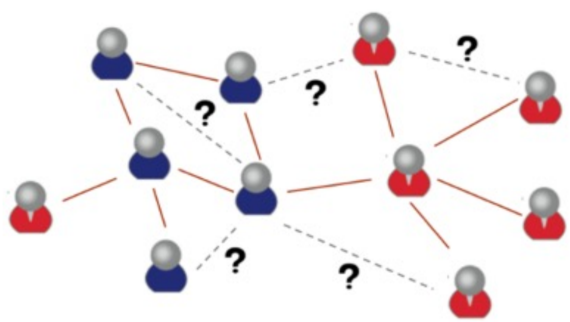
SIA-GCN: A Spatial Information Aware Graph Neural Network with 2D Convolutions for Hand Pose Estimation

Deying Kong, Haoyu Ma and Xiaohui Xie

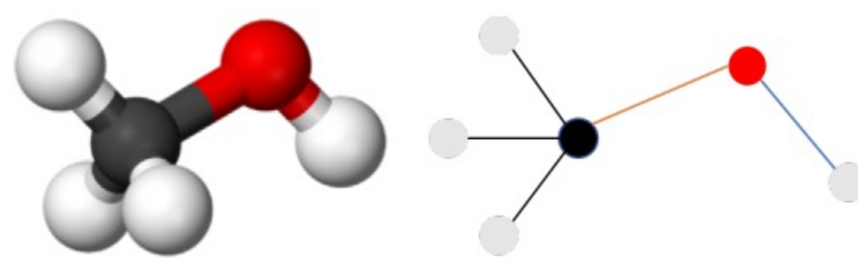
University of California, Irvine

Background

Graph Neural Networks have shown success in many application domains such as computer vision, social networks and chemistry.

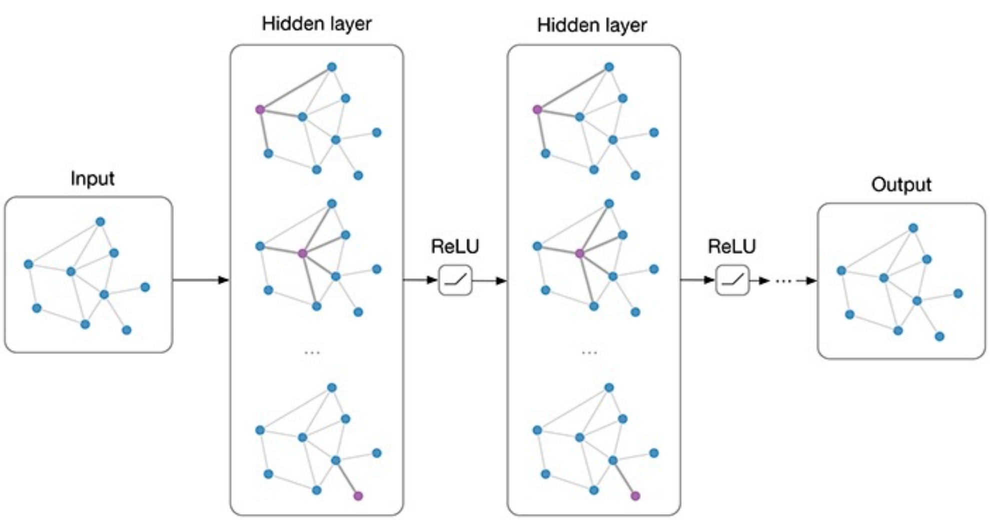


(a) Social network



(b) molecule

Graph Convolutional Network (GCN) by Thomas Kipf



$$H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right)$$

$\tilde{A}$ Adjacency matrix with self connections

$\tilde{D}$ Degree matrix

$H^{(l)} \in \mathbb{R}^{N \times M}$ Matrix of activations in the l-th layer

N Number of nodes in the graph

M Length of 1-d feature at each node

$W^{(l)}$ Trainable weight matrix of layer l

Limitations of the vanilla GCN

- Only processes 1-d feature at each node

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}\boxed{H^{(l)}}W^{(l)}\right)$$

$$H^{(l)} \in \mathbb{R}^{N\times M}$$

What if the feature at each node is 2-dimensional, e.g., 2D confidence maps?

Resize 2-d feature to 1-d feature ?

