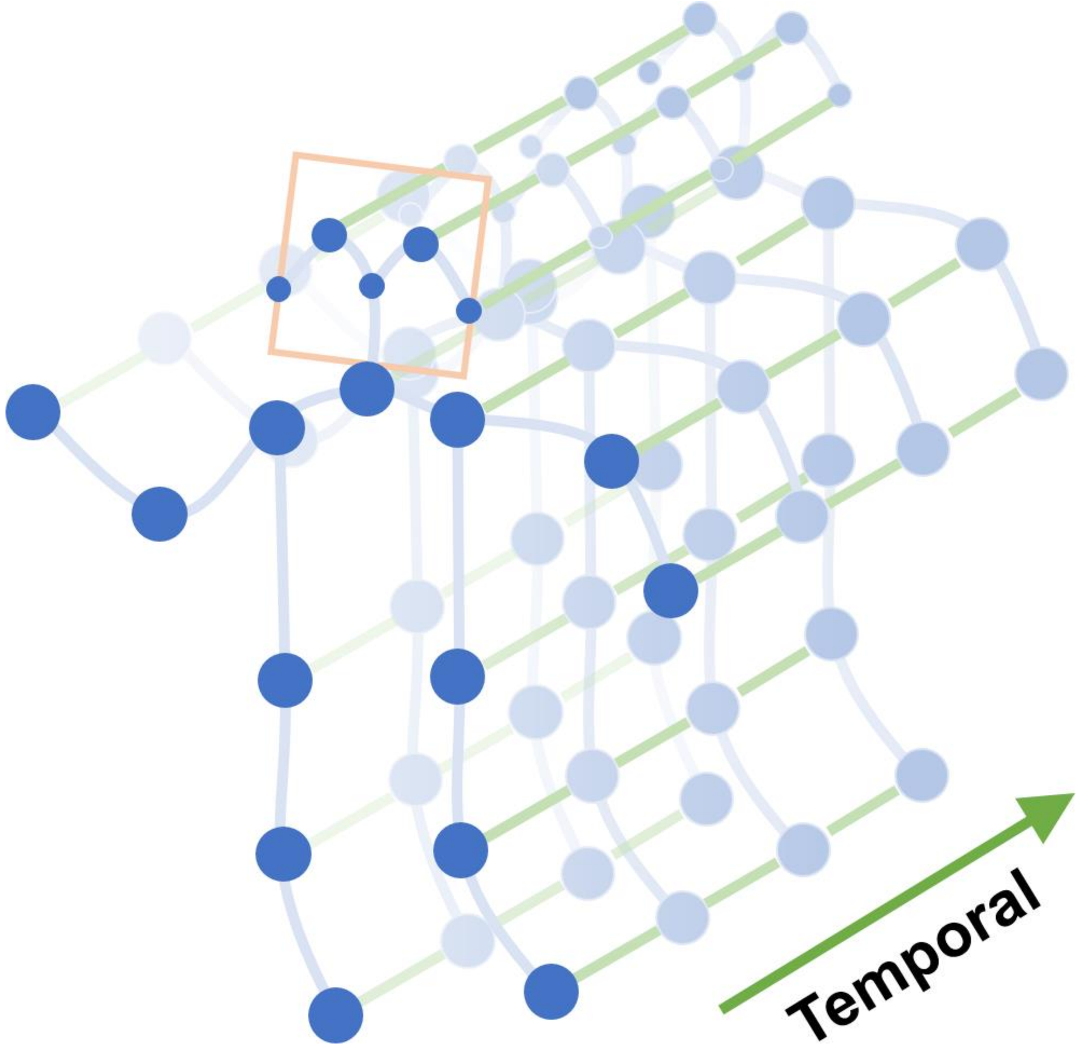
RGB and optical flow



Simonyan, Karen, and Andrew Zisserman. "Two-stream convolutional networks for action recognition in videos." Advances in neural information processing systems 27 (2014).

