



# 多任务学习及其在图像分类中的应用



# 报告的主要内容

2

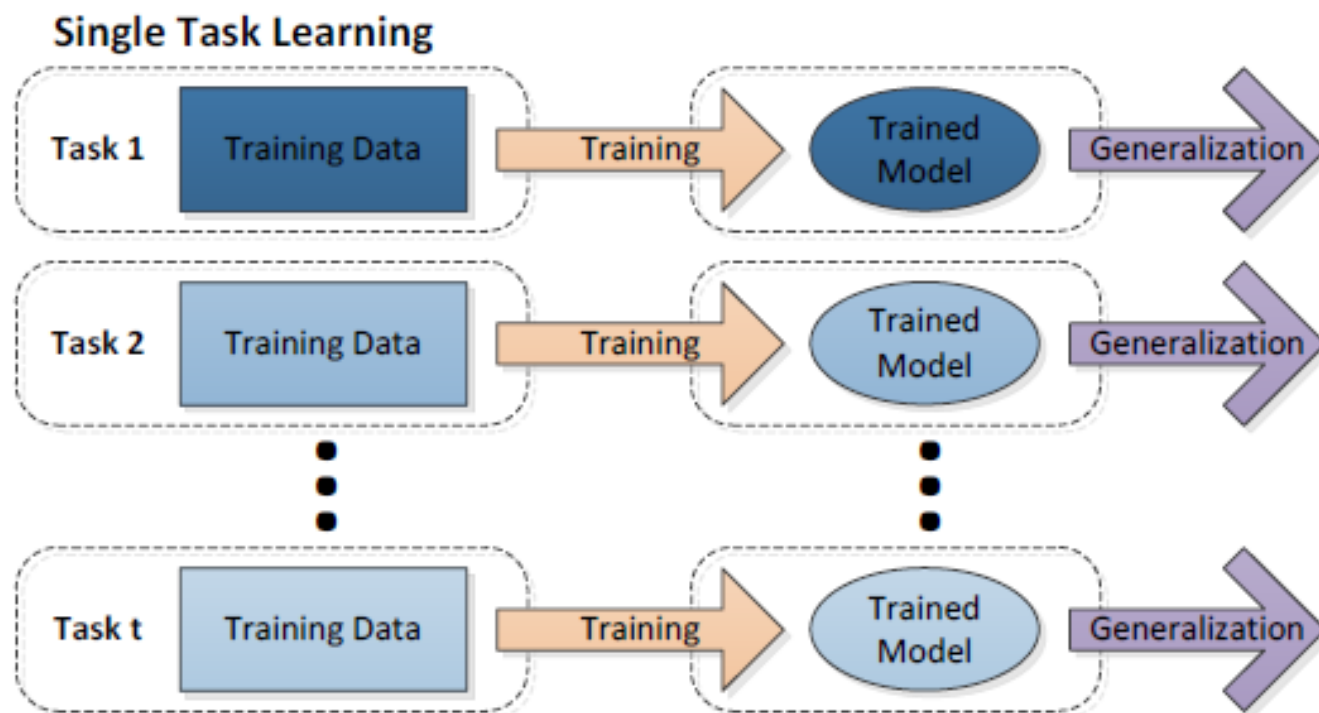
- 几个常见的多任务学习模型
- 基于多任务学习的自回归分类模型
- 未来可以挖掘的潜力



- 几个常见的多任务学习模型
- 基于多任务学习的自回归分类模型
- 未来可以挖掘的潜力

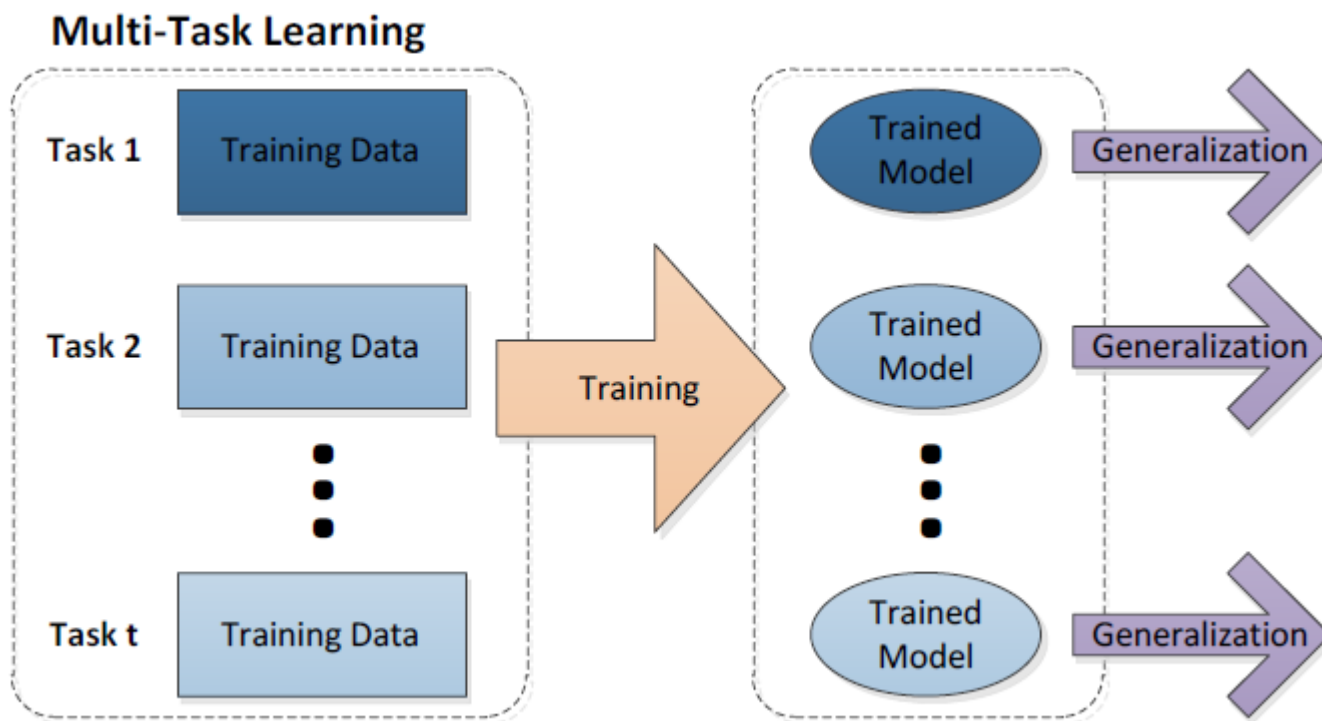


## □ 单任务学习





## □ 多任务学习





- 多任务学习的构建原则
  - ▣ 建模任务之间的相关性；
  - ▣ 同时对多个任务的模型参数进行联合学习，挖掘其中的共享信息；
  - ▣ 考虑任务之间的差异性，增强模型的适应能力；



