FUJITSU
shaping tomorrow with you

Temporal Extension Module for Skeleton-based Action Recognition

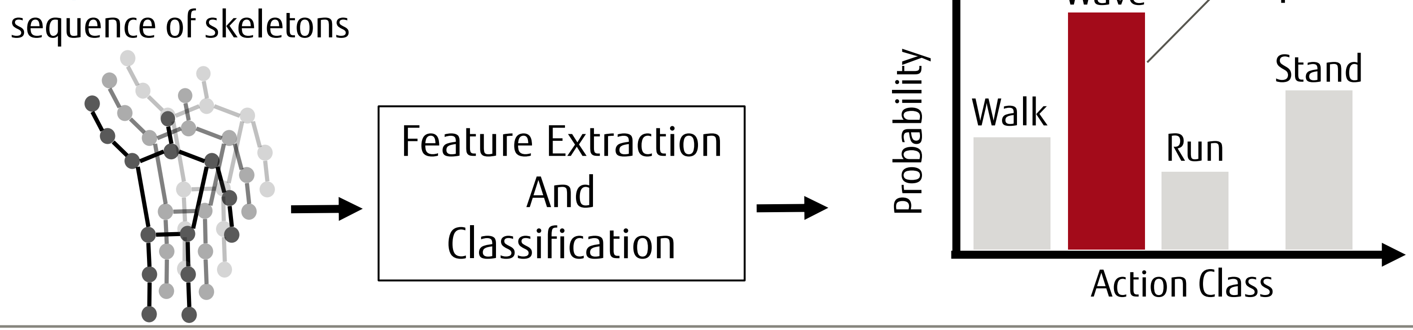
Yuya Obinata, Takuma Yamamoto
FUJITSU LABORATORIES LTD.

Agenda

- Overview
- Motivation
  - ST-GCN
    - Drawback
- Proposed Method
  - Temporal Extension Module (TEM)
- Implementation of TEM
- Ablation Study
- Comparison of State-of-the-art

Overview

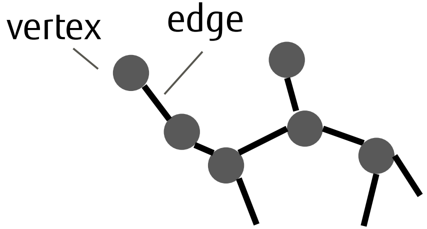
- **Expectation of Action recognition**
  - Applications in human behavior understanding and human social behavior
- **Mainstream method in action recognition**
  - Methods using RGB images
  - **Methods using sequence of skeletons**
    - Low-cost computing resources
    - Robust by noise from various illumination conditions

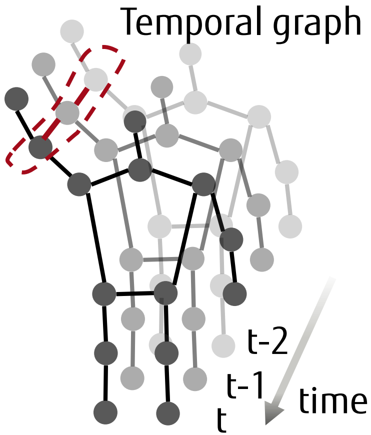


Motivation

FUJITSU

- ST-GCN (AAAI 2018)
  - The well-known method using graph convolutional network
    - Represents human joints as vertices and their natural connections in human body as edges
    - Draws attention owing of high performance
  - Temporal graph
    - Adds edges to same vertex on temporal dimension
    - Extracts feature of trajectory of same joints
  - Drawback
    - Cannot extract feature of correlative movement between each joint on the inter-frame
      - Does not add edges to another vertex on temporal graph