❌ Would lose spatial information.

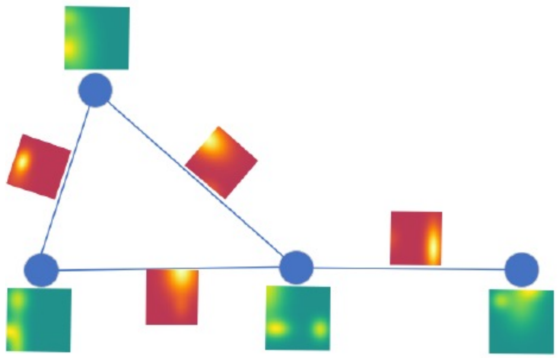
- All nodes share the same weight matrix W

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}\boxed{W^{(l)}}\right)$$

What if neighbouring nodes along different edges have different relationships?

SIA-GCN: A Spatial Information Aware Graph Neural Network with 2D Convolutions

- 2D features at each node
- 2D learnable convolution kernels along each edge
- Different 2D kernels for different edges



SIA-GCN: Propagation Rule

$$X^{(l+1)} = \sigma\left(\hat{A}\left(\left(BX^{(l)}\right)\tilde{\circledast}F^{(l)}\right)\right)$$

$\mathcal{G} = (\mathcal{V}, \mathcal{E})$: Graph

$\mathcal{V} = \{v_1, v_2, \cdots v_K\}$: The set of all nodes

K : Number of nodes in the graph

$\mathcal{E}$: The set of all edges

$\tilde{\circledast}$: Channel-wise 2D convolutional operation

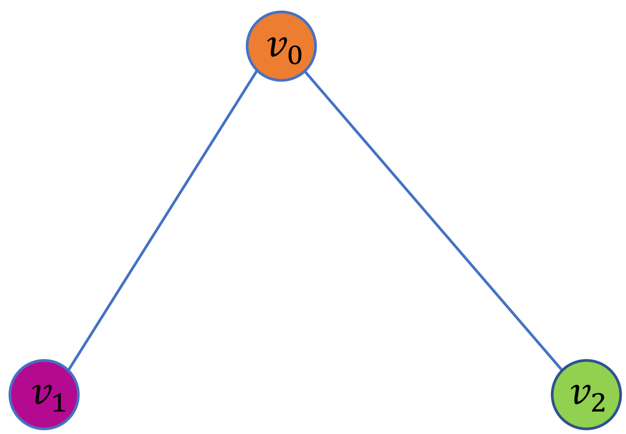
$X \in \mathbb{R}^{K\times h\times w}$: Features of all nodes

$F \in \mathbb{R}^{|\mathcal{E}|\times h'\times w'}$: Learnable kernels along all edges

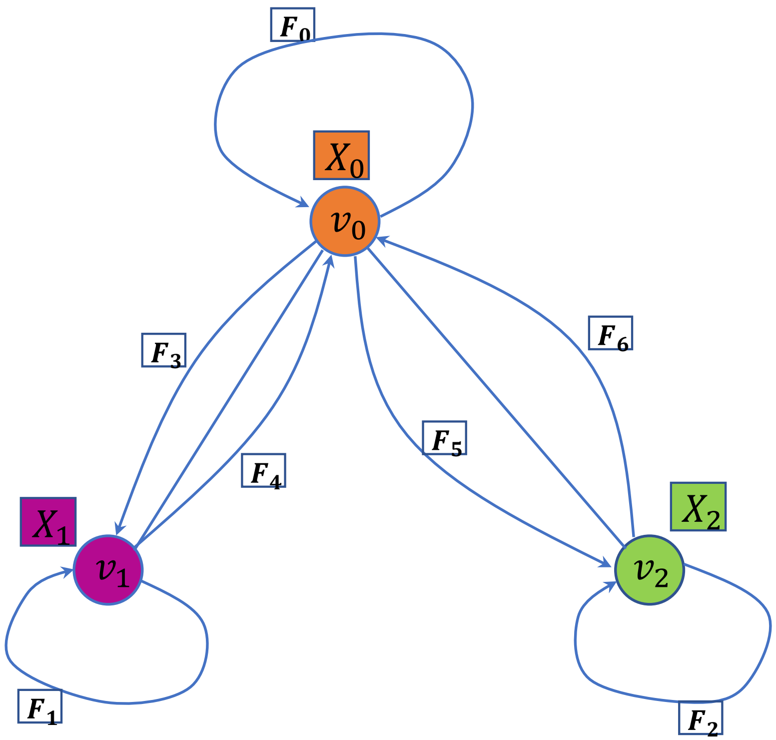
$B \in \mathbb{R}^{|\mathcal{E}|\times K}$: Broadcast matrix