## □ 多任务学习的两种主要方式

### ▣ 基于参数的共享

例如：神经网络隐层节点的共享

### ▣ 基于正则化约束的共享

例如：均值约束、联合特征学习等

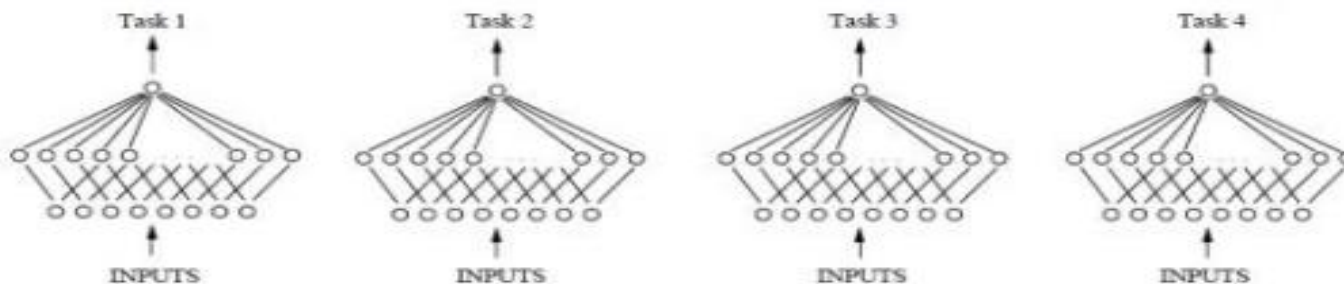


# Brief Introduction to Multi-task Learning

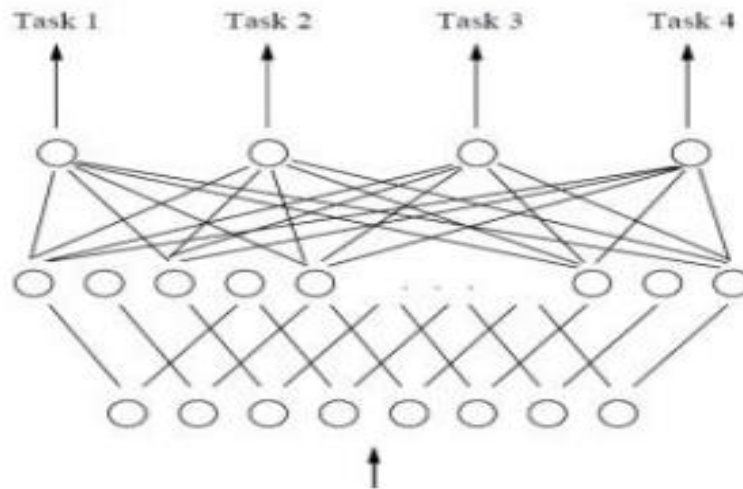
8

## □ 参数共享-神经网络节点共享

单任务学习



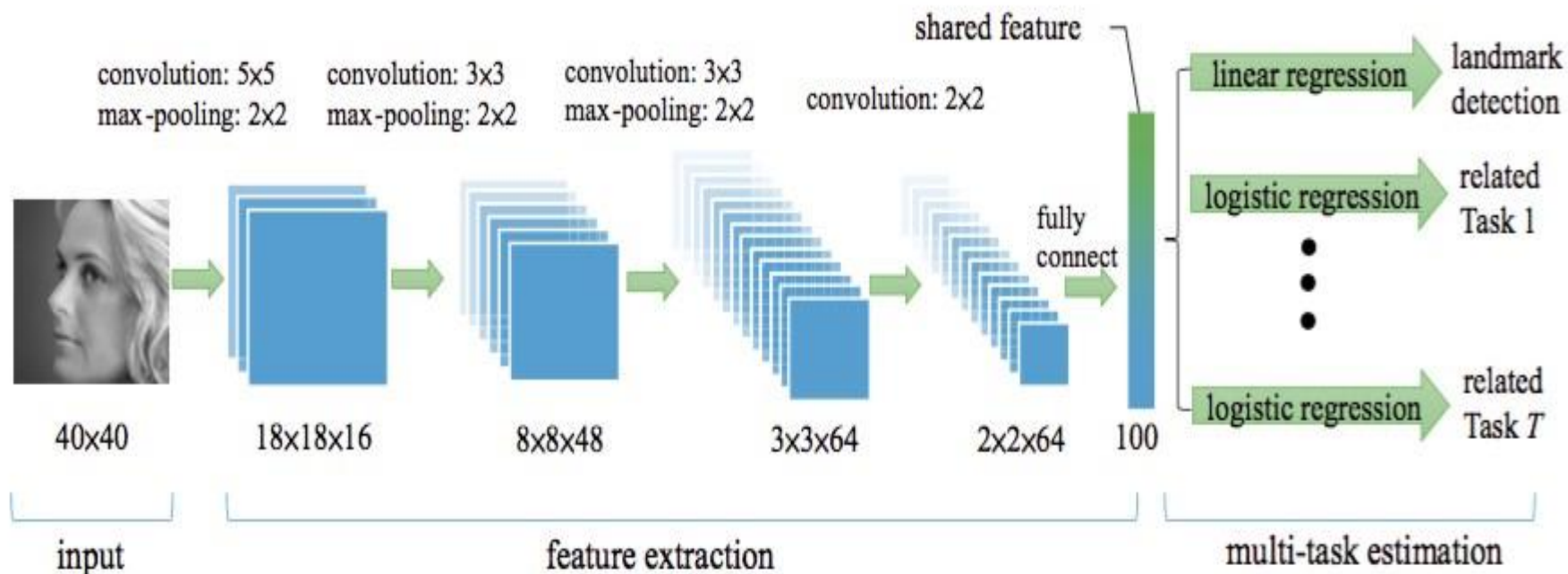
多任务学习







## □ 基于CNN的实例



[Z Zhang](#), [P Luo](#), [CL Chen](#), [X Tang](#). [Facial Landmark Detection by Deep Multi-task Learning](#), **ECCV, 2014**

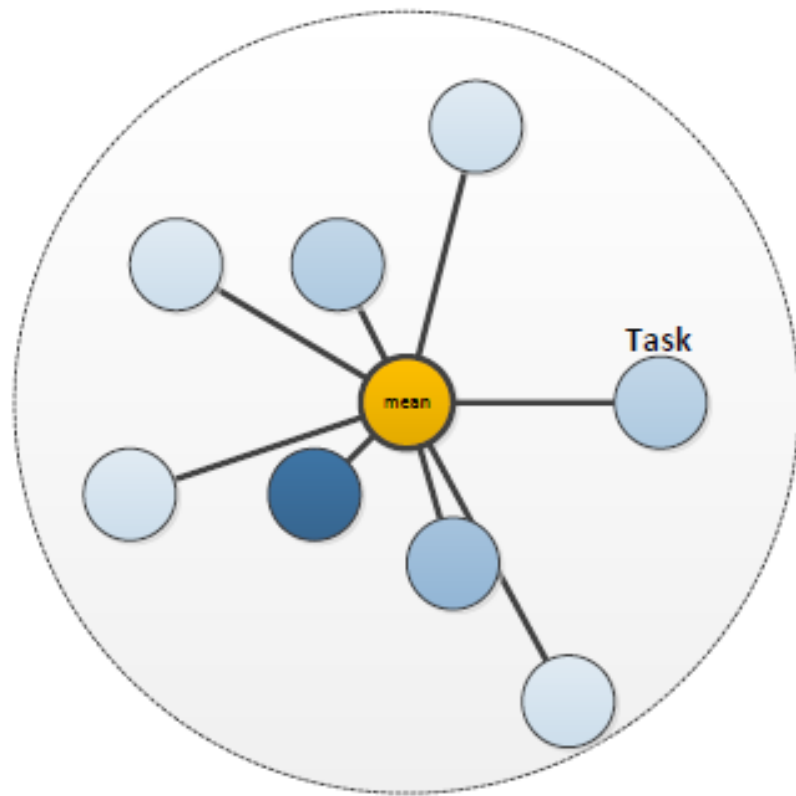
多任务学习及其在图像分类中的应用



## □ 基于正则化约束的共享

### ▣ 均值约束共享

$$\min_W \frac{1}{2} \|XW - Y\|_F^2 + \lambda \sum_{i=1}^m \left\| W_i - \frac{1}{m} \sum_{s=1}^m W_s \right\|_2^2$$