Temporal graph of elbow

Proposed Method

FUJITSU

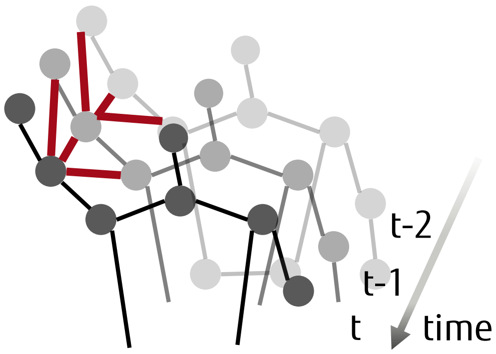
Temporal Extension Module (TEM)

Extends of temporal graph on inter-frame

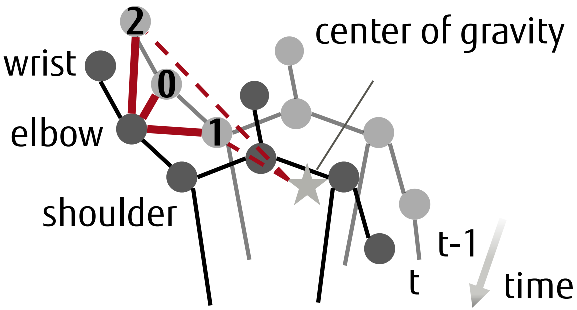
- Adds edges to not only same vertex but also neighboring multiple vertices
- Sets multiple subsets based on length to center of gravity
  - Adds info of kinematics correlation

Advantage

- Extract feature of correlative movement between each joint on inter-frame
  - Good for recognition of action that adjacent joints often move together, such as "throw."
- Apply to existing model easily



Temporal graph with TEM



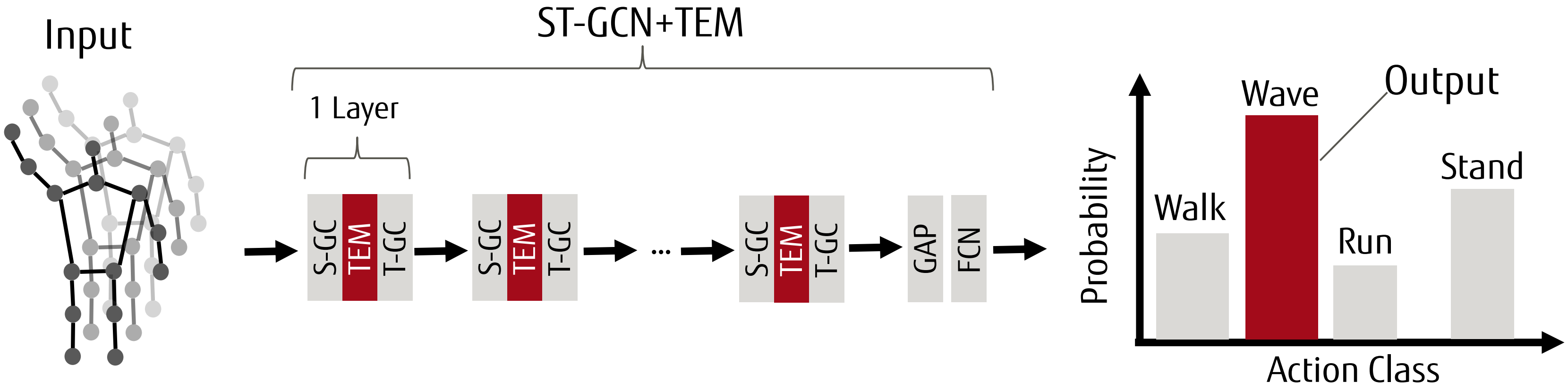
Subset on inter-frame

Implementation of TEM

ST-GCN+TEM

*Spacial Graph Convolution and Temporal Graph Convolution

- We attach our module between *S-GC and T-GC.
  - To expand sampling area for temporal dimension gradually
- We do not change the structure of conventional convolutions.
  - Our module can readily apply to many existing methods.



Ablation Study

- Methods with TEM outperform without TEM.
  - This showed effectiveness of our module.

