ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

A Two-Stream Recurrent Network for Skeleton-based Human Interaction Recognition

Qianhui Men[*], Edmond S. L. Ho[†], Hubert P. H. Shum[‡], Howard Leung[*]

[*]Department of Computer Science, City University of Hong Kong

[†]Department of Computer and Information Sciences, Northumbria University

[‡]Department of Computer Science, Durham University

CityU 香港城市大學 City University of Hong Kong

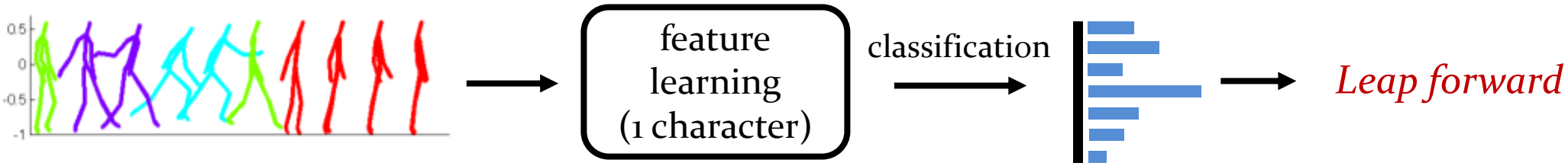
Northumbria University NEWCASTLE

Durham University

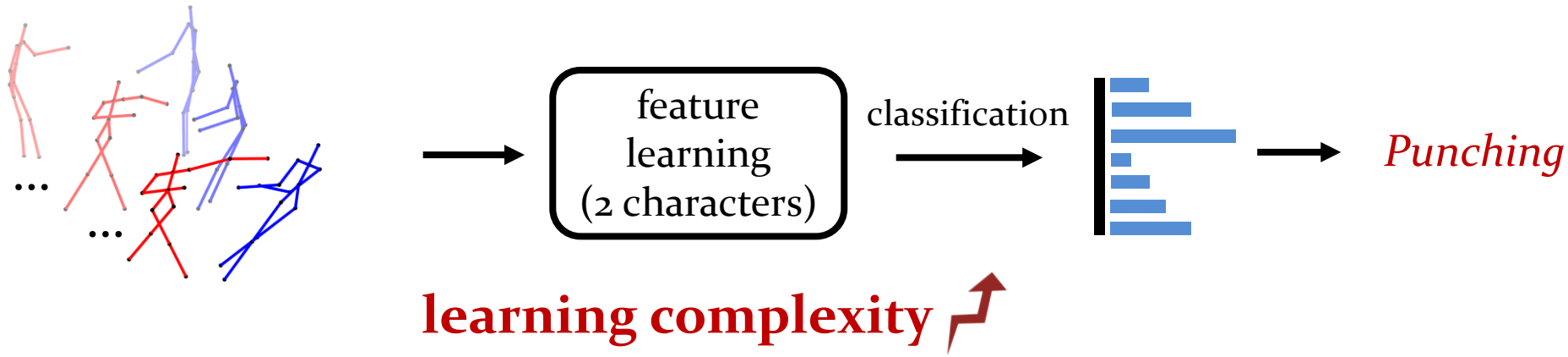
Introduction

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

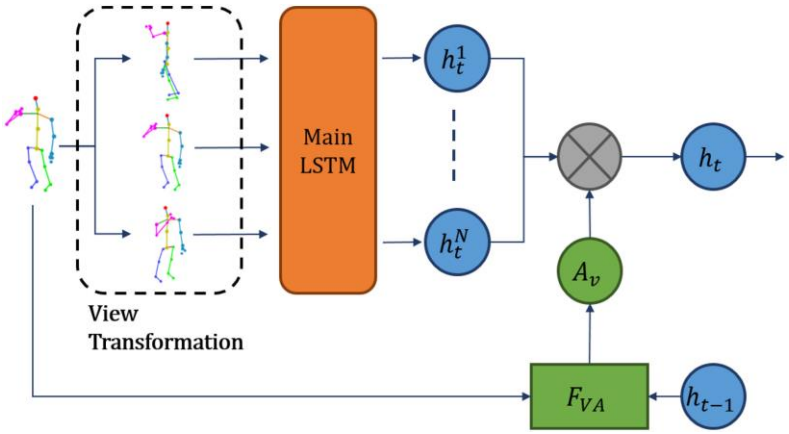
- **Skeleton-based Action Recognition of Single Character**



