Neural Collaborative Subspace Clustering

Tong Zhang, **Pan Ji**, Mehrtash Harandi,
Wenbing Huang, Hongdong Li

International Conference of Machine Learning (ICML), Long Beach, CA, June 10-15, 2019

Subspace Clustering

- Cluster data points drawn from a union of low-dimensional subspaces

- Applications: image clustering, motion segmentation, etc.

Subspace Clustering Methods

- STOA methods consist of two steps:

    1. Construct an affinity matrix for the whole dataset,

    2. Apply normalized cuts or spectral clustering.

Scalability Issue!!

- Affinity matrix construction is expensive:
  - Large memory footprint;
  - High complexity in optimization.

- Spectral clustering is expensive:
  - Computing SVD on large matrices is demanding.

Can we avoid the construction of huge affinity matrices and bypass the spectral clustering?

International Conference of Machine Learning (ICML), Long Beach, CA, June 10-15, 2019

Our Idea

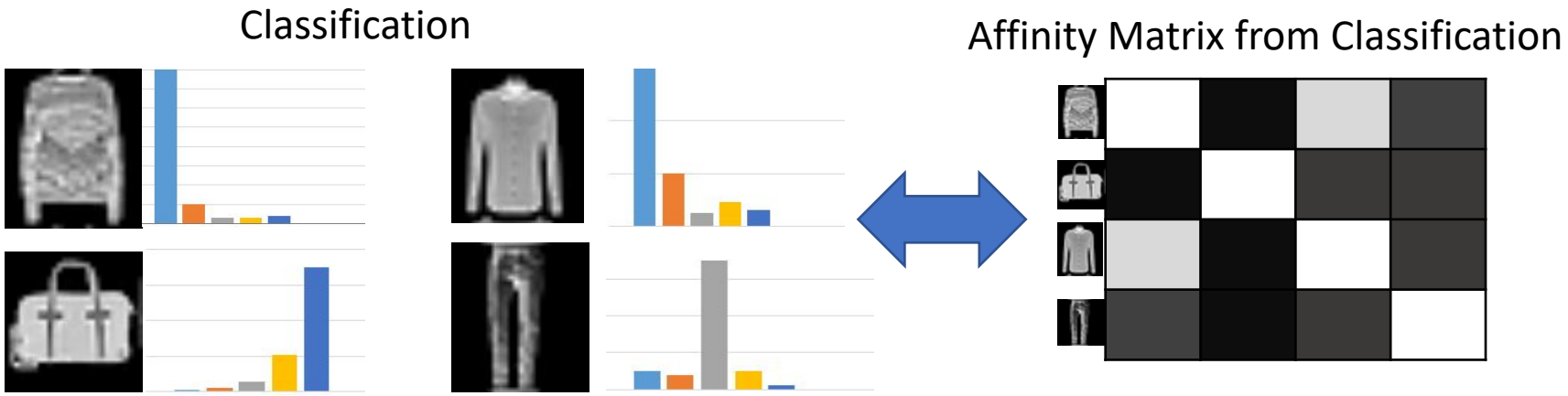
- Construct affinity matrix in a batch;
- Train a classifier using affinity matrices.

Our Idea

- Construct affinity matrix in a batch;
- Train a classifier using affinity matrices.
- **How?**

