西安电子科技大学
XIDIAN UNIVERSITY

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

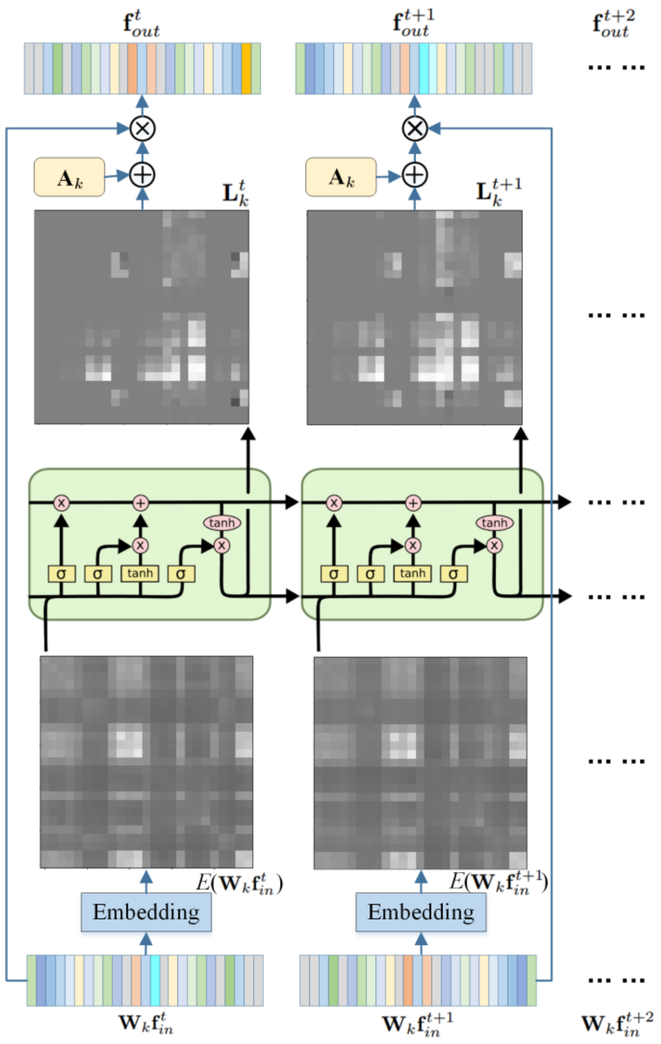
Recurrent Graph Convolutional Networks for Skeleton-based Action Recognition

Guangming Zhu, Lu Yang, Liang Zhang, Peiyi Shen, and Juan Song

嵌入式技术与视觉处理研究中心
EMBEDDED TECHNOLOGY & VISION PROCESSING RESEARCH CENTER

Embedded Technology & Vision Processing Research Center, Xidian University, China

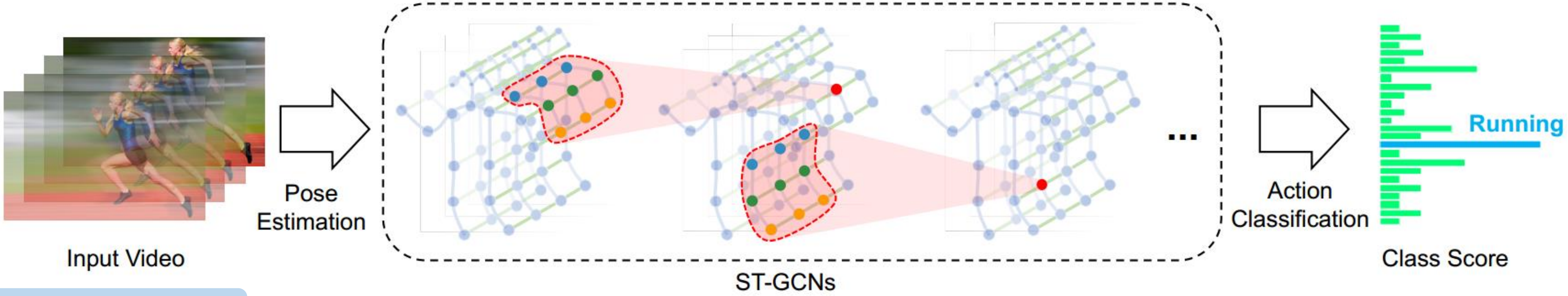
■ What have we done?



