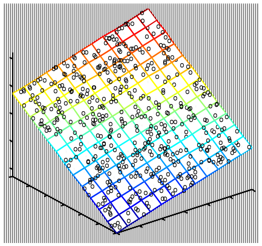
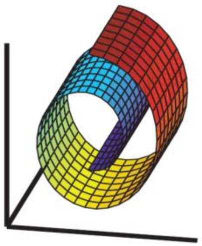
Self-Supervised Convolutional Subspace Clustering Network

张军建[1]，李春光[1]，由翀[2]，齐宪标[3]，
张洪刚[1]，郭军[1]，林宙辰[4]

[1] 北京邮电大学　[2] 加州大学伯克利分校　[3] 深圳大数据研究院　[4] 北京大学

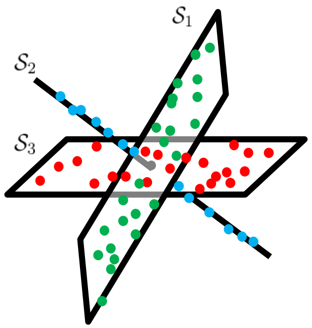
Introduction

- **High-dimensional data often reside in low-dimensional structure(s)**
  - – linear subspace in R^3
  - – nonlinear manifold in R^3

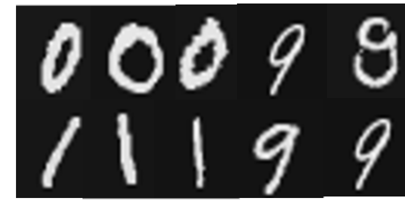
(S. Roweis & L. Saul: SCIENCE 2000)

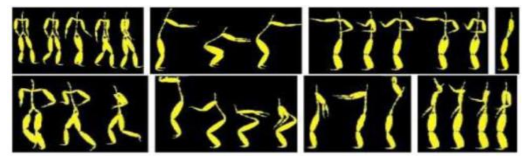
High-dimensional data with **multiple classes** ➡ **multiple** low-dimension **structures**.



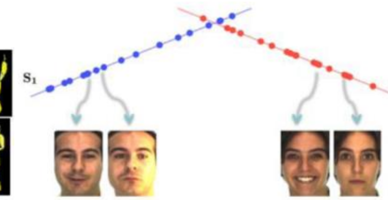
Examples

- **Motion Segmentation**
- **Facial / Handwriting Digits Image Clustering**
- **Action Segmentation in Video**
- **Planar area in 3D vision**
- **Cancer Subtypes in Gene Microarray**

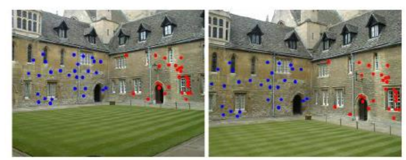




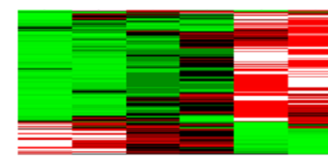
Video Temporal Segmentation



Face Recognition/Clustering

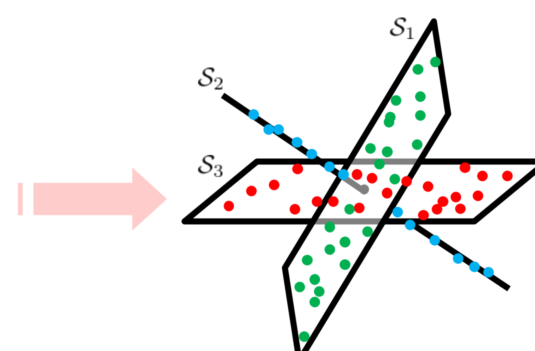


Planar Segmentation



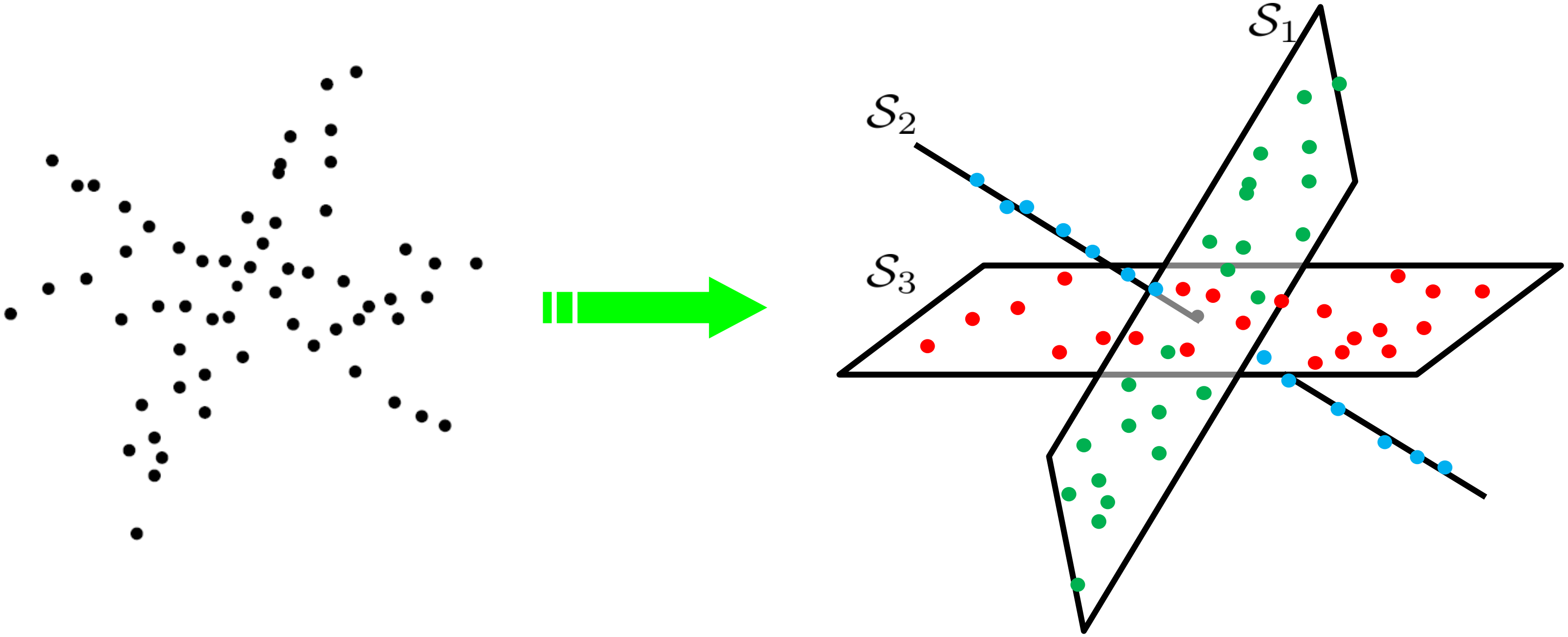
Cancer Subtypes Clustering

Union of Subspaces



Subspace Clustering

- **Task:**
  – Given data points lying in a **union of subspaces**, to **segment** the data points into each **subspace**



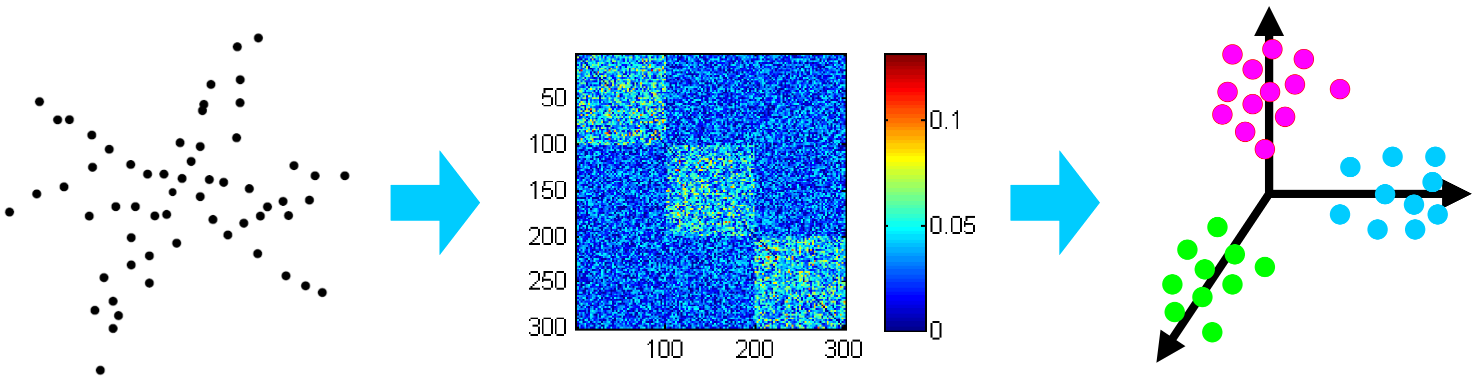
Don’t know: basis of each subspace / dimension of each subspace / number of data points per subspaces / number of subspaces

Prior Work

- **Iterative methods**
  - – k-plane, q-flats, …
- **Statistical methods**
  - – Factorization, MPPCA, …
- **Algebraic methods**
  - – GPCA, …
- **Spectral clustering based methods**
  - – SSC, LRR/LRSC, LSR, EnSC, …

