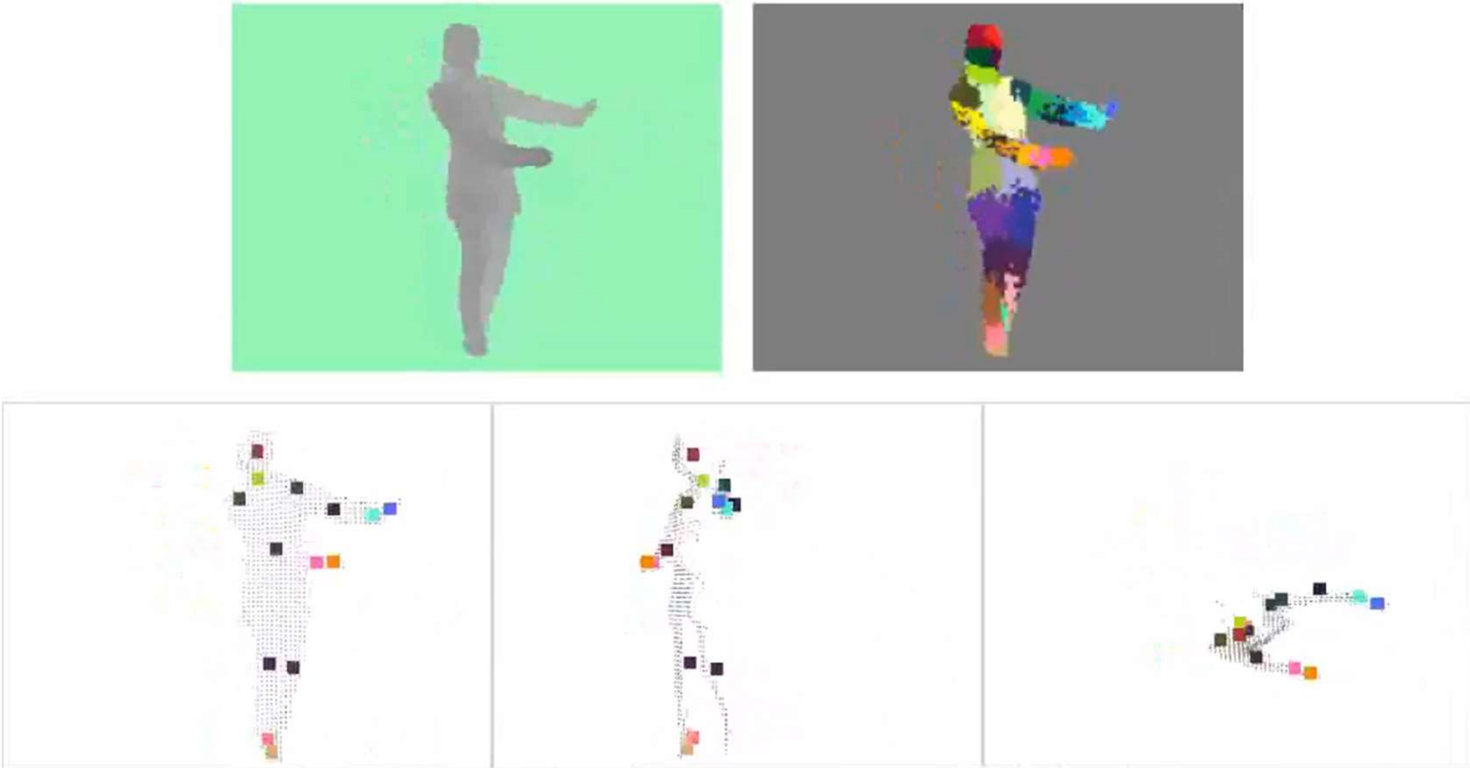
Deep Learning on Lie Groups for Skeleton-based Action Recognition