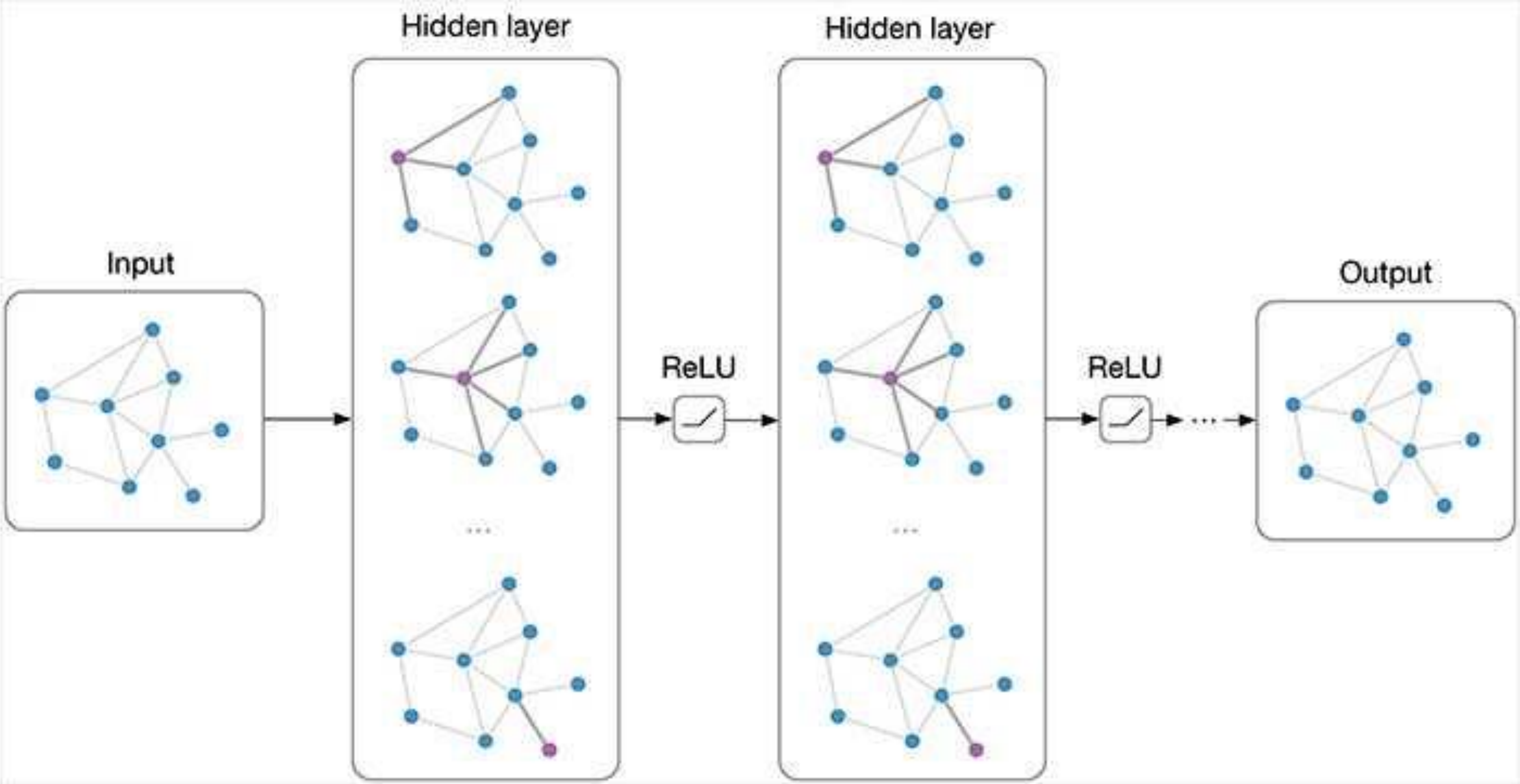
Skeleton-based action classification

Semantic features over time



Yan, Sijie, Yuanjun Xiong, and Dahua Lin. "Spatial temporal graph convolutional networks for skeleton-based action recognition." Proceedings of the AAAI conference on artificial intelligence. Vol. 32. No. 1. 2018.

Graph neural network



Zhou, Jie, et al. "Graph neural networks: A review of methods and applications." AI open 1 (2020): 57-81.