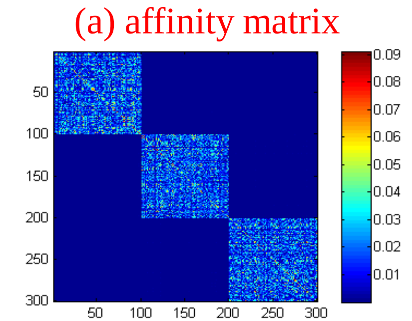
Spectral Clustering based Methods

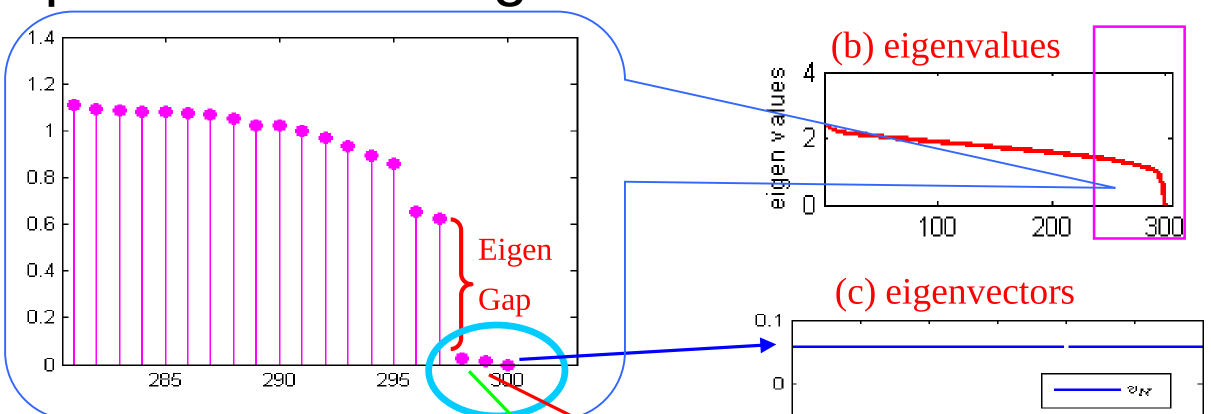
- **Two-Step Approach**
  - Step 1: build a data affinity matrix
  - Step 2: apply spectral clustering



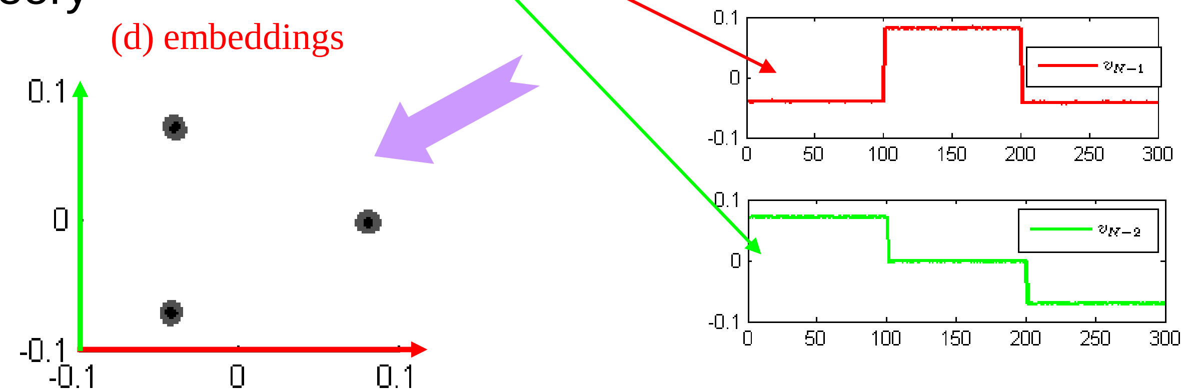
Spectral Clustering based Methods

- **Two-Step Approach**
  - Step 1: build a data affinity matrix
  - Step 2: apply spectral clustering



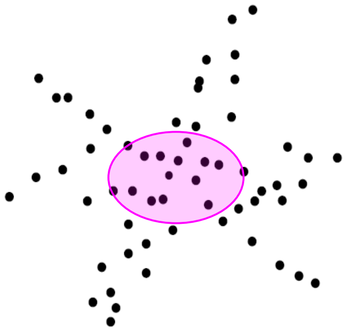


  - Spectral graph theory

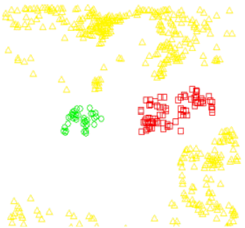


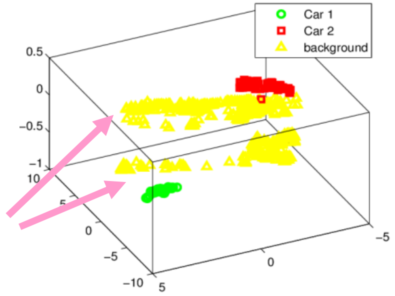
Spectral Clustering based Methods

- **Two-Step Approach**
  - Step 1: build a data affinity matrix
  - Step 2: apply spectral clustering



- **Challenges:**

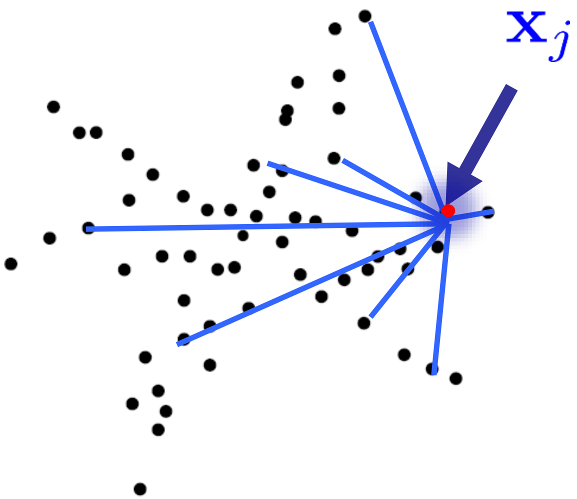




  - Distance based affinity fails
    - Two points nearby may belong to different subspaces
      - E.g., near the intersection of subspaces
    - Two points faraway may belong to same subspace
      - E.g., disjoint components, e.g. the yellow points

Spectral Clustering based Methods

- **Two-Step Approach**
  - Step 1: build a data affinity matrix
  - Step 2: apply spectral clustering



- **Self-Expressive Model:**
  - A point in a union of subspaces can be expressed as a linear combination of other data points, i.e.

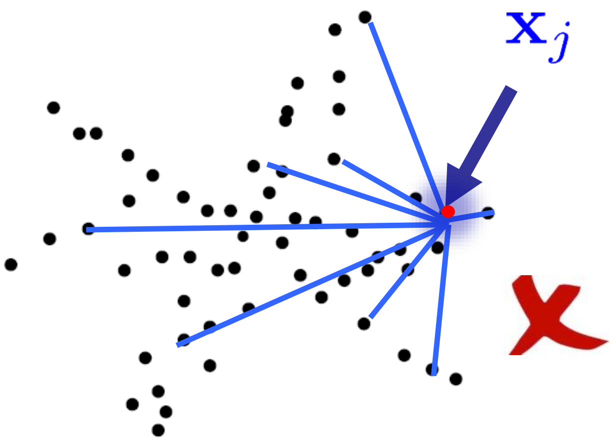
$$\mathbf{x}_j = \sum_{i \neq j}^{N} c_i \mathbf{x}_i = c_1 \mathbf{x}_1 + c_2 \mathbf{x}_2 + c_{j-1} \mathbf{x}_{j-1} + c_{j+1} \mathbf{x}_{j+1} + \ldots + c_N \mathbf{x}_N,$$

[1] E. Elhamifar & R. Vidal: "Sparse subspace clustering", CVPR 2009.
[2] E. Elhamifar & R. Vidal: "Sparse subspace clustering: Algorithm, theory, and applications", IEEE TPAMI 2013.

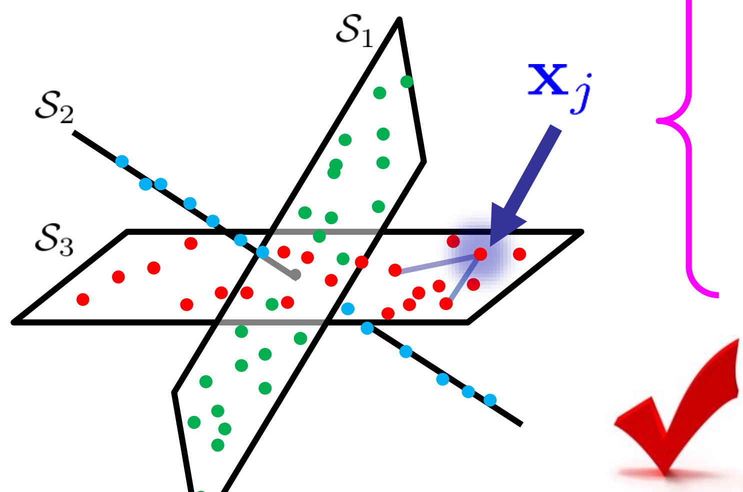
Subspace-Preserving Property

- A solution $\boldsymbol{c}$ to $\mathbf{x}_j = X\mathbf{c},\ c_j = 0$ is said to be subspace-preserving if :

> *The non-zero coefficients in* $\mathbf{c}$ *only with the data points that belong to the same subspace as* $\mathbf{x}_j$



(a) NOT Subspace-Preserving



(b) Subspace-Preserving

$$\left\{ \begin{array}{l} \|\mathbf{c}\|_0 , \|\mathbf{c}\|_1 \\ \|\mathbf{c}\|_2^2 \\ \|\mathbf{c}\|_1 + \|\mathbf{c}\|_2^2 \\ \|C\|_* \end{array} \right.$$

Sparse Subspace Clustering (SSC)

- SSC finds the sparsest linear combination in terms of other data points by solving

$$\min_{\mathbf{c}} \|\mathbf{c}\|_1 \quad \text{s.t. } \mathbf{x}_j = X\mathbf{c}, \ \ c_j = 0$$