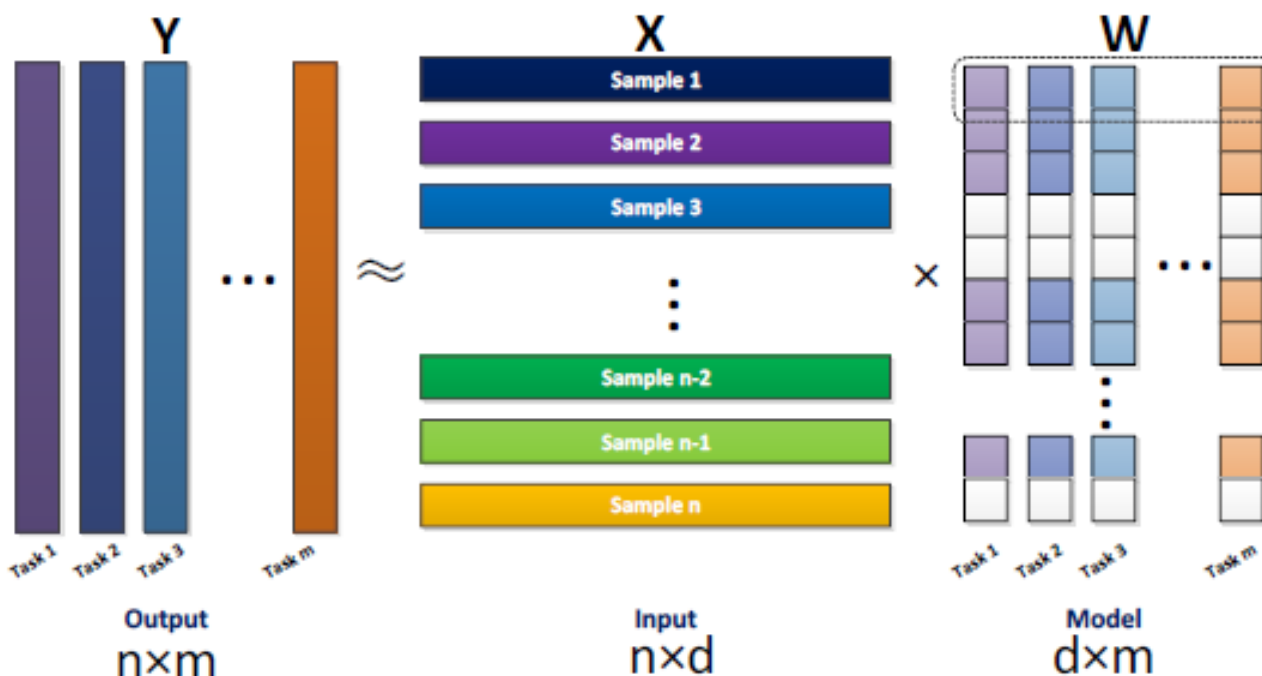


# Brief Introduction to Multi-task Learning

11

## 联合特征约束多任务学习

$$\min_W \frac{1}{2} \|XW - Y\|_F^2 + \lambda \|W\|_{1,q}$$



多任务学习及其在图像分类中的应用

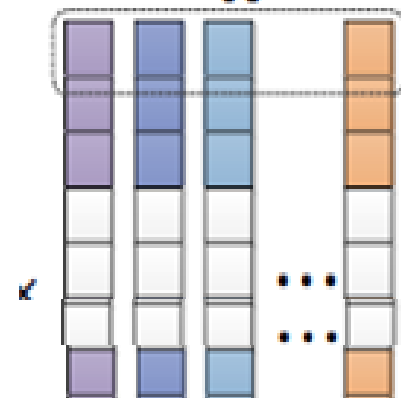


# Brief Introduction to Multi-task Learning

上次成绩	上次排名	睡觉时间	学号	姓名	身高	父母教育背景
------	------	------	----	----	----	--------

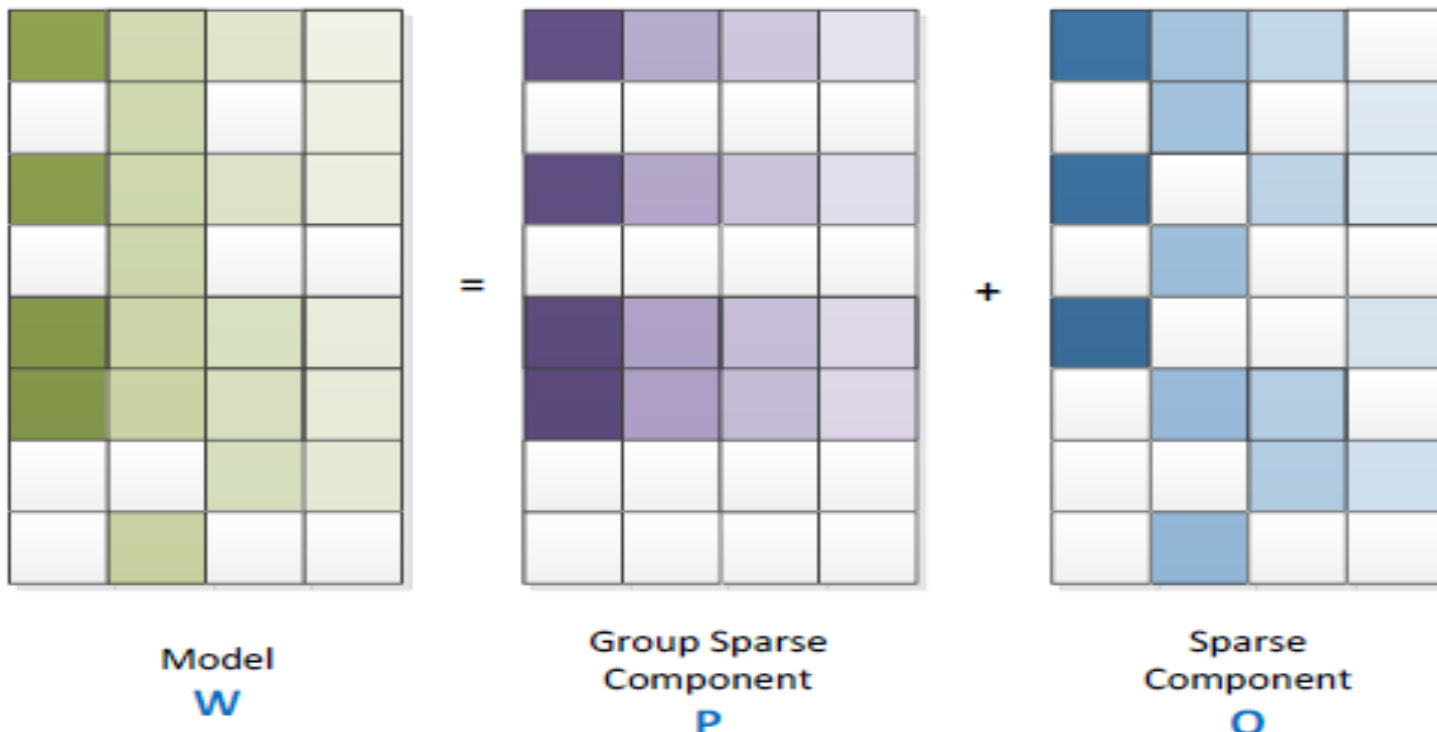
学校 1	学校 2	学校 3	学校 5
------	------	------	------