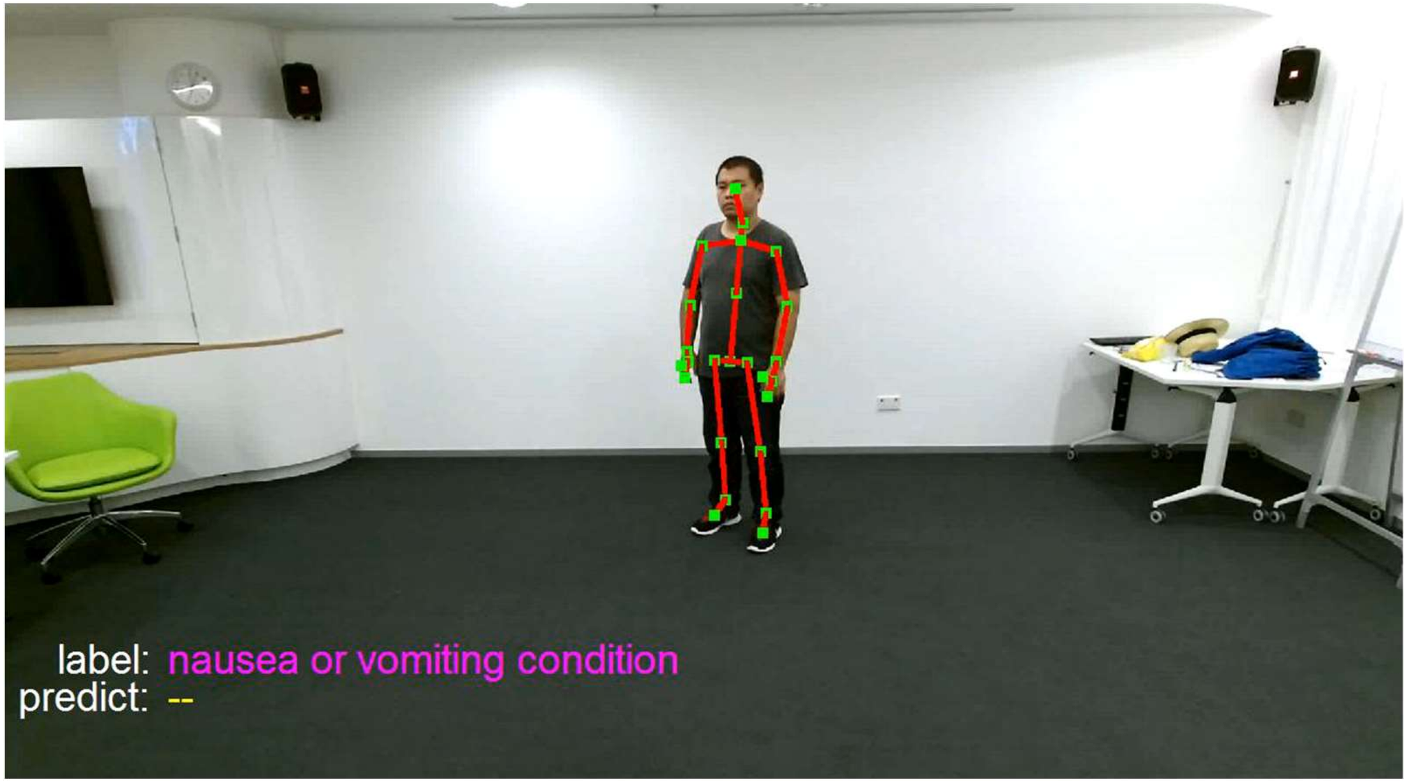
Zhiwu Huang, Chengde Wan, Thomas Probst, Luc Van Gool

Computer Vision Lab @ ETH Zurich

22 July 2017



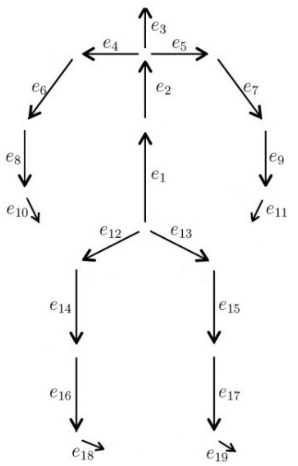
J. Shotton et al., "Real-time Human Pose Recognition in Parts From a Single Depth Image", *CVPR 2011*.