- **Skeleton-based Human Interaction Recognition**

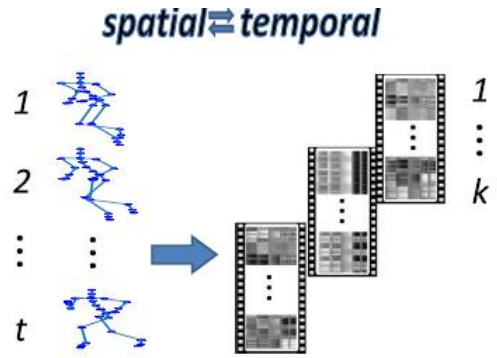


Related Work

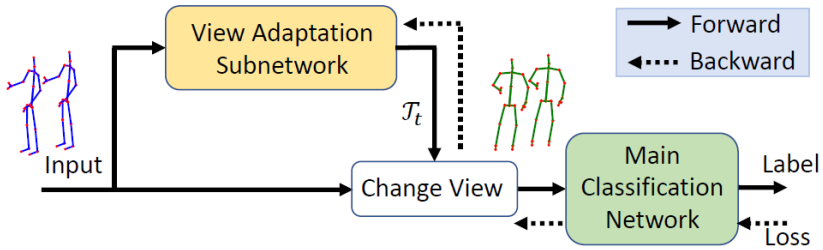


[Fan *et al*. TMM, 2019]

- Stacking joint features



[Ke *et al*. CVPR, 2017]

