요약

- ST-GCN의 단점을 개선(주변 joint간의 convolution만 하여 거리가 먼 joint간의 관계성을 학습하지 않음)하기 위해 inter-graph와 intra-graph 로 그래프를 표현하는 방법은 제안
- 중요 joint의 조합을 intra-graph로 표현하여 총 11개의 조합을 구성 (왼팔/오른팔/왼다리/오른다리/왼팔+오른팔/왼팔+왼다리/…/상반신/몸전체)
    - 사족) 사람이 정의한 intra-graph와 inter-graph를 사용하는 것이 올바른 것인지 의문
- intra-graph 간의 관계를 학습할 수 있도록 inter graph를 구성하여 structual graph convolution 진행
- ST-GCN처럼 temporal graph convolution 진행 (현재 time t를 기준으로 t-(kernel_size k/2) ~ t+(kernel_size k/2) convolution)
- NTU RGB+D, HDM05 데이터셋으로 비교/파라미터 실험

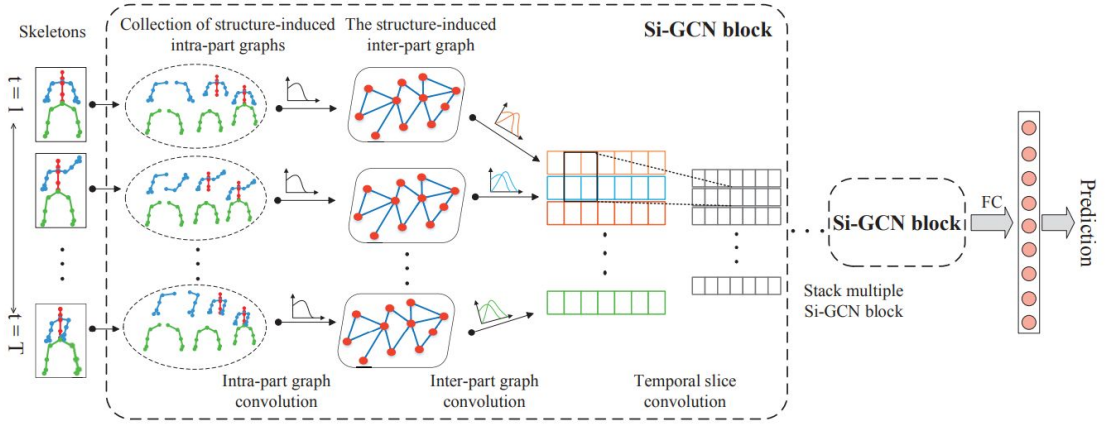

Fig. 1. Architecture of Si-GCN framework. A graph convolution is performed on the structure-induced part-graphs in the collection $\mathcal{G}_{stru}$, another graph convolution is performed on the structure-induced interactive graph, then the temporal convolution is adopted to integrate the characteristics of a sequence, finally the spatial-temporal feature is feed to pooling layer and the softmax classifier.

수식화

- ST-GCN의 단점을 개선하기 위해 intra-graph와 inter-graph 로 그래프를 표현하는 방법은 제안
  식 (5) : Si-GCN Block 묘사
- 중요 joint의 조합을 intra-graph로 표현하여 총 intra-graph num m = 11개의 조합을 구성
  (왼팔/오른팔/왼다리/오른다리/왼팔+왼다리/왼팔+오른다리/오른팔+왼다리/
  오른팔+오른다리상반신/왼팔+척추및몸통+오른팔/왼다리+오른다리/몸전체)
- joint 조합에 따라 Mask M을 만들고 몸전체를 표현하는 11번째 inter-graph에서 추출 =>식 (6)
- intra-graph에 대한 graph convolution 진행
  - 식 (7) : kenel size에 따른 j번째 hidden feature의 intra-graph convolution layer 연산을 수식화
  - 식 (8) : kenel size에 따른 intra-graph convolution layer 전체 연산 수식
  - 식 (9) : 식 (8) 요약
- intra-graph 끼리의 연관성을 가지고 inter-graph graph convolution 연산
  - 어느 intra-graph 끼리 연관성을 설정했는지에 대한 정보가 없음
  - 그 외에는 GCN과 동일 연산
- 식 (11) : 모든 프레임에 대한 GCN 연산 이후 ST-GCN과 같이 시간을 기준으로 convolution (1x1x7 사용)

