Decoupling Graph Convolutional Network with Adaptive Normalization

Haiyun Zhou[1] , Xuezhi Xiang[2]
[1]College of Public Security, Nanjing Forest Police College
[2]College of Information and Communication Engineering, Harbin Engineering University

• Problem Definition and Contribution:

➢ Goal: classify the skeleton sequences acquired by depth sensors or pose estimation.

➢ Key Contributions:

- An action recognition network model which combines the decoupling graph convolution and adaptive normalization module.
- A higher action recognition accuracy (86.8 X-sub, 88.4 X-setup) is achieved on the NTU-RGBD-120 skeleton dataset.

• DC-GCN and AN :

➢ DC-GCN: using multiple groups of trainable adjacency matrices to spatially aggregate features on different channels.

$$X' = A_1 X_1 W \,\|\, A_2 X_2 W \,\|\cdots\|\, A_n X_n W$$

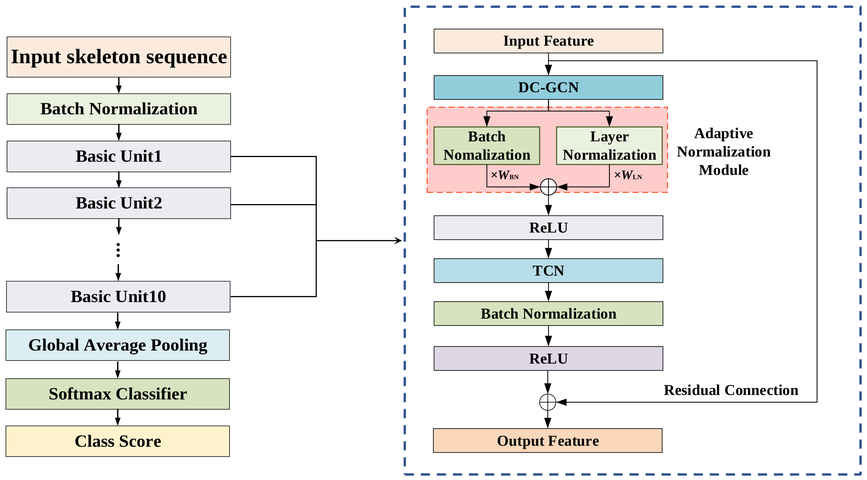
➢ AN: calculating the weighted sum from the results of BN and LN through trainable weight coefficients in different layers of the network.

$$f' = \alpha \frac{f - \sum_k w_k \mu_k}{\sqrt{\sum_k w_k' \sigma_k^2 + \varepsilon}} + \beta$$

• Proposed Model:

➢ Model:

- The overall architecture:



• Experiments:

➢ Dataset:

- NTU-RGBD-120 dataset: 120 classes of actions, 114480 action samples. The dataset is filmed by 106 volunteers and divided into 32 setups with different positions and backgrounds.
- Strategies of distinguishing training set and the validation set:

a) Cross-Subject (X-sub): distinguished by the volunteers

b) Cross-Setup (X-Setup: distinguished by the setup ID

➢ Evaluation:

- Statistical Results:

TABLE I. RESULTS ON THE TEST OF NTU-RGBD-120 DATASET

Methods	X-sub (%)	X-setup (%)
SGN	79.2	81.5
2s-AGCN	82.9	84.9
Shift-GCN	85.9	87.6
DC-GCN	86.5	88.1
Ours	**86.8**	**88.4**

- Ablation Analysis:

TABLE II. ABLATION STUDY ON THE NTU-RGBD-120 DATASET

Methods	X-sub (%)	X-setup (%)
DC-GCN	82.4	84.3
DC-GCN+SN	82.5	**84.6**
DC-GCN+AN	**82.6**	**84.6**