  – If Gaussian noise:

$$\min_{\mathbf{c}} \|\mathbf{c}\|_1 + \frac{\lambda}{2}\|\mathbf{x}_j - X\mathbf{c}\|_2^2 \quad \text{s.t. } c_j = 0$$

  – If sparse corruption (i.e. outlying entries):

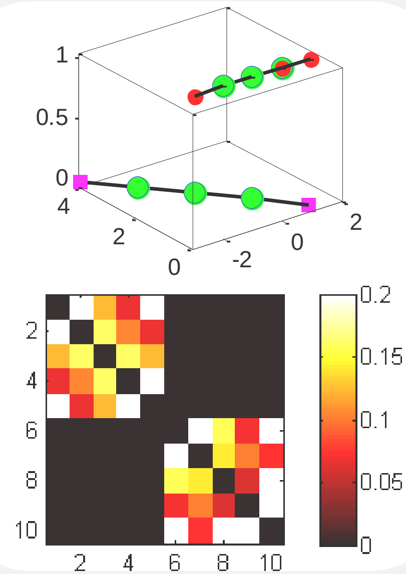
$$\min_{\mathbf{c}} \|\mathbf{c}\|_1 + \lambda\|\mathbf{x}_j - X\mathbf{c}\|_1 \quad \text{s.t. } c_j = 0$$

  – If subspaces are affine :

$$\min_{\mathbf{c}} \|\mathbf{c}\|_1 \quad \text{s.t. } \mathbf{x}_j = X\mathbf{c}, \ \ c_j = 0, \ \ \mathbf{1}^T\mathbf{c} = 1$$

Theoretical guarantees:

Independent, disjoint, even with outliers, noisy, **affine**, missing entries, and etc.

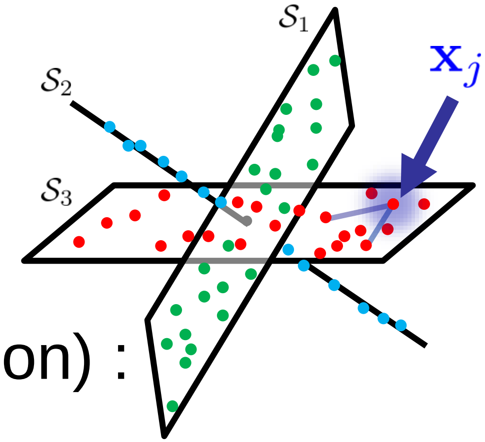


[1] E. Elhamifar & R. Vidal: CVPR'09. [2] Sol. & Candes: Annual Stat.'12. [3] E. Elhamifar & R. Vidal: TPAMI '13. [4] Y.-X. Wang & H. Xu: ICML' 13; JMLR' 16. [5] M. Tsakiris et al. ICML'18. [6] C.-G. Li, C. You, & R. Vidal, SPARS'15 / IEEE J. STSP '18.

Least Square Regression (LSR)

- LSR finds a linear combination in terms of other data points by solving

$$\min_{\mathbf{c}} \|\mathbf{c}\|_2^2 \quad \text{s.t. } \mathbf{x}_j = X\mathbf{c}, \;\; c_j = 0$$



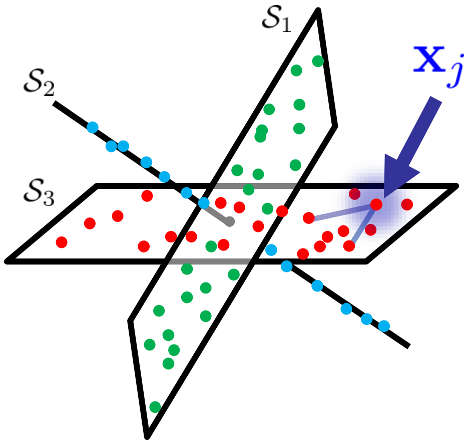
  – Theoretical guarantee (sufficient condition) :

The optimal solution is guaranteed to be subspace-preserving when subspaces are ***independent****.*

[1] Canyi Lu et al. Robust and Efficient Subspace Segmentation via Least Squares Regression, ECCV 2012.

Low-Rank Representation (LRR)

- LRR finds linear combinations collectively with a regularization on the rank of C:



$$\min_{C} \operatorname{rank}(C) \quad \text{s.t. } X = XC$$

$$\min_{C} \|C\|_* \quad \text{s.t. } X = XC$$

where $C = [\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_N]$.

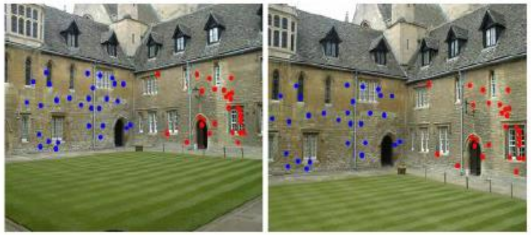
  – Theoretical guarantee (sufficient condition):

The optimal solution is guaranteed to be subspace-preserving when subspaces are ***independent***.

[1] G. Liu et al.: "Robust Subspace Segmentation by Low Rank Representation", ICML 2010.
[2] G. Liu et al.: "Robust recovery of subspace structures by low-rank representation", TPAMI 2013.
[3] P. Favaro et al.: "A closed form solution to robust subspace estimation and clustering," CVPR 2011.
[4] R. Vidal & P. Favaro, "Low rank subspace clustering (LRSC)," Pattern Recognition Letters, vol. 43, pp. 47–61, 2014.

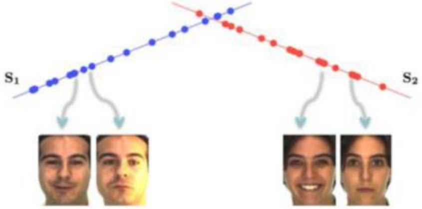
Applications are still limited …

- **Motion Segmentation**
- **Planar area in 3D vision**

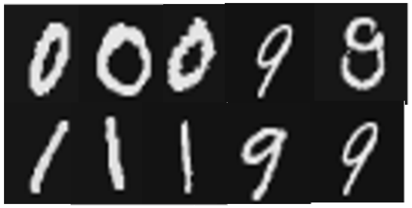


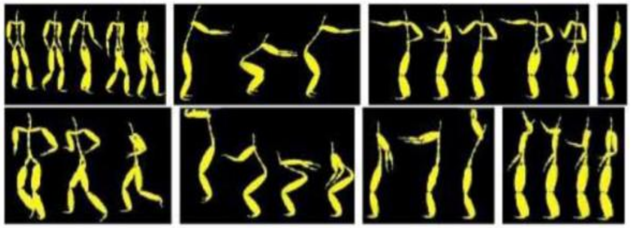
Planar Segmentation

- **Facial / Handwriting Digits Image Clustering**
- **Action Segmentation in Video**
- **…**



Face Recognition/Clustering





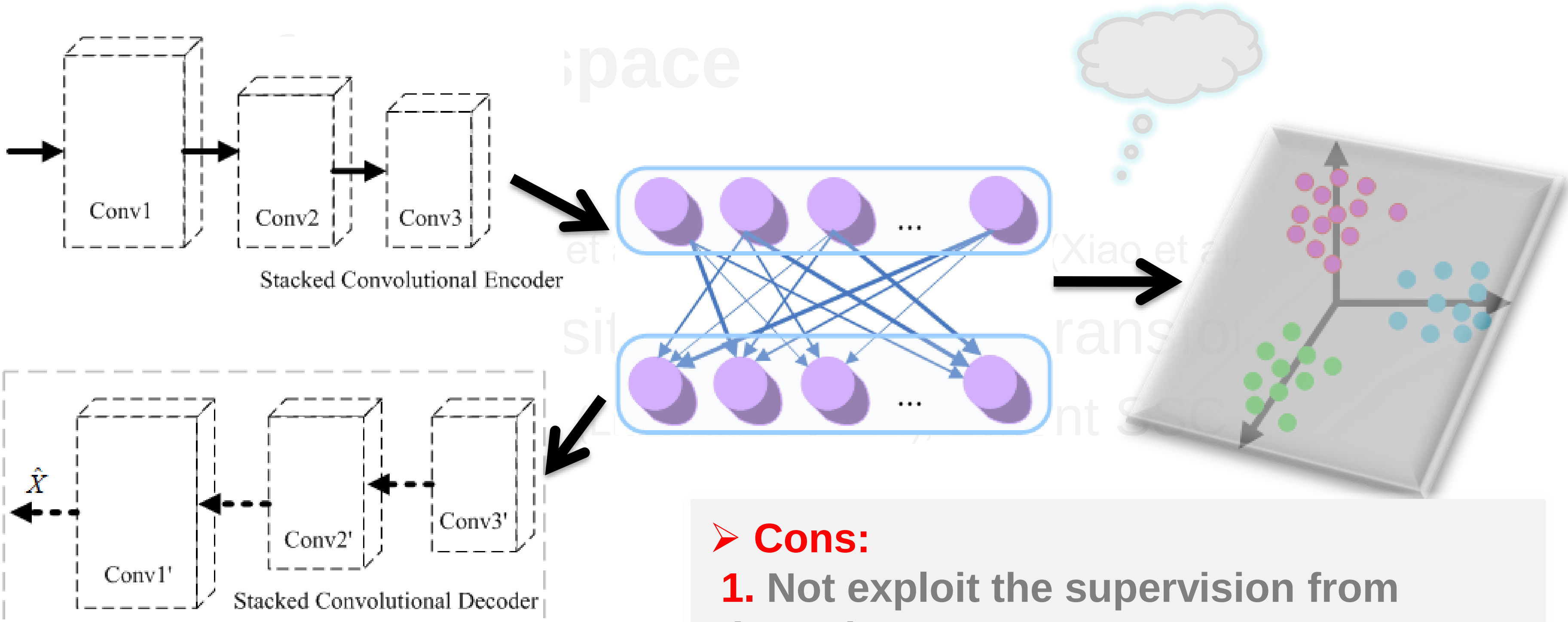
Video Temporal Segmentation

However, visual data in its raw representation DO NOT always align with a union of subspaces. A suitable feature extraction step is needed.