$\hat{A} \in \mathbb{R}^{K\times|\mathcal{E}|}$: Aggregation matrix

SIA-GCN: A simple example



SIA-GCN: A simple example



Expand undirected edges to directed edges.

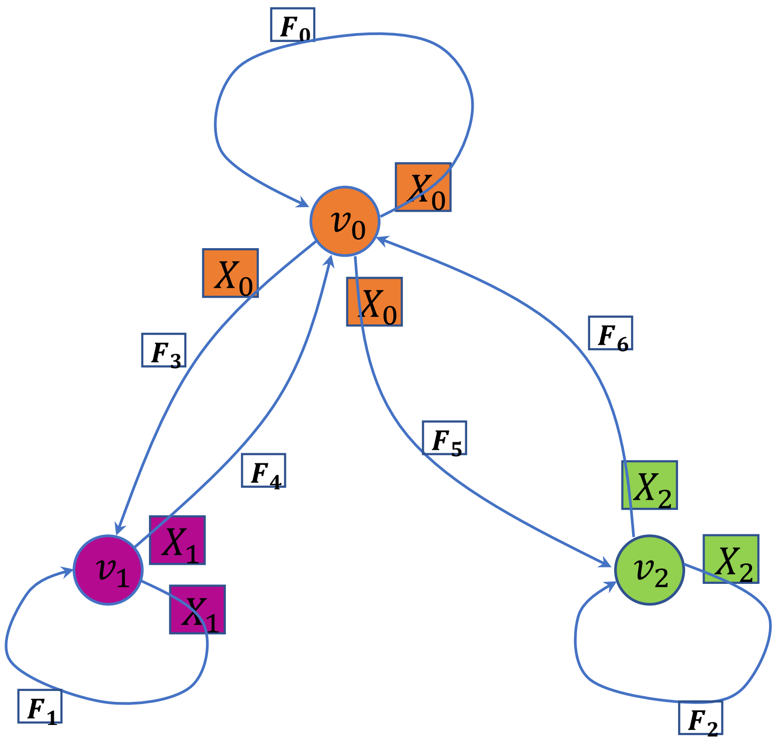
Add self connections

X_0 2D feature at node 0

F_0 2D convolution kernel along edge 0

We omit the superscript "l" in the drawing.

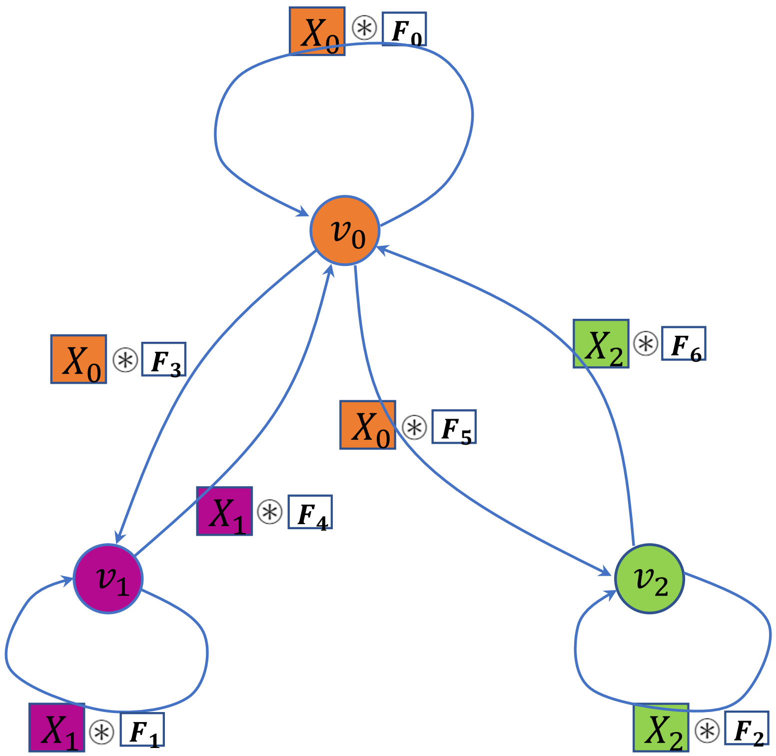
SIA-GCN: A simple example



$$X^{(l+1)} = \sigma\left(\hat{A}\left(\left(BX^{(l)}\right)\tilde{\circledast}F^{(l)}\right)\right)$$

