机器学习

7. PCA进阶

主要内容

- 低秩矩阵分解；低秩正则
- PCA的MAP理解
- 误差建模

- PCA的主成分方向需要限定正交吗?

三个相关联的问题：

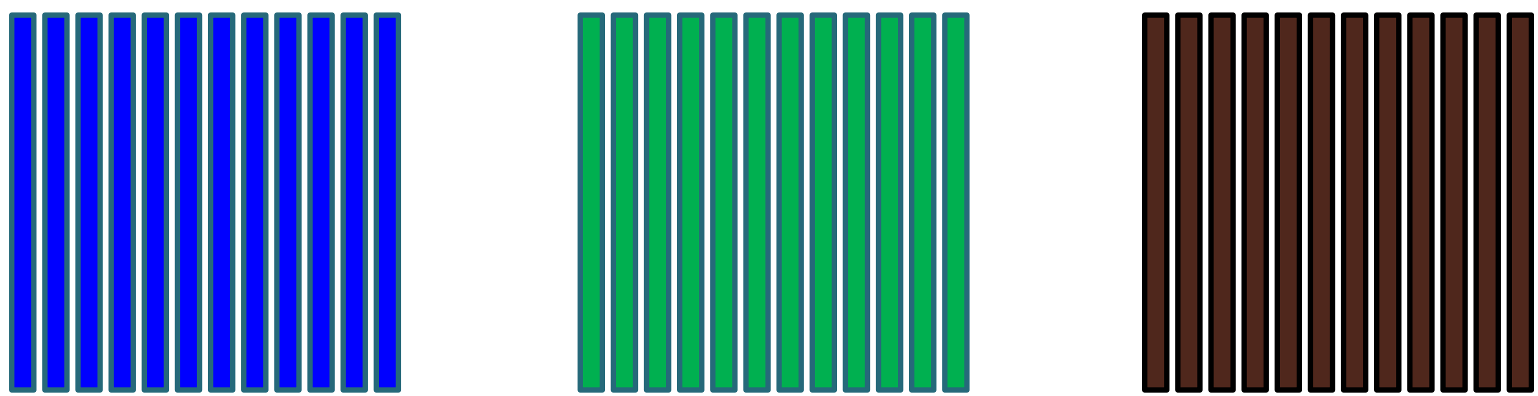
- 低秩矩阵分解
- 低秩矩阵逼近
- 低秩正则

主要内容

- 低秩矩阵分解；低秩正则
- PCA的MAP理解
- 误差建模

数据表达：

$$Y = X + E$$



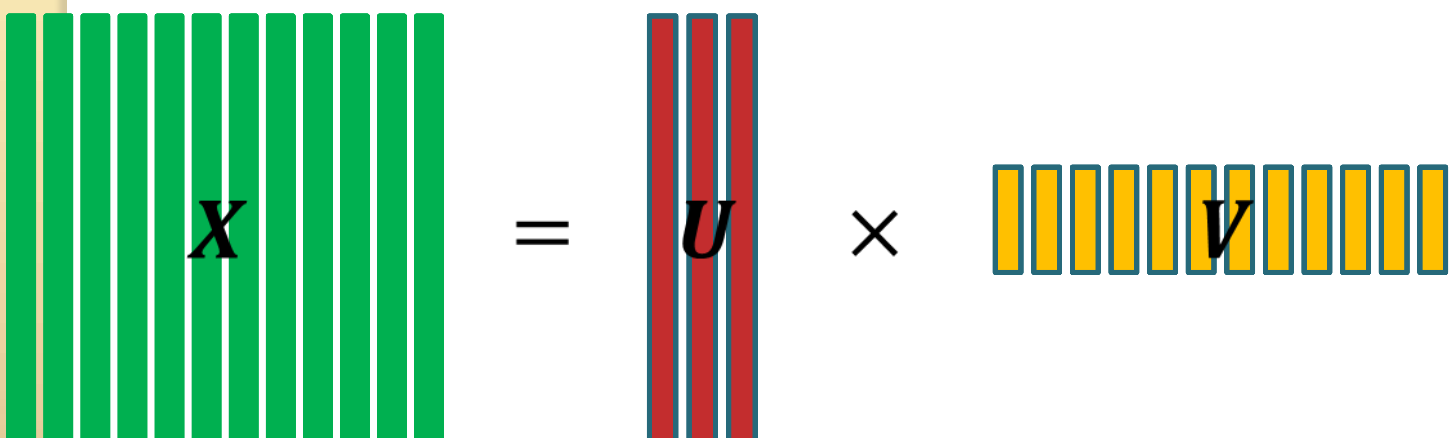
Each column corresponds to a sample

The desired latent low-rank matrix

The residual matrix

How to formulate $\boldsymbol{X}$?

- Formulation:
  - ✓ $\boldsymbol{X} \in \mathbb{R}^{d\times n}$; d: dimensionality; n: number of samples
  - ✓ Basis matrix: $\boldsymbol{U} \in \mathbb{R}^{d\times r}$ and $r(\ll d, n)$ is rank
  - ✓ Coefficient matrix: $\boldsymbol{V} \in \mathbb{R}^{n\times r}$



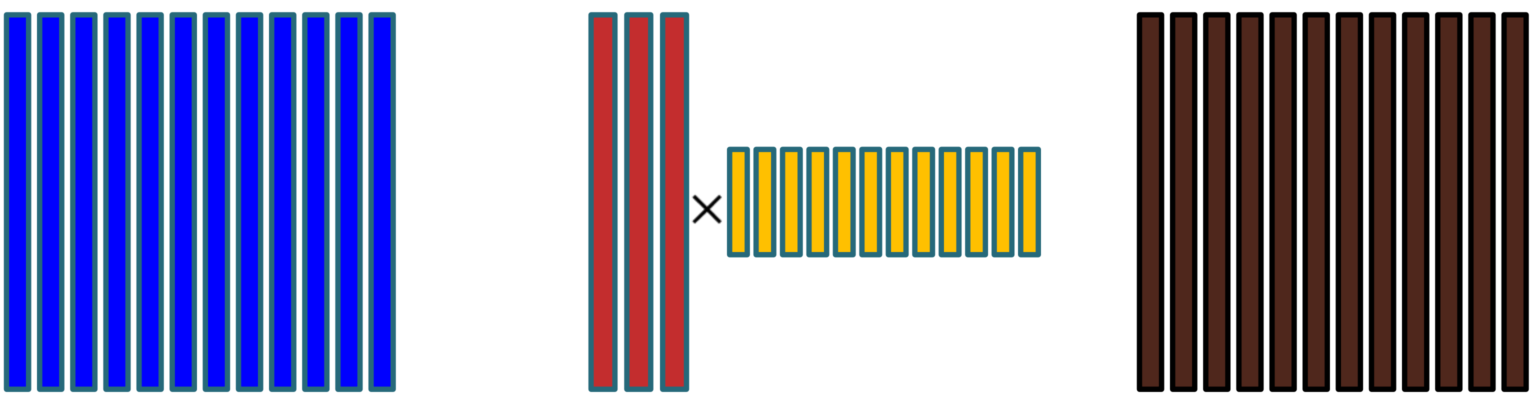
Each column corresponds to a basis vector

Each column corresponds to a coefficient vector of a sample

The problem

- Low rank matrix factorization (LRMF) aims to:
  - Given $\boldsymbol{Y}$, find its best low-rank approximation $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{V}^{T}$

$$\boldsymbol{Y} \quad = \quad \boldsymbol{U}\boldsymbol{V}^{T} \quad + \quad \boldsymbol{E}$$



The approximation residual

- The residual $\boldsymbol{E}$

In matrix form $$\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{V}^{\boldsymbol{T}} + \boldsymbol{E}$$

In element form $$y_{ij} = \boldsymbol{u}_{i,:}\boldsymbol{v}_{:,j} + \varepsilon_{ij}$$

- a_{ij} means the element of matrix $\boldsymbol{A}$ at its i^{th} row and j^{th} column
- $\boldsymbol{a}_{i,:}$ means the i^{th} row of the matrix $\boldsymbol{A}$
- $\boldsymbol{a}_{:,j}$ means the j^{th} column of the matrix $\boldsymbol{A}$