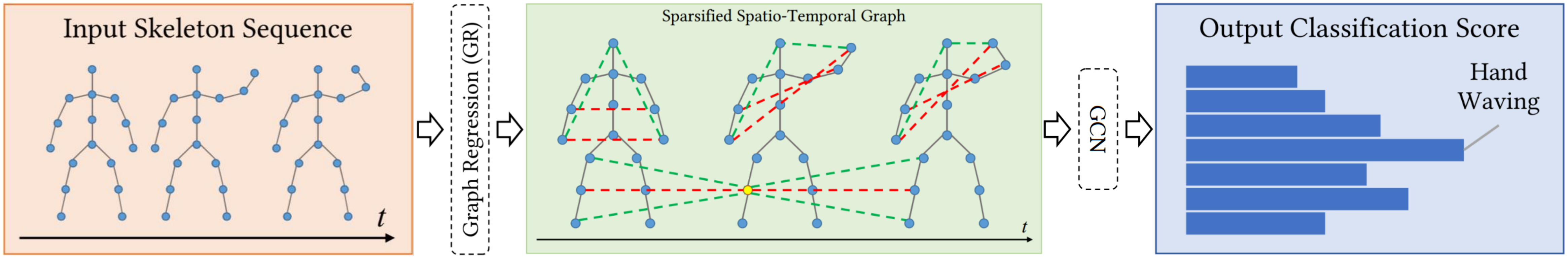
- Propose a **recurrent graph convolutional network (R-GCN)**
- Learn the graph topologies for **different layers**, **different time steps** and **different kinds of actions**
- Evolve the graph topology **across the consecutive frames adaptively**
- Capture the **interactions of different subsets of joints** beyond corresponding neighboring joints in the neighboring frames