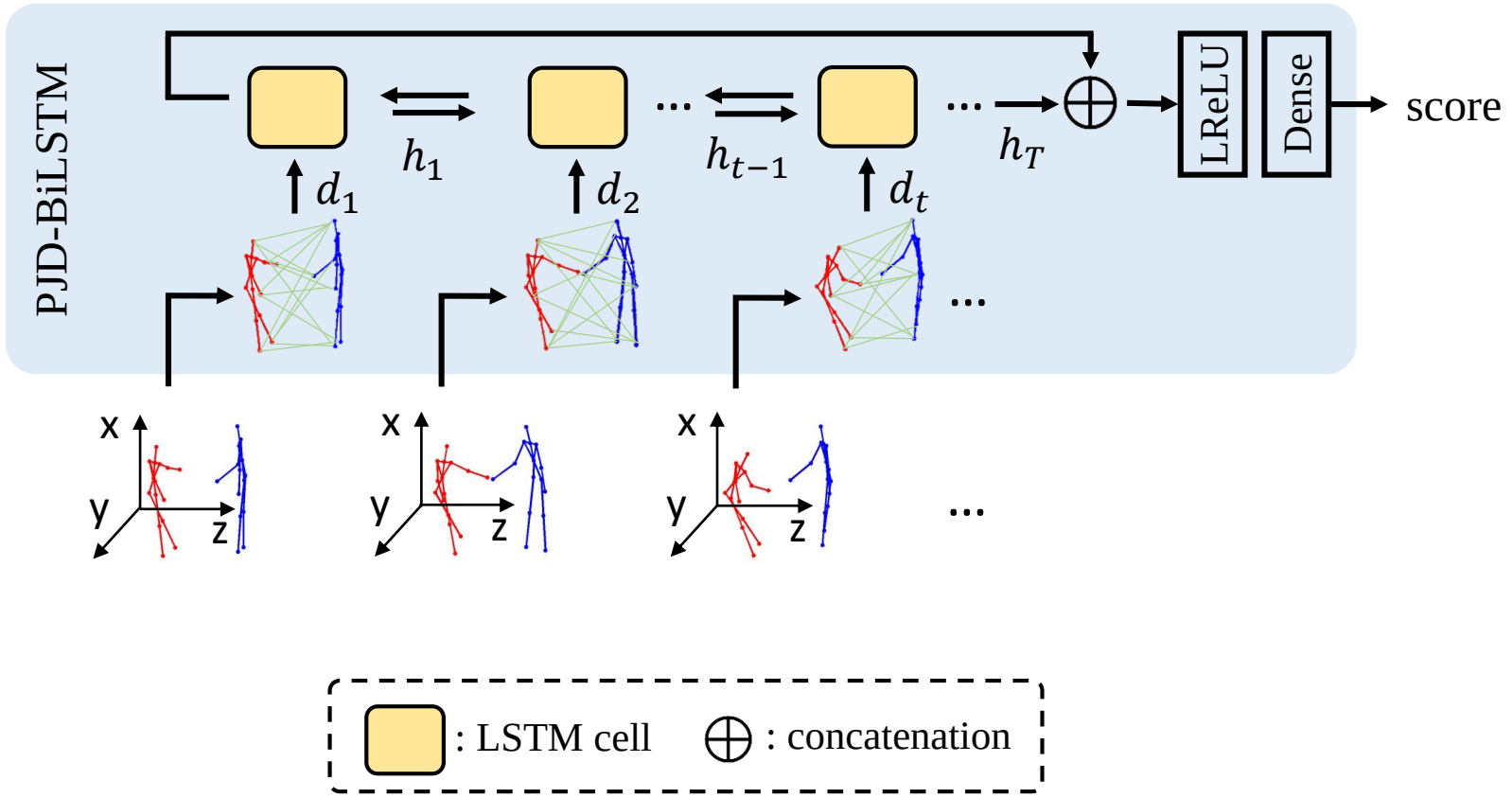


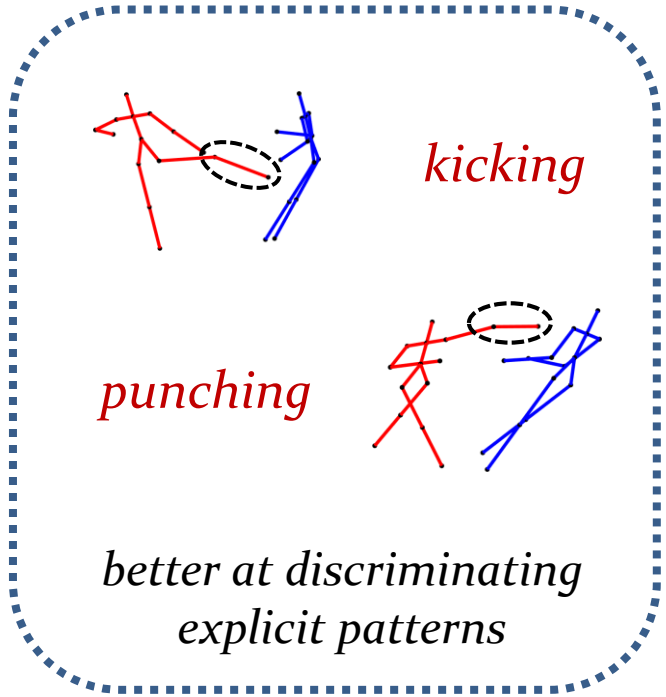
[Zhang *et al*. TPAMI, 2019]