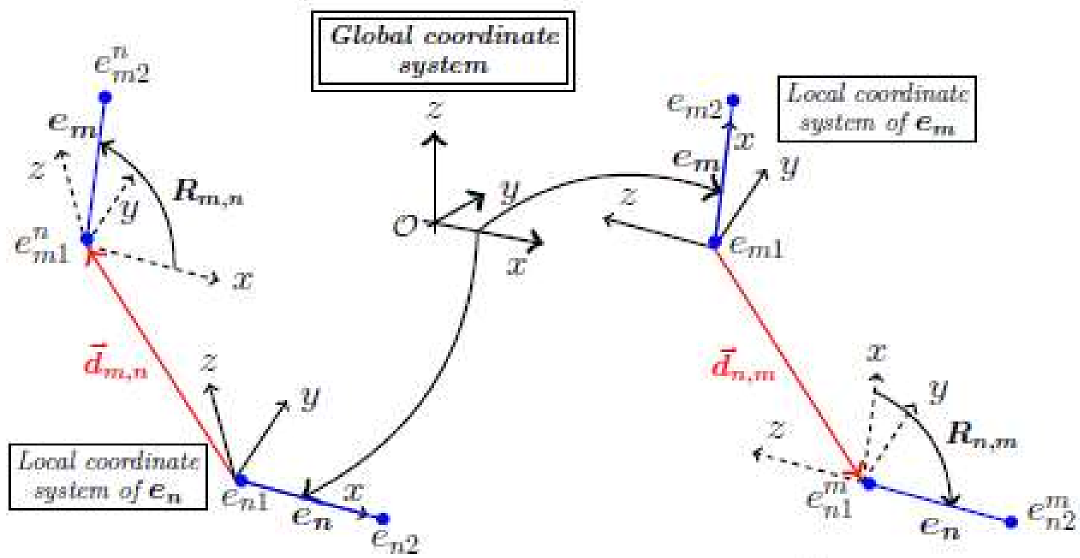
We develop a manifold network (LieNet) to deeply learn Lie group representations for robust action recognition based on skeletal data of human movement

Demo of the proposed LieNet on the NTU-RGB+D dataset released by *A. Shahroudy et al., CVPR 2016*