Subspace Clustering in Feature Space

- **Latent feature space**
  - Kernel trick
    - e.g. KSSC (Patel et al. JSTSP'15), KLRR (Xiao et al. TNNLS'16)
  - Matrix decomposition or learned transform
    - e.g. Latent LRR (Liu et al. ICCV'11), Latent SSC (Patel et al. ICCV'13)
- **Explicit feature space**
  - Designed features manually
    - e.g. SIFT, HOG, Scatter Transform, …
  - Learned via neural networks
    - e.g. MLP+SSC (Peng et al. IJCAI'16), DSCNet (Ji et al. NIPS'17)

Subspace Clustering in Feature Space



➢ **Cons:**
1. Not exploit the supervision from clustering
2. Not an end-to-end trainable framework

- e.g. SIFT, HOG, Scatter Transform (You et al. CVPR'16)

– Learned

- e.g. MLP+SSC (Peng et al. IJCAI'16), DSCNet (Ji et al. NIPS'17)

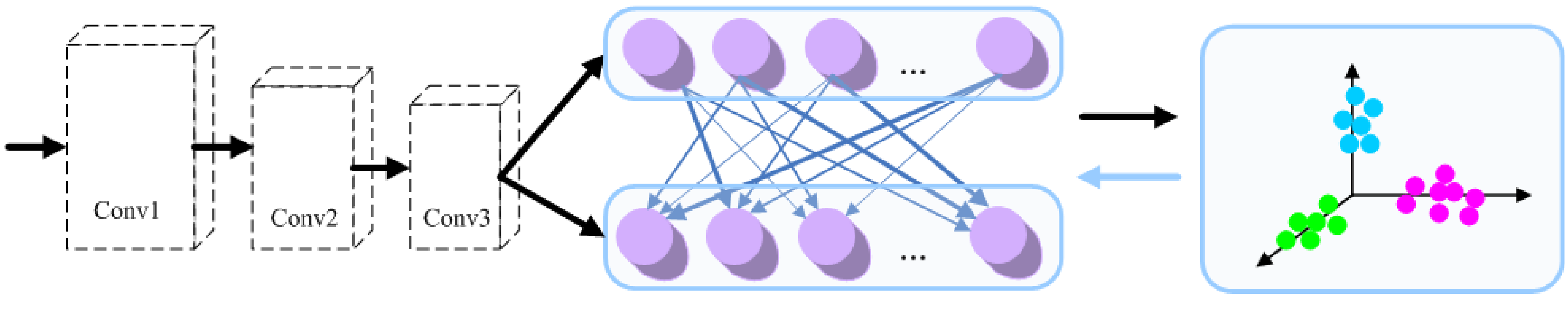
Pan Ji, et al.: Deep Subspace Clustering Network, NIPS 2017.

Outline

- Introduction
- Related Work
- Our Proposal
- Experimental Results
- Summary

Our Goal

- **Conv. Feature + Subspace Clustering → Jointly Trainable Framework**
  - Conv. Feature Extraction
  - Self-expressive Model
  - Spectral Clustering

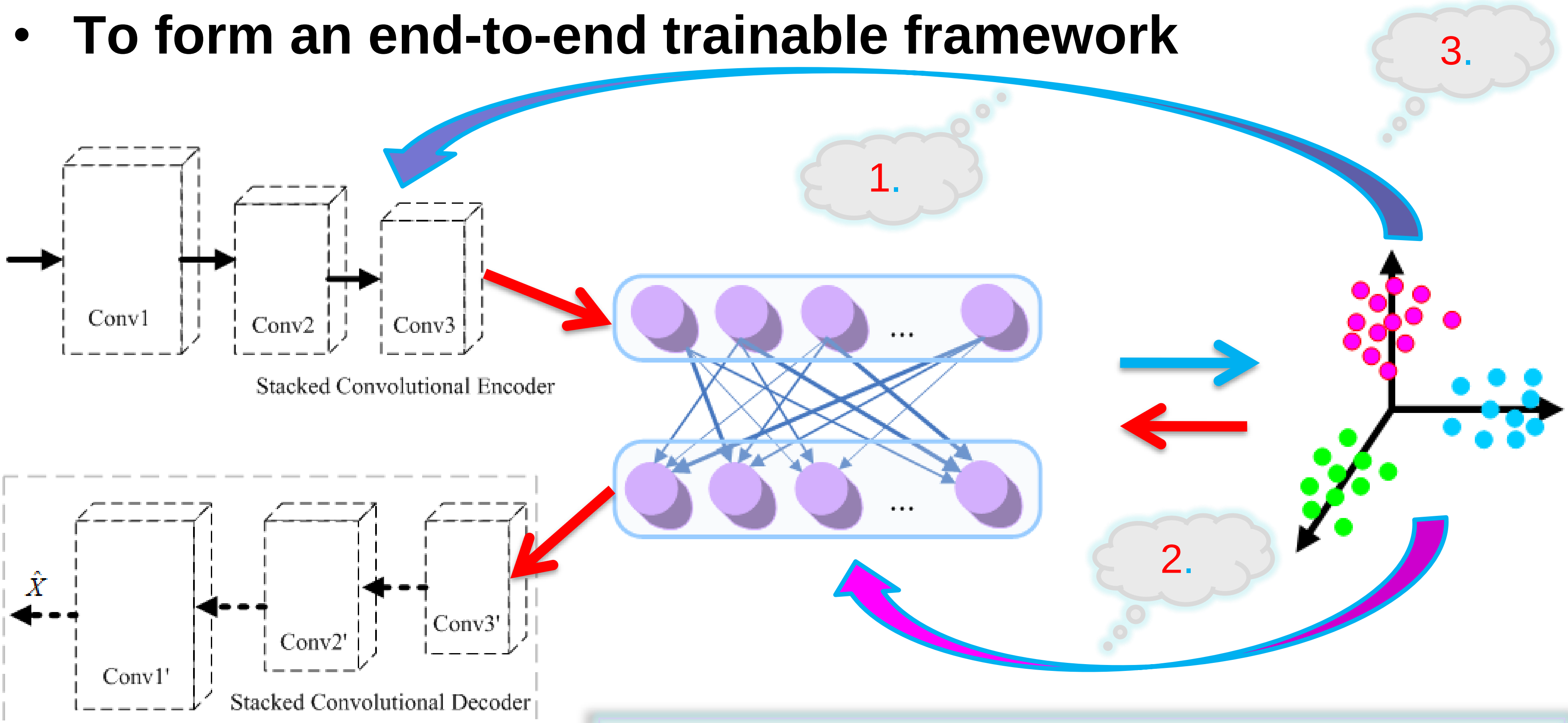


(a) Feature Extraction Module (b) Self-Expression Module (c) Spectral Clustering Module

1. To exploit the supervision info. from clustering.
2. To build an end-to-end jointly trainable framework.

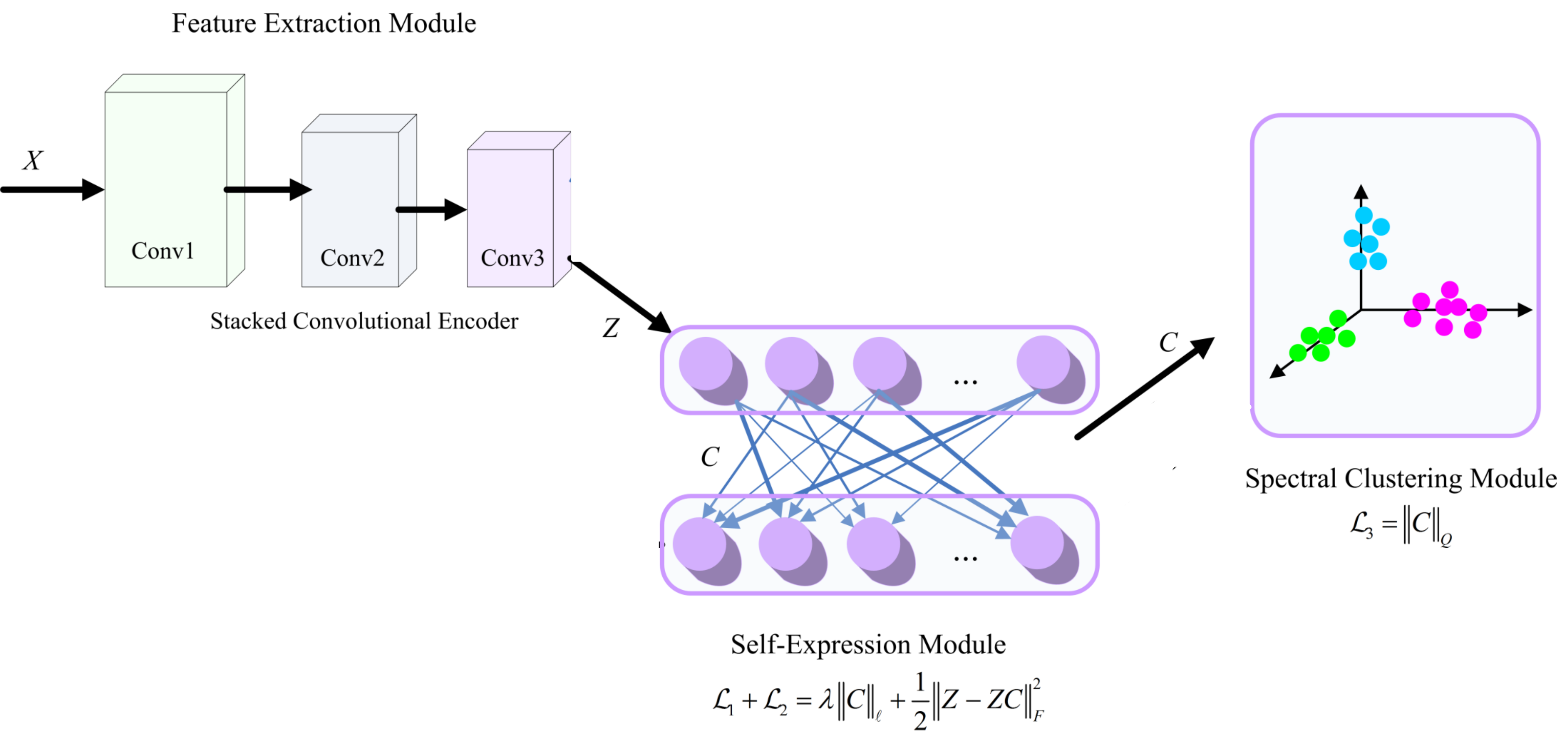
Conv. Feat. Extraction + Subspace Clustering