F-norm LRMF

- Assuming noise is i.i.d. Gaussian

$$\boxed{y_{ij} = \boldsymbol{u}_{i,:}\boldsymbol{v}_{:,j} + \varepsilon_{ij}} \qquad \boxed{p(\varepsilon_{ij}) \sim N(\varepsilon_{ij}|0, \sigma^2)}$$

- Likelihood function:

$$p(\boldsymbol{Y}) = \prod_{i,j} N(y_{ij}|\boldsymbol{u}_{i,:}\boldsymbol{v}_{:,j}, \sigma^2)$$

- Maximum likelihood estimation (MLE):

$$max_{U,V} \sum_{i,j} \log N(y_{ij}|\boldsymbol{u}_{i,:}\boldsymbol{v}_{:,j}, \sigma^2) \Longrightarrow min_{U,V} \sum_{i,j} (y_{ij} - \boldsymbol{u}_{i,:}\boldsymbol{v}_{:,j})^2$$

$$\boxed{min_{\boldsymbol{U},\boldsymbol{V}} \|\boldsymbol{Y} - \boldsymbol{U}\boldsymbol{V}^T\|_F^2}$$

主要内容

- 低秩矩阵分解；低秩正则
- PCA的MAP理解
- 误差建模

低秩矩阵分解

- 已知数据： $\mathbf{X} \in \mathbf{R}^{d \times n}$
- 计算两个低秩矩阵

$$\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k] \in \mathbf{R}^{d \times k}, \mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k] \in \mathbf{R}^{n \times k}, \mathbf{k} < \mathbf{d}, \mathbf{n}$$

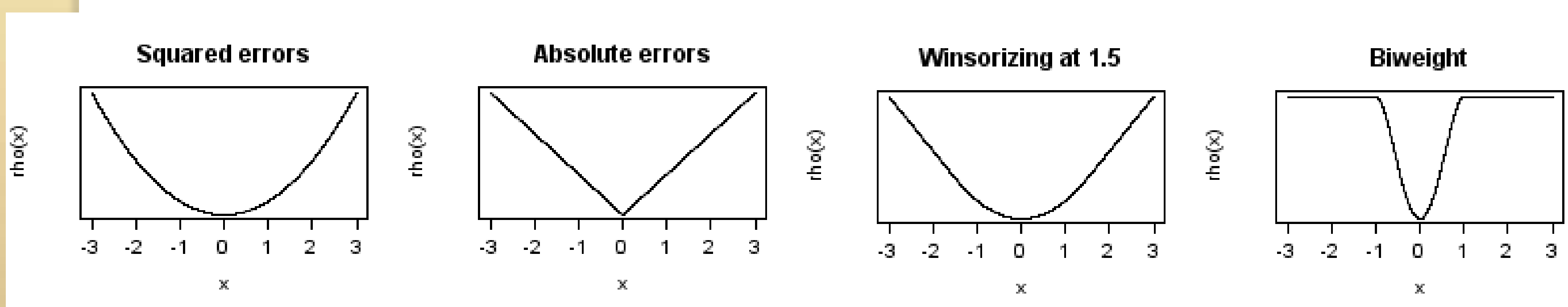
- 目标：使低秩矩阵乘积尽可能重建原矩阵

$$\boxed{X} \longrightarrow \boxed{UV^T}$$

低秩矩阵分解

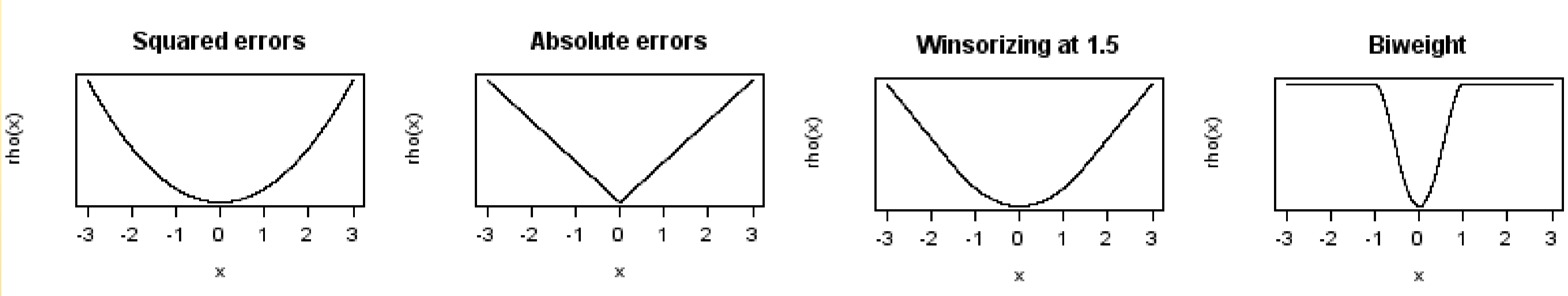
- 关键问题：
  - 如何度量原数据与重建数据偏差?

$X \longrightarrow UV^T$



低秩矩阵分解

- 性能?



- 最常见误差函数选择:

$$\| (\mathbf{X} - \mathbf{UV}) \|_F \qquad \| (\mathbf{X} - \mathbf{UV}) \|_1$$

The approximation residual

- The residual $\boldsymbol{E}$

In matrix form $\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{V}^{\boldsymbol{T}} + \boldsymbol{E}$

In element form $y_{ij} = \boldsymbol{u}_{i,:}\boldsymbol{v}_{:,j} + \varepsilon_{ij}$

- a_{ij} means the element of matrix $\boldsymbol{A}$ at its i^{th} row and j^{th} column
- $\boldsymbol{a}_{i,:}$ means the i^{th} row of the matrix $\boldsymbol{A}$
- $\boldsymbol{a}_{:,j}$ means the j^{th} column of the matrix $\boldsymbol{A}$

L_1-norm LRMF

- Assuming noise is i.i.d. Laplacian

$$\boxed{y_{ij} = \boldsymbol{u}_{i,:}\boldsymbol{v}_{:,j} + \varepsilon_{ij}}$$

$$\boxed{p(\varepsilon_{ij}) \sim Laplacian(\varepsilon_{ij}}$$

- Likelihood function:

$$p(\boldsymbol{Y}) = \prod_{i,j} Laplacian(y_{ij}|\boldsymbol{u}_{i,:}\boldsymbol{v}_{:,j}, \eta)$$

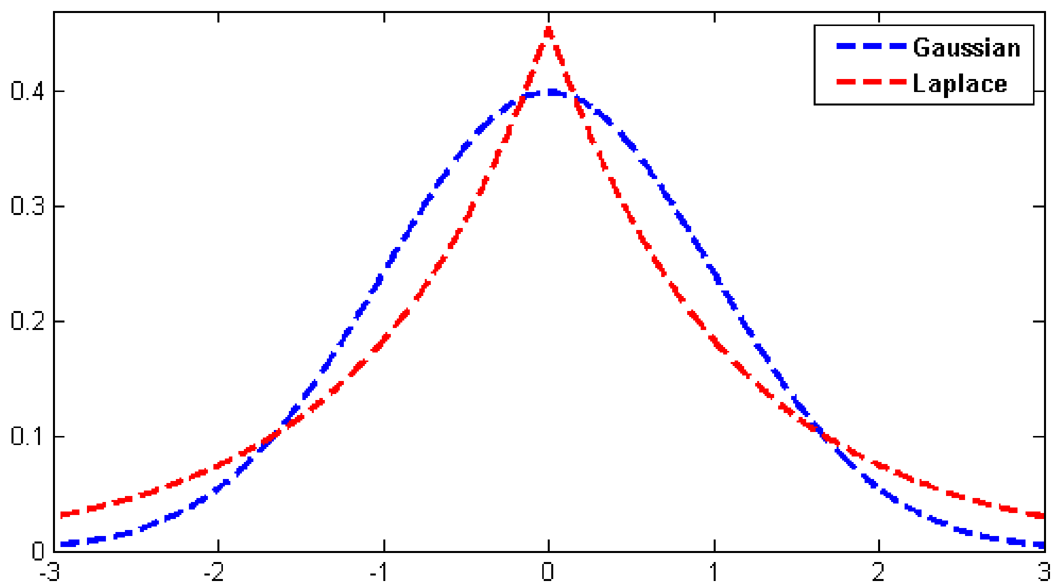
- Maximum likelihood estimation (MLE):

$$max_{U,V} \sum_{i,j} \log Laplacian(y_{ij}|\boldsymbol{u}_{i,:} \quad \Rightarrow \quad min_{U,V} \sum_{i,j} |y_{ij} - \boldsymbol{u}_{i,:}\boldsymbol{v}_{:,j}|$$

$$\boxed{min_{\boldsymbol{U},\boldsymbol{V}} \left\| \boldsymbol{Y} - \boldsymbol{U}\boldsymbol{V}^T \right\|_1}$$

L_1-norm LRMF

- Laplacian noise modeling is more suitable for outliers/heavy noises than Gaussian.
  - Heavy/long tailed distribution, which better adapts the noises with large values.



