SEPARABLE-SPARSE GRAPH CONVOLUTIONAL NETWORK

Sampieri A. - Sapienza, University of Rome
sampieri@diag.uniroma1.it

ICVSS 2022
Sicily ~ 10-16 July
International Computer Vision Summer School

Abstract

A new Separable-Sparse Graph Convolutional Network (SeS-GCN) for pose forecasting is proposed. SeS-GCN bottlenecks the interaction of the spatial, temporal and channel-wise dimensions in GCNs. It learns sparse adjacency matrices by a teacher-student framework. SeS-GCN improves the SoA in human pose forecasting by 3.6% on Human3.6M and it achieves the best performance in CHICO. It reaches an average error of 53.0mm with a run time of 2.3msec, enabling cobots to be aware of human operators.

Applications

- *Human-Robot Collaboration*: the study of collaborative processes where human and robot agents work together to achieve shared goals and avoid collisions;
- *Game Forecasting*: predicting the trajectory of players on playing fields such as football and basketball;
- *Human Pose Forecasting*, *Anomaly Detection* and *Action Recognition*.

Datasets & Results

- We obtain SoA results on an established dataset such as *Human3.6M*, improving performance by 3.6%;
- Our model outperforms recent SoA techniques on CHICO (Cobots and Humans in Industrial COllaboration), with an average error of 53.0mm.

	$H36M$	$CHICO$	$RunTime$
$W.Mao, ECCV20$	112.1	76.4	9.1×10^{-3}
$L.Dang, ICCV21$	114.1	136.5	25×10^{-3}
$STSGCN[1]$	75.6	59.0	2.3×10^{-3}
$SeS-GCN$	72.9	53.0	2.3×10^{-3}

Separable & Sparse Graph Convolutional Network

SeS-GCN is an accurate, memory efficient and fast GCN by bridging three diverse research directions: **I.** Space-time separable adjacency matrices; **II.** Depth-wise separable graph convolutions; **III.** Sparse adjacency matrices. This results in an all-Separable and Sparse GCN encoder for the human body kinematics, from which the future frames are forecast by a Temporal Convolutional Network.

I. Space-Time Separability

A Space-Time separable GCN [1] factorizes the adjacency matrix A of a GCN into two terms A_s and A_t.

$$\mathcal{X}^{(l+1)} = \sigma\left(A_s^{(l)} A_t^{(l)} \mathcal{X}^{(l)} W^{(l)}\right)$$

It reduces the memory-footprint of a GCN by approx. *4x*.