UniGe | MaLGa

Practical example - BABEL

Babel

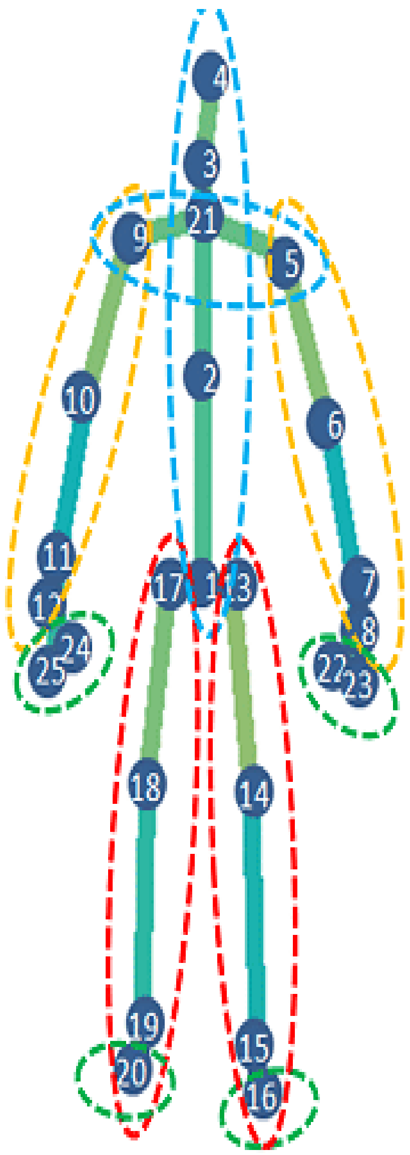
Samples of 3D human poses while performing actions.

