JT-MGCN: Joint-temporal Motion Graph Convolutional Network for Skeleton-Based Action Recognition

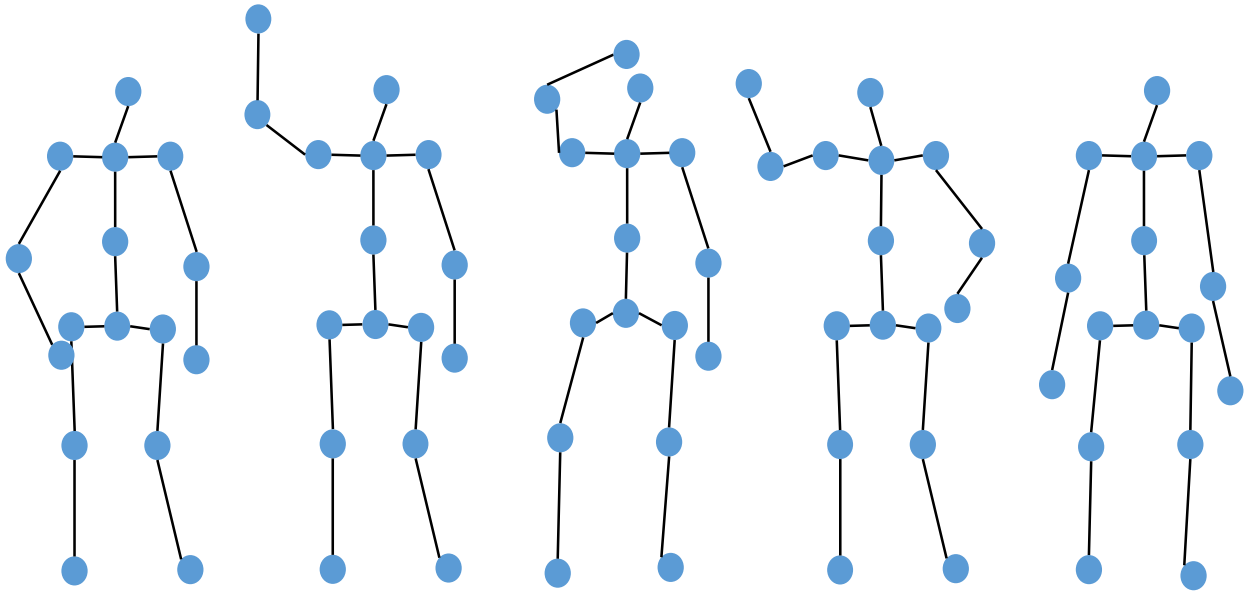
Suekyeong Nam, Seungkyu Lee*

Perception & Computer Vision Lab
Kyung Hee Unviersity

01. 21. 2021

Introduction

Skeleton-Based Action Recognition

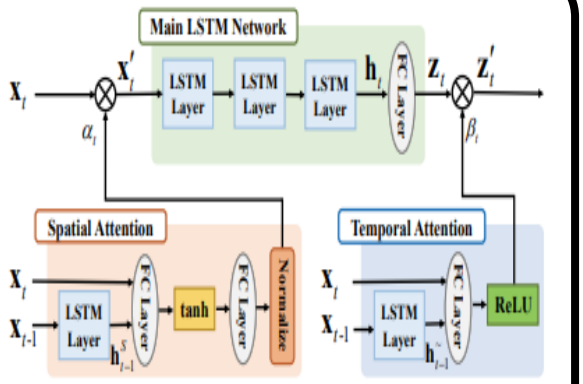


- Without other different features
- Obtaining from RGB images or Depth cameras

Introduction

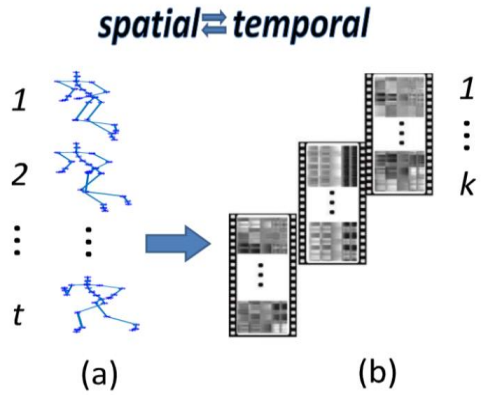
PerCV
Perception & Computer Vision Lab.

RNN based algorithm



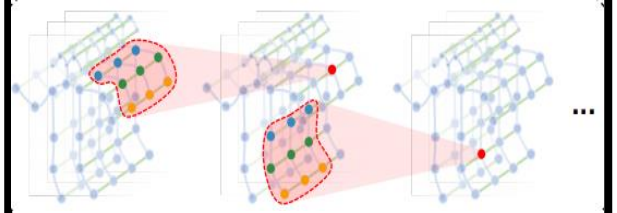
Song, Sijie, et al. "An end-to-end spatio-temporal attention model for human action recognition from skeleton data." AAAI 2017

CNN based algorithm



Ke, Qiuhong, et al. "Learning clip representations for skeleton-based 3D action recognition." IEEE Transactions on Image Processing

