**Introduction**

- We model the image set as a graph and formulate image set classification as the graph matching task.
- We build the first end-to-end graph convolutional network, the Deep SetNet, to learn the graph structure of an image set.
- We impose the $\ell_{1,2}$-norm based sparsity constraint to select vertex features in the set graph to improve the model generalization capability.

**Deep SetNet**

- Convolutional layers and fully-connected layers.
- Input a set of images with different cardinality into the DCNN each time with variant batch size.

**Deep GCN**

- The graph convolutional operation for the $G = (A, Z)$

$$\hat{Z} = g(D^{-1}AZW)$$

- Stacking to the deep GCN

$$Z^{t+1} = g(D^{-1}AZ^tW^t)$$

**Graph Pooling Layers**

- Vertex features are aggregated to the graph feature

$$v = \frac{1}{m} \sum_k Z^k$$

**Joint Sparsity for Vertex Selection**

- Impose the $\ell_{1,2}$-norm constraint on the vertex matrix $Z$

**The Objective**

$$\frac{1}{n} \sum_{v \in X} y \log(\hat{y}) + \lambda \sum_{k=1}^{t} ||Z^k||_{1,2}$$

**Results**

<table>
<thead>
<tr>
<th>Method</th>
<th>UCSD Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discriminant Canonical Correlations (DCC) [14]</td>
<td>91.5 ± 3.4</td>
</tr>
<tr>
<td>Manifold Discriminant Analysis (MDA) [25]</td>
<td>92.7 ± 3.6</td>
</tr>
<tr>
<td>Grassmann Discriminant analysis (GDA) [17]</td>
<td>92.5 ± 2.6</td>
</tr>
<tr>
<td>Covariance Discriminative learning (CDL) [11]</td>
<td>91.7 ± 0.9</td>
</tr>
<tr>
<td>SPD Manifold Learning (SPD-ML) [15]</td>
<td>92.1 ± 1.5</td>
</tr>
<tr>
<td>Log-Euclidean metric learning (LEML) [16]</td>
<td>92.5 ± 2.9</td>
</tr>
<tr>
<td>Discriminant Grassmann kernels (DGK) [2]</td>
<td>96.5 ± 1.7</td>
</tr>
</tbody>
</table>

Deep SetNet (ours) 98.3 ± 0.8

**Conclusion**

- We propose the first end-to-end graph convolutional networks, the Deep SetNet, for image-set classification.
- Our model can match the image set in an efficient way.
- Extensive experiments and analysis show the great effectiveness of our model for image-set classification.