$$Y = \begin{bmatrix} 87 & 12 & 22.8 & 76 & 张三 & 1.72 & 博士 \\ 93 & 5 & 20.5 & 105 & 李四 & 1.65 & 本科 \\ 80 & 35 & 21.2 & 35 & 王五 & 1.60 & 硕士 \\ 52 & 102 & 19 & 57 & 赵六 & 1.83 & 初中 \\ 65 & 78 & 20 & 82 & 宋七 & 1.78 & 高中 \end{bmatrix}$$





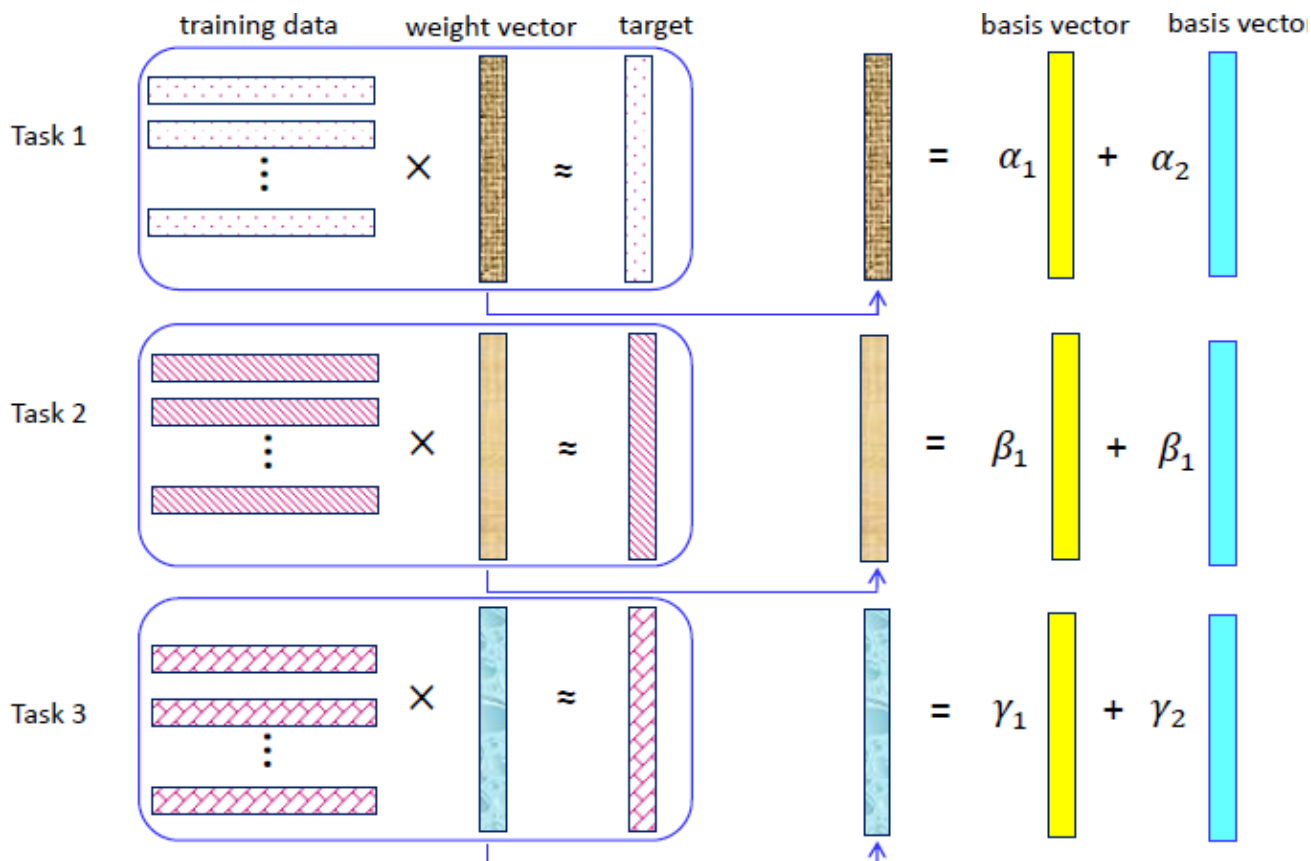
## Dirty Model 多任务学习



$$\min_{P, Q} \|Y - X(P + Q)\|_F^2 + \lambda_1 \|P\|_{1, q} + \lambda_2 \|Q\|_1$$



