Constrained Spectral Clustering Network with Self-Training

Xinyue Liu, Shichong Yang, Linlin Zong

Dalian University of Technology

25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

Outline

- Motivation
- Proposed Approach
- Experiment
- Conclusion

Motivation

Motivation

- Deep Clustering
  - Combine feature learning and cluster assignment.

- Deep Constrained Clustering
  - Pairwise constraints: must-link and cannot-link,
  - Most existing methods do not make use of unlabeled data.

Motivation

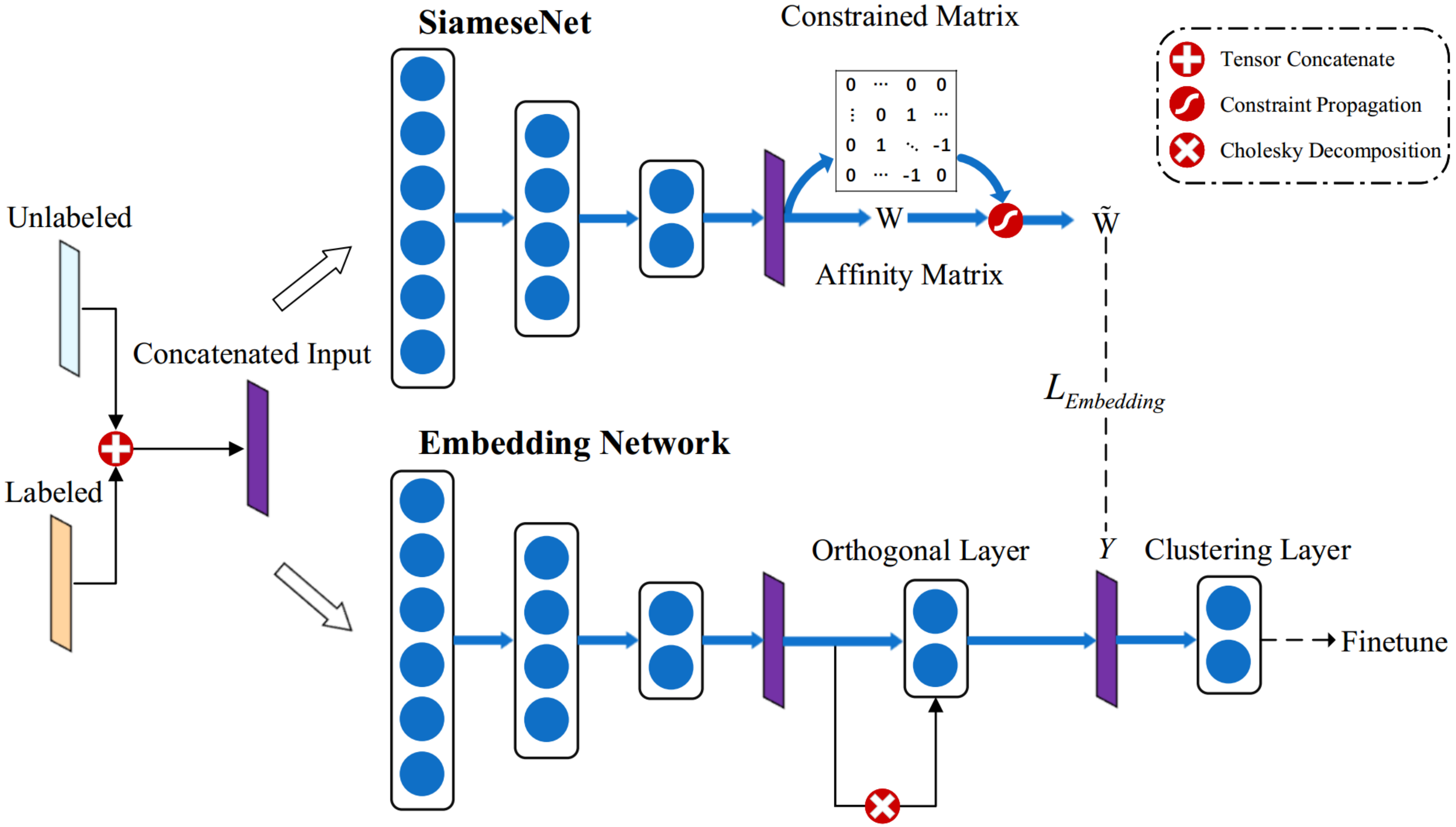
- Spectral Clustering
  - Handle non-convex clusters.