- Feature extraction from individual characters

Methodology

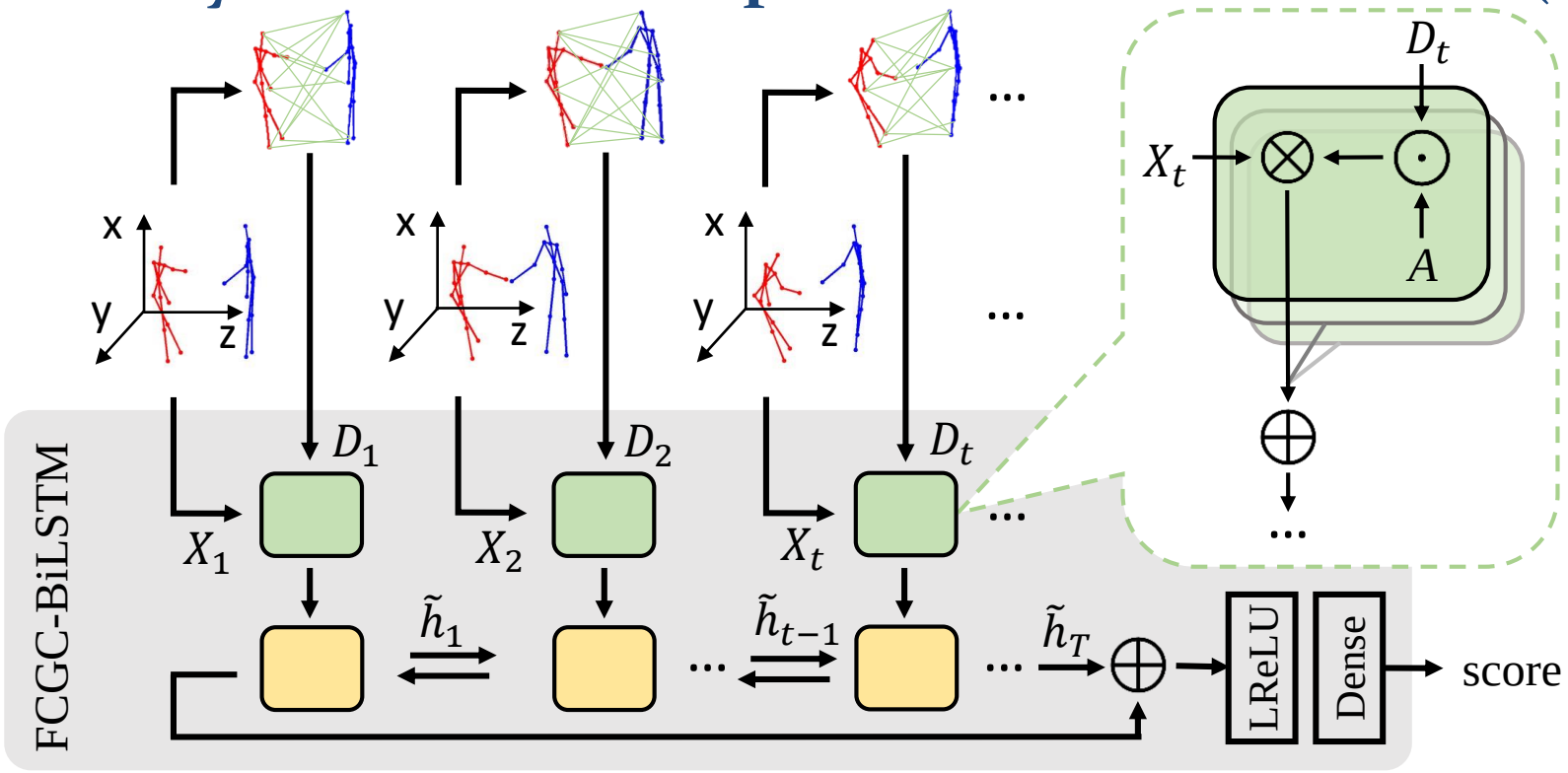
- **Pairwise Joint Distance BiLSTM (PJD-BiLSTM)**



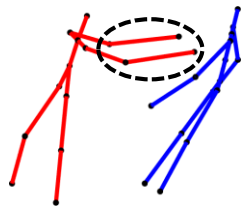


Methodology

- **Fully-Connected Graph Convolution BiLSTM (FCGC-BiLSTM)**

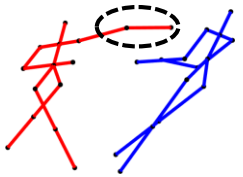


$$W \star_g \mathbf{X}_t = \sigma((\bigoplus_{c=1}^{C}(A \odot D_t)\mathbf{X}_t^C)W)$$