GCN based algorithm

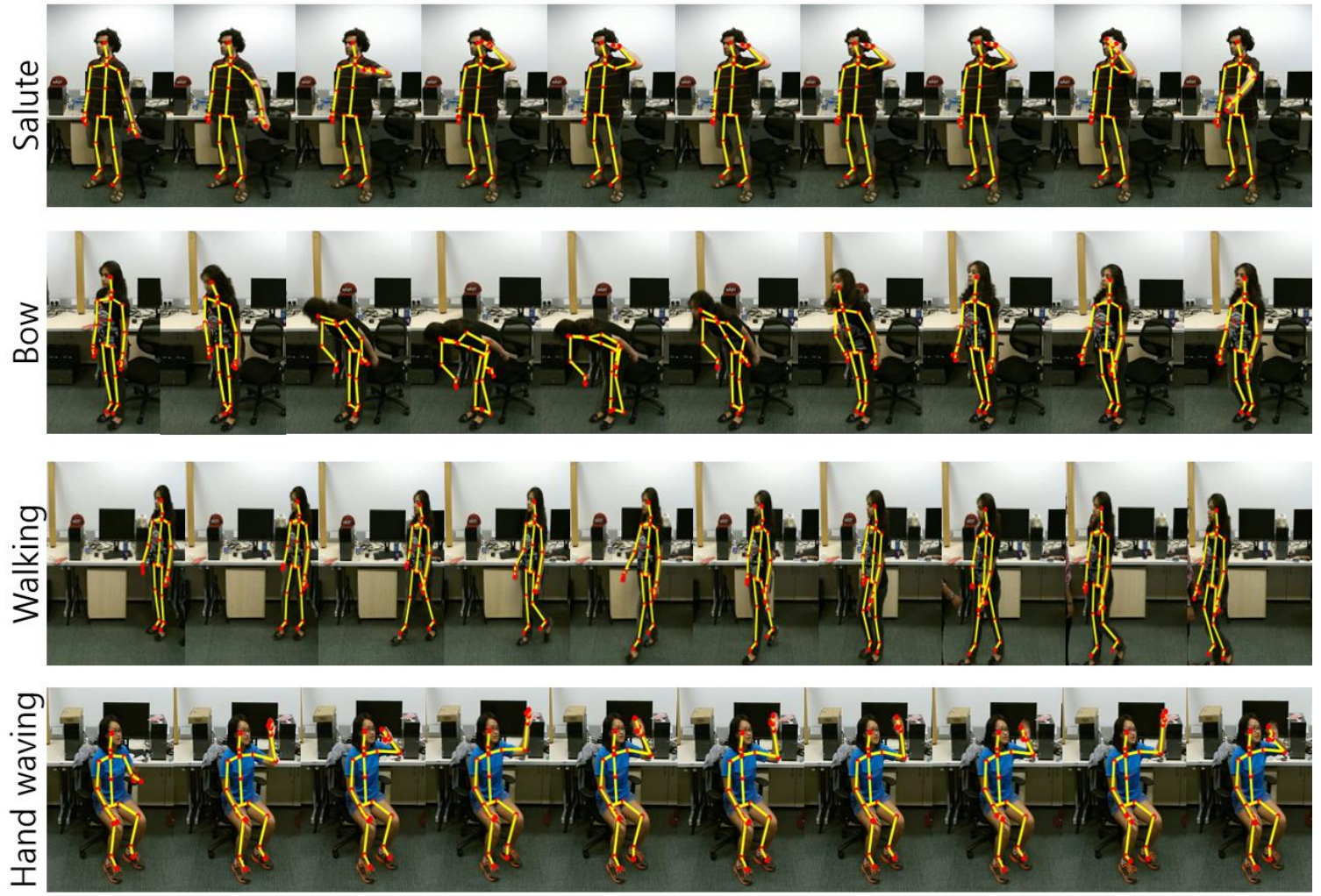


Yan, Sijie, Yuanjun Xiong, and Dahua Lin. "Spatial temporal graph convolutional networks for skeleton-based action recognition." AAAI2018.

	HBRNN[2]	PA-LSTM[17]	ST-LSTM+TS[12]	GCA-LSTM[13]	VA-LSTM[26]	STA-LSTM[22]	2S 3D CNN[11]	MTLN[4]	Res-TCN[6]	MS deep CNN[7]	MS CNN[14]	Motion+Trans[8]	ST-NBMIM[24]	DPRN+GCNN[23]	ARRN-LSTM[28]	SR-TSL[21]	IndRNN[10]	(P+C)Net[1]	AE-LSTM[16]	MTCNN[5]	ST-GCN[25]	AS-GCN[9]	AGC-LSTM[20]	2S-AGCN[19]	DGNN[18]	BGC-LSTM[27]
RNN	v	v	v	v	v	v								v		v	v	v	v	v				v			v
CNN							v	v	v	v	v	v	v							v	v						
GCN															v	v						v	v	v	v	v	v
		2016		2017										2018									2019				

Problem

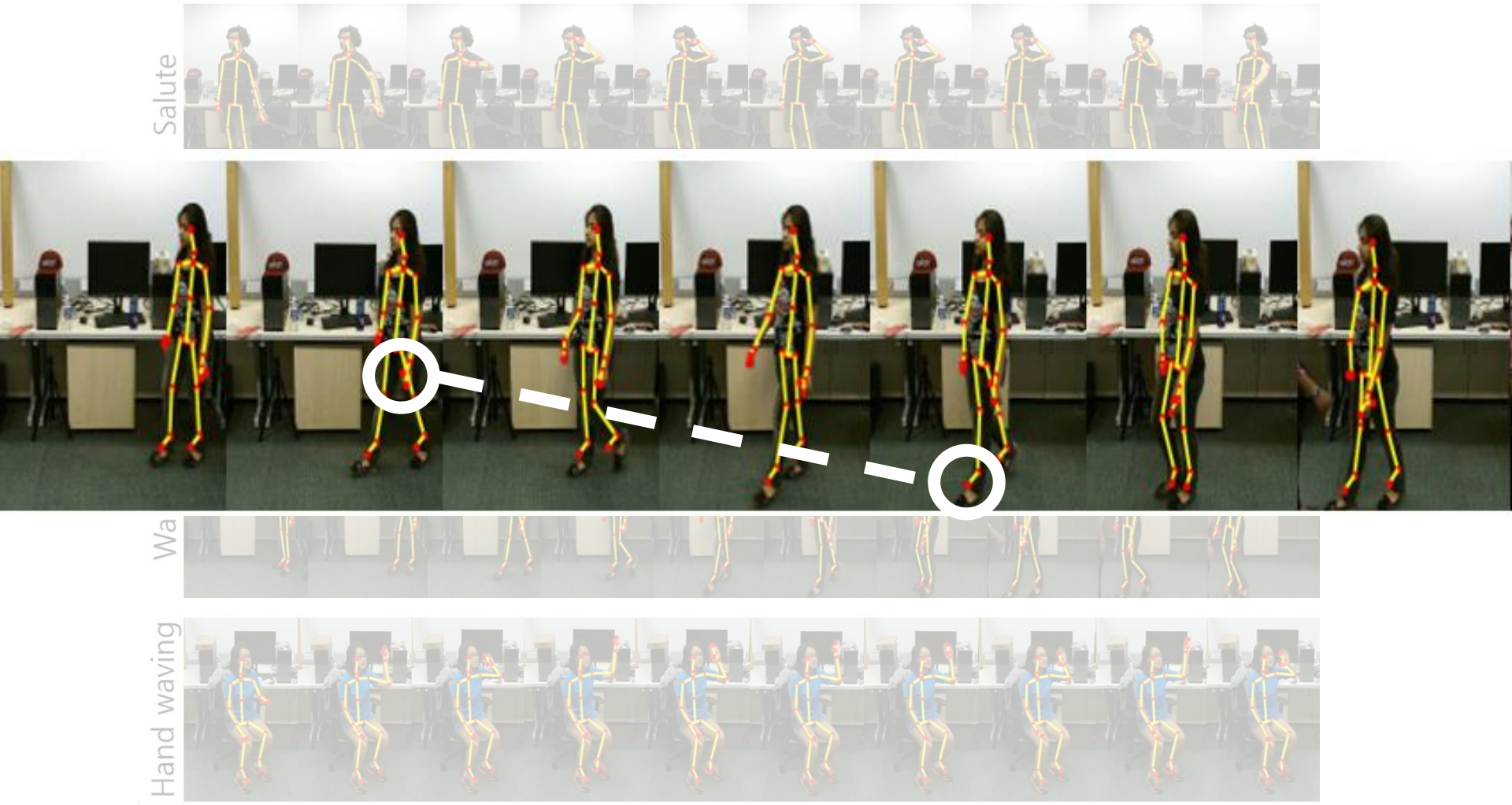
There are relationship between time and joint