- To exploit the supervision info. from clustering
- To form an end-to-end trainable framework

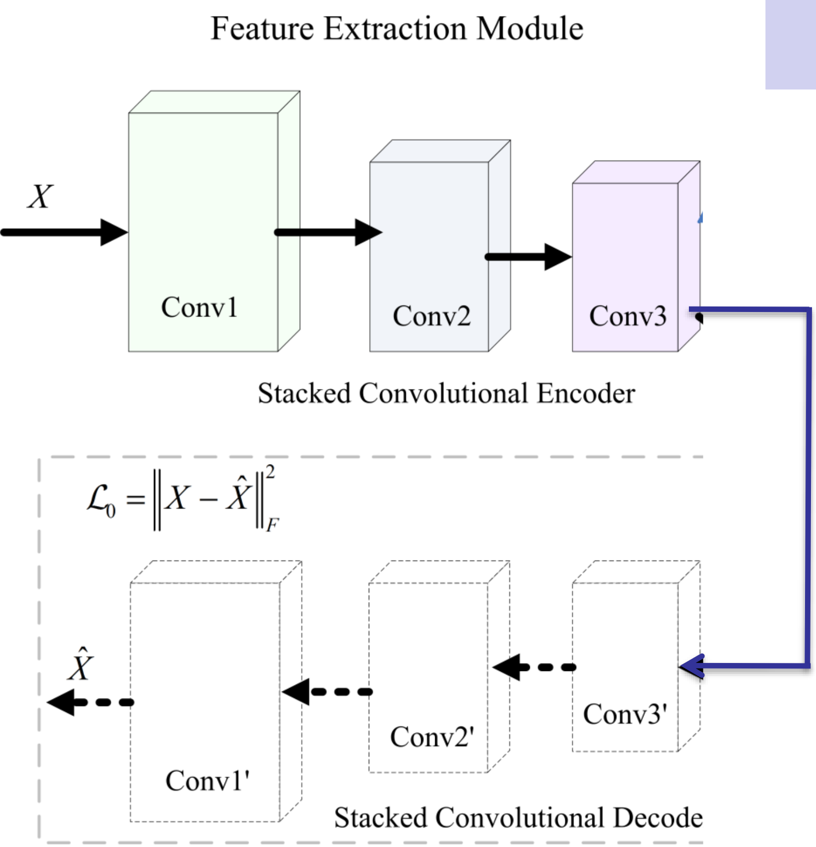


1. How to supervise conv. feature extraction ?
2. How to supervise self-expressive model ?
3. How to build an end-to-end trainable framework ?

Our Proposal: Self-Supervised Conv. Subspace Clustering Network (S^{2}ConvSCN)