Existing graph convolutional networks

- **ST-GCN**



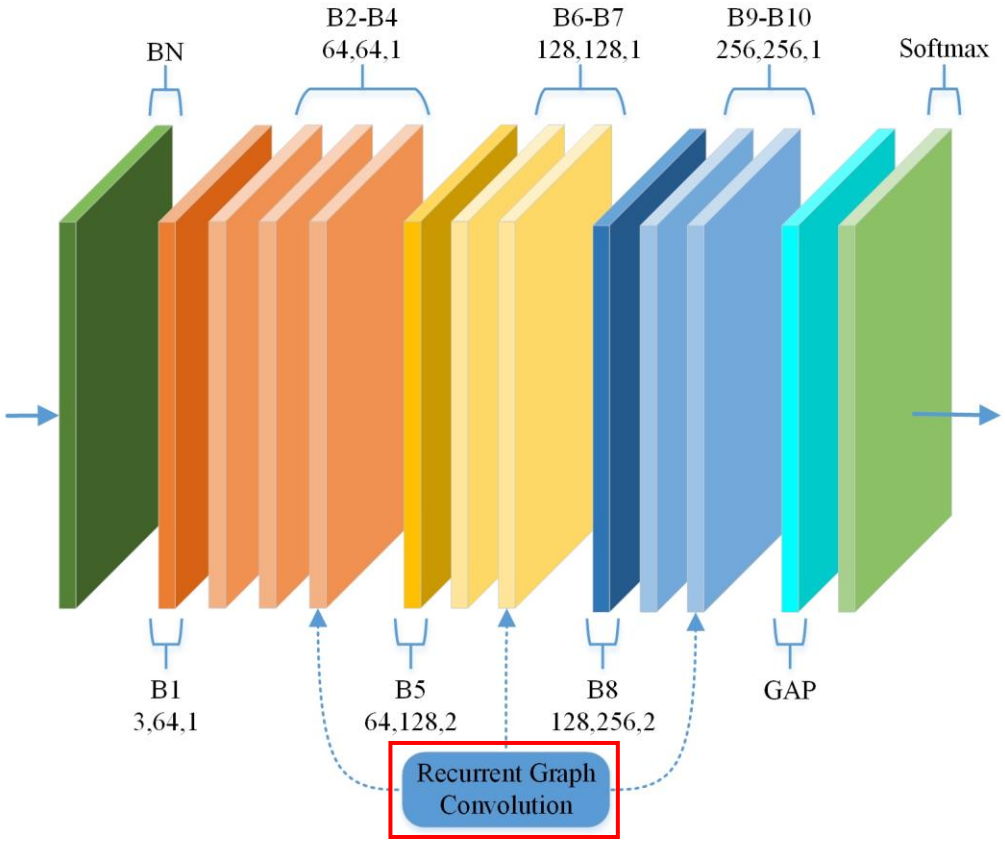
- **GR-GCN**



[1] S. Yan, Y. Xiong, D. Lin, and xiaoou Tang, "Spatial temporal graph convolutional networks for skeleton-based action recognition," in AAAI Conference on Artificial Intelligence (AAAI), 2018, pp. 7444–7452.

[2] X. Gao, W. Hu, J. Tang, P. Pan, J. Liu, and Z. Guo, "Generalized graph convolutional networks for skeleton-based action recognition." arXiv preprint arXiv:1811.12013, 2018.

■ Architecture Overview



- Apply the adaptive graph convolution and the recurrent graph convolution into ST-GCN
- "**Graph Convolution** – BN – ReLU – **Temporal Convolution** – BN - ReLU" structure for each block (B1 – B10)
- A residual connection for each block

Experimental Results

- **Ablation Study**

Networks	BP on Embedding	Natural Connectivity	Accuracy(%)	Parameter Size
R-GCN + Dot-Prod		√	93.8	3.2M
			93.5	
	√	√	**94.2**	
	√		93.8	
R-GCN + Gaussian		√	93.9	3.2M
	√	√	94.1	
R-GCN + EmbedGaussian		√	93.9	3.8M
	√	√	**94.2**	
R-GCN* + Dot-Prod		√	93.7	3.3M
			93.2	
	√	√	93.9	
	√		93.7	

■ Experimental Results

● Comparison with the State-of-the-art

Networks	Stream	NTU RGB+D		Kinetics-Skeleton	
		X-Sub(%)	X-View(%)	Top-1(%)	Top-5(%)
2s-AGCN [9]	Joint	-	93.2	35.1	57.1
	Bone	-	93.2	33.3	55.7
	Both	88.5	95.1	36.1	58.7
AGC-LSTM [21]	Joint	87.5	93.5	-	-
	Part	87.5	93.8	-	-
	Both	**89.2**	95.0	-	-
R-GCN	Joint	86.3	94.2	35.1	57.5
	Bone	87.6	94.1	34.6	57.1
	Both	**89.2**	**95.6**	**36.9**	**59.8**