TABLE I. COMPARISONS OF THE RECOGNITION ACCURACY THE MODELS WITH TEM AND WITHOUT TEM

Methods	NTU-RGB+D		Kinetics-Skeleton	
	CS (%)	*CV (%)*	*Top-1 (%)*	*Top-5 (%)*
ST-GCN	82.6	88.7	32.5	54.9
ST-GCN+TEM	85.2	90.2	34.5	56.7
2s-AGCN	88.6	95.2	36.7	59.8
2s-AGCN+TEM	88.7	95.8	38.6	61.6
MS-AAGCN	90.3	96.1	37.4	60.6
MS-AAGCN+TEM	91.0	96.5	38.0	61.4

Comparison of State-of-the-art

FUJITSU

- Our model achieves state-of-the-art performance.

TABLE II. Comparisons of the Recognition Accuracy with MS-AAGCN+TEM and Current State-of-the-Art Methods on NTU RGB+D Dataset

Methods	NTU-RGB+D	
	CS (%)	*CV (%)*
ST-LSTM (Tree Traversal) + Trust Gate [19]	69.2	77.7
TSRJI (Late Fusion) [27]	73.3	80.3
STA-LSTM [16]	73.4	81.2
ESV (Synthesized+Pre-trained) [26]	80.0	87.2
VA-LSTM [20]	79.6	87.6
ST-GCN [7]	81.5	88.3
Si-GCN [34]	84.2	89.1
CNN-based [24]	83.2	89.3
ARRN-LSTM [23]	81.8	89.6
DPRL+GCNN [8]	83.5	89.8
multi-scale network (ResNet152 + 3scale) [25]	85.0	92.3
Complete GR-GCN model [33]	87.5	94,3
2s-AGCN [9]	88.5	95.1
GCN-NAS [11]	89.4	95.7
DGNN [10]	89.9	96.1
MS-AAGCN [12]	90.0	96.2
BAGCN [22]	90.3	96.3
Sym-GNN [13]	90.1	96.4
MS-AAGCN+TEM(Ours)	**91.0**	**96.5**

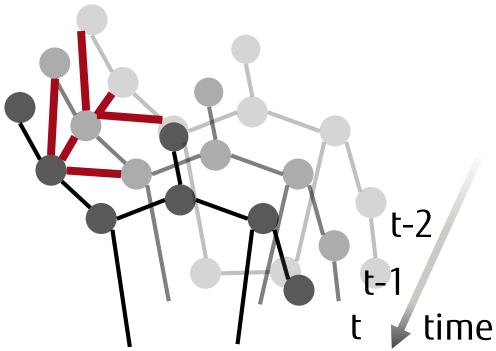
TABLE III. Comparisons of the Recognition Accuracy with 2s-AGCN+TEM and Current State-of-the-Art Methods on the Kinetics-Skeleton

Methods	Kinetics-Skeleton	
	Top-1 (%)	***Top-5 (%)***
ST-GCN [7]	30.7	52.8
2s-AGCN [9]	36.1	58.7
DGNN [10]	36.9	59.6
GCN-NAS [11]	37.1	60.1
Sym-GNN [13]	37.2	58.1
BAGCN [22]	37.3	60.2
MS-AAGCN [12]	37.8	61.0
2s-AGCN+TEM (Ours)	**38.6**	**61.6**

Summary

FUJITSU

- **Proposed method: Temporal Extension Module (TEM)**
  - Extend temporal graph on inter-frame
    - Adds edges to not only same joints but also neighboring multiple joints
  - Advantage
    - Extract feature of correlative movement between each joint on inter-frame
    - Apply to existing model easily



Temporal graph with TEM

- **Experiments on NTU RGB +D and Kinetics-Skeleton dataset**
  - Showed effectiveness of our module
  - Showed that model with our module achieves state-of-the-art performance

shaping tomorrow with you