Affinity from Classification

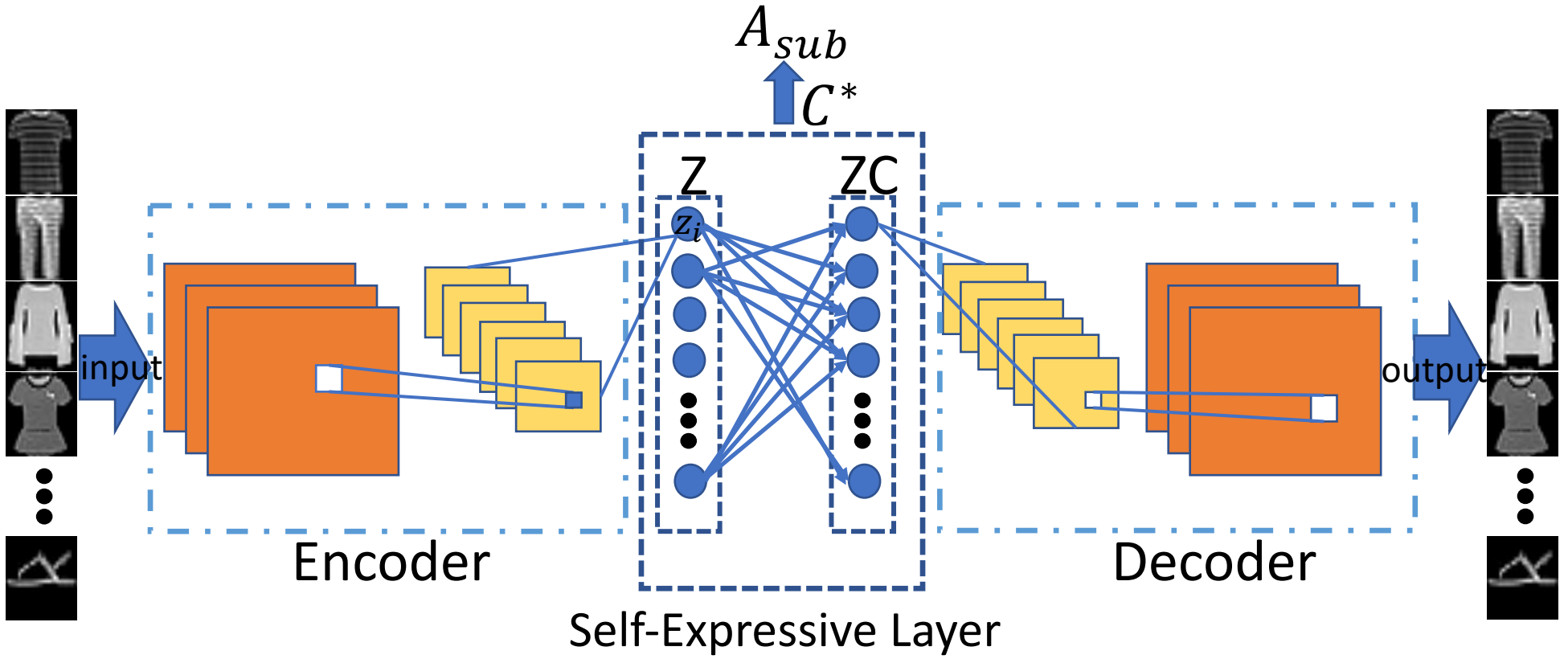
- Build connection between clustering and classification via affinity matrices