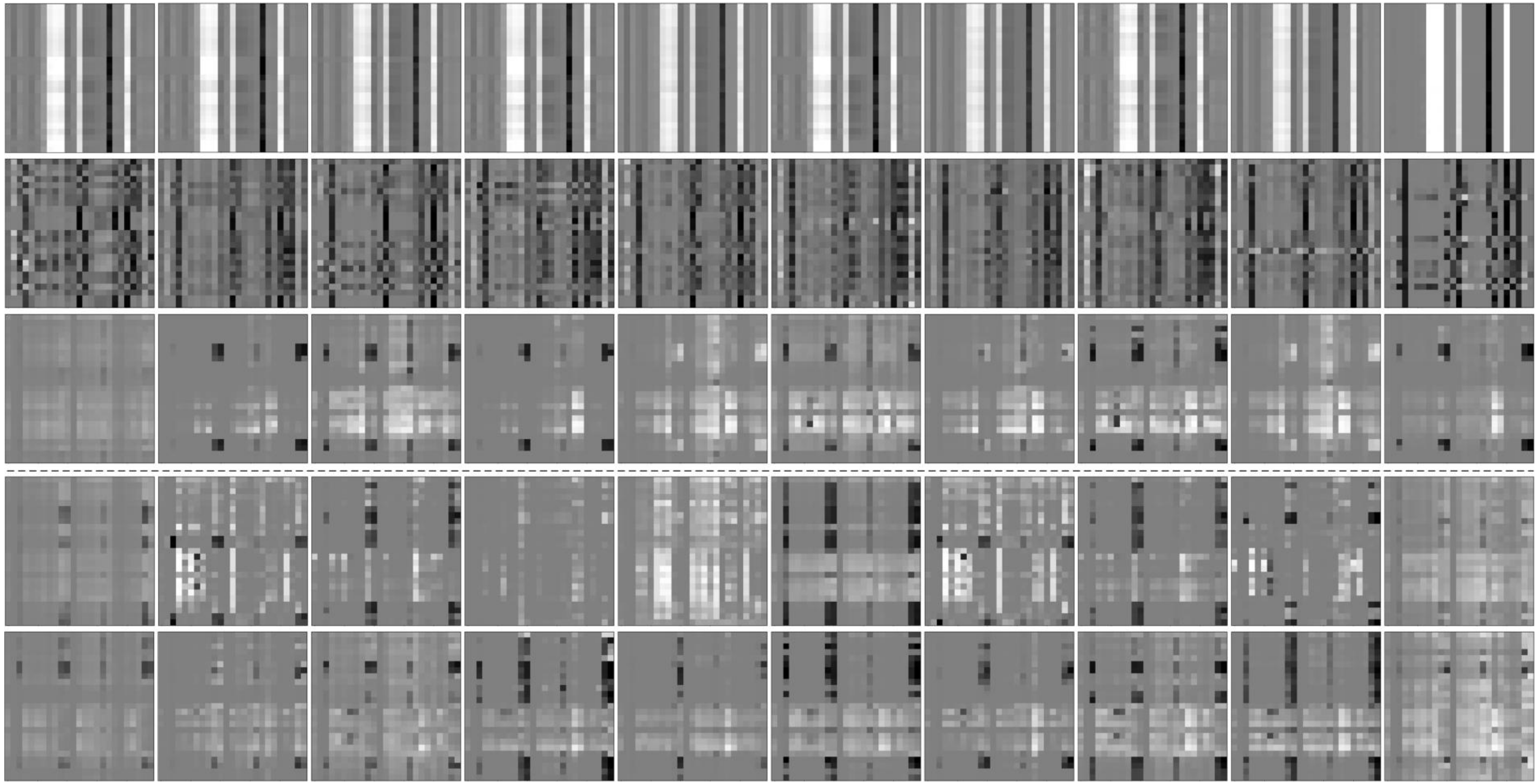
Experimental Results

Comparison with the State-of-the-art

Methods	X-Sub(%)	X-View(%)
Lie Group [3]	50.1	82.8
HBRNN [4]	59.1	64.0
Deep LSTM [5]	60.7	67.3
ST-LSTM [6]	69.2	77.7
STA-LSTM [26]	73.4	81.2
VA-LSTM [27]	79.2	87.8
ARRN-LSTM [28]	80.7	88.8
Ind-RNN [29]	81.8	88.0
Two-Stream 3DCNN [30]	66.8	72.6
TCN [31]	74.3	83.1
Clips+CNN+MTLN [7]	79.6	84.8
Synthesized CNN [32]	80.0	87.2
CNN+Motion+Trans [33]	83.2	89.3
3Scale ResNet152 [34]	85.0	92.3
PA-GCN [16]	80.4	82.7
ST-GCN [8]	81.5	88.3
DPRL+GCNN [35]	83.5	89.8
BPLHM [36]	85.4	91.1
3s RA-GCN [37]	85.9	93.5
AS-GCN [10]	86.8	94.2
PB-GCN [15]	87.5	93.2
2s-ASGCN [13]	88.3	95.4
2s-AGCN [9]	88.5	95.1
AGC-LSTM [21]	**89.2**	95.0
Two-stream R-GCN	**89.2**	**95.6**

Methods	Top-1(%)	Top-5(%)
Feature Enc. [38]	14.9	25.8
Deep LSTM [5]	16.4	35.3
TCN [31]	20.3	40.0
ST-GCN [8]	30.7	52.8
BPLHM [36]	33.4	56.2
2s-ASGCN [13]	34.5	56.9
AS-GCN [10]	34.8	56.5
2s-AGCN [9]	36.1	58.7
Two-stream R-GCN	**36.9**	**59.8**

■ Visualization of the Evolved Graph Topologies



西安電子科技大學
XIDIAN UNIVERSITY

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

Thanks for listening!

嵌入式技术与视觉处理研究中心
EMBEDDED TECHNOLOGY & VISION PROCESSING RESEARCH CENTER

Embedded Technology & Vision Processing Research Center, Xidian University, China