The residual model

- The best approximation depends on the assumption on the distribution of residual (or noise)
  - I.I.D. Gaussian noise
  - I.I.D. Laplacian noise (outliers and sparse noise)
  - Mixture of Gaussians
  - More complex noise

误差函数的本质

$$\boxed{D = f(W) + E, \quad e \sim p(e)}$$

$$\min_{f \in F} \mathrm{L}(f(W), \mathrm{D}) + \mathrm{R}(W)$$

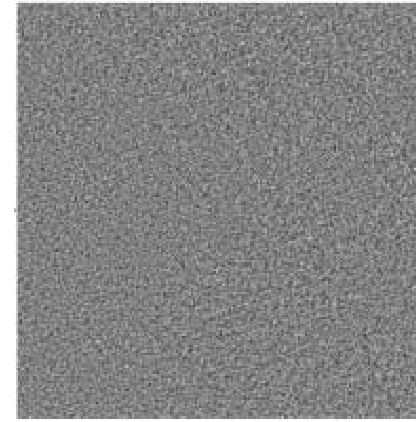
$$\mathrm{L}(f(W), \mathrm{D}) = \mathrm{L}(\mathrm{E})$$

$$\mathrm{p}(\mathrm{W}|\mathrm{D}) \sim Likelihood(\mathrm{D}|\mathrm{W})\mathrm{p}(\mathrm{W})$$

$$Likelihood(\mathrm{D}|\mathrm{W}) = \prod p(e_i)$$

$$=$$

$$+$$



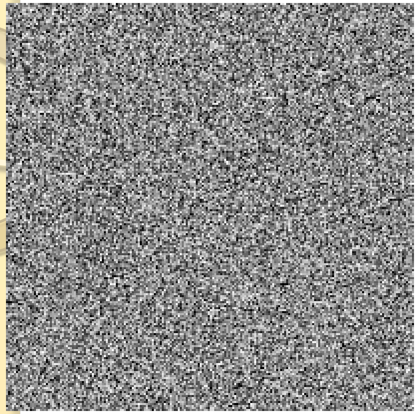
$\mathbf{X}$ $\quad$ $f(\mathbf{W})$ $\quad$ $\mathbf{E}:\ e \sim p(e)$

$$\min_{f \in \Omega} \quad \mathcal{L}(f(\mathbf{W}), \mathbf{X}) + \mathrm{R}(\mathbf{W})$$

Stochastic Knowledge

Deterministic Knowledge

What we assume noise as:

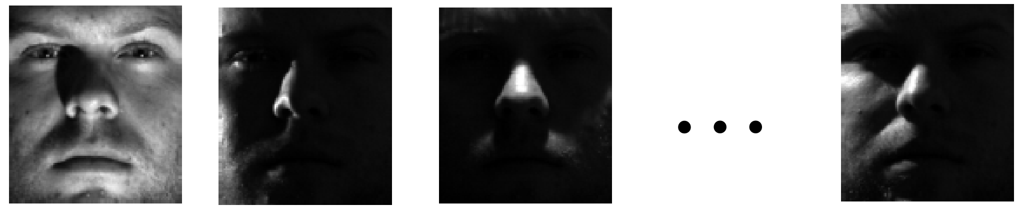


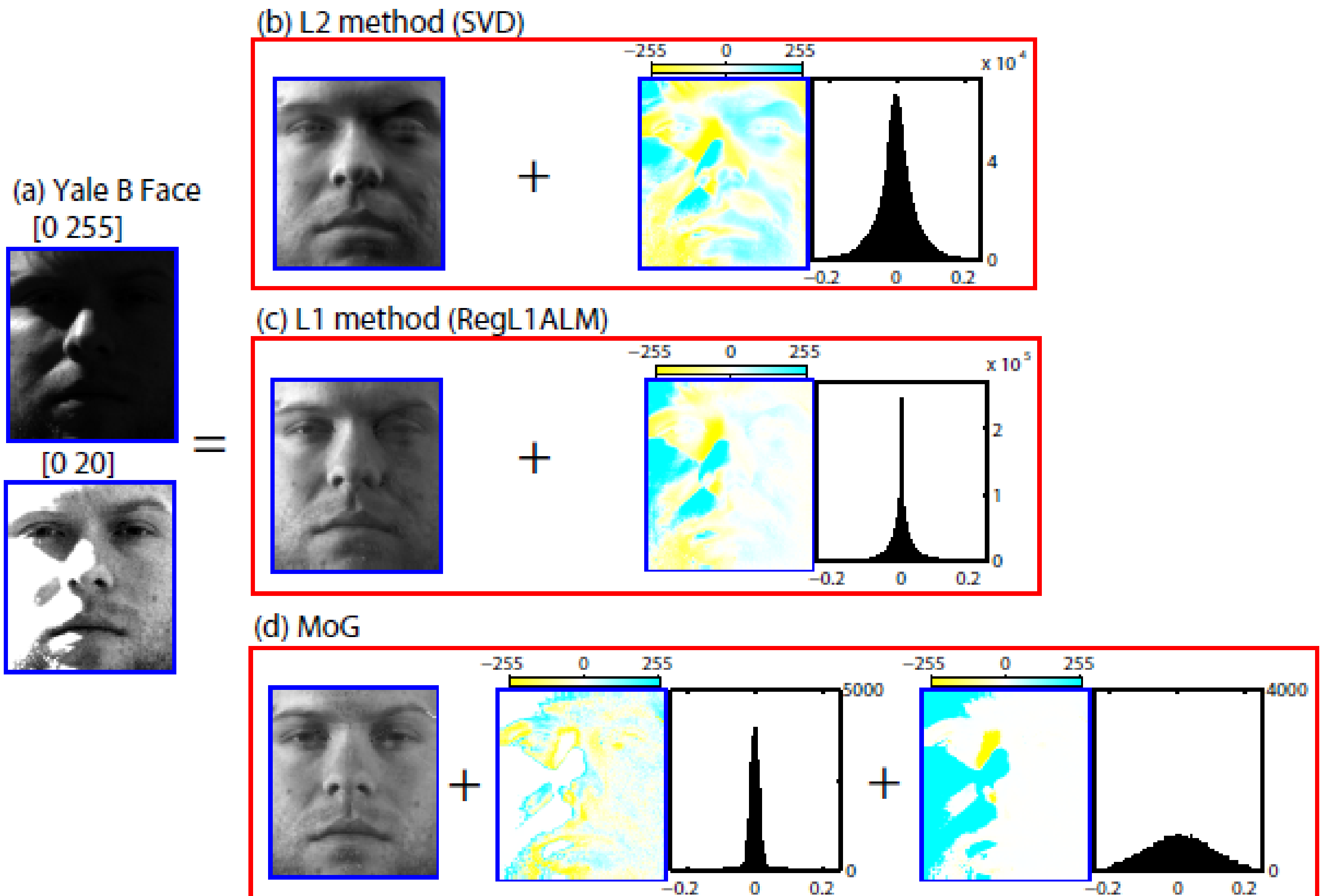
But the real noise is:

Robust Problem

$$\min_{f\in\Omega} \boxed{\mathcal{L}(f(\mathbf{W}),\mathbf{X})} + \mathrm{R}(\mathbf{W})$$

Yale B faces:





影像噪音

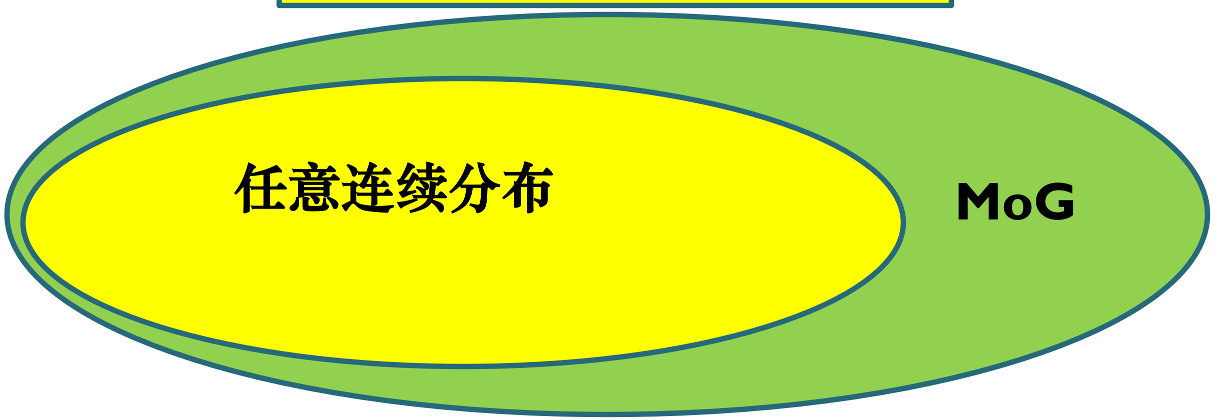
过饱和/阴影噪音

➢ 一种解决方案：**混合高斯！！**

$$x_{ij} = \widetilde{\mathbf{u}}_i^T \widetilde{\mathbf{v}}_j + \varepsilon_{ij}$$

$$p(\varepsilon) \sim \sum_{k=1}^{K} \pi_k \mathcal{N}(\varepsilon | 0, \sigma_k^2)$$

MoG的万有逼近性



(Maz'ya and Schmidt, 1996)

➢ 如：拉普拉斯分布可被等价表达为一个尺度化后的MoG

(Andrews and Mallows, 1974)

最大似然模型：

$$x_{ij} = \widetilde{\mathbf{u}}_i^T \widetilde{\mathbf{v}}_j + \varepsilon_{ij}$$

$$p(\varepsilon) \sim \sum_{k=1}^{K} \pi_k \mathcal{N}(\varepsilon|0, \sigma_k^2)$$

$$\begin{aligned} p(\mathbf{X}|\mathrm{U}, \mathbf{V}, \mathbf{\Pi}, \mathbf{\Sigma}) &= \prod_{i,j\in\Omega} p(x_{ij}|(\mathbf{u}^i)^T\mathbf{v}^j, \mathbf{\Pi}, \mathbf{\Sigma}) \\ &= \prod_{i,j\in\Omega} \sum_{k=1}^{K} \pi_k \mathcal{N}(x_{ij}|(\mathbf{u}^i)^T\mathbf{v}^j, \sigma_k^2) \end{aligned}$$

$$\max_{\mathrm{U},\mathbf{V},\mathbf{\Pi},\mathbf{\Sigma}} \mathcal{L}(\mathrm{U}, \mathbf{V}, \mathbf{\Pi}, \mathbf{\Sigma}) = \sum_{i,j\in\Omega} \log \sum_{k=1}^{K} \pi_k \mathcal{N}(x_{ij}|(\mathbf{u}^i)^T\mathbf{v}^j, \sigma_k^2)$$

- **EM 算法!**

E 步:

$$E(z_{ijk}) = \gamma_{ijk} = \frac{\pi_k \mathcal{N}(x_{ij}|(\mathbf{u}^i)^T\mathbf{v}^j, \sigma_k^2)}{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_{ij}|(\mathbf{u}^i)^T\mathbf{v}_j^j, \sigma_k^2)}$$

$$\boxed{\begin{aligned} &E_{\mathbf{Z}}\ \log p(\mathbf{X}, \mathbf{Z}|\mathbf{U}, \mathbf{V}, \mathbf{\Pi}, \mathbf{\Sigma}) = \\ &\sum_{i,j\in\Omega} \sum_{k=1}^{K} \gamma_{ijk} \left(\log \pi_k - \log \sqrt{2\pi}\sigma_k - \frac{(x_{ij} - (\mathbf{u}^i)^T\mathbf{v}^j)^2}{2\pi\sigma_k^2} \right) \end{aligned}}$$

M 步:

$$N_k = \sum_{i,j} \gamma_{ijk}, \quad \pi_k = \frac{N_k}{N},$$

$$\sigma_k^2 = \frac{1}{N_k} \sum_{i,j} \gamma_{ijk} (x_{ij} - (\mathbf{u}^i)^T\mathbf{v}^j)^2$$

$$\sum_{i,j\in\Omega} \sum_{k=1}^{K} \gamma_{ijk} \left(-\frac{(x_{ij} - (\mathbf{u}^i)^T\mathbf{v}^j)^2}{2\sigma_k^2} \right) = -\sum_{i,j\in\Omega} \left(\sum_{k=1}^{K} \frac{\gamma_{ijk}}{2\sigma_k^2} \right) (x_{ij} - (\mathbf{u}^i)^T\mathbf{v}^j)^2$$

$$\Longleftrightarrow$$

$$\|\mathbf{W} \odot (\mathbf{X} - \mathbf{UV})\|_F^2$$

EM算法

- M步骤

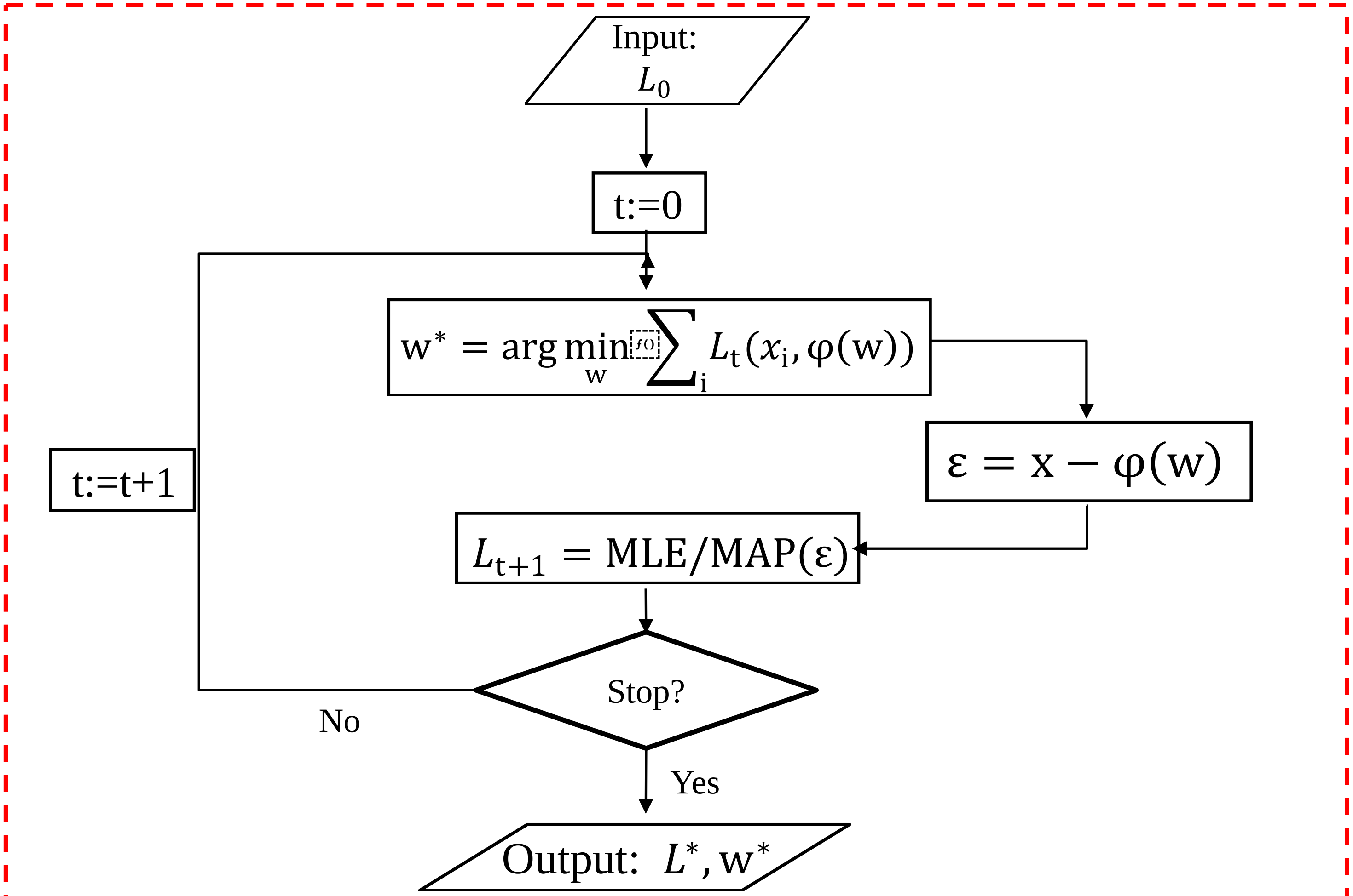
$$Q(\theta^t|\theta^{t-1}) = \sum_{j=1}^{n}\sum_{i=1}^{K} P(y_j = i|x_j, \theta^{t-1}) \log P(x_j, y_j = i|\theta^t)$$