babel60 & babel120 → 60 or 120 actions labels

Each sample is composed by 150 “frames” and 25 3D keypionts

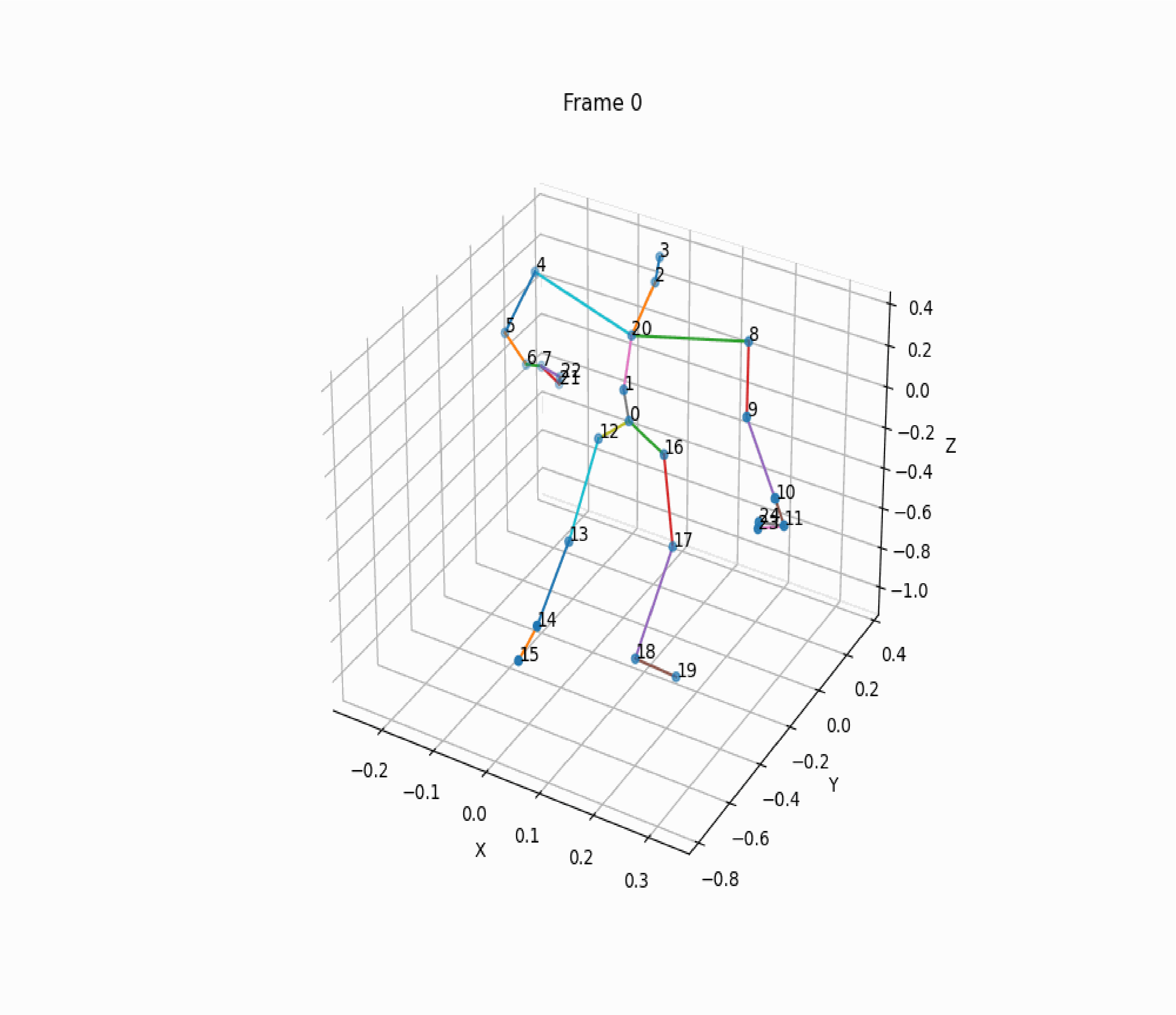
Babel60 → 45473 samples

Babel120 → 48978 samples



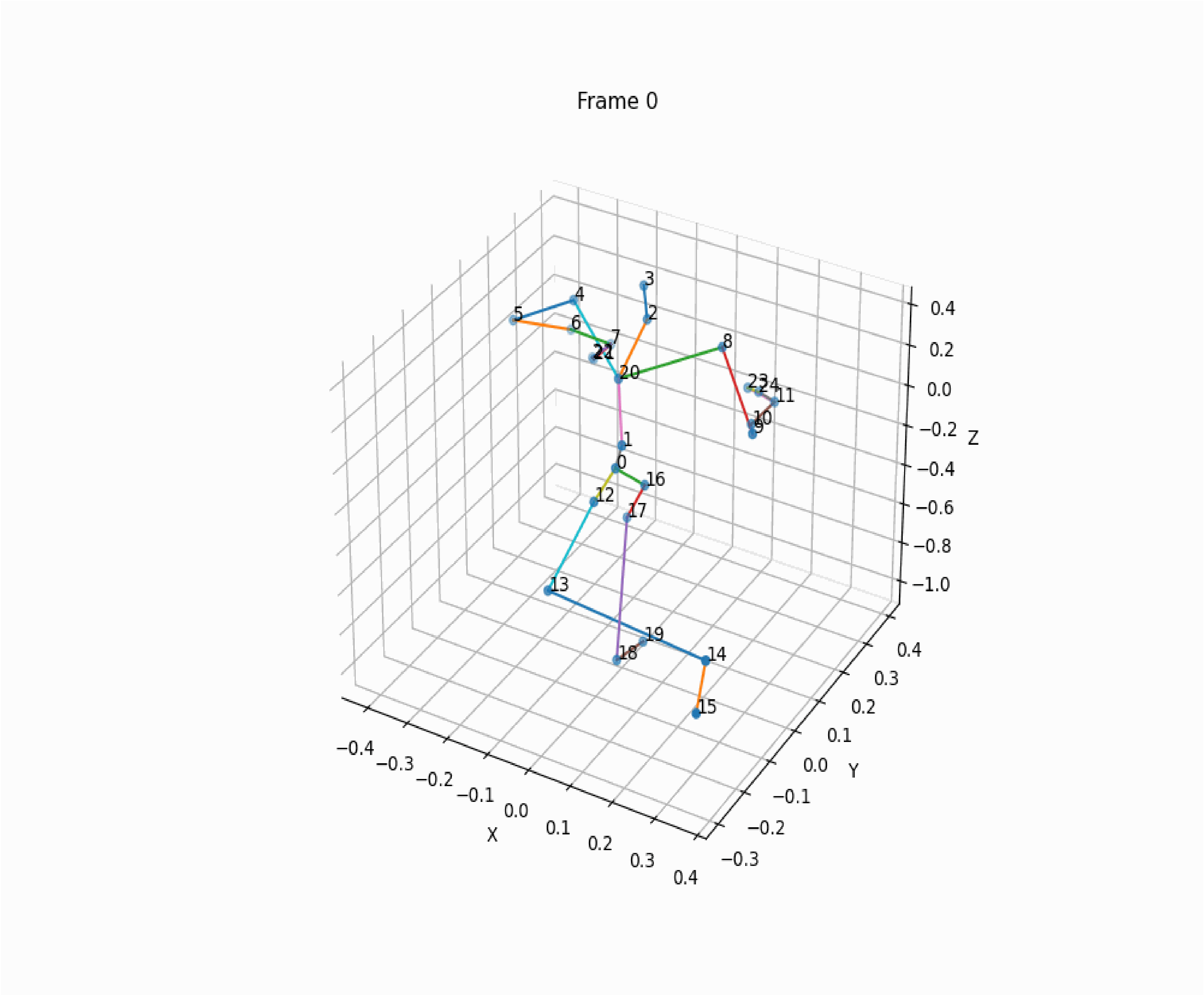
Punnakkal, Abhinanda R., et al. "BABEL: Bodies, action and behavior with english labels." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