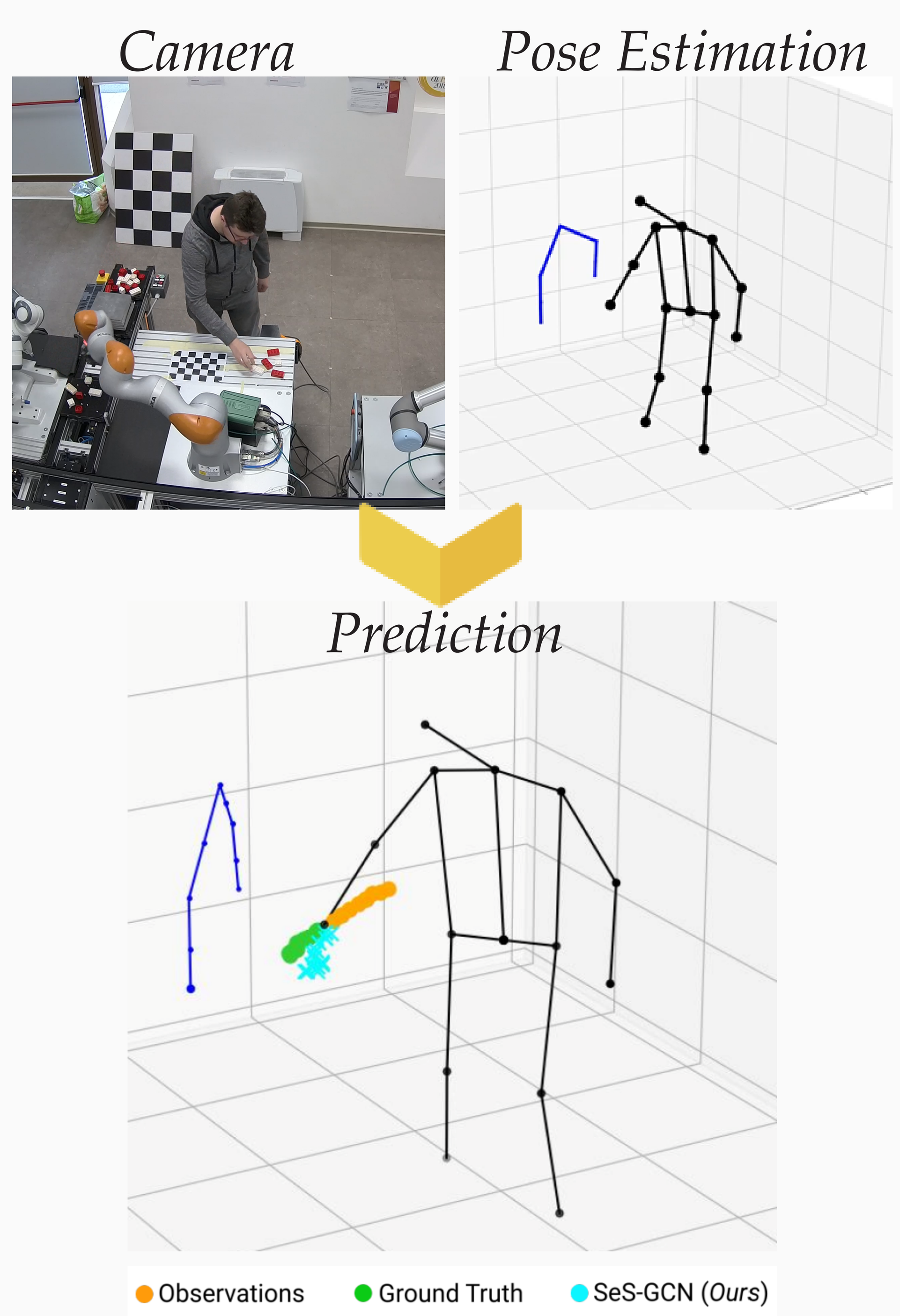
II. Depth-wise Convolution

Depth-wise (DW) graph convolution for pose forecasting is considered. Its formulation bottlenecks the interplay of space and time with the channels of the graph convolution.
The resulting all-separable model is formulated as such:

$$\mathcal{H}^{(l)} = \gamma\left(A_s^{(l)} A_t^{(l)} \mathcal{X}^{(l)} W_{DW}^{(l)}\right)$$
$$\mathcal{X}^{(l+1)} = \sigma\left(\mathcal{H}^{(l)} W_{\text{MLP}}^{(l)}\right)$$

Pose Forecasting

The following is the process involved, from the acquisition of the sequences to the human pose forecasting.



The sequences were acquired by the *ICE Lab* at the University of Verona

III. Sparse GCN

Sparsification is used to improve the efficiency (memory and runtime) of neural networks. The learned masks $\mathcal{M}$ perform a pruning of the adjacency matrices.
Then SeS-GCN:

$$\mathcal{H}^{(l)} = \gamma\left((\mathcal{M}_s^{(l)} \odot A_s^{(l)})(\mathcal{M}_t^{(l)} \odot A_t^{(l)})\mathcal{X}^{(l)} W_{DW}^{(l)}\right)$$
$$\mathcal{X}^{(l+1)} = \sigma\left(\mathcal{H}^{(l)} W_{\text{MLP}}^{(l)}\right)$$

Take Home Message

- SeS-GCN is an all-separable framework and achieves SoA on human pose forecasting;
- It is an efficient network and requires only 2.3msec at inference time.

References & Acknowledge

[1] Sofianos T.*, Sampieri A.*, Franco L, Galasso F., Space-Time-Separable Graph Convolutional Network for Pose Forecasting, in *ICCV*, 2021
[2] Chollet F., Xception: Deep learning with depthwise separable convolutions, in *CVPR*, 2017
[3] LeCun V., Denker J., Solla S.: Optimal brain damage, in *NIPS*, 1989

Special thanks to Team Verona for creating the dataset, to Guido D'Amely for sharing his time with me, and to professor Fabio Galasso for showing us the way.

PINLAB

Feel free to contact us!
Code, Video and other...