Problem

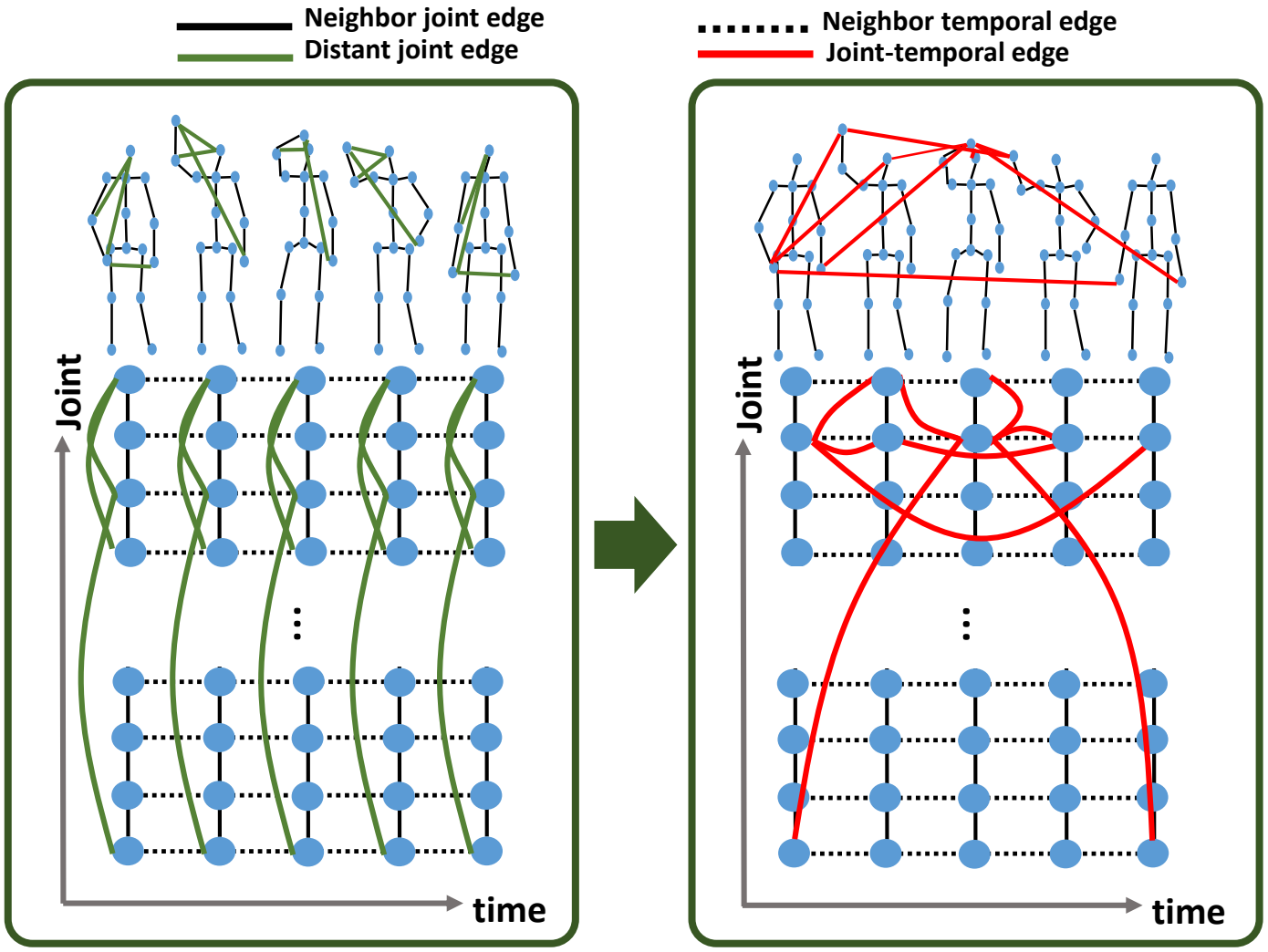
PerCV
Perception & Computer Vision Lab.

There are relationship between time and joint



Proposed algorithm

PerCV
Perception & Computer Vision Lab.



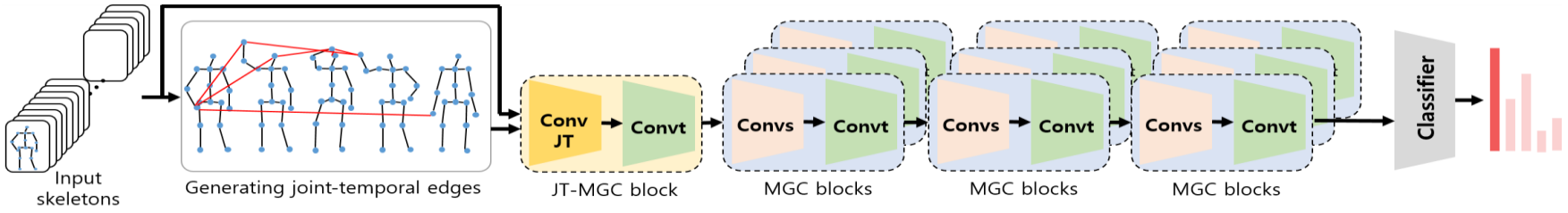
2S-AGCN, Shi et al. *CVPR* 2019.

Our proposed algorithm

Proposed algorithm

PerCV
Perception & Computer Vision Lab.