Broadcast 2D features of each node to their outgoing edges

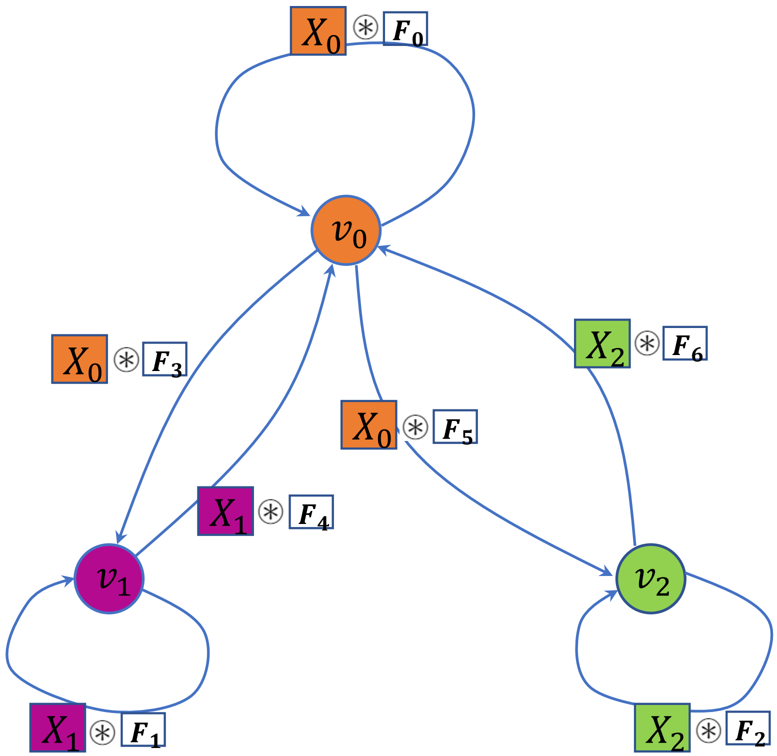
SIA-GCN: A simple example



$$X^{(l+1)} = \sigma\left(\hat{A}\left(\left(BX^{(l)}\right)\tilde{\circledast}F^{(l)}\right)\right)$$

Perform 2D convolutions along each edge.