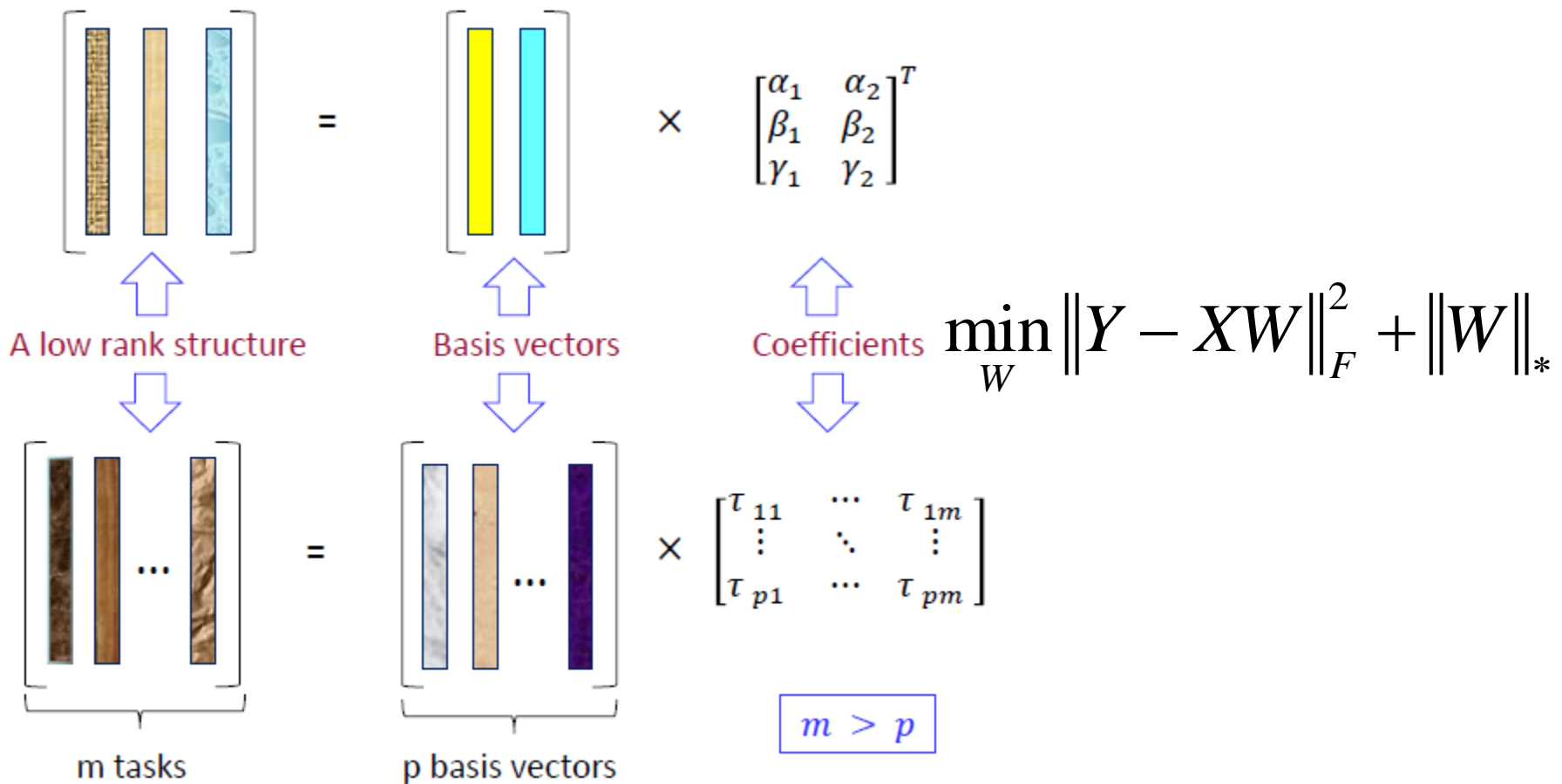
## 低秩约束的多任务学习





# Brief Introduction to Multi-task Learning

15



多任务学习及其在图像分类中的应用



## □ 存在的问题

- 认为任务之间的相关性仅与模型参数有关；
- 忽略了样本原始特征的差异性；





- 交替结构优化(Alternating Structure Optimization, ASO)的多任务学习

$$\mathbf{u}_i = \mathbf{w}_i + \boldsymbol{\theta} \mathbf{v}_i$$

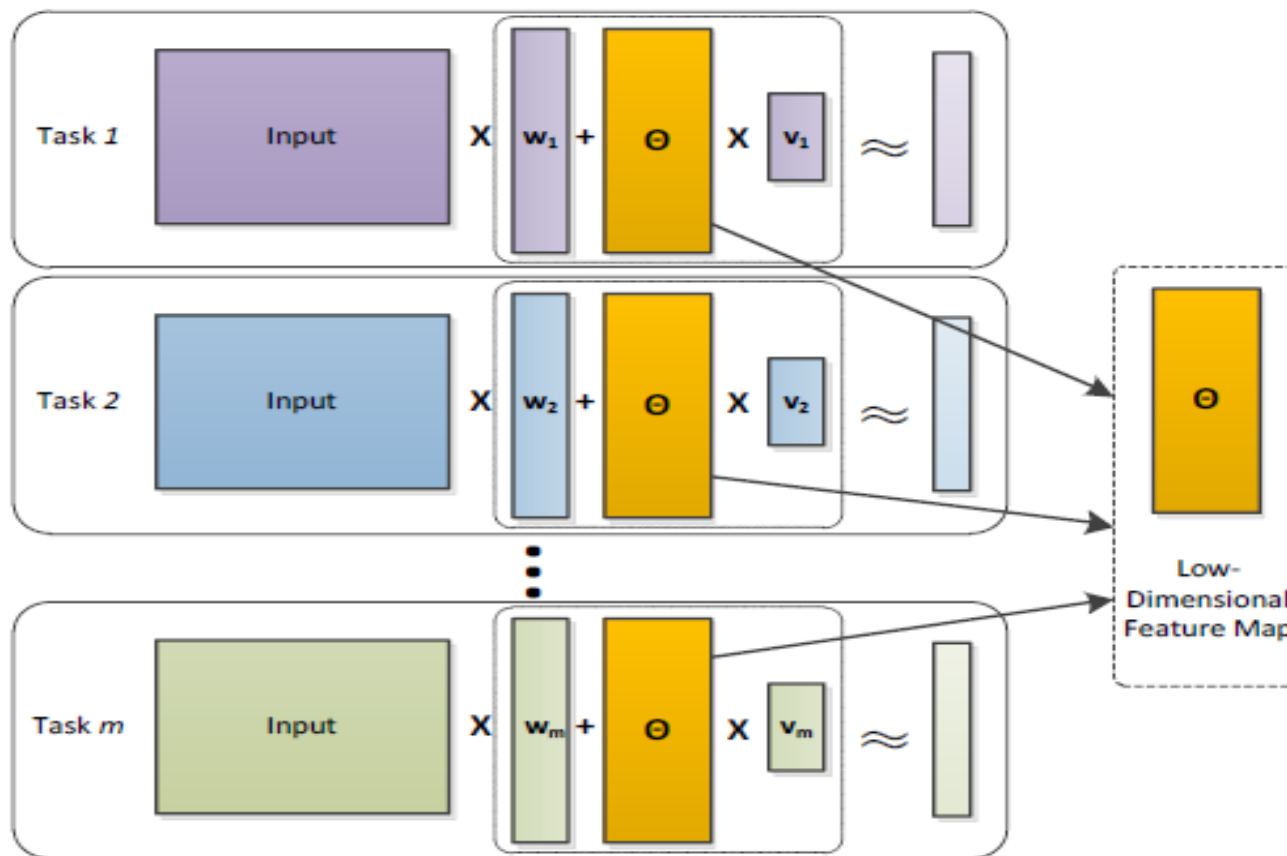
$$\mathbf{u}_i^T x = (\mathbf{w}_i^T + \mathbf{v}_i^T \boldsymbol{\theta}^T) x$$

$$\mathbf{U} = \mathbf{W} + \boldsymbol{\theta} \mathbf{V}$$



# Brief Introduction to Multi-task Learning

18



基于多任务学习的图像分类



# Brief Introduction to Multi-task Learning

19

- Loss function:

$$\mathcal{L}_i(X_i(\Theta v_i + w_i), y_i) = \|X_i(\Theta v_i + w_i) - y_i\|^2$$

- ASO:

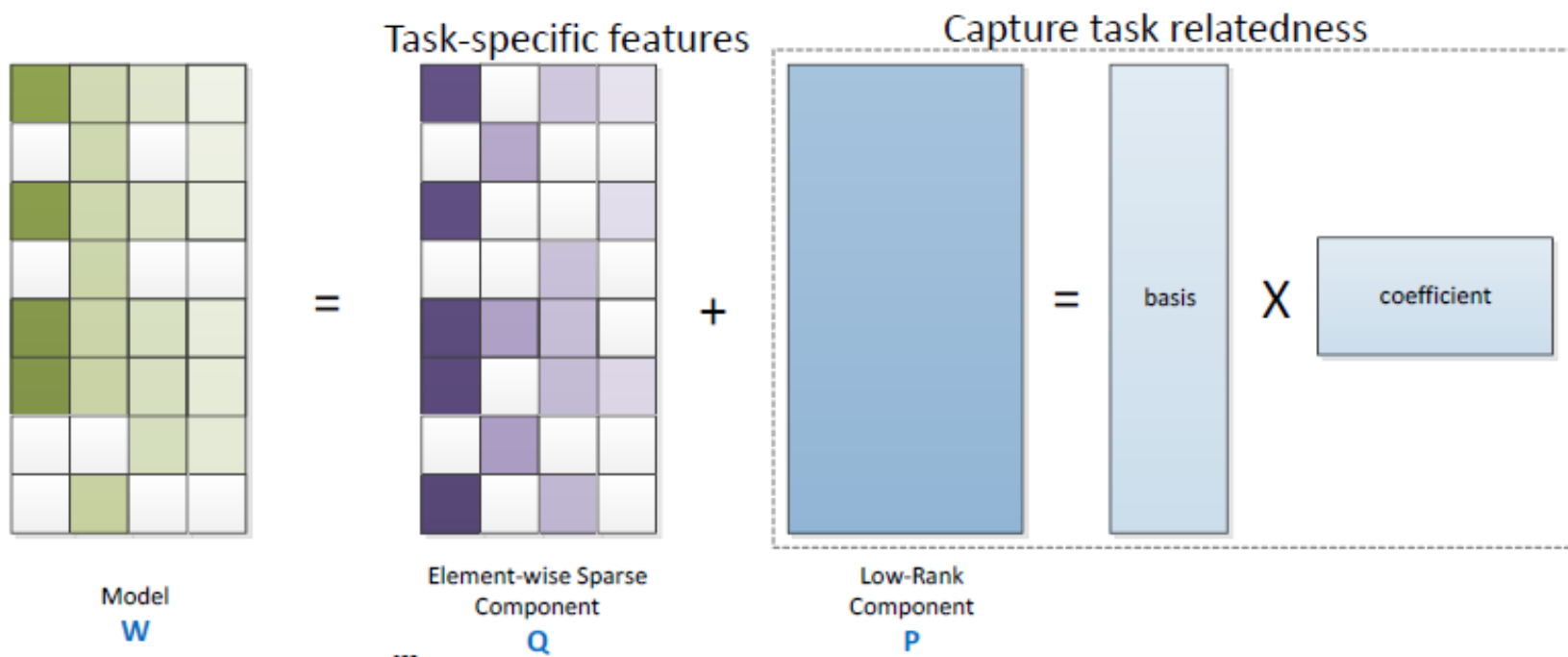
$$\begin{aligned} & \min_{\Theta, \{v_i, w_i\}} \sum_{i=1}^m \{\mathcal{L}_i(X_i(\Theta v_i + w_i), y_i) + \alpha \|w_i\|^2\} \\ & \text{subject to} \quad \Theta^T \Theta = I \end{aligned}$$

- iASO

$$\begin{aligned} & \min_{\Theta, \{v_i, w_i\}} \sum_{i=1}^m \{\mathcal{L}_i(X_i(\Theta v_i + w_i), y_i) + \alpha \|\Theta v_i + w_i\|^2 + \beta \|w_i\|^2\} \\ & \text{subject to} \quad \Theta^T \Theta = I \end{aligned}$$



## □ Incoherent Low Rank and Sparse Structure



$$\min_{P, Q} \sum_{i=1}^m \mathcal{L}_i(X_i(P_i + Q_i), y_i) + \lambda \|Q\|_1$$

subject to  $\|P\|_* \leq \eta$



- 多任务学习的构建原则
  - ▣ 如何建模任务之间的相关性；
  - ▣ 如何提取任务间的共享信息进行联合学习；
  - ▣ 如何体现任务之间的差异性；



- 几个常见的多任务学习模型
- 基于多任务学习的自回归分类模型
- 未来可以挖掘的潜力



传统多任务学习框架特点：

$$\min_{P, Q} \sum_{i=1}^m \mathcal{L}_i(X_i(P_i + Q_i), y_i) + \lambda \|Q\|_1$$

subject to  $\|P\|_* \leq \eta$

1. 能够用于解决多种机器学习问题：

(1) 股票价格预测 (2) 学生成绩预测 ...

2. 实际上是一个特征提取的过程，还需训练额外的分类器；



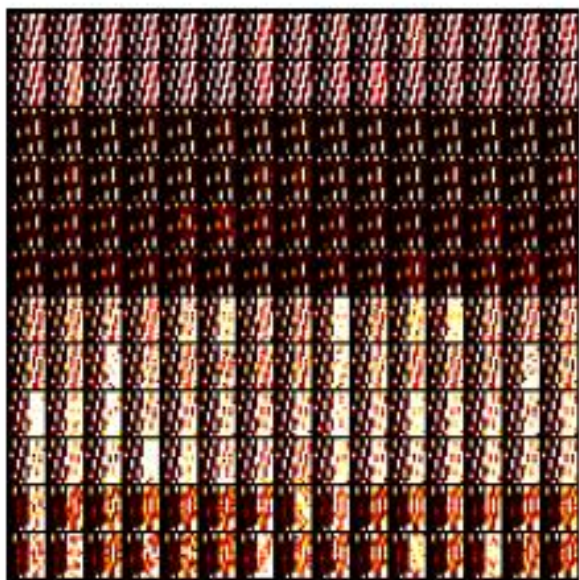
- 我们的工作思路：
  - 针对图像的结构，寻找任务之间的相关性来设计一种有效的多任务学习框架
  - 一次性的进行特征提取和分类器训练，将两个独立的过程统一到一个框架中；
  - 如兼顾的考虑共享特征与差异性，并给出适当的数值求解算法