Special Rotation Group (Lie Group) representation for one skeleton



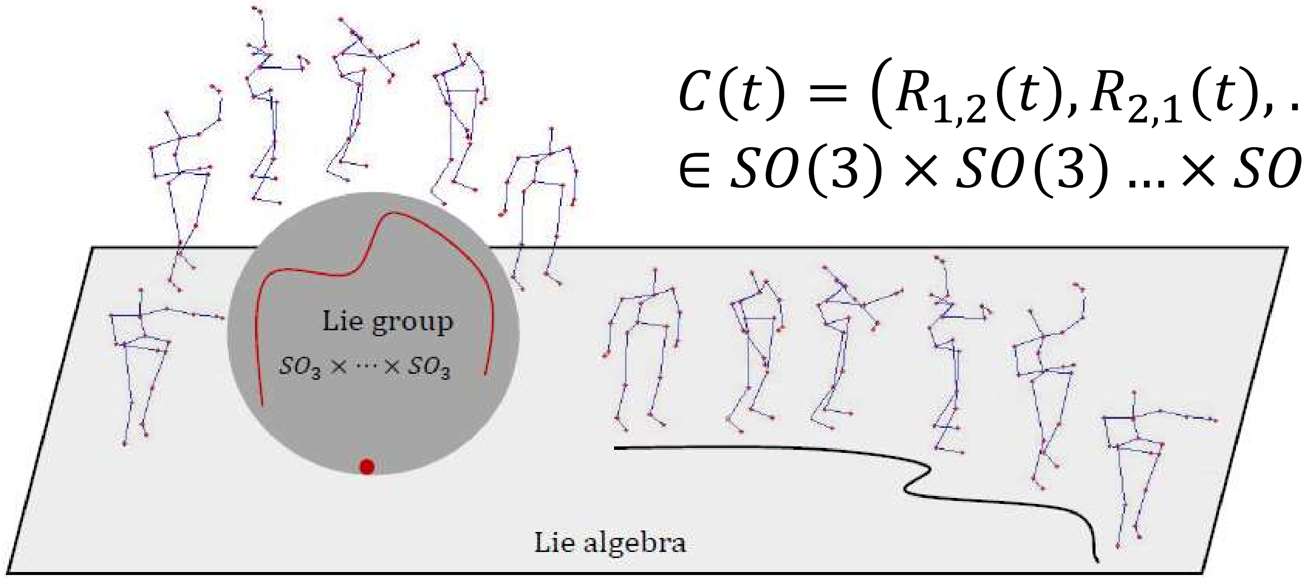
Skeleton with body bones



$$C = \left(R_{1,2}, R_{2,1}, \dots, R_{M,N}, R_{N,M}\right)$$
$$\in SO(3) \times SO(3) \dots \times SO(3)$$

R.Vemulapalli et al., CVPR 2014, CVPR 2016

Lie Group curve representation for one moving skeleton

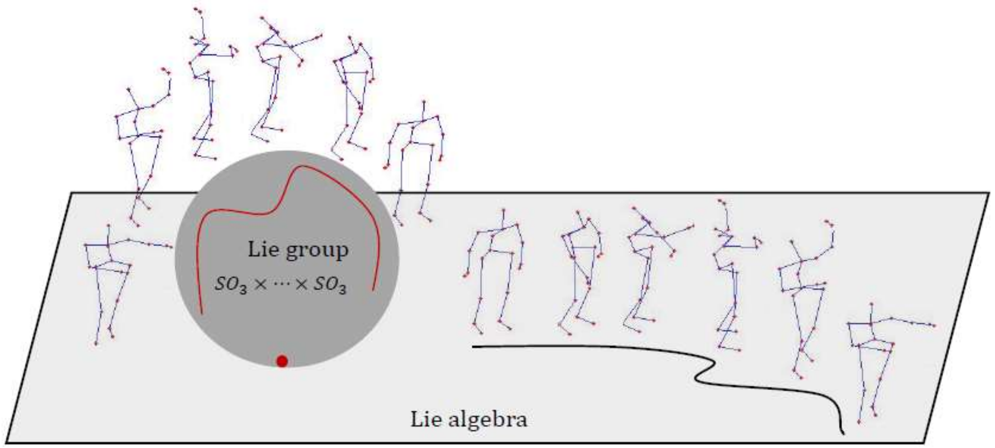


$$C(t) = \left(R_{1,2}(t), R_{2,1}(t), \ldots, R_{M,N}(t), R_{N,M}(t), \right)$$
$$\in SO(3) \times SO(3) \ldots \times SO(3)$$

R.Vemulapalli et al., CVPR 2014, CVPR 2016

Motivation A

- Speed variations (Temporal misalignment)
  - Compute a nominal curve and warp all the curves to this nominal using dynamic time warping (DTW) *[M. Muller, 2007]*