$$A_{cla}(i,j) = f_i f_j^T$$

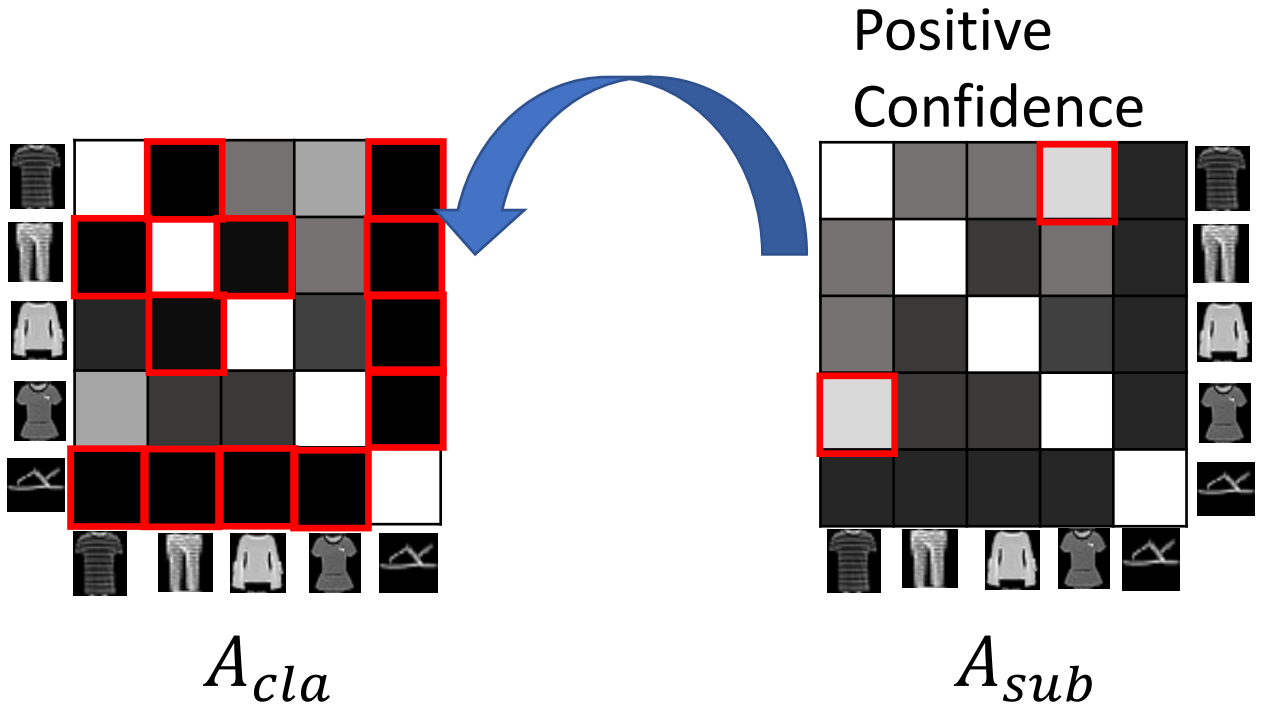
Affinity from Subspace

- Subspace affinity A_{sub} from self-expressiveness



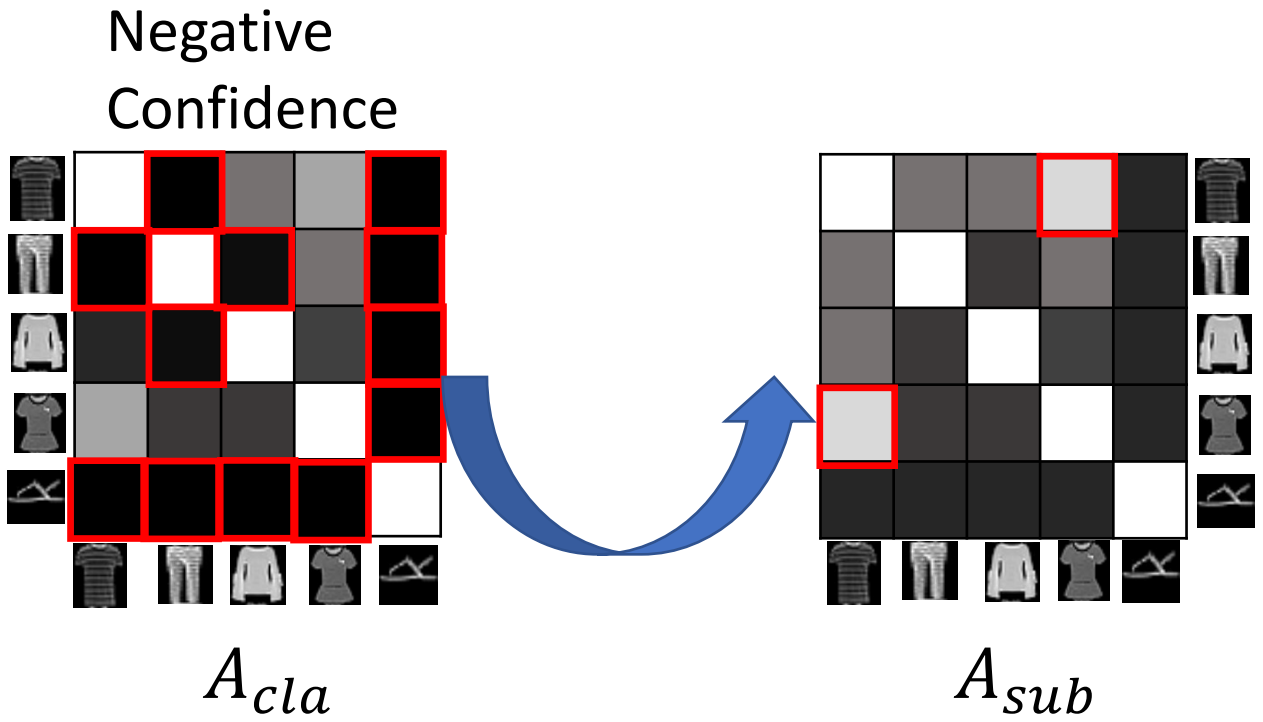
Collaborative Learning

- Subspace affinity is more confident of identifying samples from the same class.