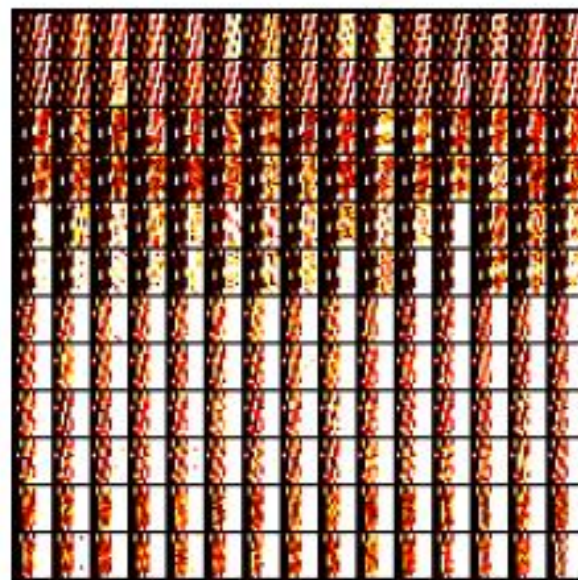


# Our work

25



(a) Gaussian Channel



(b) Rayleigh Channel

<https://cyclostationary.blog>



□ 提出的损失函数:

$$L(x_n^i, W_S, W_i) = \|A(x_n) - (W_S + W_i) \Theta(P_i x_n)\|_2^2$$

$$\text{令: } X = [X_1 \quad X_2 \quad \cdots \quad X_C], \quad X_i \in R^{d \times n_i}$$

$$L(X, W_S, W_i) = \sum_i \sum_{n=1}^{n_i} L(x_n^i, W_S, W_i) = \sum_{i=1,2,\dots,C} \|A(X_i) - (W_S + W_i) \Theta(P_i X_i)\|_F^2$$



- The proposed objective function

$$\min_{\{W_i, P_i\}_{i=1}^C, W_S} \sum_{i=1}^C \left( \| \mathcal{A}(X_i) - (W_S + W_i) \Theta(P_i X_i) \|_F^2 \right. \\ \left. + \varphi(P_i) + \lambda \|W_i\|_1 \right) + \lambda_S \|W_S\|_F^2$$

$$\varphi(P_i) = \gamma \|P_i\|_F^2 + \theta \|P_i \bar{X}_i\|_F^2, \quad \Theta_1(t) = \frac{t}{1 + e^{-t}}$$

$$\Theta_2(t) = \text{sign}(t) \text{Shrink}_\tau(t)$$



## □ Numerical Solution

$$\begin{aligned} \min_{\{W_i, P_i, \Phi_i\}, W_S} & \sum_{i=1}^C \left( \| \mathcal{A}(X_i) - (W_S + W_i) \Phi_i \|_F^2 \right. \\ & \left. + \varphi(P_i) + \lambda \|W_i\|_1 \right) + \lambda_S \|W_S\|_F^2 \\ \text{s.t.} & \quad \Phi_i = \Theta(P_i X_i), i = 1, 2, \dots, C \end{aligned}$$



□ Solve the following minimization problems iteratively:

$$\textcircled{1} \quad \min_{W_S} \sum_{i=1}^C \|\mathcal{A}(X_i) - (W_S + W_i) \Phi_i\|_F^2 + \lambda_S \|W_S\|_F^2$$

$$\textcircled{2} \quad \min_{W_i} \|\mathcal{A}(X_i) - (W_S + W_i) \Phi_i\|_F^2 + \lambda_i \|W_i\|_1$$

$$\textcircled{3} \quad \min_{P_i, \Phi_i} \|\mathcal{A}(X_i) - (W_S + W_i) \Phi_i\|_F^2 + \varphi(P_i)$$

*s.t.*  $\Phi_i = \Theta(P_i X_i)$



## □ Minimization - 1

$$\min_{W_S} \sum_{i=1}^C \|\mathcal{A}(X_i) - (W_S + W_i) \Phi_i\|_F^2 + \lambda_S \|W_S\|_F^2$$

$$\frac{\partial J_{W_S}}{\partial W_S} = 0 \quad \Rightarrow$$

$$W_S^* = \left( \sum_{i=1}^C (\mathcal{A}(X_i) - W_i \Phi_i) \Phi_i^T \right) \left( \sum_{i=1}^C \Phi_i \Phi_i^T + \lambda_S I \right)^{-1}$$



## □ Minimization - 2

$$\min_{W_i} \|\mathcal{A}(X_i) - (W_S + W_i) \Phi_i\|_F^2 + \lambda_i \|W_i\|_1$$

Suppose:

$$f(W_i) = \|A(X_i) - (W_s + W_i) \Phi_i\|_F^2$$

$$\hat{W}_i^{(k+1)} = \arg \min \left\| W_i - \left( W_i^{(k)} - \nabla f(W_i^{(k)}) \delta_i \right) \right\|_F^2 + \lambda \|W_i\|_1$$



## □ Minimization - 3

$$\begin{aligned} \min_{P_i, \Phi_i} & \|\mathcal{A}(X_i) - (W_S + W_i) \Phi_i\|_F^2 + \varphi(P_i) \\ \text{s.t.} & \Phi_i = \Theta(P_i X_i) \end{aligned}$$

$$\begin{aligned} \min_{P_i, \Phi_i} & \|\mathcal{A}(X_i) - (W_S + W_i) \Phi_i\|_F^2 + \varphi(P_i) \\ \text{s.t.} & \Phi_i = \Theta(Q_i), Q_i = P_i X_i \end{aligned}$$





- The augmented Lagrangian function:

$$\begin{aligned} \min_{P_i, \Phi_i, Q_i, U_1, U_2} & \|\mathcal{A}(X_i) - (W_S + W_i) \Phi_i\|_F^2 + \varphi(P_i) \\ & + \frac{\rho}{2} \|\Phi_i - \Theta(Q_i)\|_F^2 + \frac{\rho}{2} \|Q_i - P_i X_i\|_F^2 \\ & + \langle U_1, \Phi_i - \Theta(Q_i) \rangle + \langle U_2, Q_i - P_i X_i \rangle \end{aligned}$$



- With the ADMM, the minimization can be solved by the following three subproblems:

$$\begin{aligned}\Phi_i^{(t+1)} = \arg \min_{\Phi_i} & \|\mathcal{A}(X_i) - (W_S + W_i) \Phi_i\|_F^2 \\ & + \frac{\rho}{2} \|\Phi_i - \Theta(Q_i^{(t)}) + \frac{1}{\rho} U_1^{(t)}\|_F^2\end{aligned}$$

$$\begin{aligned}P_i^{(t+1)} = \arg \min_{P_i} & \frac{\rho}{2} \|Q_i^{(t)} - P_i X_i + \frac{1}{\rho} U_2^{(t)}\|_F^2 \\ & + \gamma \|P_i\|_F^2 + \theta \|P_i \bar{X}_i\|_F^2\end{aligned}$$

$$\begin{aligned}Q_i^{(t+1)} = \arg \min_{Q_i} & \frac{\rho}{2} \|\Phi_i - \Theta(Q_i) + \frac{1}{\rho} U_1^{(t)}\|_F^2 \\ & + \frac{\rho}{2} \|Q_i - P_i X_i + \frac{1}{\rho} U_2^{(t)}\|_F^2\end{aligned}$$



## □ Overall Algorithm

---

**Input:** Training data set  $\{X_i\}_{i=1}^C$ , self-projected operator  $\mathcal{A}$  regularization parameters  $\gamma, \theta, \lambda, \lambda_S, \rho$  initial random  $W_S^{(0)}, \{W_i^{(0)}, P_i^{(0)}, \Phi_i^{(0)}\}_{i=1}^C, U_1^{(0)} = U_2^{(0)} = \mathbf{0}$ , iteration number  $T$ , initial  $k = 0$ .