SCAE in S²ConvSCN for Conv. Feature Extraction

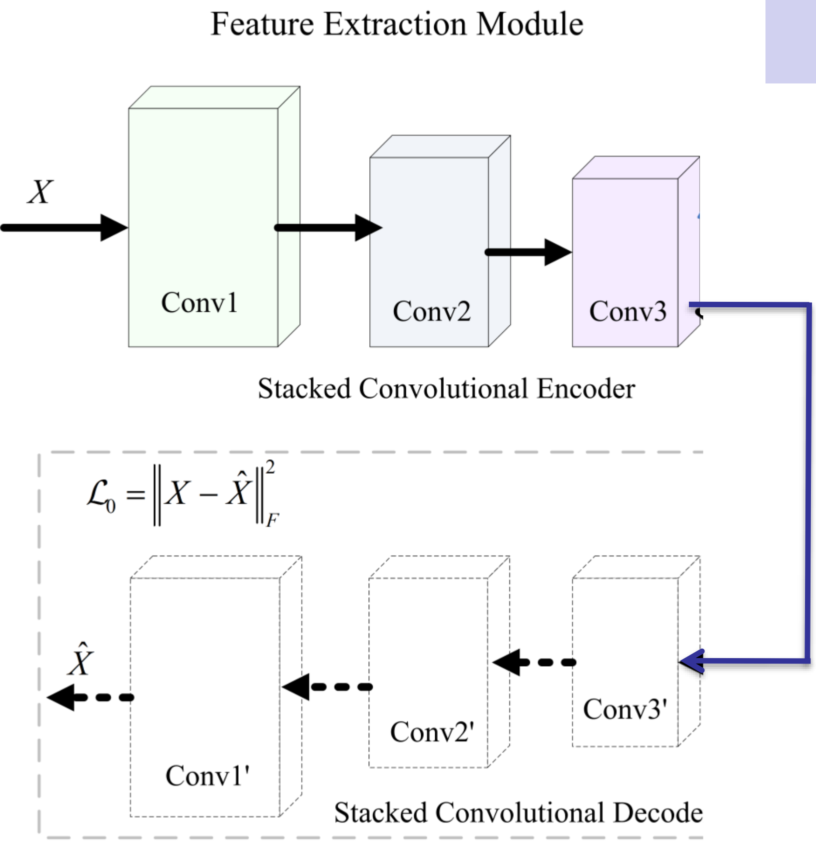


$$W = 0$$

$$Q = 0$$

$$C = \mathrm{I}$$

SCAE for Conv. Feature Extraction



$$W = 0$$

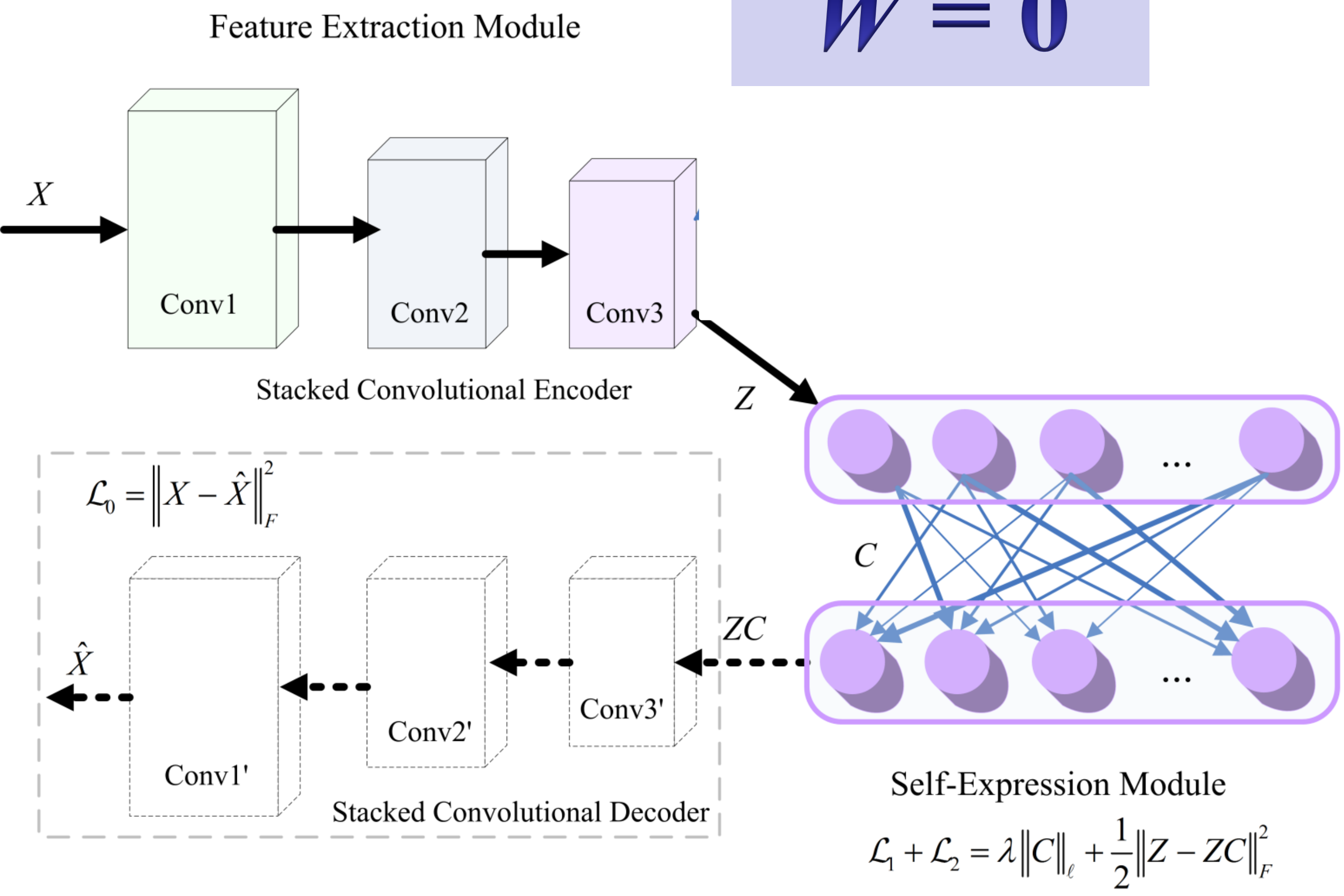
$$Q = 0$$

$$C = \mathrm{I}$$

SCAE + Self Expression

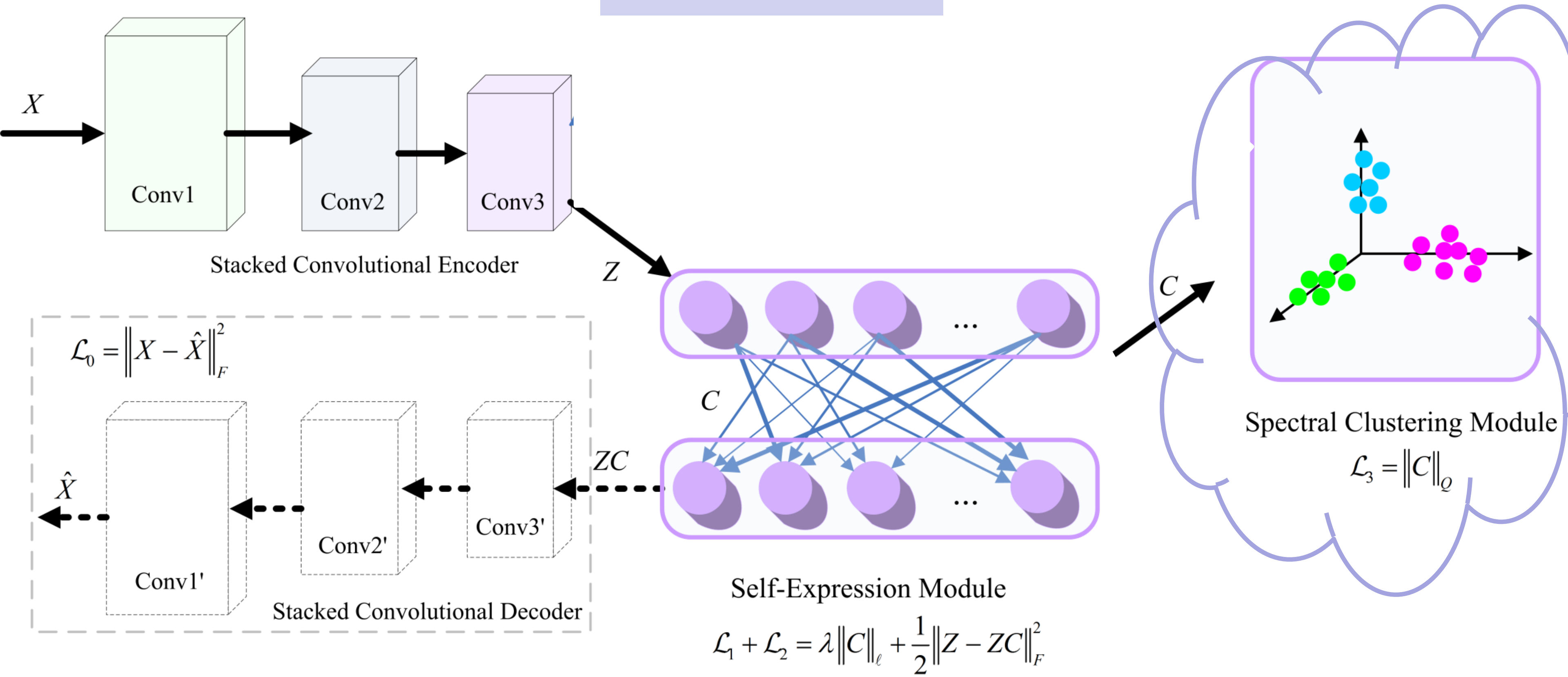
$$W = 0$$

$$Q = 0$$



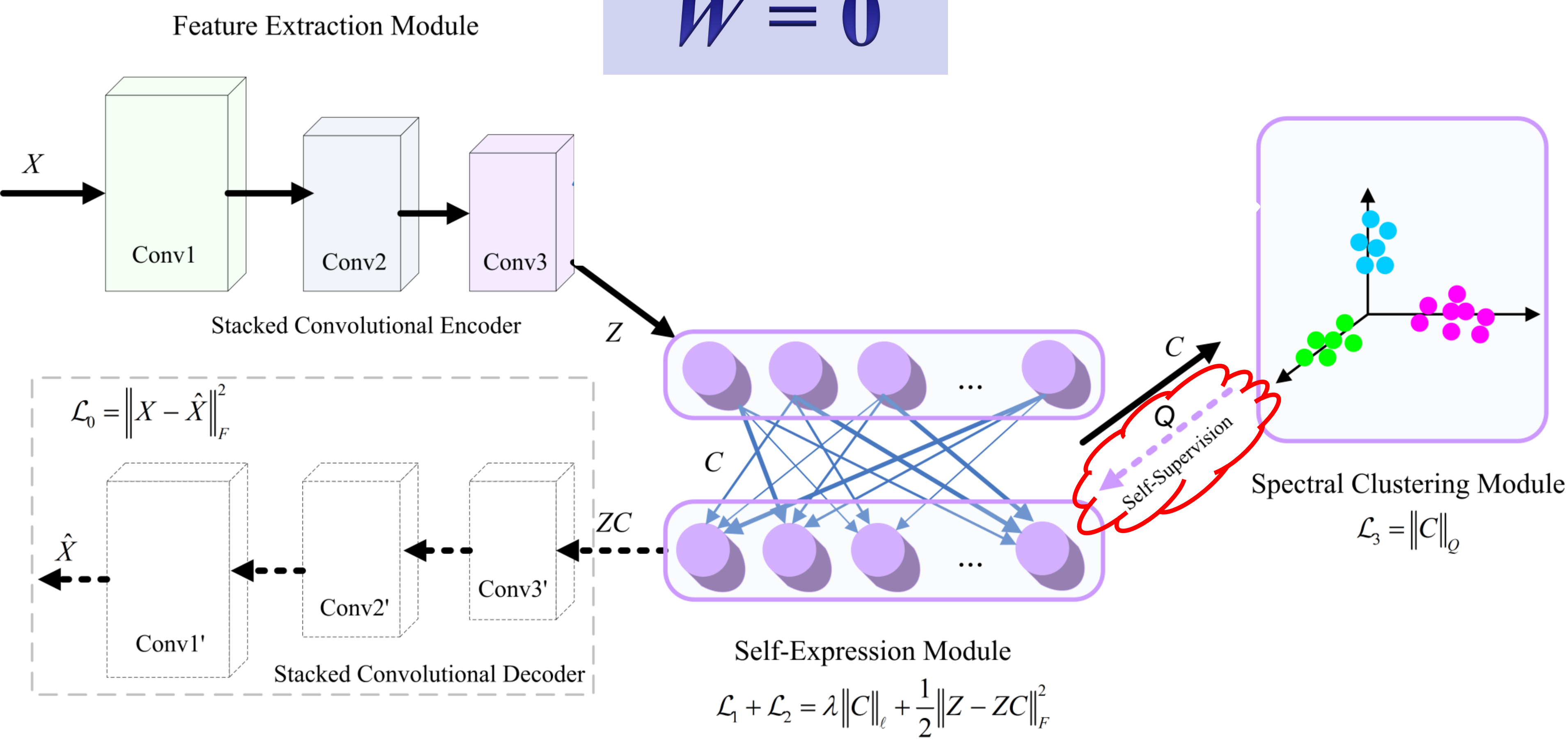
SCAE + Self Expression + Spectral Clustering

$$W = 0$$



SCAE + Self Expression + Spectral Clustering

$$W = 0$$



Reformulate Spectral Clustering

- Spectral clustering

$$\operatorname{trace}\left(\mathrm{Q}^T \cdot L \cdot \mathrm{Q}\right) = \operatorname{trace}\left(\mathrm{Q}^T \cdot (D - A) \cdot \mathrm{Q}\right) = \sum_{i,j=1}^{N} \left|C_{ij}\right| \cdot \frac{1}{2} \left\|\mathbf{q}^{(i)} - \mathbf{q}^{(j)}\right\|_2^2$$

where $A = \left(\left|C\right| + \left|C^T\right|\right)/2,\ \mathrm{Q}^T = \left[\mathbf{q}^{(1)}, \mathbf{q}^{(2)}, \ldots, \mathbf{q}^{(N)}\right].$