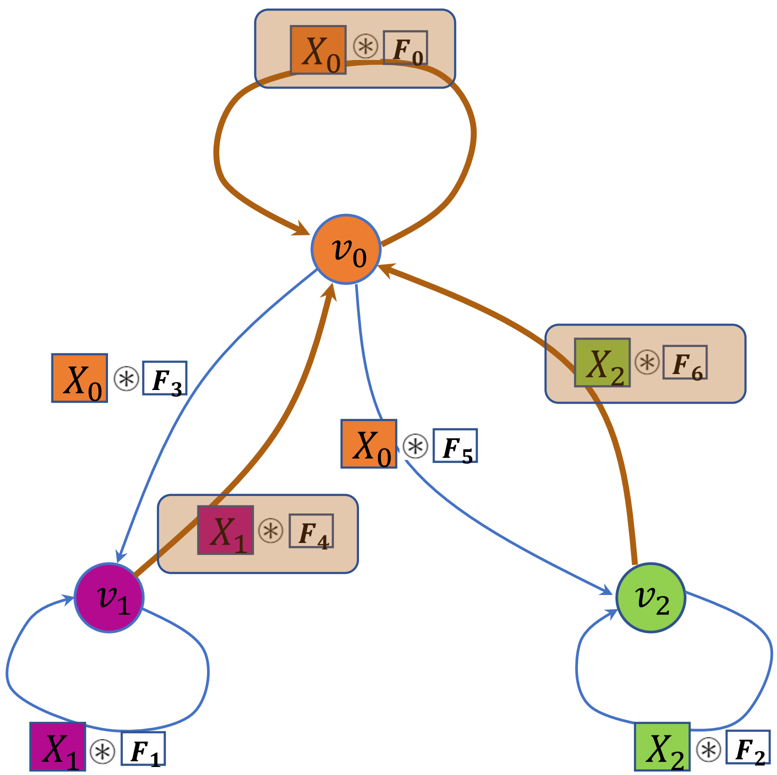
SIA-GCN: A simple example



$$X^{(l+1)} = \sigma\left(\hat{A}\left((BX^{(l)})\tilde{\circledast}F^{(l)}\right)\right)$$

Information aggregation.

SIA-GCN: A simple example

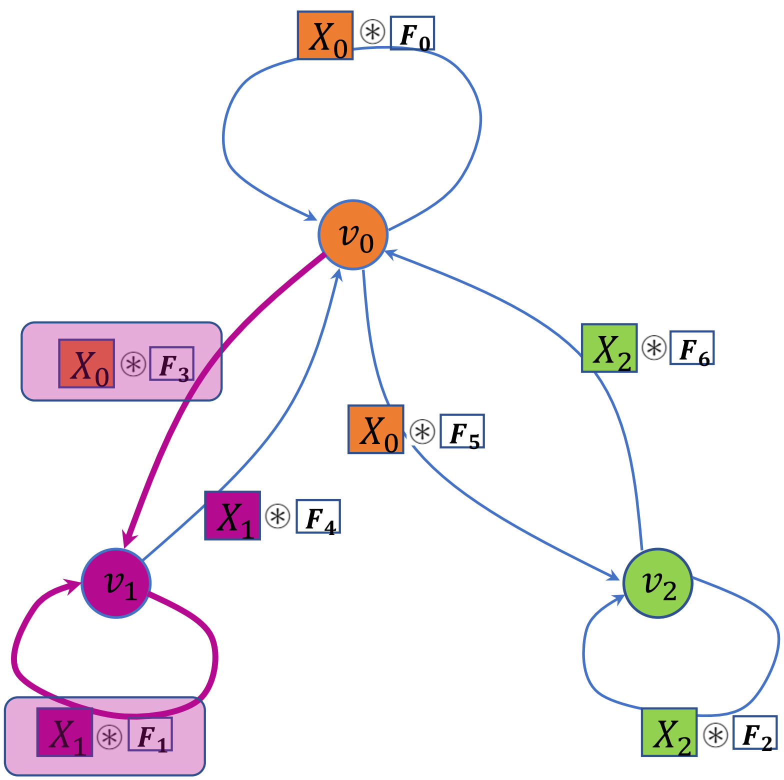


$$X^{(l+1)} = \sigma\left(\hat{A}\left((BX^{(l)})\tilde{\circledast}F^{(l)}\right)\right)$$

Information aggregation.

$$X_0^{\mathrm{new}} = \frac{1}{3}\left(X_0 \circledast F_0 + X_1 \circledast F_4 + X_2 \circledast F_6 \right)$$

SIA-GCN: A simple example



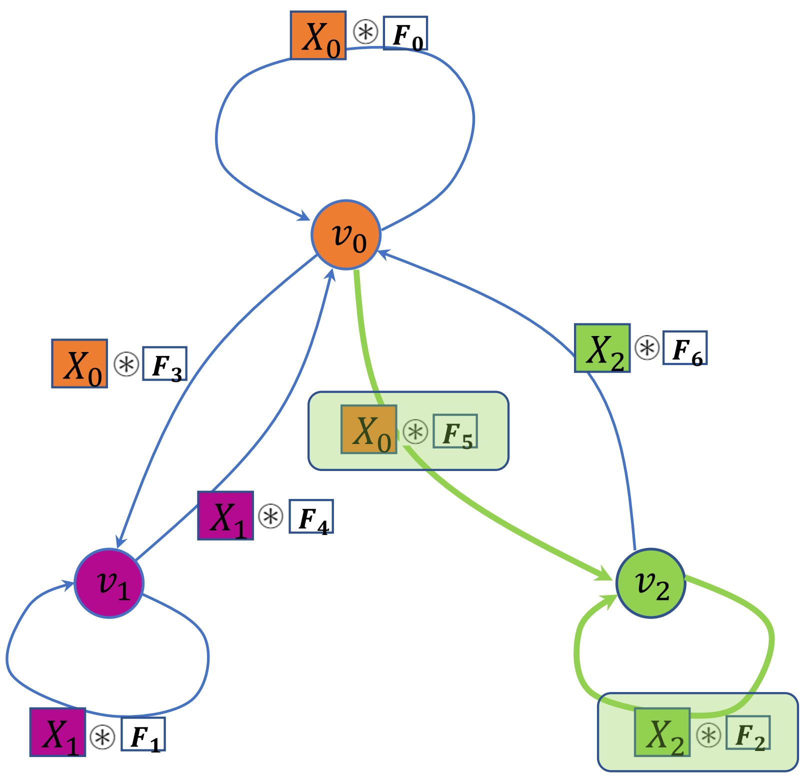
$$X^{(l+1)} = \sigma\left(\hat{A}\left((BX^{(l)})\tilde{\circledast}F^{(l)}\right)\right)$$