- Deep Spectral Clustering
  - The affinity learned by the SpectralNet is not guaranteed to be consistent with local invariance and thus hurts the final clustering performance.

Proposed Approach

Proposed Approach

- Constrained Spectral Clustering Network (CSCN)



Proposed Approach

- The Unsupervised Building Block SpectralNet
  - Use the SiameseNet to measure similarity

$$
\begin{aligned}
L_c = & P\|g_\theta(x_i) - g_\theta(x_j)\|^2 + \\
& (1-P)\max(c - \|g_\theta(x_i) - g_\theta(x_j)\|, 0)^2
\end{aligned}
$$

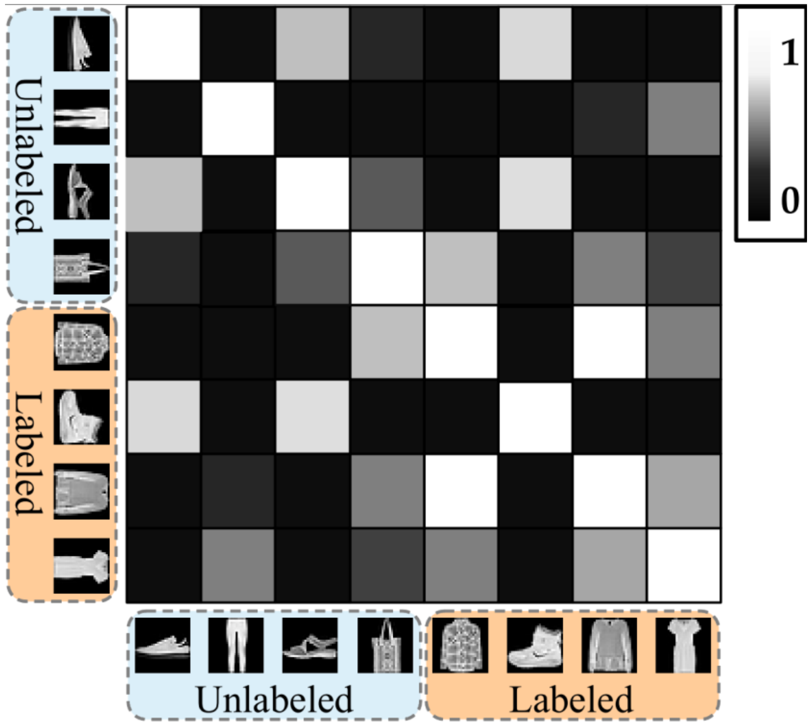
  - Spectral embedding

$$
L_{Embedding} = \frac{1}{m^2} \sum_{i,j=1}^{m} W_{i,j} \|y_i - y_j\|^2
$$

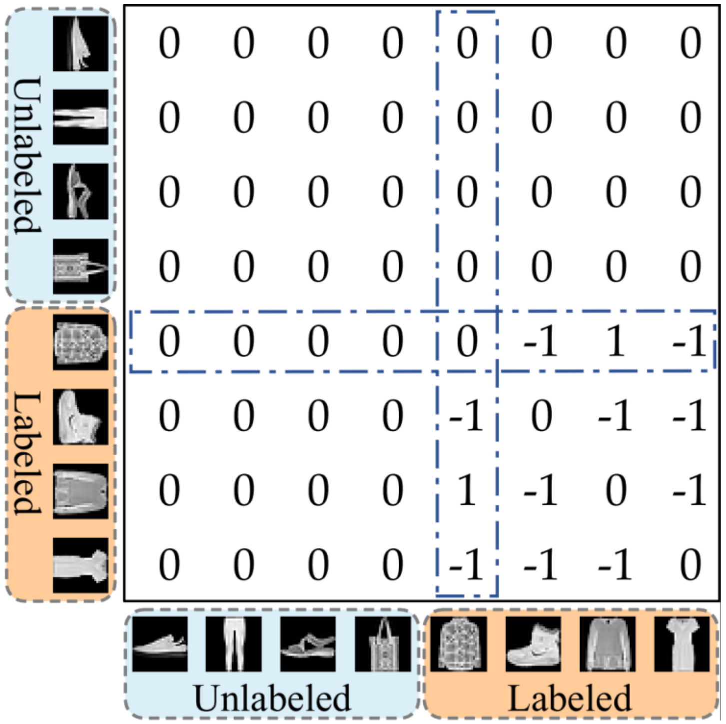
Proposed Approach

- Constraints and their Propagation

Affinity Matrix



Constraint Matrix



Proposed Approach

- Constraint propagation

$$F_{ij}(t+1) = \sum_s P_{is} F_{sj}(t),$$