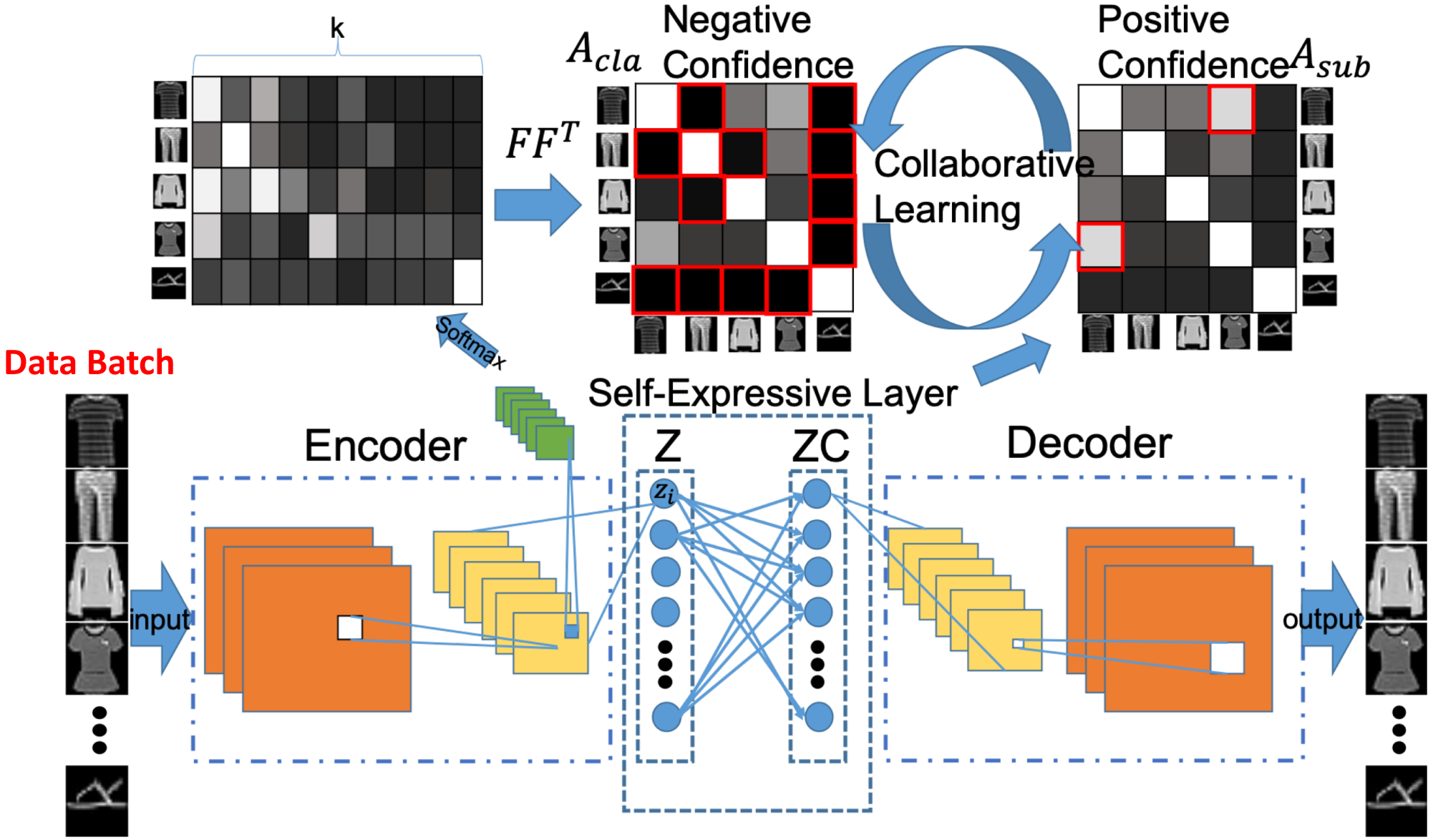


Collaborative Learning (cont'd)

- Classification affinity is more confident of identifying samples from different classes.



Our Framework: Collaborative Learning



International Conference of Machine Learning (ICML), Long Beach, CA, June 10-15, 2019

Clustering via Classifier

- Output the clustering simply through the classification part (bypass the spectral clustering):

$$s_i = argmax_h(f_{ih}), h = 1, \cdots, k,$$

where k is number of clusters.

Experiments:

- MNIST

	ACC(%)	NMI(%)	ARI(%)
SAE-KM	81.29	73.78	67
CAE-KM	51	44.87	33.52
K-means	53	50	37
PCA-KS	68.53	64.17	54.17
DEC	84.3	80	75
DCN	83.31	80.86	74.87
kSCN	87.14	78.15	
Ours	**94.09**	**86.12**	**87.52**

- Fashion-MNIST

	ACC(%)	NMI(%)	ARI(%)
SAE-KM	54.35	58.53	41.86
CAE-KM	39.84	39.80	25.93
K-means	47.58	51.24	34.86
PCA-KS	53.41	57.5	41.17
DEC	59	60.1	44.6
DCN	58.67	59.4	43.04
kSCN	63.78	62.04	48.04
ClusterGAN	63.0	64.0	50.0
InfoGAN	61.0	59.0	44.0
DAC	61.5	63.2	50.2
Ours	**72.14**	**68.60**	**59.17**

Conclusion

- Subspace is a powerful tool to represent data in high-dimensional space.
- Introduced a *collaborative learning* paradigm for clustering.
- Made subspace clustering scalable through batch-wise training.

Thank You

Poster time: Room Pacific Ballroom
06:30-09:00 pm, June 13th

We're hiring! For more details, visit http://www.nec-labs.com/research-departments/media-analytics.

International Conference of Machine Learning (ICML), Long Beach, CA, June 10-15, 2019