$$\{\mathcal{G}_{t1}^{intra}, \mathcal{G}_{t2}^{intra}, ..., \mathcal{G}_{tm}^{intra}\} \text{ with } m = 11, t = 1, \cdots, T$$

$$\widetilde{\mathbf{X}} = (g_{inter} * (g_{intra} * \mathbf{X}_{N\times d\times T})) * f_{temp}, \quad (5)$$

$$f(l_j) = \begin{cases} 1, & l_j \in \mathcal{G}_{ti}^{intra}, \\ 0, & l_j \notin \mathcal{G}_{ti}^{intra}. \end{cases} \quad (6)$$

$$\mathbf{y}_{tij} = \sum_{l=1}^{d}\sum_{k=0}^{K-1} \omega_{tijkl} T_k(\widetilde{\mathbf{L}}_i)\mathbf{x}_{tl}, \quad (7)$$

$$\mathbf{Y}_{ti} = \sum_{k=0}^{K-1} T_k(\widetilde{\mathbf{L}}_i)\mathbf{X}_t\mathbf{W}_{tik}^{intra} = [T_0(\widetilde{\mathbf{L}}_i)\mathbf{X}_t, \cdots, T_{K-1}(\widetilde{\mathbf{L}}_i)\mathbf{X}_t] \times [(\mathbf{W}_{ti0}^{intra})^\intercal, \cdots, (\mathbf{W}_{tiK-1}^{intra})^\intercal]^\intercal, \quad (8)$$

$$\mathbf{Z}_{ti}^{intra} = \mathbf{A}_{ti}\mathbf{W}_{ti}^{intra}. \quad (9)$$

$$\mathbf{L}^{inter} \in \mathbb{R}^{m\times m} \qquad \mathbf{Z}_{ti}^{intra} \in \mathbb{R}^{N\times d'}$$

$$\mathbf{Z}_t^{inter} = \sum_{k'=0}^{K'-1} (\widetilde{\mathbf{L}}^{inter})\mathbf{Z}_t^{intra}\mathbf{W}_k^{inter}, \quad (10)$$

$$\mathbf{Z} = \mathbf{Z}^{inter} * f, \quad (11)$$

실험

- NTU RGB+D, HDM05 데이터셋으로 실험
  - HDM05 (2007)
    - skeleton-based dataset으로 joint 24개의 mocap data
- 표 1) NTU RGB+D dataset의 Cross-Subject/Cross-View 정확도를 타 모델과 비교
- 표 2) HDM05 정확도를 타 모델과 비교
- 표 3) NTU RGD+D dataset에서 논문이 제안한 intra-graph/inter graph 사용/비사용에 따른 정확도 실험
- 표 4) Convolution 시 kernel size 변화에 따른 정확도 실험
- 그림 2) NTU RGB+D dataset에서 cross-subject에 대한 confusion matrix
- 그림 3) HMD05 dataset에 대한 confusion matrix

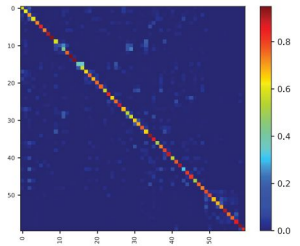

Fig. 2. confusion matrix of our Si-GCN in the CS bechmark on NTU D+RGB dataset.

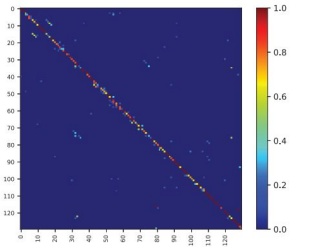

Fig. 3. Confusion matrix of our Si-GCN on HMD05 dataset.