$$\begin{aligned} F_{ij}(t+1) &= \sum_s W_{is} F_{sj}(t) / \sqrt{D_{ii} D_{ss}} \\ &= (SF(t))_{ij}. \end{aligned}$$

- Affinity adjustment

$$\tilde{W}_{ij} = \begin{cases} 1 - (1 - F^*_{ij})(1 - W_{ij}), & F^*_{ij} \geq 0; \\ (1 + F^*_{ij}) W_{ij}, & F^*_{ij} < 0. \end{cases}$$

Proposed Approach

- Clustering Oriented Fine-tuning

  - Clustering soft assignment

$$q_{ij} = \frac{(1 + \|y_i - \mu_j\|^2)^{-1}}{\sum_j (1 + \|y_i - \mu_j\|^2)^{-1}}$$

  - Auxiliary target distribution

$$p_{ij} = \frac{q_{ij}^2 / \sum_i q_{ij}}{\sum_j (q_{ij}^2 / \sum_i q_{ij})}$$

  - KL divergence based clustering objective

$$L_f = KL(P\|Q) = \sum_i \sum_j p_{ij} log \frac{p_{ij}}{q_{ij}}$$

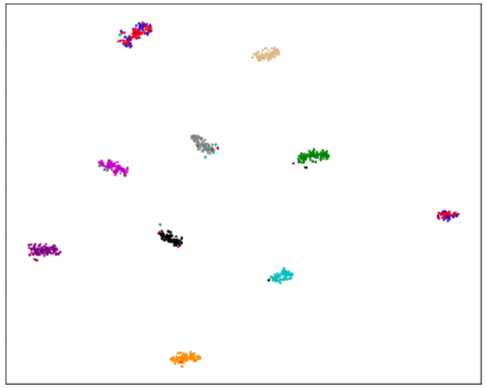
Experiment

Experiment

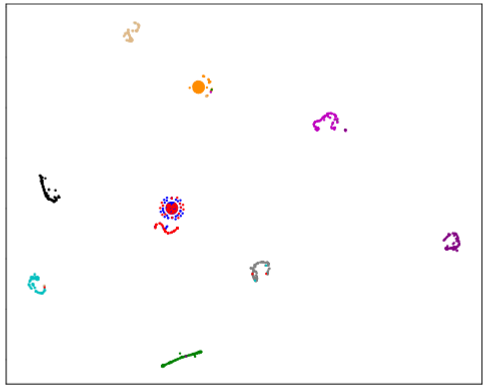
TABLE I
COMPARISON OF CLUSTERING PERFORMANCE IN TERMS OF ACC AND NMI. THE BEST RESULTS ARE IN BOLD.