Babel - walk



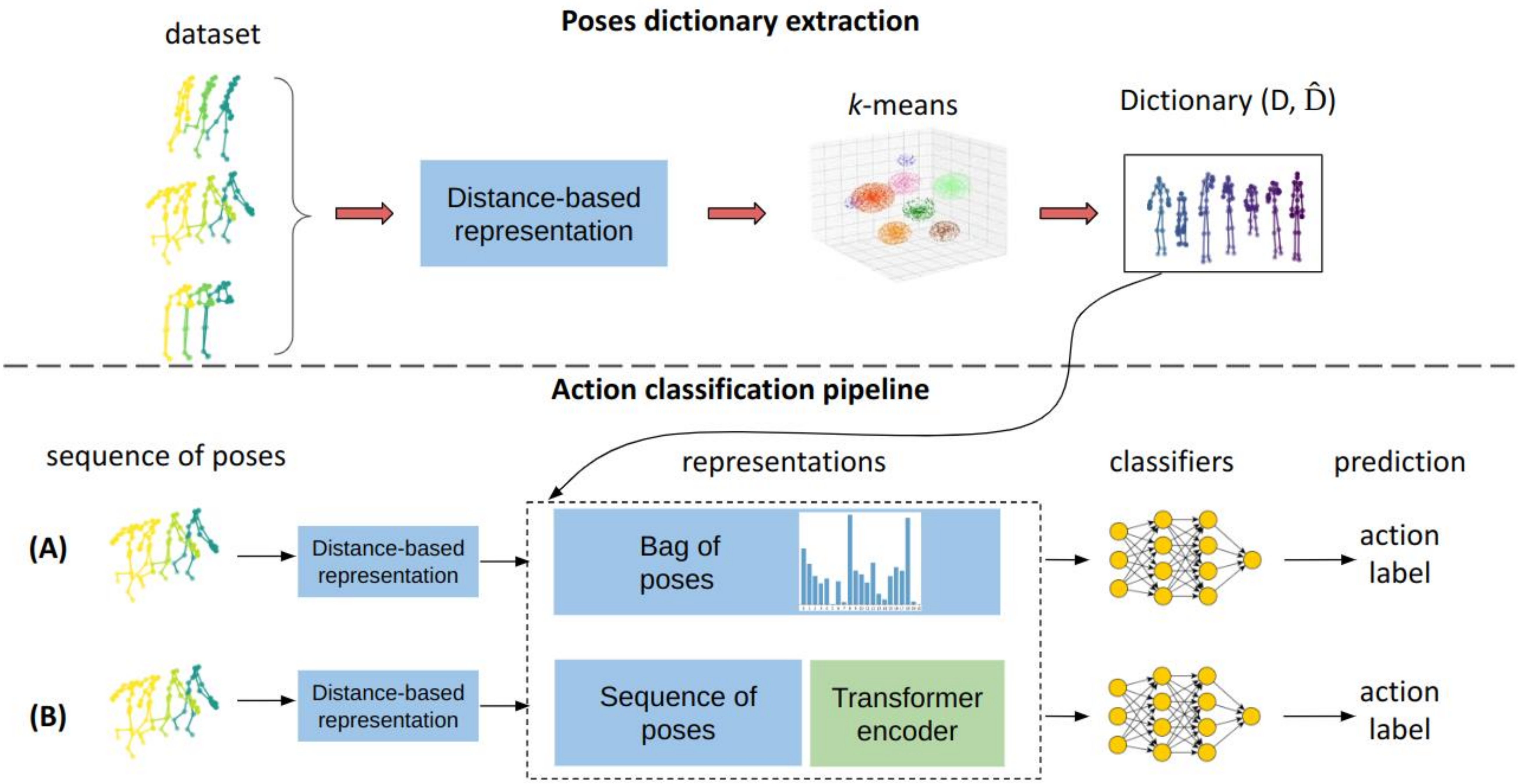
Punnakkal, Abhinanda R., et al. "BABEL: Bodies, action and behavior with english labels." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