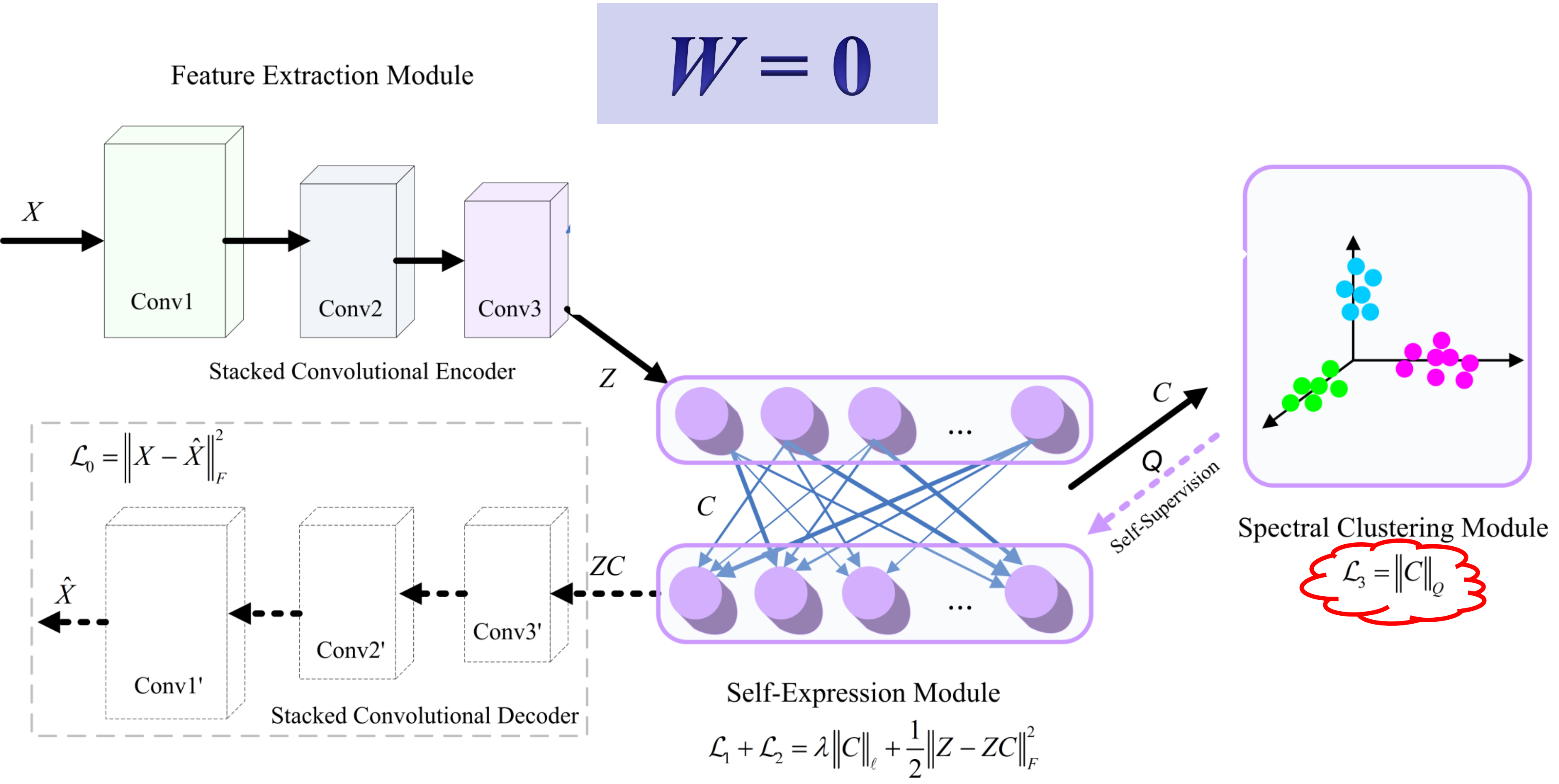
  – Connection to coefficients matrix of C $w.\ r.\ t.$ Q :

$$\operatorname{trace}\left(\mathrm{Q}^T \cdot L \cdot \mathrm{Q}\right) = \sum_{i,j=1}^{N} \left|C_{ij}\right| \cdot \underbrace{\frac{1}{2} \left\|\mathbf{q}^{(i)} - \mathbf{q}^{(j)}\right\|_2^2} = \left\|C\right\|_{\mathrm{Q}}$$

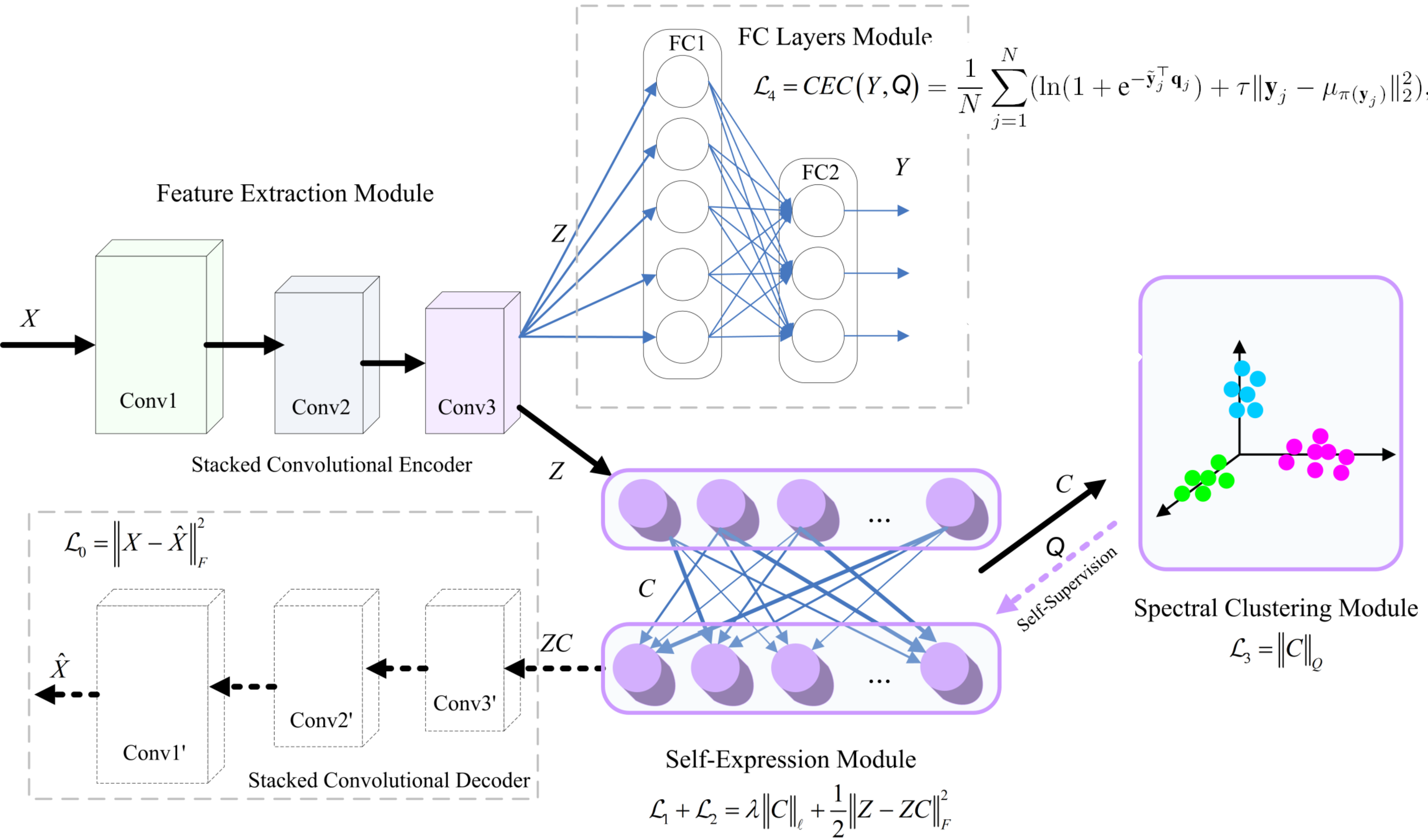
$$\begin{cases} 0, & \mathbf{q}^{(i)} = \mathbf{q}^{(j)} \text{ : in same subspace} \\ 1, & \mathbf{q}^{(i)} \neq \mathbf{q}^{(j)} \text{ : in different subspaces} \end{cases}$$

[1] C.-G. Li, and R. Vidal, "Structured Sparse Subspace Clustering: A Unified Optimization Framework", In: IEEE CVPR 2015.

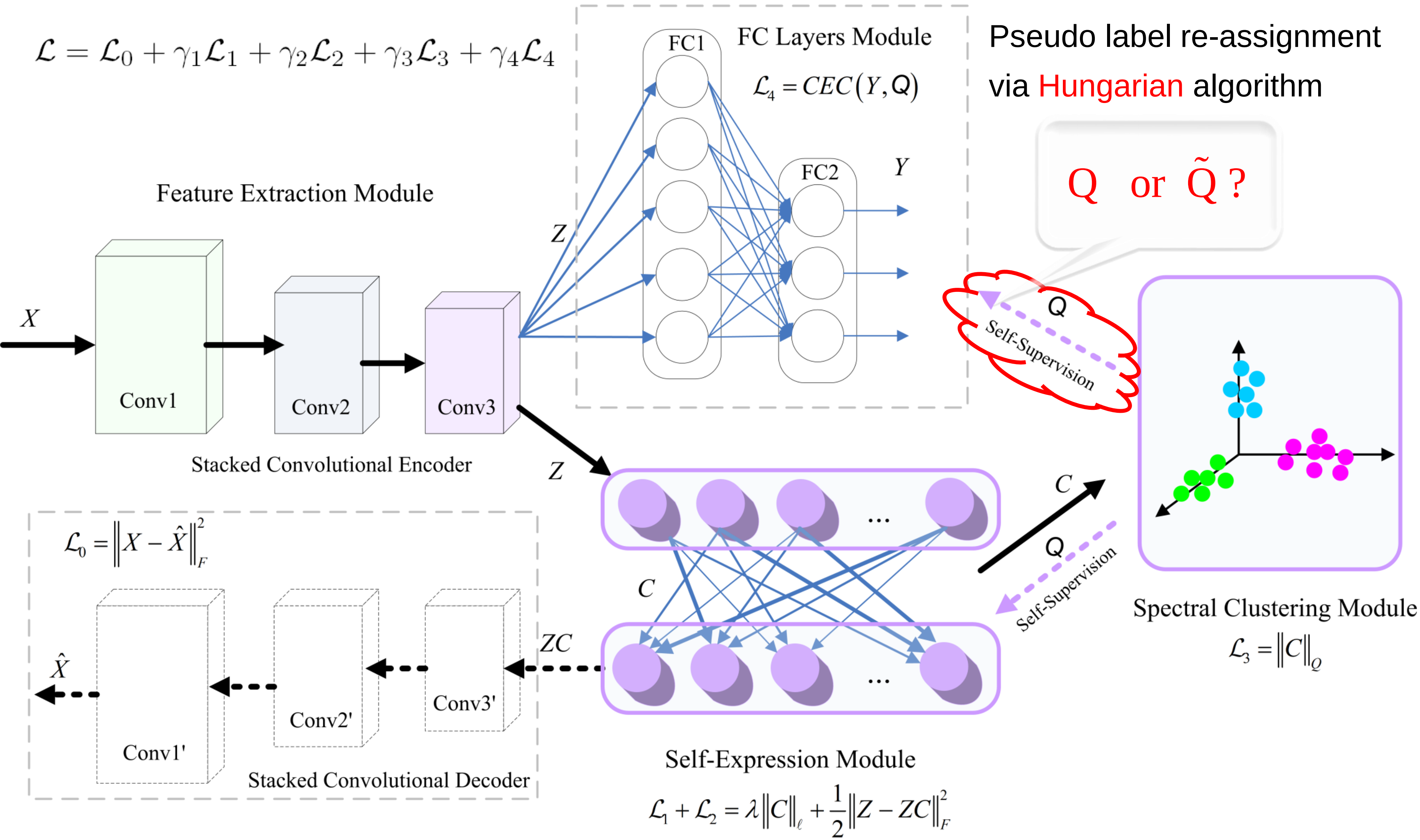
Spectral Clustering + Classification Block