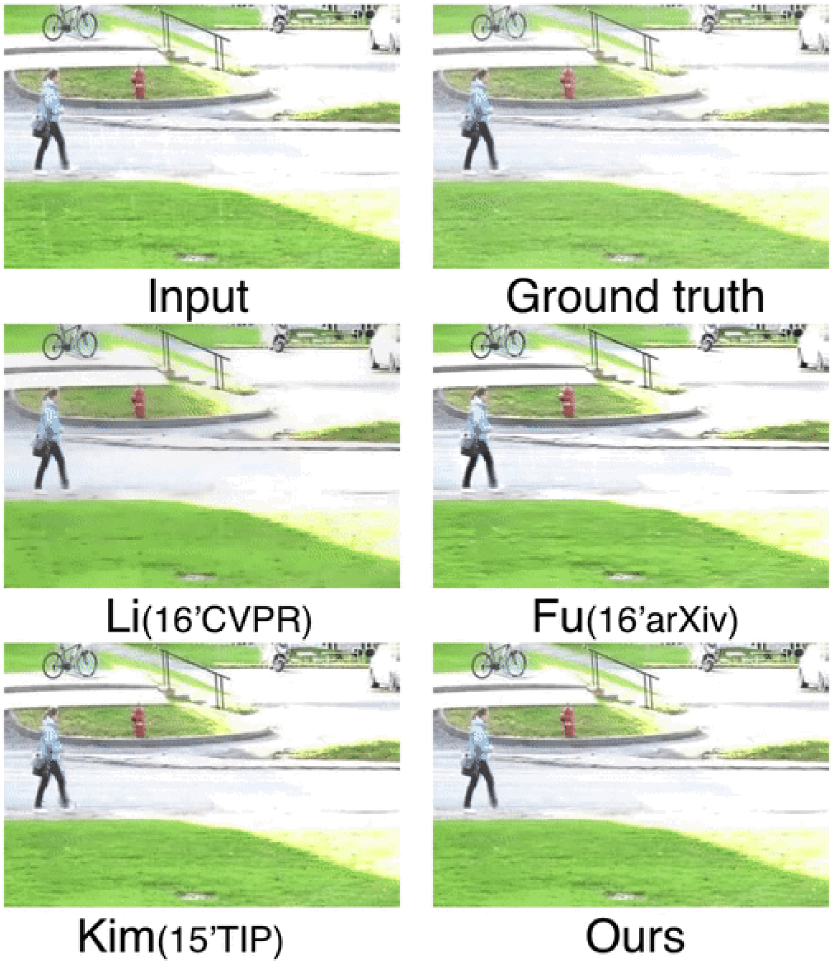
$$= \sum_{j=1}^{n}\sum_{i=1}^{K} P(y_j = i|x_j, \theta^{t-1})\underbrace{[\log \underbrace{P(x_j|y_j = i, \theta^t)}_{\propto \exp(-\frac{1}{2\sigma^2}\|x_j - \mu_i^t\|^2)} + \log \underbrace{P(y_j = i|\theta^t)}_{\pi_i}]}_{\text{联合分布}}$$

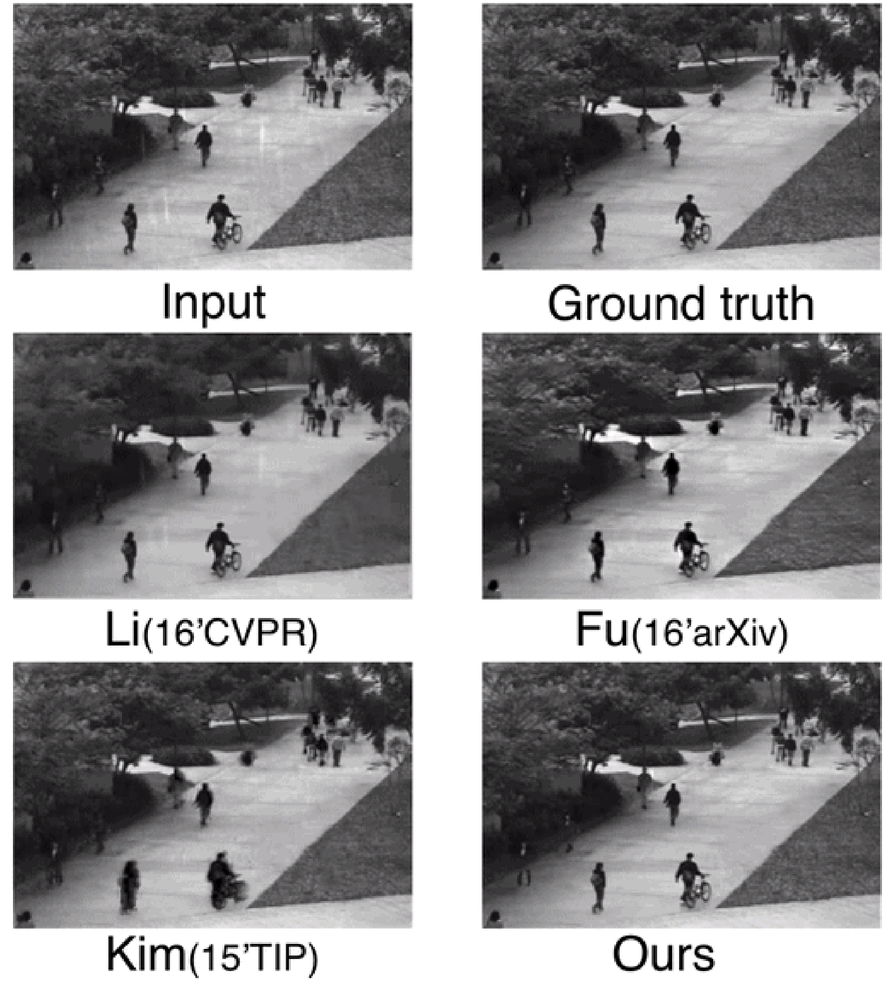
E步已获得

Loss Modeling Principle

- Assume the loss function L_θ containing certain parameters θ
- Learn L_θ from data