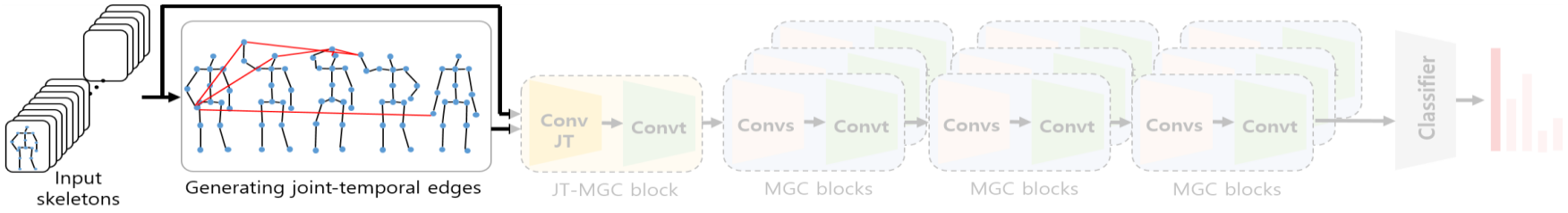
Overall frameworks



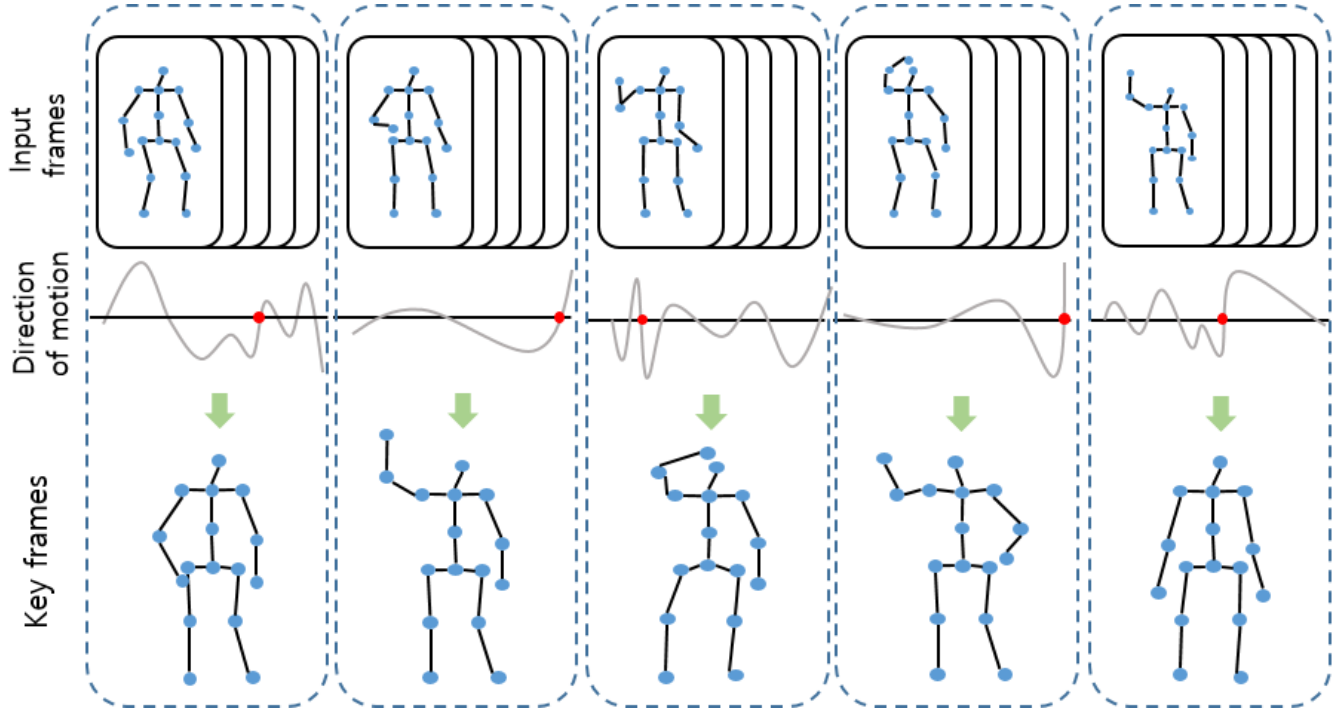
Proposed algorithm

PerCV
Perception & Computer Vision Lab.

Overall frameworks



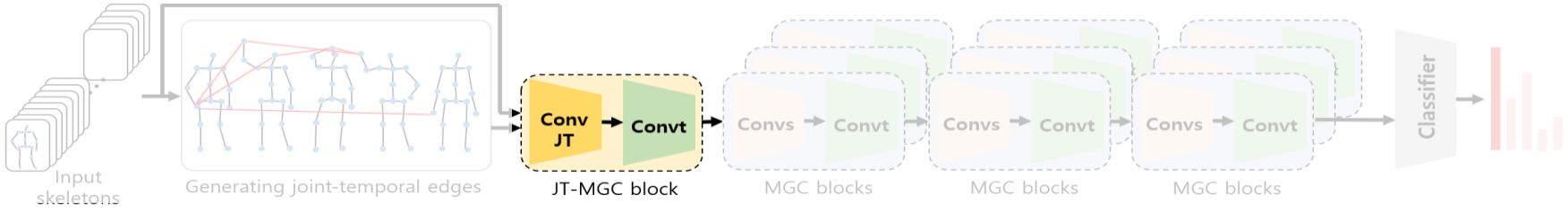
Frame Sampling



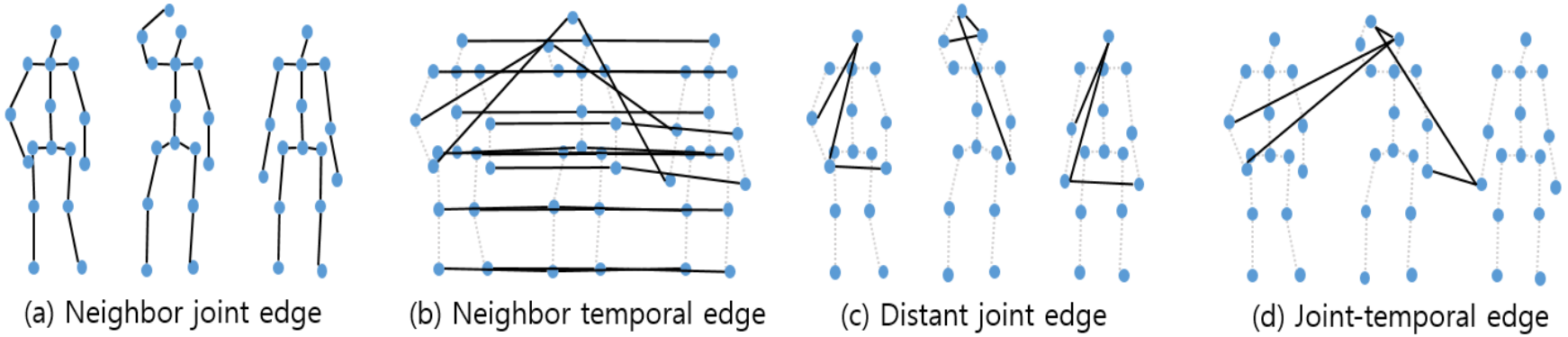
Reducing computational cost!

Proposed algorithm(2)

Overall frameworks



Generating various edges



(a) Neighbor joint edge (b) Neighbor temporal edge (c) Distant joint edge (d) Joint-temporal edge