pushing

punching

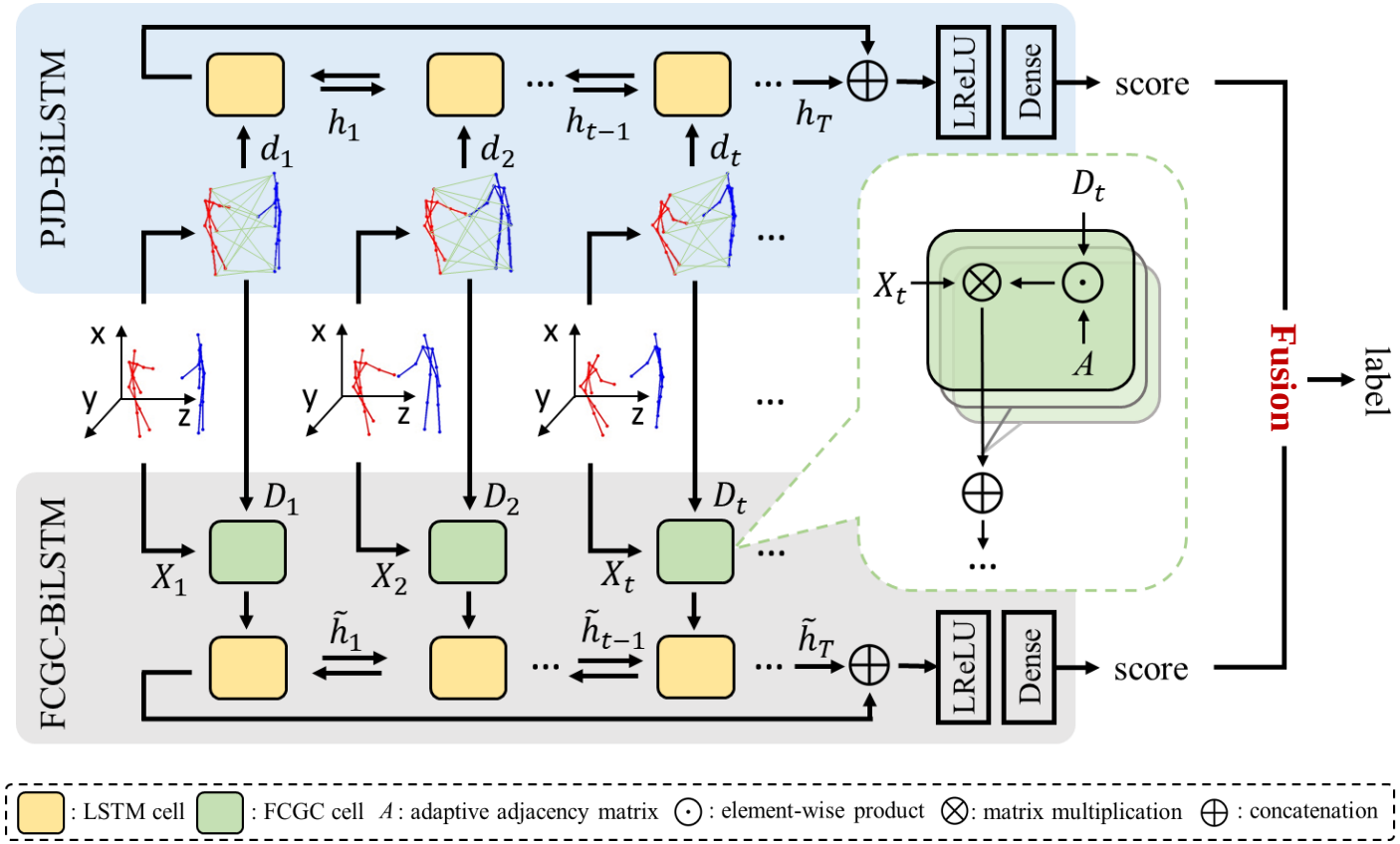


better at discriminating implicit patterns

: LSTM cell : FCGC cell A: adaptive adjacency matrix $\odot$: element-wise product $\otimes$: matrix multiplication $\oplus$: concatenation