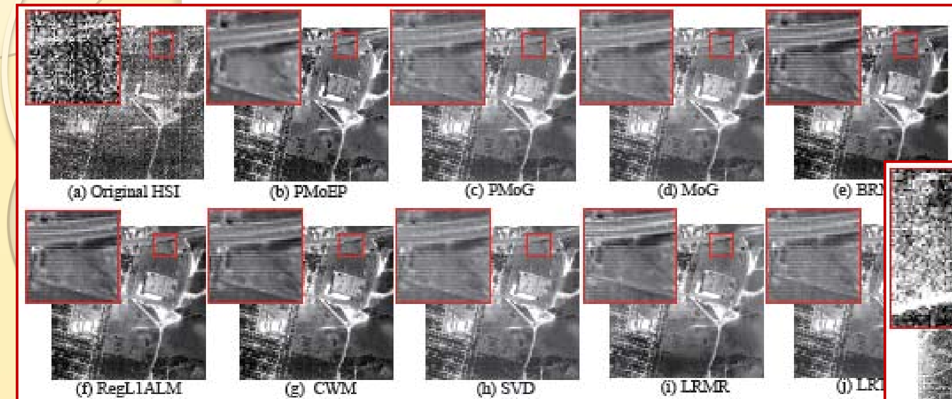


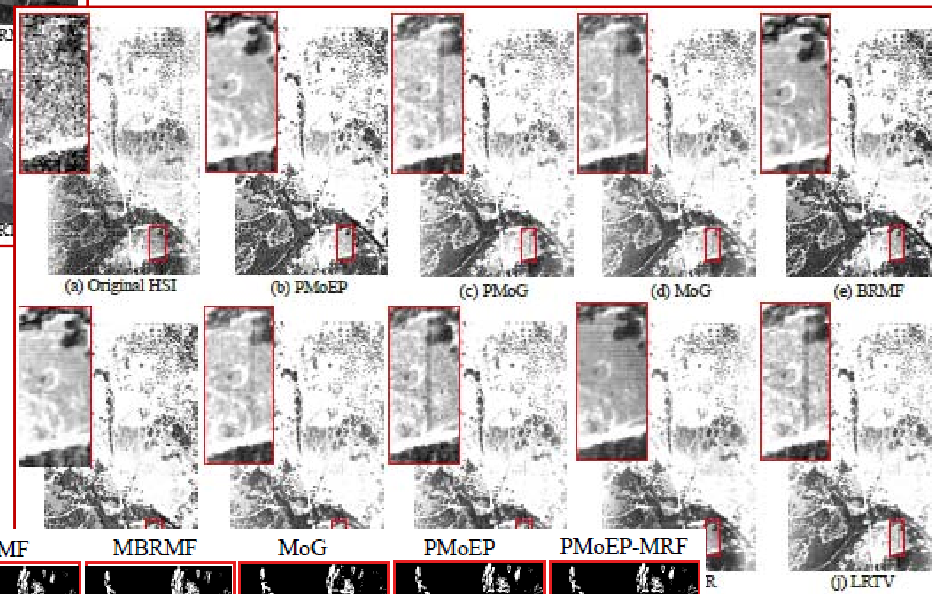


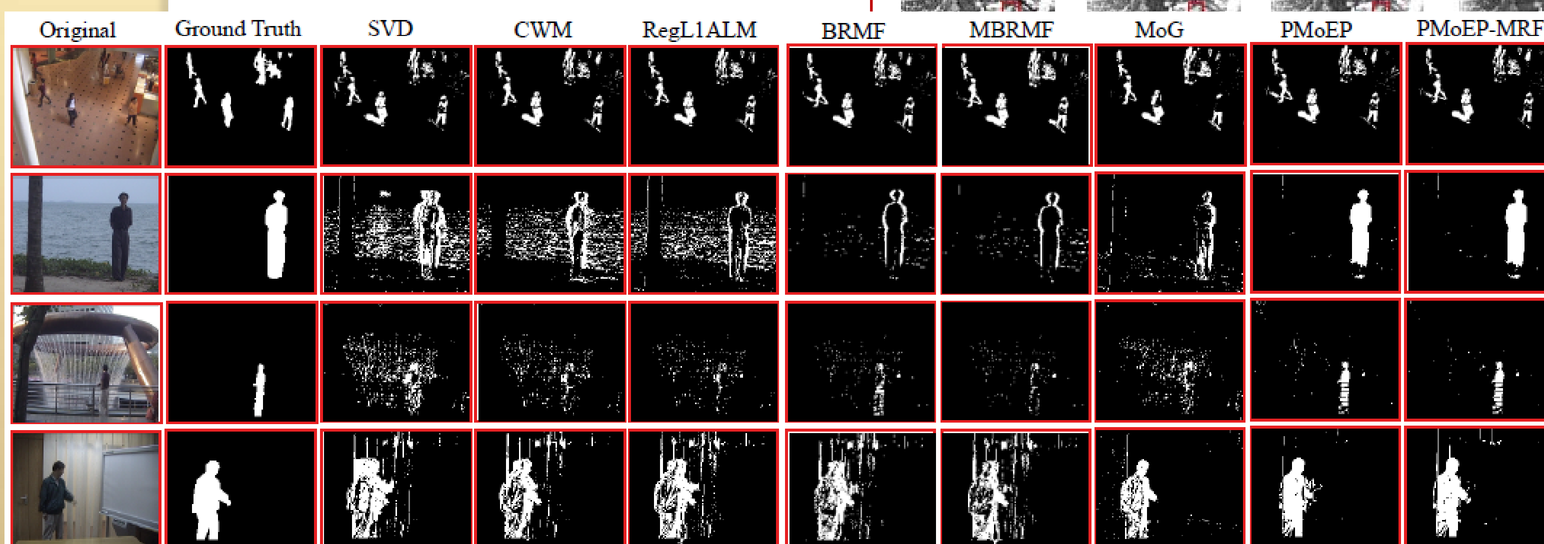


Dataset	Dataset 1				Dataset 2				Dataset 3				Dataset			
Metrics	VIF	SSIM	FSIM	UQI	VIF	SSIM	FSIM	UQI	VIF	SSIM	FSIM	UQI	VIF	SSIM	FSIM	UQI
Input	0.846	0.981	0.991	0.934	0.731	0.950	0.975	0.927	0.591	0.877	0.935	0.816	0.717	0.917	0.970	0.763
Fu [10]	0.696	0.956	0.968	0.847	0.673	0.948	0.971	0.923	0.530	0.887	0.933	0.812	0.670	0.935	0.967	0.808
Garg [14]	0.862	0.984	0.990	0.949	0.745	0.961	0.979	0.944	0.712	0.935	0.969	0.887	0.707	0.920	0.972	0.772
Kim [17]	0.810	0.981	0.987	0.941	0.642	0.949	0.968	0.933	0.666	0.943	0.967	0.907	0.589	0.912	0.960	0.758
Ours	**0.904**	**0.993**	**0.993**	**0.969**	**0.786**	**0.977**	**0.985**	**0.968**	**0.757**	**0.960**	**0.980**	**0.952**	**0.768**	**0.949**	**0.981**	**0.891**

W Wei, LX Yi, DY Meng, et al., ICCV 2017







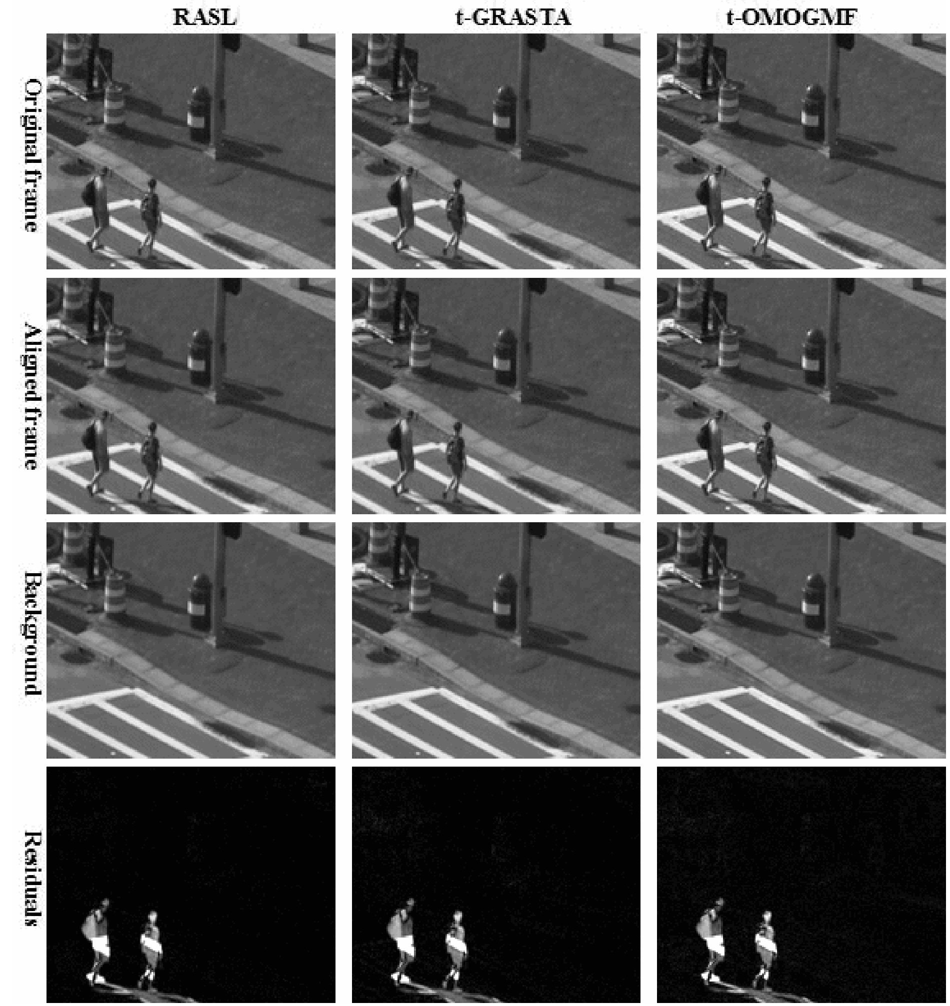
XY Cao, Q Zhao, DY Meng, et al., TIP 2016