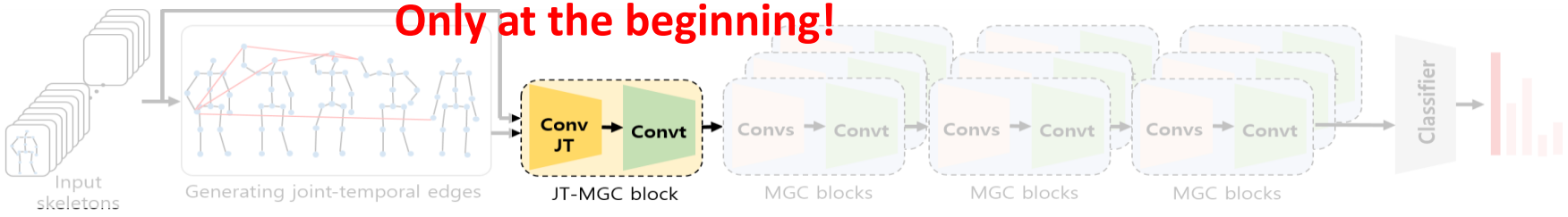
Way to make JT edges

1) Learnable parameter 2) Encoder-Decoder model

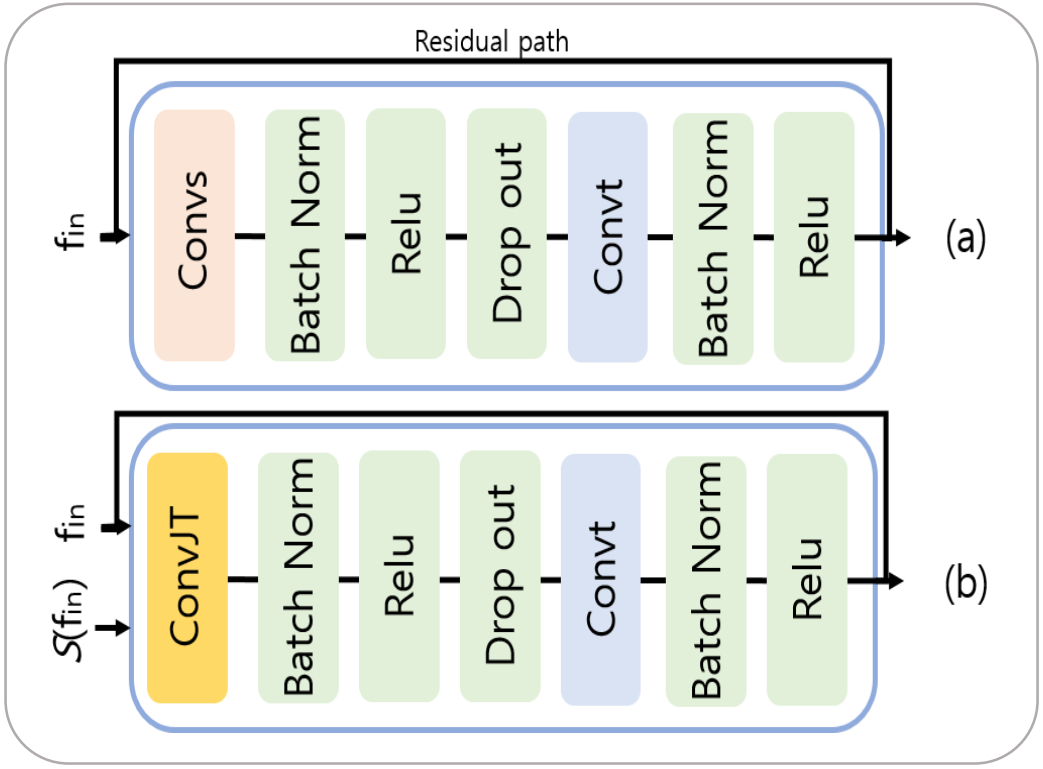
Proposed algorithm

PerCV
Perception & Computer Vision Lab.

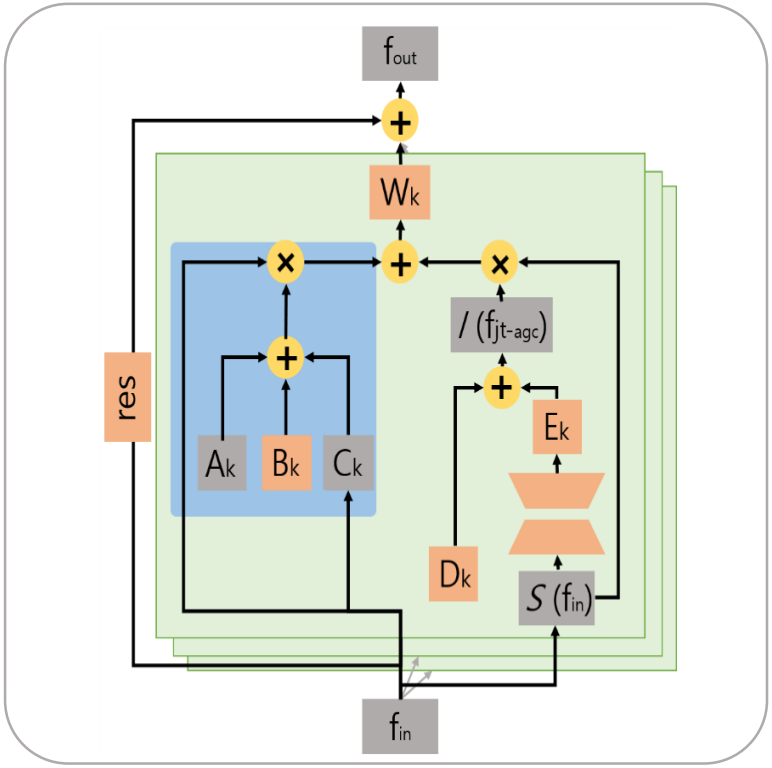
Overall frameworks



JT-MGC block



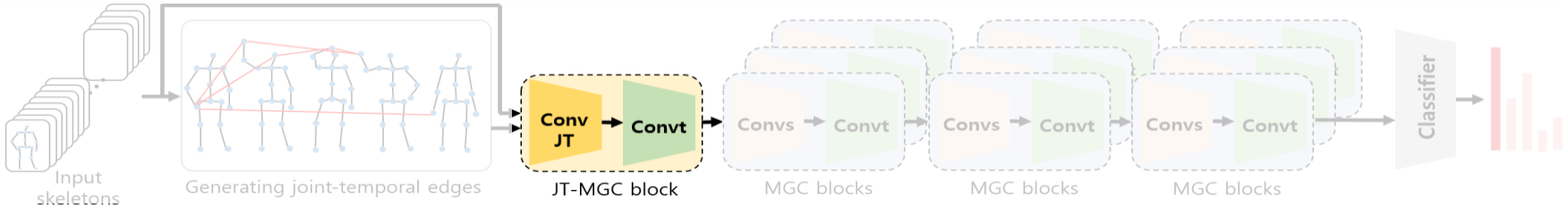
ConvJT



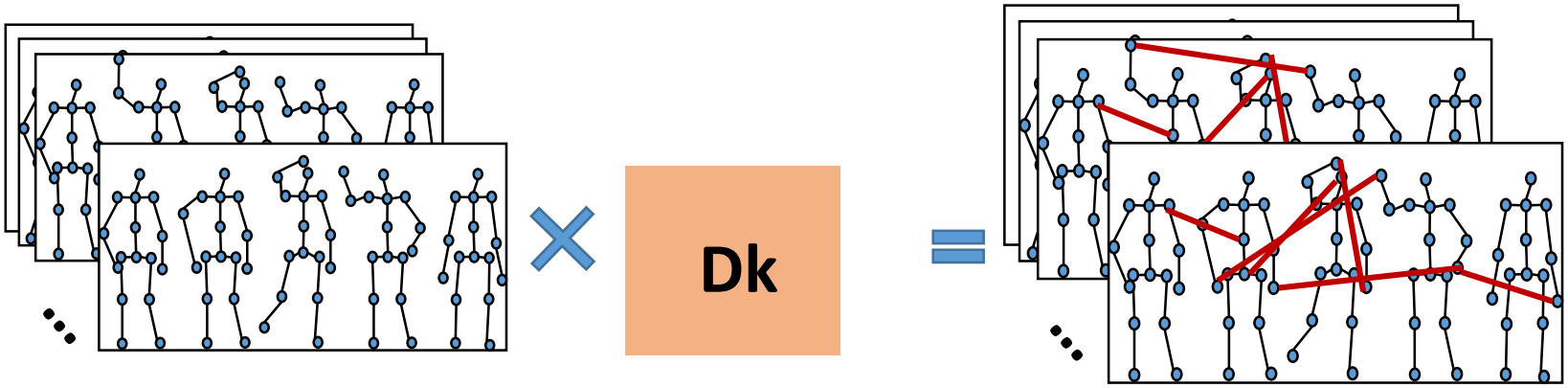
Proposed algorithm

PerCV
Perception & Computer Vision Lab.