Babel - throw

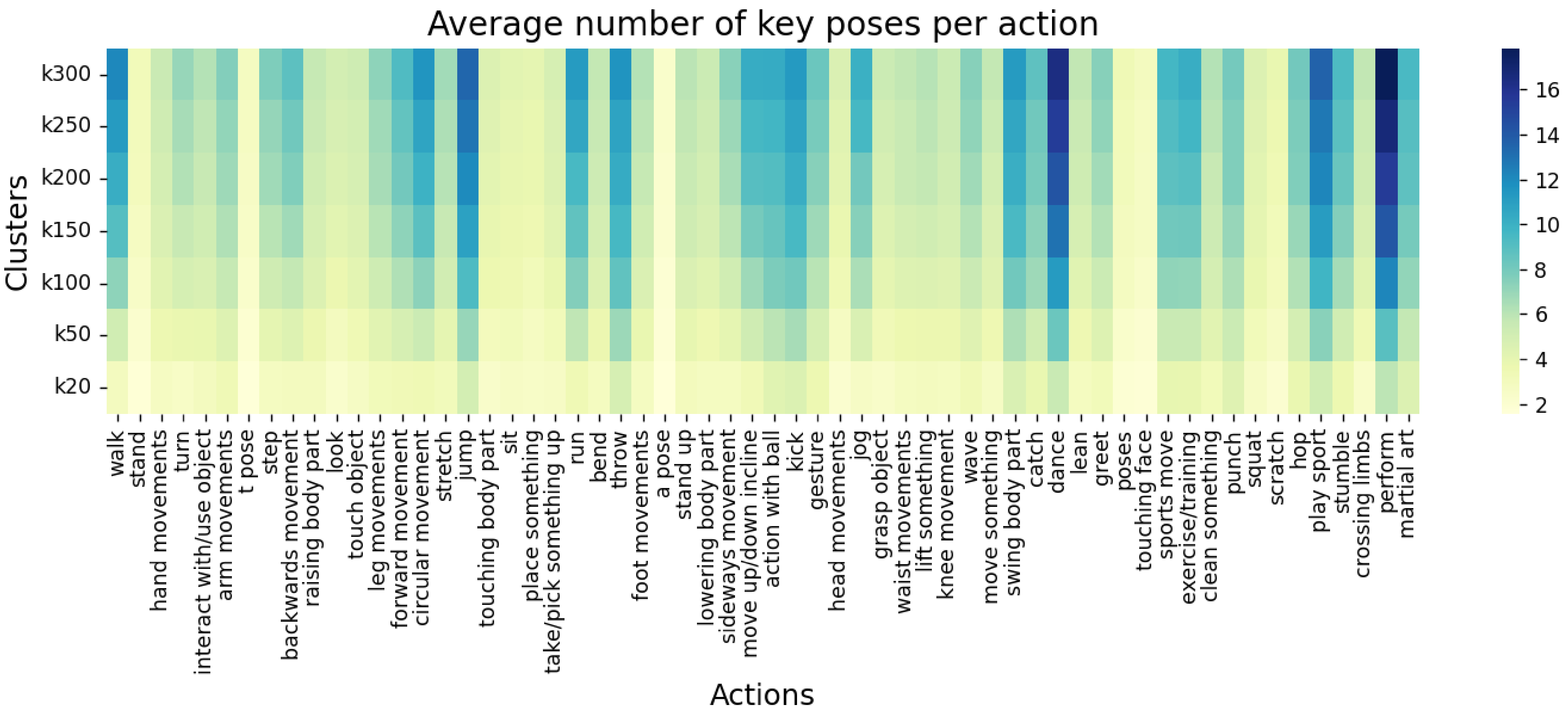


Punnakkal, Abhinanda R., et al. "BABEL: Bodies, action and behavior with english labels." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

Pipeline


Poses dictionary extraction
dataset
Distance-based representation
k-means
Dictionary (D, $\hat{\text{D}}$)
Action classification pipeline
sequence of poses
representations
classifiers
prediction
(A)
Distance-based representation
Bag of poses
action label
(B)
Distance-based representation
Sequence of poses
Transformer encoder
action label

Number of poses



UniGe | MaLGa

UniGe

MaLGa