Methods	data									
	air.	*boo.*	*sho.*	*lob.*	*esc.*	*cur.*	*cam.*	*wat.*	*fou.*	Average
RPCA [16]	71.11	**67.67**	**72.79**	78.12	64.09	81.65	44.56	65.56	72.39	68.66
GODEC [19]	62.69	58.39	70.71	73.29	57.42	59.84	43.71	48.79	66.01	60.09
RegL1 [29]	65.63	62.46	71.97	75.27	60.95	62.69	44.42	57.86	73.17	63.82
PRMF [17]	65.87	62.29	71.99	75.32	60.20	65.17	44.04	61.95	72.98	64.42
OPRMF [17]	66.17	61.82	71.95	73.99	60.12	70.86	42.89	61.89	71.80	64.61
GRASTA [21]	61.87	58.07	71.47	60.98	57.26	68.20	44.53	75.88	69.23	63.05
incPCP [38]	59.84	62.47	71.28	75.83	45.59	61.10	44.55	74.94	70.49	62.90
PracReProCS [37]	70.01	*63.71*	71.61	61.89	56.08	77.74	42.28	*87.53*	62.76	65.96
OMoGMF	*74.08*	59.87	71.80	78.01	61.42	86.08	44.48	87.34	71.78	70.54
DECOLOR [18]	63.98	59.97	65.37	68.93	*75.93*	*89.56*	**77.14**	64.03	**86.76**	*72.41*
GOSUS [22]	65.80	61.95	72.12	*80.97*	**86.27**	68.26	51.30	84.37	73.15	71.35
OMoGMF+TV	**77.20**	61.17	*72.43*	**83.47**	66.37	**92.54**	*65.88*	**93.14**	*82.53*	**77.19**

➢ ***Better foreground object detection***

Video	*esc.*	*air.*	*sho.*
Frame Size	130× 160	144× 176	256× 320
OPRMF [17]	0.5	0.4	0.1
PracReProCS [37]	1.5	1.2	0.2
GOSUS [22]	3.8	2.7	0.6
OMoGMF+TV	18.5	14.8	3.5
OMoGMF	99.6	63.0	5.2
GRASTA [21]	*166.9*	*123.9*	*28.7*
incPCP [38]	**274.5**	**220.8**	**85.2**
GRASTA&1%SS	303.2	246.7	65.5
OMoGMF&1%SS	**332.0**	**263.6**	**104.7**

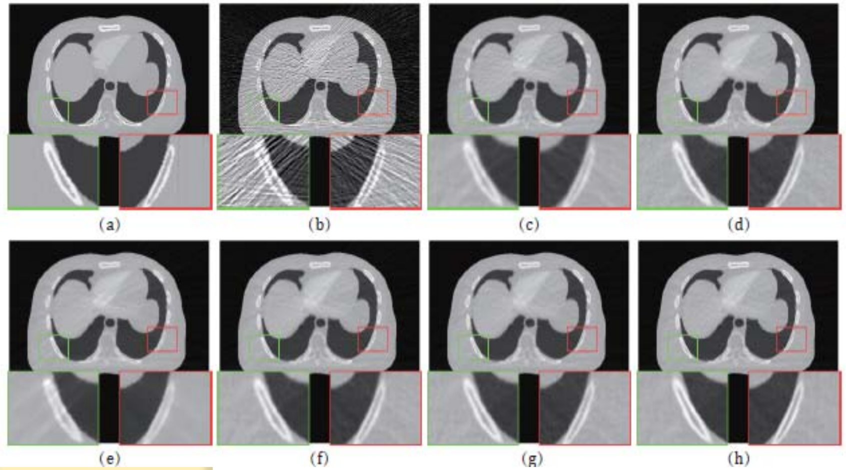
➢ ***Faster computational speed***

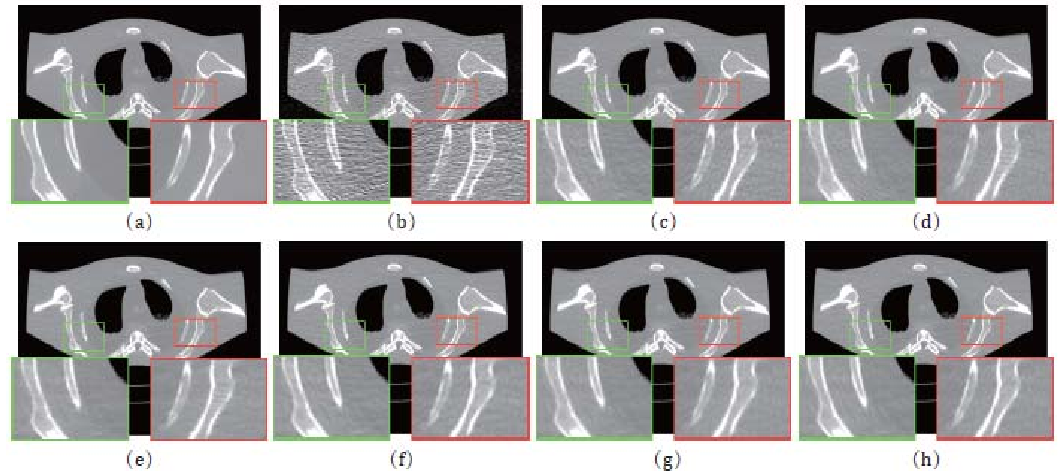


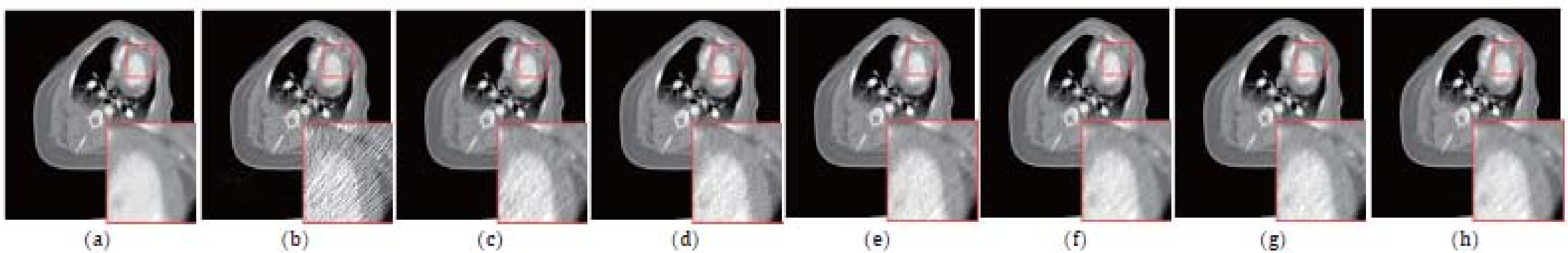
➢ ***Better background scene subtraction***