Overall frameworks



1) Multiply Learnable Parameters - Dk



Different subjects

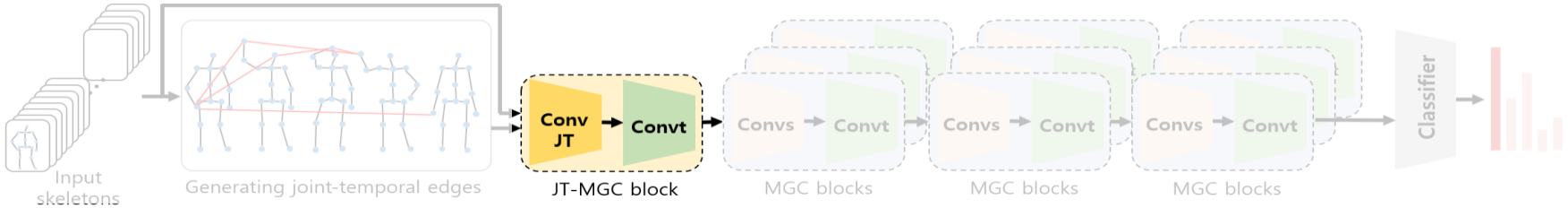
Metrix for common JT edges

Same JT edges for every subjects

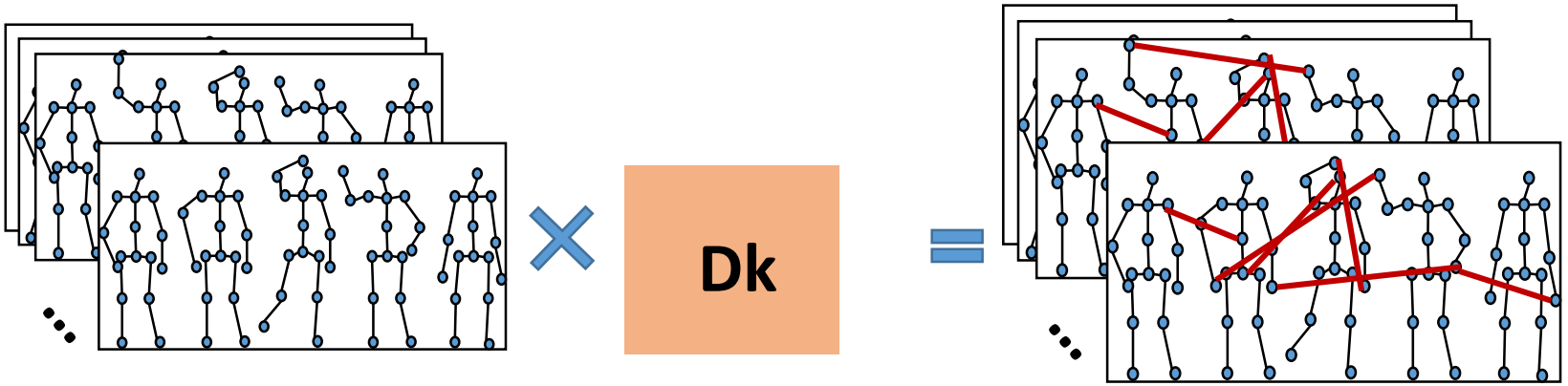
Proposed algorithm

PerCV
Perception & Computer Vision Lab.