Spectral Clustering + Classification Block



Spectral Clustering + Classification Block



Procedure for Training S²ConvSCN

$$\mathcal{L} = \mathcal{L}_0 + \gamma_1\mathcal{L}_1 + \gamma_2\mathcal{L}_2 + \gamma_3\mathcal{L}_3 + \gamma_4\mathcal{L}_4$$

dure for training S^2ConvSCN

Require: Input data, tradeoff parameters, maximum iteration T_{max}, T_0, and t=1.

1. Pre-train the stacked convolutional module via stacked CAE.
2. (Optional) Pre-train the stacked convolutional module with the self-expressive layer.
3. Initialize the FC layers.
4. Run self-expressive layer.
5. Run spectral clustering layer to get the segmentation Q.
6. **while** $t \leq T_{max}$ **do**
   Fixed Q, update the other parts T_0 epoches.
   Run spectral clustering once to update Q and set $t \leftarrow t+1$.
7. **end while**

Ensure: trained S^2ConvSCN and Q.

Outline

- Introduction
- Related Work
- Our Proposal
- Experimental Results
- Summary

Details in Conv. Module

	Extended Yale B		ORL	
Layers	kernel size	channels	kernel size	channels
encoder-1	5×5	10	3×3	3
encoder-2	3×3	20	3×3	3
encoder-3	3×3	30	3×3	5
decoder-1	3×3	30	3×3	5
decoder-2	3×3	20	3×3	3
decoder-3	5×5	10	3×3	3

  Table 1. Network settings for Extended Yale B and ORL.

	COIL20		COIL100	
Layers	kernel size	channels	kernel size	channels
encoder-1	3×3	15	5×5	50
decoder-1	3×3	15	5×5	50

  Table 4. Network settings for COIL20 and COIL100.

  - Keep the same as in DSCNet (Ji et al. NIPS'17)

Experimental Results

Methods	LRR	LRSC	SSC	AE+ SSC	KSSC	SSC-OMP	EDSC	AE+ EDSC	DSCNet-ℓ_1	DSCNet-ℓ_2	S^{2}ConvSCN-ℓ_2	S^{2}ConvSCN-ℓ_1
10 subjects												
Mean	22.22	30.95	10.22	17.06	14.49	12.08	5.64	5.46	2.23	1.59	**1.18**	**1.18**
Median	23.49	29.38	11.09	17.75	15.78	8.28	5.47	6.09	2.03	1.25	**1.09**	**1.09**
15 subjects												
Mean	23.22	31.47	13.13	18.65	16.22	14.05	7.63	6.70	2.17	1.69	1.14	**1.12**
Median	23.49	31.64	13.40	17.76	17.34	14.69	6.41	5.52	2.03	1.72	**1.14**	**1.14**
20 subjects												
Mean	30.23	28.76	19.75	18.23	16.55	15.16	9.30	7.67	2.17	1.73	1.31	**1.30**
Median	29.30	28.91	21.17	16.80	17.34	15.23	10.31	6.56	2.11	1.80	1.32	**1.25**
25 subjects												
Mean	27.92	27.81	26.22	18.72	18.56	18.89	10.67	10.27	2.53	1.75	1.32	**1.29**
Median	28.13	26.81	26.66	17.88	18.03	18.53	10.84	10.22	2.19	1.81	1.34	**1.28**
30 subjects												
Mean	37.98	30.64	28.76	19.99	20.49	20.75	11.24	11.56	2.63	2.07	1.71	**1.67**
Median	36.82	30.31	28.59	20.00	20.94	20.52	11.09	10.36	2.81	2.19	1.77	**1.72**
35 subjects												
Mean	41.85	31.35	28.55	22.13	26.07	20.29	13.10	13.28	3.09	2.65	1.67	**1.62**
Median	41.81	31.74	29.04	21.74	25.92	20.18	13.10	13.21	3.10	2.64	1.69	**1.60**
38 subjects												
Mean	34.87	29.89	27.51	25.33	27.75	24.71	11.64	12.66	3.33	2.67	1.56	**1.52**
Median	34.87	29.89	27.51	25.33	27.75	24.71	11.64	12.66	3.33	2.67	1.56	**1.52**

Table 2. Clustering Error (%) on Extended Yale B. The best results are in bold and the second best results are underlined.

No. Subjects / Losses	10 subjects		15 subjects		20 subjects		25 subjects		30 subjects		35 subjects		38 subjects	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
$\mathcal{L}_0+\mathcal{L}_1+\mathcal{L}_2$(DSC [12])	2.23	2.03	2.17	2.03	2.17	2.11	2.53	2.19	2.63	2.81	3.09	3.10	3.33	3.33
$\mathcal{L}_0+\mathcal{L}_1+\mathcal{L}_2+\mathcal{L}_3$	1.58	1.25	1.63	1.55	1.67	1.57	1.61	1.63	2.74	1.82	2.64	2.65	2.75	2.75
$\mathcal{L}_0+\mathcal{L}_1+\mathcal{L}_2+\mathcal{L}_4$	1.32	**1.09**	1.31	1.30	1.54	1.48	1.48	1.98	1.87	**1.61**	1.82	1.84	1.92	1.92
$\mathcal{L}_0+\mathcal{L}_1+\mathcal{L}_2+\mathcal{L}_3+\mathcal{L}_4$	**1.18**	**1.09**	**1.12**	**1.14**	**1.30**	**1.25**	**1.29**	**1.28**	**1.67**	1.72	**1.62**	**1.60**	**1.52**	**1.52**

Table 3. Ablation Study on S^{2}ConvSCN-ℓ_1 on Extended Yale B.

EDSC	27.25	14.86	38.13
AE+EDSC	26.25	14.79	38.88
DSC-ℓ_2	14.00	5.42	30.96
DSC-ℓ_1	14.25	5.65	33.62
S^{2}ConvSCN-ℓ_2	11.25	2.33	27.83
S^{2}ConvSCN-ℓ_1	**10.50**	**2.14**	**26.67**

Table 5. Clustering Error (%) on ORL, COIL20 and COIL100.

More ablation study

-

Loss	ORL	COIL20	COIL100
$\mathcal{L}_0$	15.25	7.92	34.94
$\mathcal{L}_0 + \mathcal{L}_1 + \mathcal{L}_2$(DSC)	14.25	5.65	33.62
$\mathcal{L}_0 + \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3$	12.75	3.42	31.14
$\mathcal{L}_0 + \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_4$	12.25	3.27	28.53
$\mathcal{L}_0 + \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3 + L_4$	**10.50**	**2.14**	**26.67**

Table 1. Ablation Study of S^{2}ConvSCN-ℓ_1.

-

Loss	ORL	COIL20	COIL100
$\mathcal{L}_0$	15.40	7.92	32.63
$\mathcal{L}_0 + \mathcal{L}_1 + \mathcal{L}_2$(DSC)	14.25	5.65	30.96
$\mathcal{L}_0 + \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3$	12.25	3.15	31.79
$\mathcal{L}_0 + \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_4$	12.00	2.75	28.17
$\mathcal{L}_0 + \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3 + L_4$	**11.25**	**2.33**	**27.83**

Table 2. Ablation Study of S^{2}ConvSCN-ℓ_2.

Summary

✓ S^2ConvSCN := Convolution feature extraction

+ Self-expression

+ Spectral clustering

Thank you!

- This work is jointly done with:

Junjian Zhang

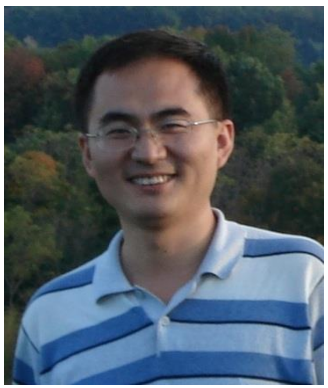
Xianbiao Qi

Chong You

Zhouchen Lin

Jun Guo

Honggang Zhang



> **For more information, please visit my homepage:**
> **http://www.pris.net.cn/teacher/lichunguang**