TABLE I
COMPARISONS OF SKELETON-BASED ACTION RECOGNITION PERFORMANCE ON THE NTU RGB+D DATASET.

Method	Accuracy (%)	
	CS	CV
Lie Group [25]	50.1	52.8
VA-LSTM [39]	79.4	87.6
CNN + MTLN [40]	79.6	84.8
Deep STGC [20]	74.9	86.3
STGCN [19]	81.5	88.3
Si-GCN	**84.15**	**89.05**

TABLE II
COMPARISONS OF SKELETON-BASED ACTION RECOGNITION PERFORMANCE ON THE HDM05 DATASET.

Method	Accuracy(%)
SPDNet [41]	61.45 ± 1.12
Lie Group [25]	70.26 ± 2.89
LieNet [42]	75.78 ± 2.26
P-LSTM [21]	73.42 ± 2.05
Deep STGC [20]	85.29 ± 1.33
STGCN [19]	82.13 ± 2.39
Si-GCN	85.45 ± 2.78

TABLE III
PERFORMANCE COMPARISONS WITH DIFFERENT GRAPH CONVOLUTIONS OF SI-GCN ON RGB+D DATASET.

Method	Accuracy (%)	
	CS	CV
Si-GCN only with intra-part convolution	81.23	85.65
Si-GCN only with inter-part convolution	76.78	81.02
Si-GCN	**84.15**	**89.10**

intra + inter 다 같이 써야 좋다.

TABLE IV
PERFORMANCE COMPARISONS WITH DIFFERENT ORDER OF SI-GCN ON RGB+D DATASET.

Method	Accuracy (%)	
	CS	CV
one	80.41	86.52
two	79.88	87.68
three	**84.15**	**89.05**
four	80.72	88.67

hop 수를 너무 키우면 안됨

실험(계속)

- 이전에 정리한 논문들에 있는 같은 데이터셋 기준의 실험결과 표 비교
  - 왼쪽 표1)NTU RGB+D에 대한 본 논문의 실험결과
  - 오른쪽 표 4) NTU RGB+D에 대한 AS-GCN의 실험결과

TABLE I
COMPARISONS OF SKELETON-BASED ACTION RECOGNITION PERFORMANCE ON THE NTU RGB+D DATASET.

Method	Accuracy (%)	
	CS	CV
Lie Group [25]	50.1	52.8
VA-LSTM [39]	79.4	87.6
CNN + MTLN [40]	79.6	84.8
Deep STGC [20]	74.9	86.3
STGCN [19]	81.5	88.3
Si-GCN	**84.15**	**89.05**

TABLE II
COMPARISONS OF SKELETON-BASED ACTION RECOGNITION PERFORMANCE ON THE HDM05 DATASET.

Method	Accuracy(%)
SPDNet [41]	61.45 ± 1.12
Lie Group [25]	70.26 ± 2.89
LieNet [42]	75.78 ± 2.26
P-LSTM [21]	73.42 ± 2.05
Deep STGC [20]	85.29 ± 1.33
STGCN [19]	82.13 ± 2.39
Si-GCN	85.45 ± 2.78

Table 4. Comparison of action recognition performance on NTU-RGB+D. The classification accuracies on both Cross-Subject and Cross-View benchmarks are presented.

Methods	Cross Subject	Cross View
Lie Group [27]	50.1%	52.8%
H-RNN [6]	59.1%	64.0%
Deep LSTM [22]	60.7%	67.3%
PA-LSTM [22]	62.9%	70.3%
ST-LSTM+TS [20]	69.2%	77.7%
Temporal Conv [14]	74.3%	83.1%
Visualize CNN [21]	76.0%	82.6%
C-CNN+MTLN [13]	79.6%	84.8%
ST-GCN [29]	81.5%	88.3%
DPRL [26]	83.5%	89.8%
SR-TSL [23]	84.8%	92.4%
HCN [18]	86.5%	91.1%
AS-GCN (Ours)	**86.8%**	**94.2%**