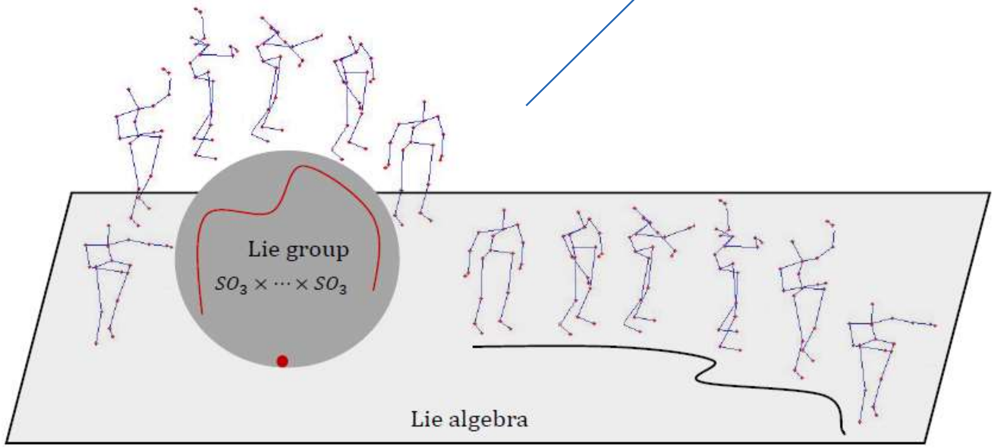
- Additional time cost
- Two-step system

Motivation B

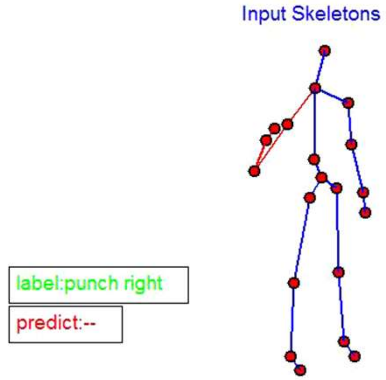
- Lie group representations tend to be extremely high-dimensional
  - Adopt PCA-like method to learn compact and discriminative features