Information aggregation.

$$X_0^{\text{new}} = \frac{1}{3}\left(X_0 \circledast F_0 + X_1 \circledast F_4 + X_2 \circledast F_6 \right)$$

$$X_1^{\text{new}} = \frac{1}{2}\left(X_0 \circledast F_3 + X_1 \circledast F_1 \right)$$

SIA-GCN: A simple example



$$X^{(l+1)} = \sigma\left(\hat{A}\left(\left(BX^{(l)}\right)\tilde{\circledast}F^{(l)}\right)\right)$$

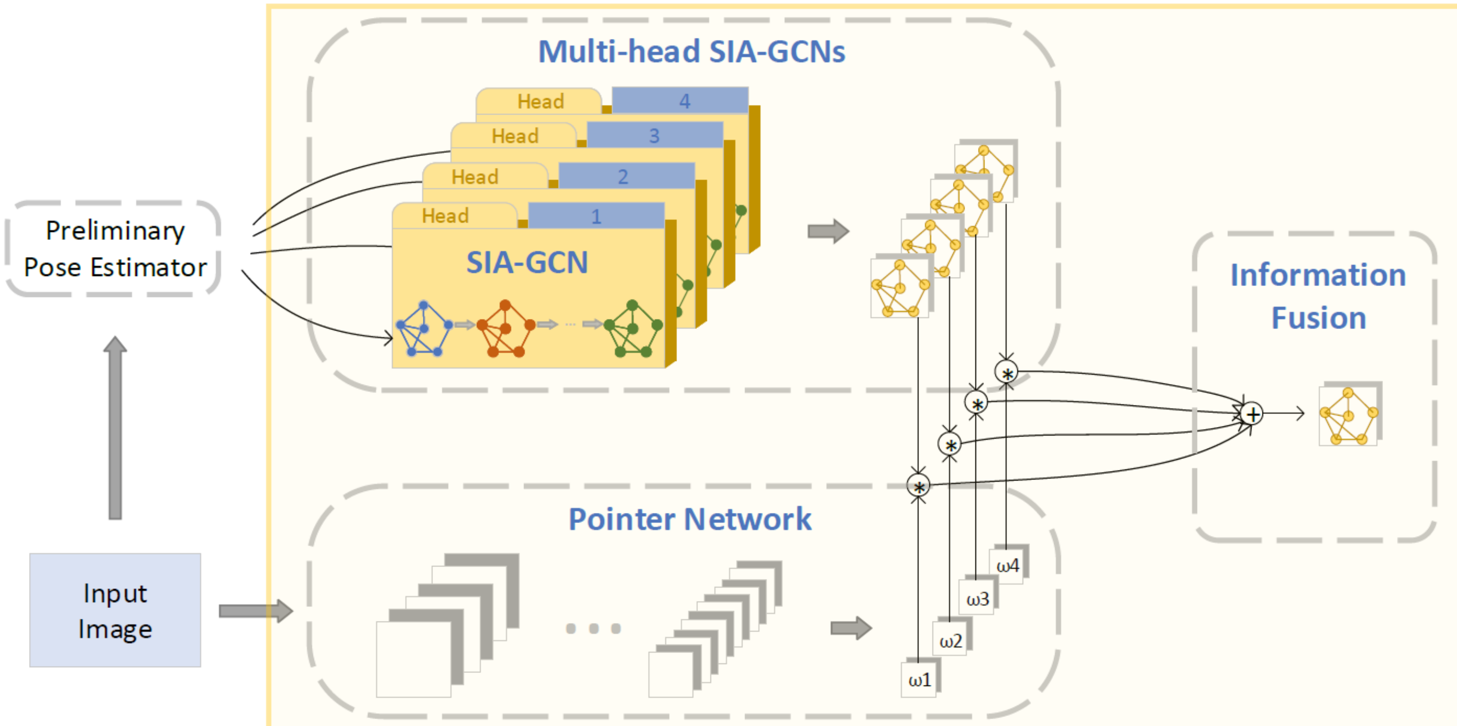
Information aggregation.