Model	MNIST		Fashion-MNIST		REUTERS-10K	
	ACC	NMI	ACC	NMI	ACC	NMI
AE+k-means	0.7548	0.7297	0.5260	0.5630	0.6728	0.3885
COP-Kmeans	0.7913	0.7392	0.5555	0.5804	0.7145	0.4358
PCKmeans	0.7758	0.7349	0.5514	0.5737	0.7117	0.4378
DEC	0.8653	0.8370	0.5720	0.6274	0.7047	0.4571
IDEC	0.8810	0.8652	0.5955	0.6341	0.7536	0.4956
SpectralNet	0.8309	0.8929	0.6577	0.7034	0.7195	0.5135
SDEC	0.8651	0.8367	0.5985	0.6328	0.7368	0.4976
DCC	0.8958	0.8782	0.6389	0.6309	0.7870	0.6386
CSCN*	0.9486	0.9286	0.6836	0.7092	0.8504	0.6447
CSCN	**0.9500**	**0.9303**	**0.6883**	**0.7127**	**0.8528**	**0.6484**

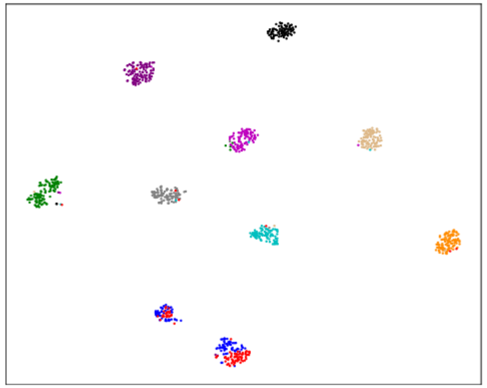
Experiment



(a) SDEC



(b) SpectralNet

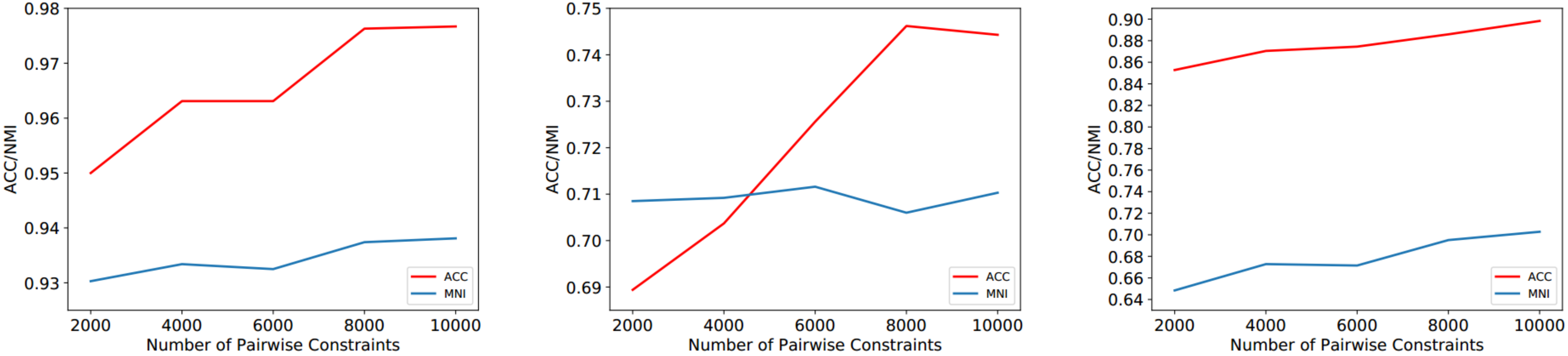


(c) DCC

(d) CSCN*

Fig. 3. Visualization of clustering results on subset of MNIST. Different colors represent different clusters. Note clusters colored by blue and red (digits 4 and 9), they are totally mixed together in other methods while still clearly separable in our CSCN*.

Experiment



(a) MNIST (b) FASHION-MNIST (c) REUTERS-10K

Fig. 4. Effect of the number of pairwise constraints on clustering performance.

Conclusion

Conclusion

- We propose Constrained Spectral Clustering Network (CSCN), which incorporates pairwise constraints and clustering oriented fine-tuning.
- Constraints are propagated to protect local invariance and guide spectral embedding.
- The clustering loss is optimized to finetune the feature space and perform cluster assignments simultaneously.

Thanks