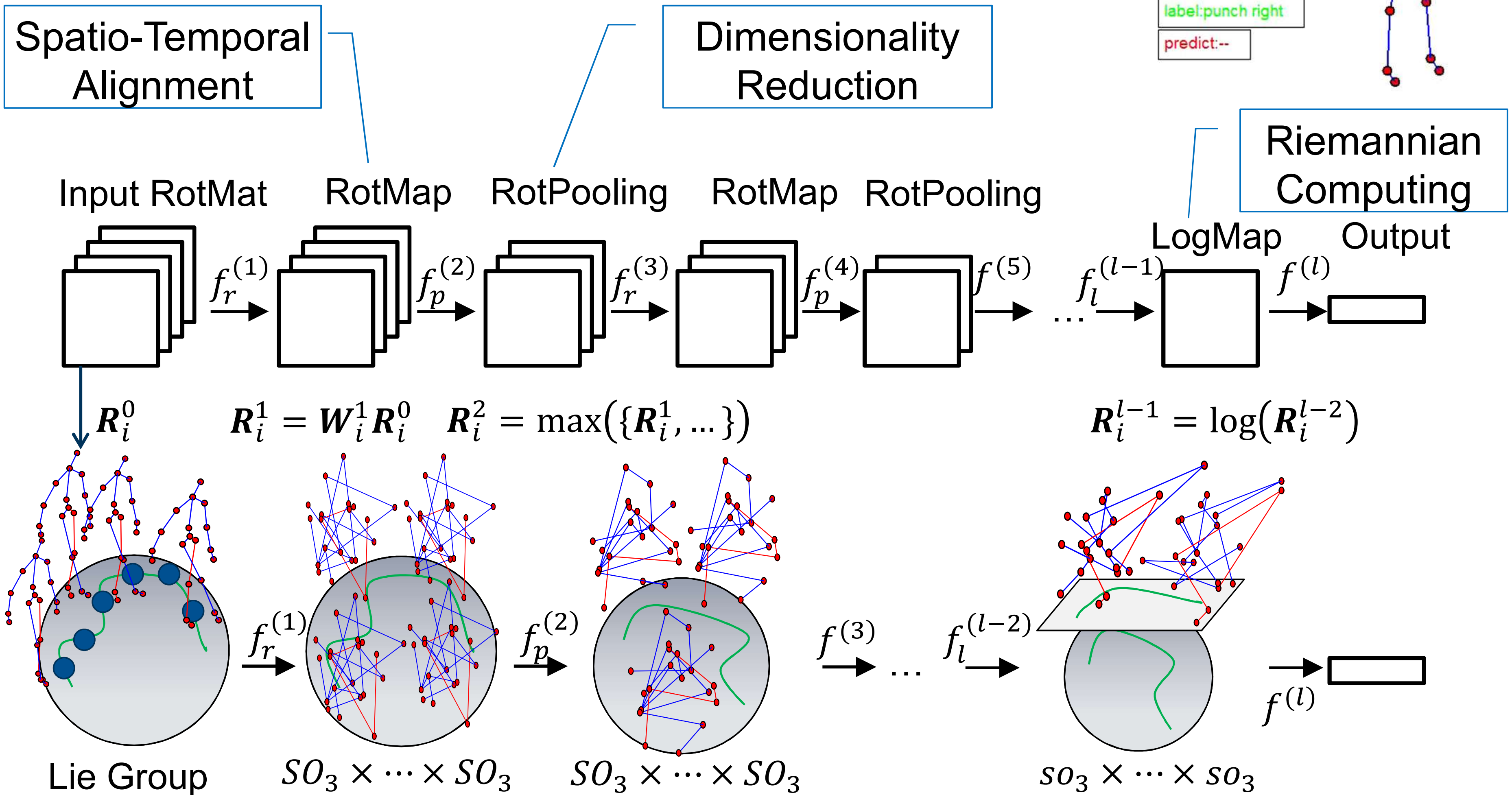
$$C(t) = \left(R_{1,2}(t), R_{2,1}(t), \ldots, R_{M,N}(t), R_{N,M}(t), \right)$$
$$\in SO(3) \times SO(3) \ldots \times SO(3)$$