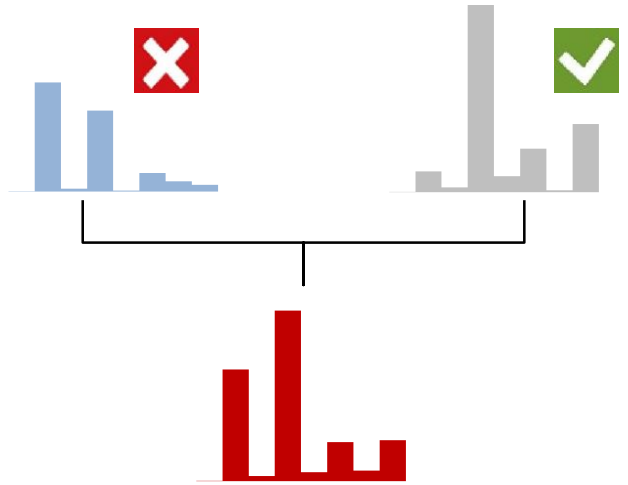
Methodology

- **Overall framework**



- Late Fusion

  Principle of Maximum Entropy

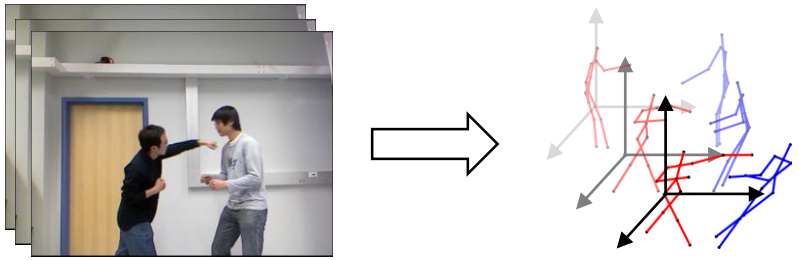


Larger entropy indicates a discriminative classification.

Experiments

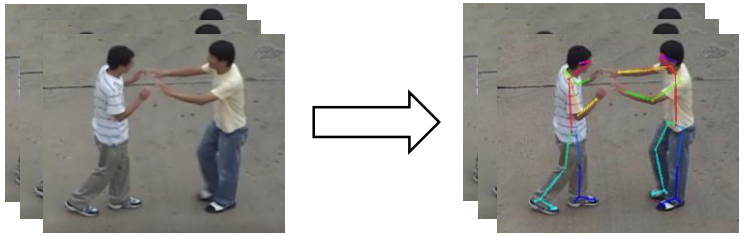
ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

- **SBU Interaction Dataset**
  - **3D Skeleton Video**



Method	Acc.(%)
Raw Skeleton [18]	49.7
Joint feature [18]	80.3
Co-occurrence LSTM [20]	90.4
ST-LSTM+Trust Gate [23]	93.3
Clips+CNN+MTLN [16]	93.5
SI and JD features [5]	93.9
GCA-LSTM [7]	94.1
CNN+Kernel Feature maps [25]	94.3
Two-stream RNN [9]	94.8
LSTM+FA+VF [8]	95.0
PJD-BiLSTM	94.0
FCGC-BiLSTM	95.1
PJD+FCGC	**96.8**

- **UT Interaction Dataset**
  - **2D Key joints in RGB Video**



Modality	Method	Acc.(%)
RGB	DBoW [33]	85.0
	MSSC [34]	83.3
	HR [35]	88.4
	IP [36]	91.6
	PKM [37]	93.3
RGB+skeleton	PA-DRL [38]	**96.7**
skeleton	PJD-BiLSTM	91.9
	FCGC-BiLSTM	92.7
	PJD+FCGC	94.4

Summary

- A pairwise joint distance BiLSTM network (PJD-BiLSTM) that models the *explicit interaction patterns* from the discriminative geometric features.
- A fully-connected graph convolution BiLSTM network (FCGC-BiLSTM) that quantifies the spatial proximity of interaction from joint positions to extract the *implicit correlations* among joints.
- A *late fusion* algorithm is defined to boost the recognition accuracy from probability outputs of both streams.
- State-of-the-art recognition performance on 3D interaction dataset. Can be easily extended to 2D key joint recognition with comparable results.

Thank you for watching!

Questions / Comments please contact: qianhumen2-c@my.cityu.edu.hk
Our Team:

Qianhui Men

Edmond S. L. Ho

Hubert P. H. Shum

Howard Leung

CityU 香港城市大學 City University of Hong Kong

Northumbria University NEWCASTLE

Durham University