Overall frameworks



1) Multiply Learnable Parameters - Dk



Different subjects

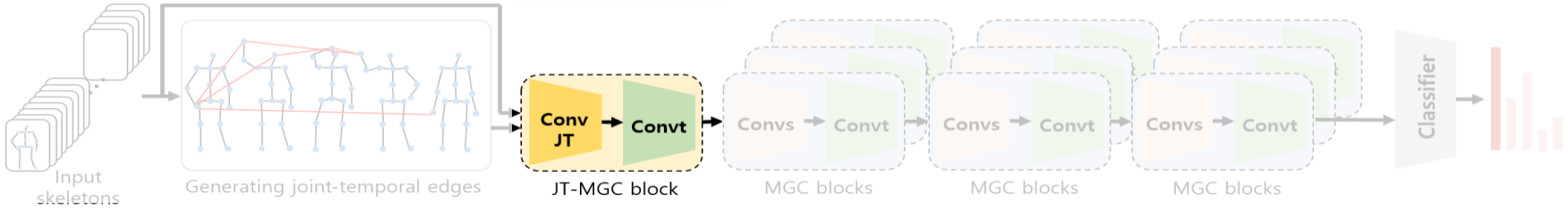
Metrix for common JT edges

Same JT edges for every subjects

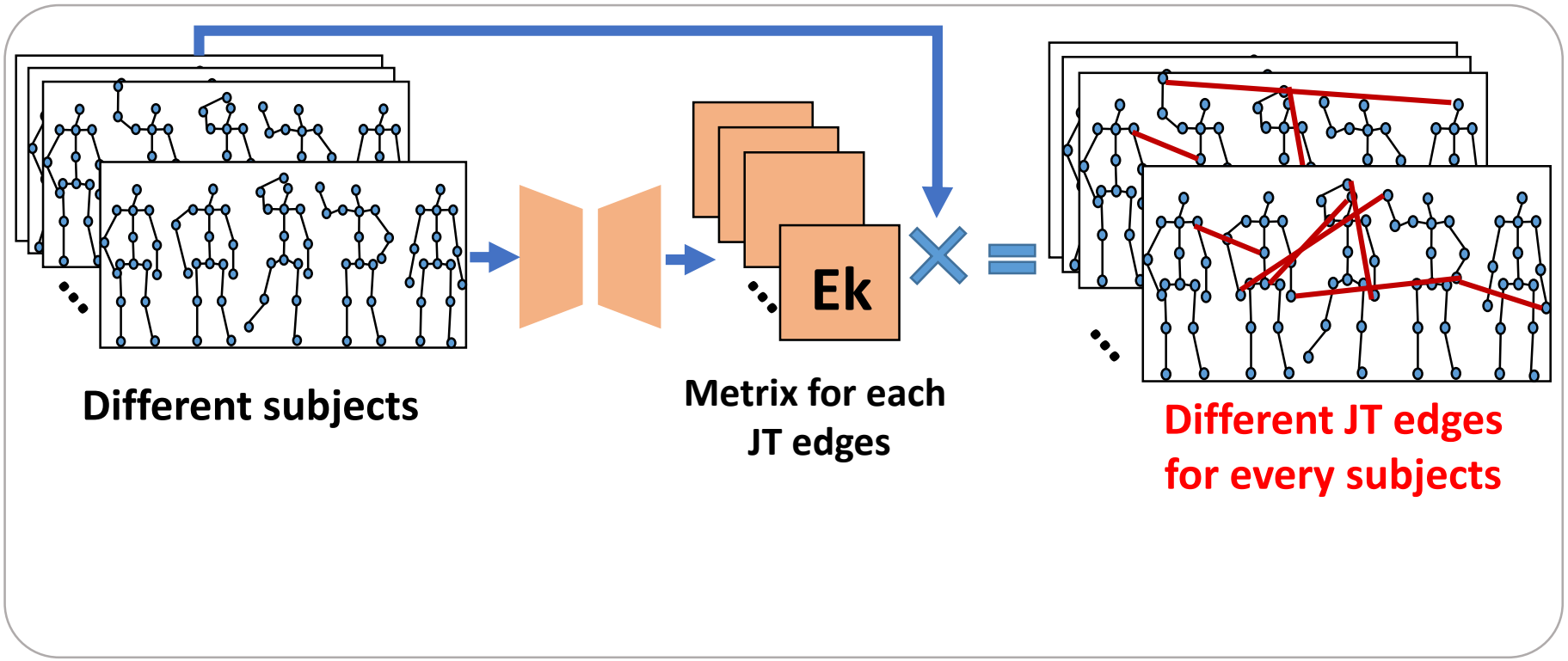
Updating weights based on outputs

Proposed algorithm

Overall frameworks



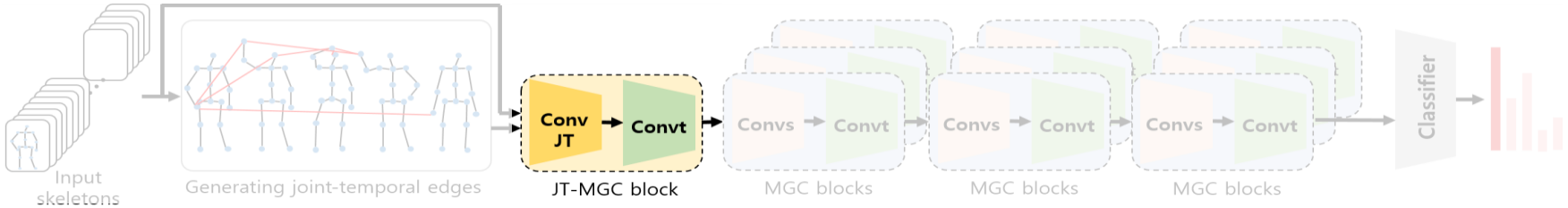
2) Multiply matrix through encoder-decoder - Ek



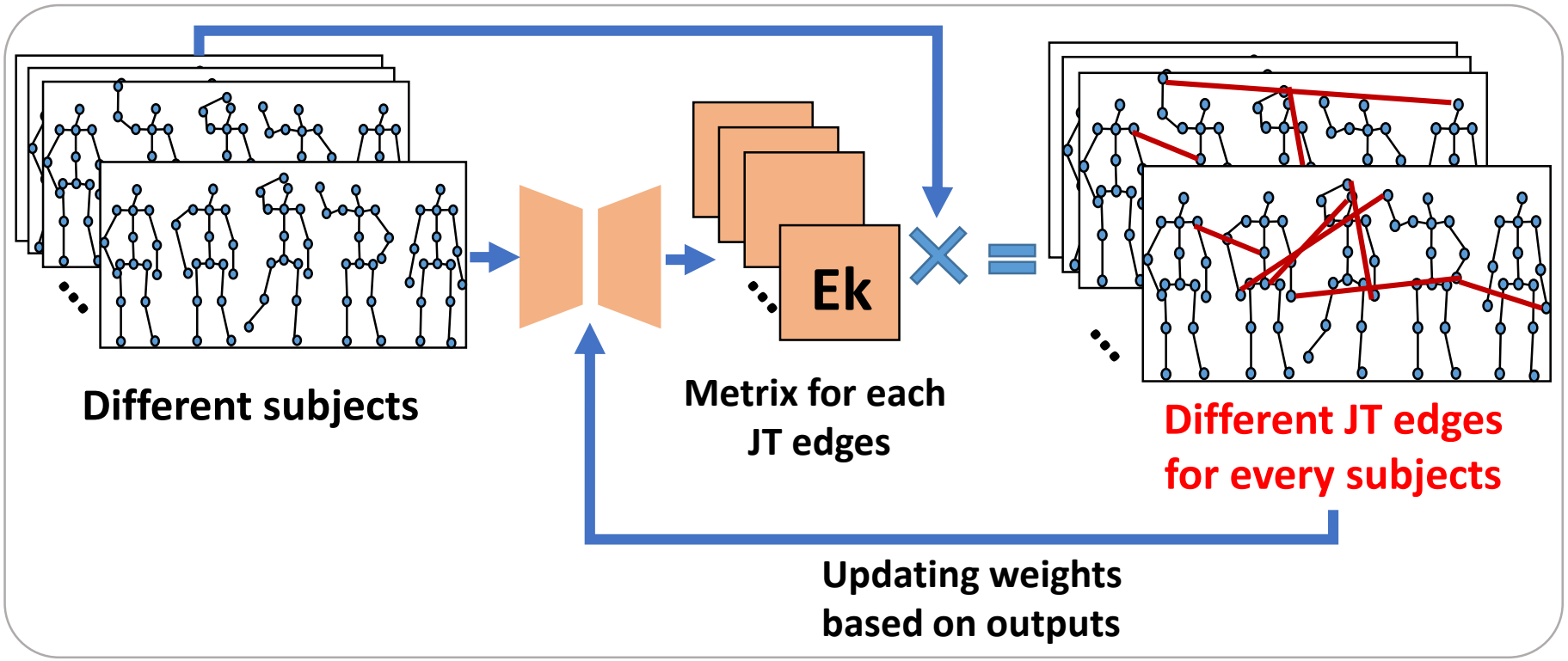
Proposed algorithm

PerCV
Perception & Computer Vision Lab.