- High dimensionality
- Shallow learning



LieNet Architecture





Quantitative evaluation

- Accuracies on the G3D-Gaming, HDM05 and NTU RGB-D datasets

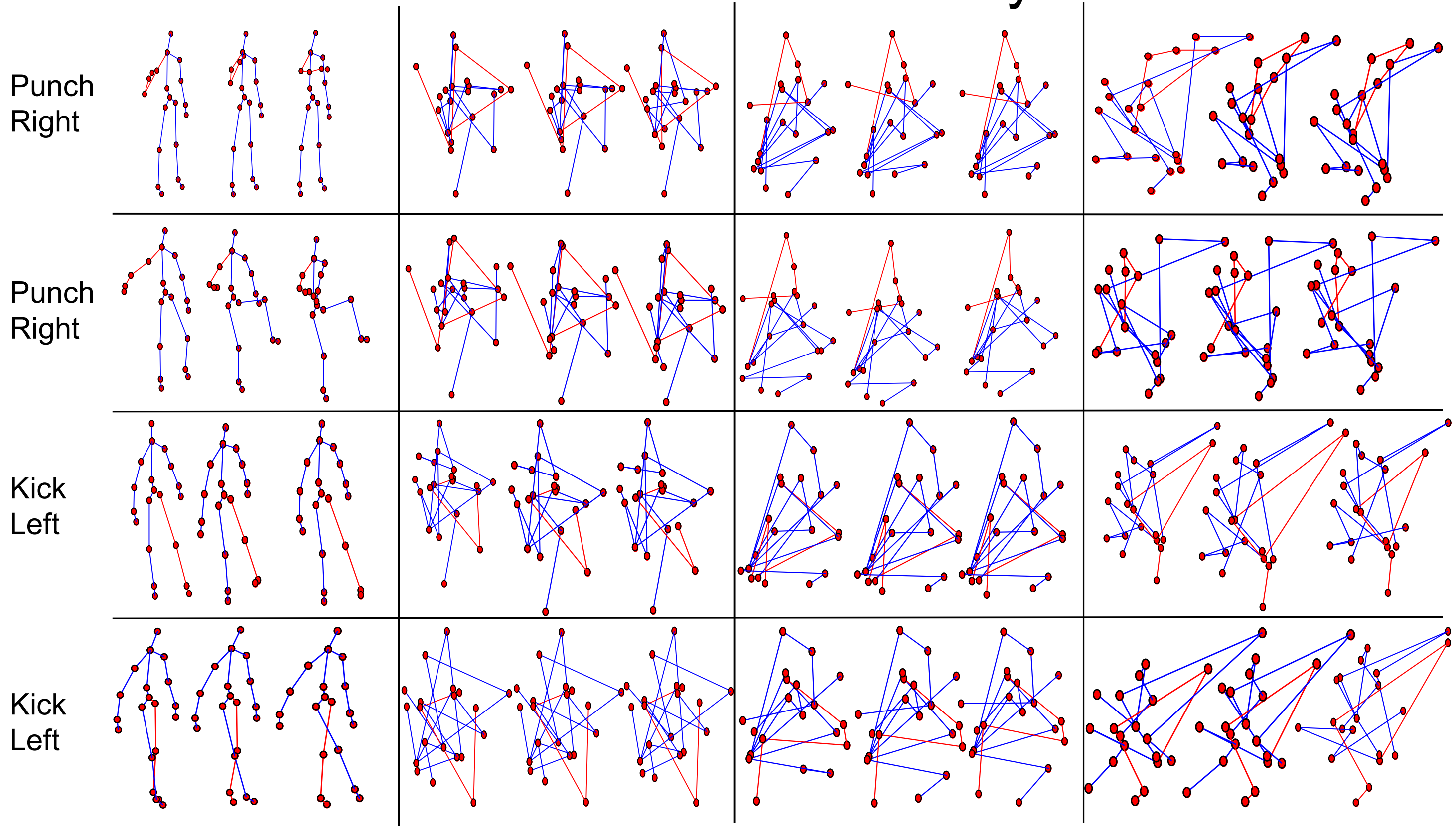
Method	G3D-Gaming
RBM+HMM [32]	86.40%
SE [41]	87.23%
SO [42]	87.95%
LieNet-0Block	84.55%
LieNet-1Block	85.16%
LieNet-2Blocks	86.67%
LieNet-3Blocks	**89.10%**

Method	RGB+D-subject	RGB+D-view
HBRNN [13]	59.07%	63.97%
Deep RNN [37]	56.29%	64.09%
Deep LSTM [37]	60.69%	67.29%
PA-LSTM [37]	62.93%	70.27%
ST-LSTM [26]	**69.2%**	**77.7%**
SE [41]	50.08%	52.76%
SO [42]	52.13%	53.42%
LieNet-0Block	53.54%	54.78%
LieNet-1Block	56.35%	60.14%
LieNet-2Blocks	58.02%	62.52%
LieNet-3Blocks	61.37%	66.95%

Method	HDM05
SPDNet [18]	61.45%±1.12
SE [41]	70.26%±2.89
SO [42]	71.31%±3.21
LieNet-0Block	71.26%±2.12
LieNet-1Block	73.35%±1.14
LieNet-2Blocks	**75.78%±2.26**

Qualitative analysis

- Reconstruction of different LieNet layers for four action sequences



(a) Input (b) 1st layer: RotMap (c) 2nd layer: RotPooling··· (d) 7th layer: LogMap

The patterns for specific motion classes become more discriminative when arriving at the output layer

Deep Learning on Lie Groups for Skeleton-based Action Recognition

Thank you for your time and attention!