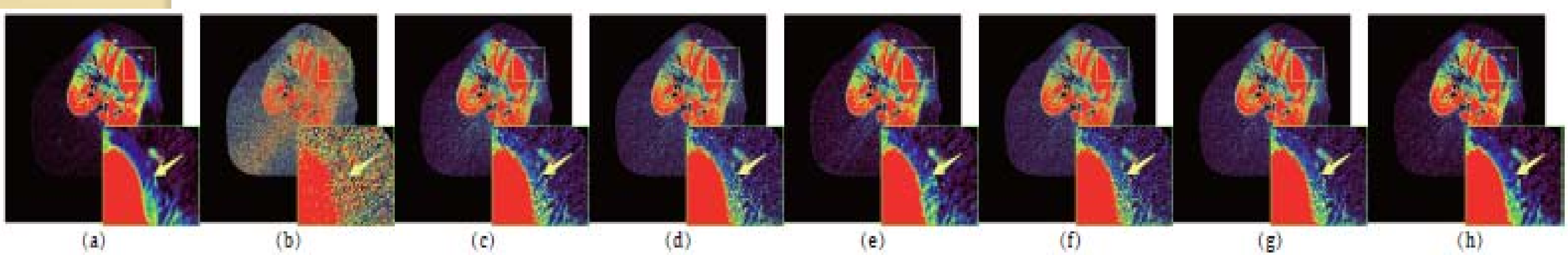
Please see more demos in http://gr.xjtu.edu.cn/web/dymeng/7.

HW Yong, DY Meng, et al., TPAMI 2017

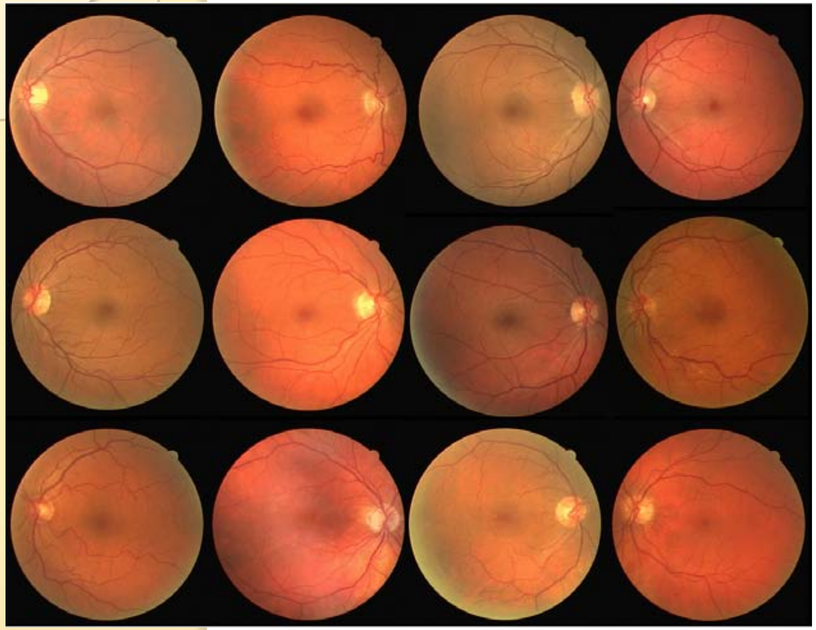


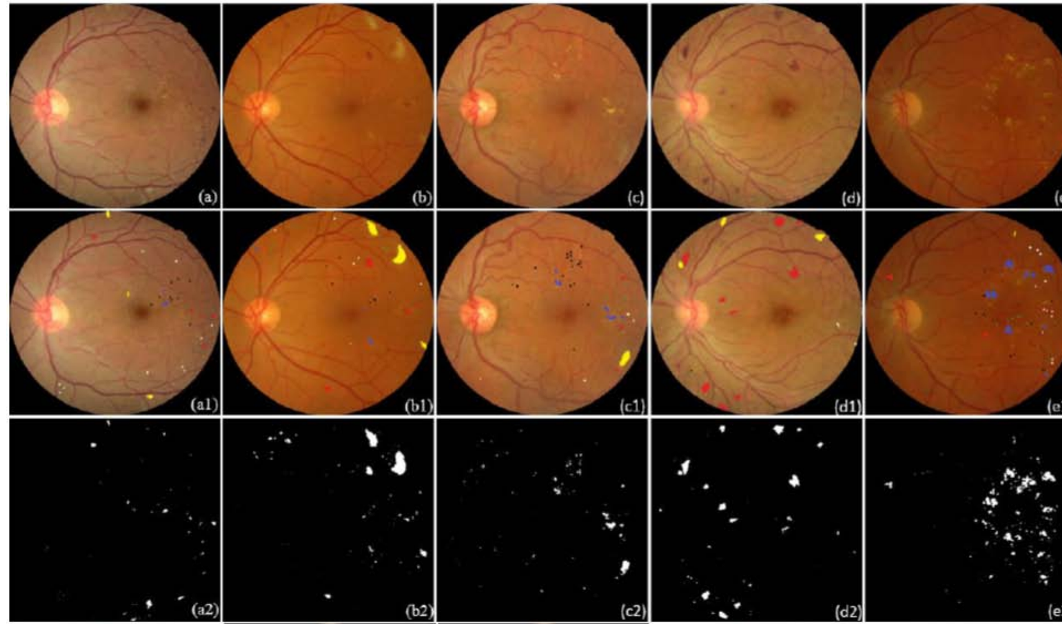






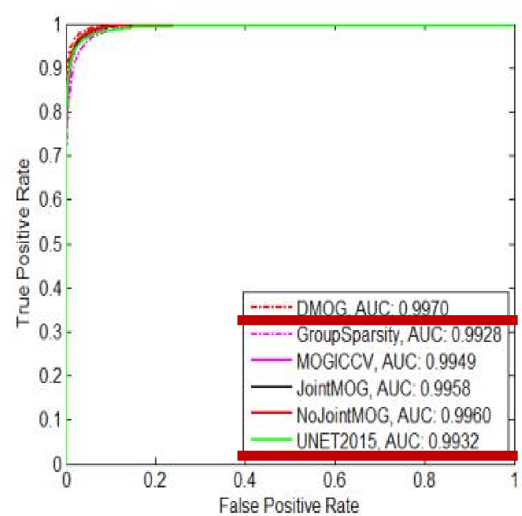
Q Xie, D Zeng, Q Zhao et al., TMI, 2017

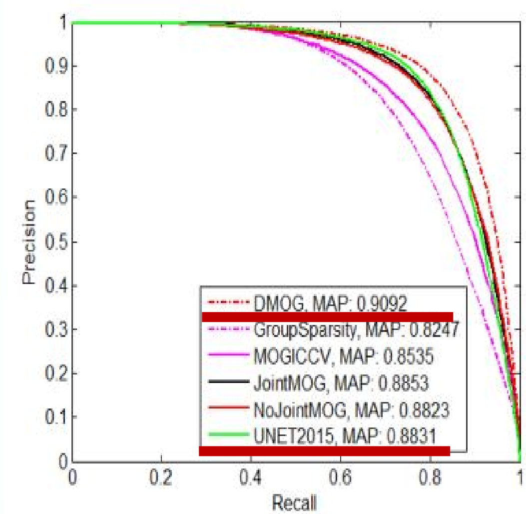




The key problem for proper modeling: **Carefully encoding noise configuration**

- ***Normal images contain a small Gaussian noise***
- ***Lesion images contain several noise components***
  - ✓ ***One similar to Gaussian in normal images***
  - ✓ ***Others represent different extents of lisions***





要求

1. L2与L1低秩矩阵分解基本问题
2. 怎样将MoG的方法论扩充至其它领域

阅读：

①MoG：Deyu Meng, Fernando De la Torre. Robust Matrix Factorization with Unknown Noise. ICCV, 2013.

SFM数据下载：
http://www.robots.ox.ac.uk/~vgg/
http://vasc.ri.cmu.edu/idb/

Yale B人脸图像下载：
http://vision.ucsd.edu/~leekc/ExtYaleDatabase/ExtYaleB.html

背景抽取数据下载：
http://perception.i2r.a-star.edu.sg/bk_model/bk_index