$$X_0^{\text{new}} = \frac{1}{3}\left(X_0 \circledast F_0 + X_1 \circledast F_4 + X_2 \circledast F_6 \right)$$

$$X_1^{\text{new}} = \frac{1}{2}\left(X_0 \circledast F_3 + X_1 \circledast F_1 \right)$$

$$X_2^{\text{new}} = \frac{1}{2}\left(X_0 \circledast F_5 + X_2 \circledast F_2 \right)$$

SIA-GCN: Application on 2D hand pose estimation



System diagram of the SiaPose, utilizing SIA-GCN.

SIA-GCN: Application on 2D hand pose estimation

Experiments:

- Datasets

  CMU Panoptic Hand Dataset

  Largescale Multiview 3D Hand Pose Dataset

  MPII+NZSL Hand Dataset

- Metric

  PCK (Percentage of Correct Keypoints):

  the percentage of detections that fall within a

  normalized distance of the ground truth.

- Baselines

  Convolutional Pose Machine (CPM)

  Stacked Hourglass (SHG)

SIA-GCN: Application on 2D hand pose estimation

Some results:

Table 1: SHG based SiaPose on Panoptic Dataset.

PCK@	0.01	0.02	0.03	0.04	0.05	0.06	mPCK
SHG Baseline	35.85	71.47	83.15	88.21	91.10	92.92	77.12
SharedWeight GCN	34.76	69.66	81.33	86.19	89.14	90.95	75.34
1-head SiaPose	35.78	71.16	83.57	88.98	92.00	93.84	77.55
5-head SiaPose	37.53	73.07	84.60	89.51	92.14	93.85	78.45
10-head SiaPose	37.97	73.53	84.95	89.70	92.26	93.91	78.72
Improvement	2.12	2.06	1.80	1.49	1.16	0.99	1.60
10-head R-SiaPose	39.46	77.22	88.45	92.97	94.85	96.09	81.48
Improvement	3.61	5.75	5.30	4.76	3.75	3.17	4.36

SIA-GCN: Application on 2D hand pose estimation

Some results:

Table 3: Comparison to state-of-the-art methods.

PCK@	0.01	0.02	0.03	0.04	0.05	0.06	mPCK
CMU Panoptic Hand Dataset							
R-MGMN [14]	23.67	60.12	76.28	83.14	86.91	89.47	69.93
AGMN [13]	23.90	60.26	76.21	83.70	87.72	90.27	70.34
R-SiaPose (Ours)	24.94	62.08	77.83	84.91	88.78	91.34	71.65
Large-scale Multiview 3D Hand Pose Dataset (MHP)							
R-MGMN [14]	41.51	85.97	93.71	96.33	97.51	98.17	85.53
AGMN [13]	41.38	85.67	93.96	96.61	97.77	98.42	85.63
R-SiaPose (Ours)	41.27	85.89	93.82	96.43	97.61	98.29	85.56

SIA-GCN: Application on 2D hand pose estimation

Qualitative results:

Qualitative results of baseline (top) and our model (bottom) on Panoptic and MPII.

Takeaways

- We proposed SIA-GCN, which can

  a) process graphs with 2D features at each node, and

  b) capture different spatial relationships for neighbouring nodes along different edges.

- We demonstrated its efficacy by

  a) implementing a network for the task of hand pose estimation, and

  b) achieving state-of-the-art performance.

Thanks!