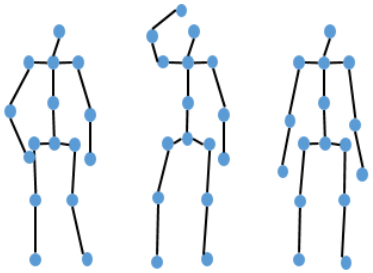
Overall frameworks



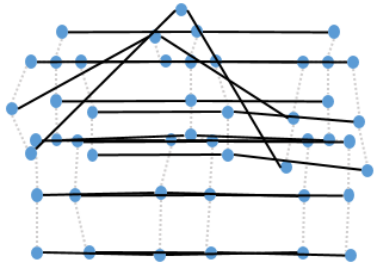
2) Multiply matrix through encoder-decoder - Ek



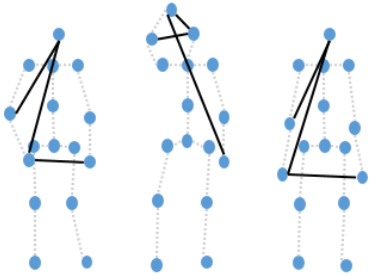
Results



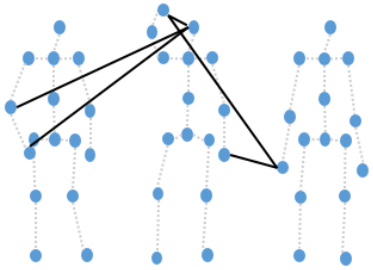
(a) Neighbor joint edge



(b) Neighbor temporal edge



(c) Distant joint edge

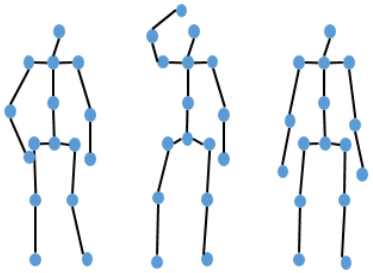


(d) Joint-temporal edge

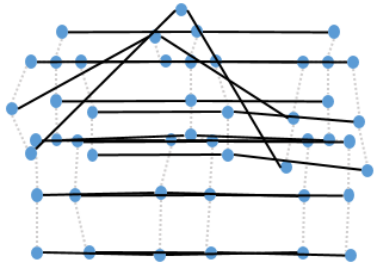
method	CS(%)
(a)+(b)	81.5
(a)+(b)+(c)	88.5
(a)+(b)+(c)+(d)	**90.4**

Method	Year	CS(%)	CV(%)
PA-LSTM [18]	2016	62.9	70.3
ST-LSTM+TS [13]	2016	69.2	77.7
STA-LSTM [23]	2017	73.4	81.2
VA-LSTM [27]	2017	79.4	87.6
MTLN [5]	2017	79.6	84.8
ST-NBMIM [25]	2018	80.0	84.2
MTCNN [6]	2018	81.1	87.4
ST-GCN [26]	2018	81.5	88.3
DPRL+GCNN[24]	2018	83.5	89.8
SR-TSL [22]	2018	84.8	92.4
AS-GCN [10]	2019	86.8	94.2
2S-AGCN [20]	2019	88.5	95.1
DGNN [19]	2019	89.9	96.1
JT-MGCN(proposed)	**2019**	**90.40**	**95.78**

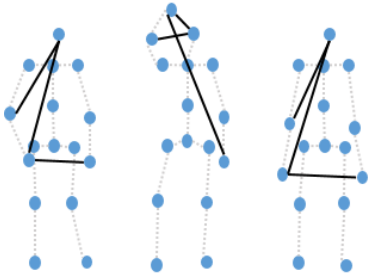
Results



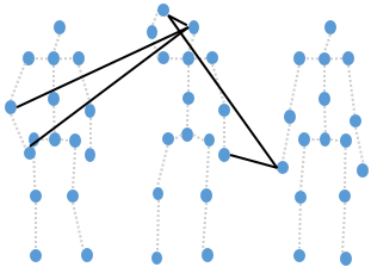
(a) Neighbor joint edge



(b) Neighbor temporal edge



(c) Distant joint edge



(d) Joint-temporal edge

	Method	CS(%)
Without shuffling	(a)+(b)+(c)	88.5
	(a)+(b)+(c)+(d)	90.4
With shuffling	(a)+(b)+(c)	13.4
	(a)+(b)+(c)+(d)	**8.8**

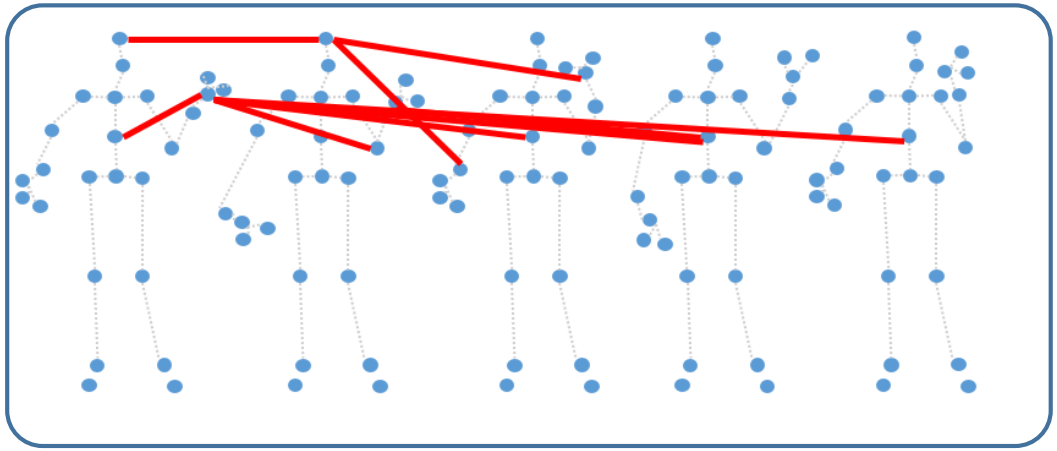
JT edges give More Meaningful features!

Results

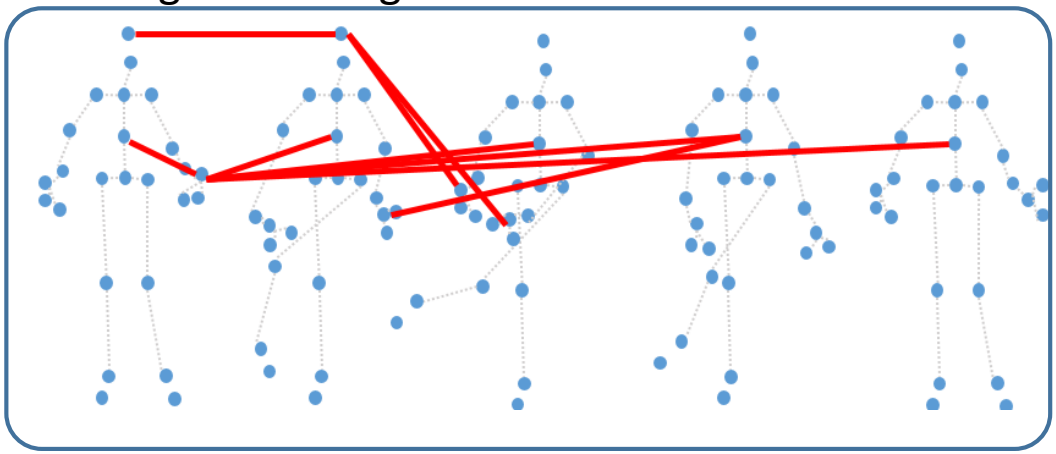
PerCV
Perception & Computer Vision Lab.

Visualization of Joint-Temporal edges

Drink water



Kicking Something



Independent edges are stronger than others!

Q & A