**Repeat**

1. Calculate matrix  $W_S^{(k+1)}$  with Eq.(10);
2. Calculate matrix  $W_i^{(k+1)}$  for each task with Eq.(12);
3. **for**  $t = 0$  to  $T - 1$  **do**, ( $i = 1, 2, \dots, C$ )

Update  $\Phi_i^{(t+1)}$  in subproblem in Eq.(16);

Update  $P_i^{(t+1)}$  in subproblem in Eq.(17);

Update  $Q_i^{(t+1)}$  in subproblem in Eq.(18);

Update Lagrangian multipliers:

$$U_1^{(t+1)} = U_1^{(t)} + \rho \left( \Phi_i^{(t+1)} - \Theta \left( Q_i^{(t+1)} \right) \right)$$

$$U_2^{(t+1)} = U_2^{(t)} + \rho \left( Q_i^{(t+1)} - P_i^{(t+1)} X_i \right)$$

**end for**

4. For each  $i$ ,  $\Phi_i^{(k+1)} \leftarrow \Phi_i^T, P_i^{(k+1)} \leftarrow P_i^T,$

$$Q_i^{(k+1)} \leftarrow Q_i^T$$

5.  $k \leftarrow k + 1;$

**Until convergence**

---

**Output:**  $W_S^*, \{W_i^*, P_i^*\}_{i=1}^C$

---



## □ Classification Criterion

$$\hat{l}_t = \arg \min_i \|\mathcal{A}(x_t) - (W_S + W_i) \Theta(P_i x_t)\|_F^2$$

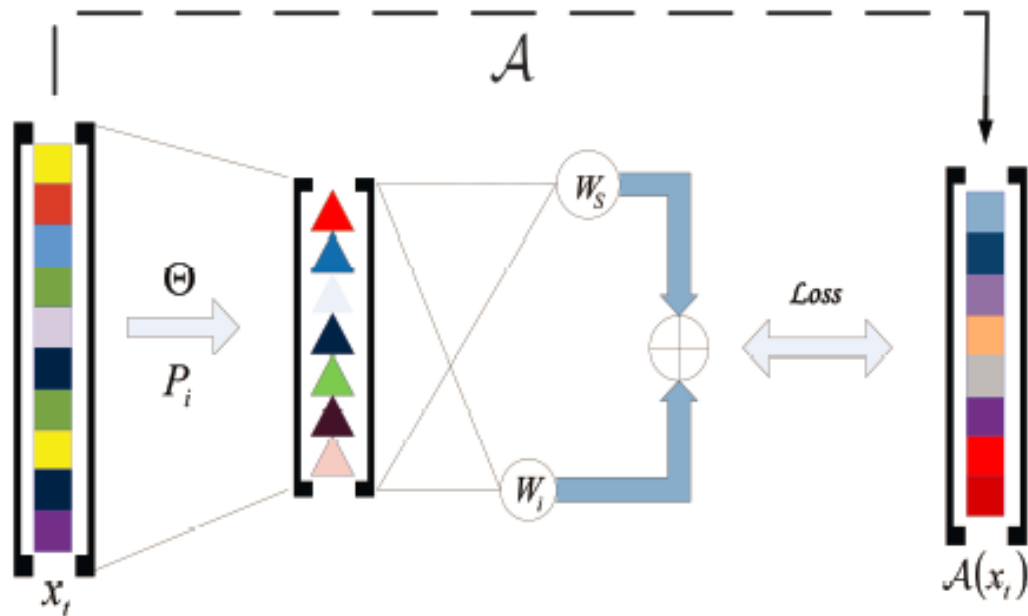
$$\begin{aligned} p(x_n | y_n, \{W_i\}_{i=1}^C, W_S) &= \sum_i y_{ni} \exp\left\{-\|A(x_n) - (W_S + W_i)x_n\|_2^2\right\} \\ -\log P(X | Y, \{W_i\}_{i=1}^C, W_S) &= -\log \prod_n p(x_n | y_n) \\ &= \sum_i \sum_n y_{ni} \|A(x_n) - (W_S + W_i)S(P_i x_n)\|_2^2 \end{aligned}$$



# Our work

37

## □ Structure for each task

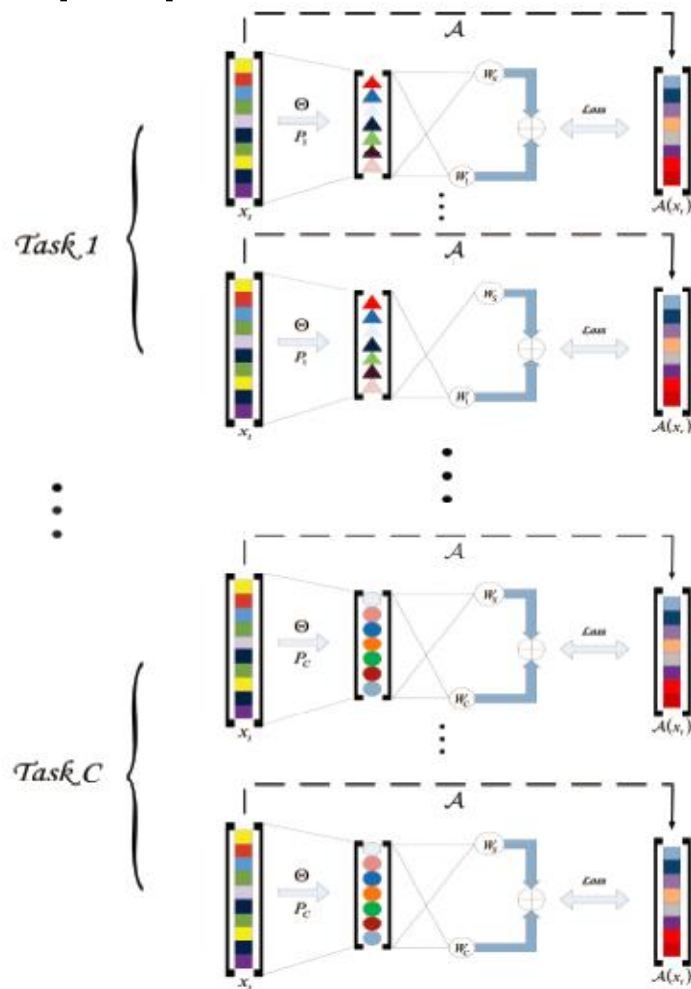




# Our work

38

## □ Structure of our proposed model



基于多任务学习的图像分类



## □ Experimental results

	Extended YaleB	WSU-HOS-G	WSU-HOS-R
SRC	95.42±0.47	92.87±1.56	79.85±2.79
CRC	97.23±0.35	96.42±0.66	83.62±0.93
ProCRC	93.25±1.28	95.87±1.93	88.89±0.62
NNGR	92.32±1.43	95.47±1.40	82.86±1.33
Ours with $\Theta_1$	96.98±0.28	98.63±1.29	<b>90.63±1.23</b>
Ours with $\Theta_2$	<b>97.08±0.48</b>	<b>98.79±0.23</b>	89.80±0.88



## □ Scene Classification

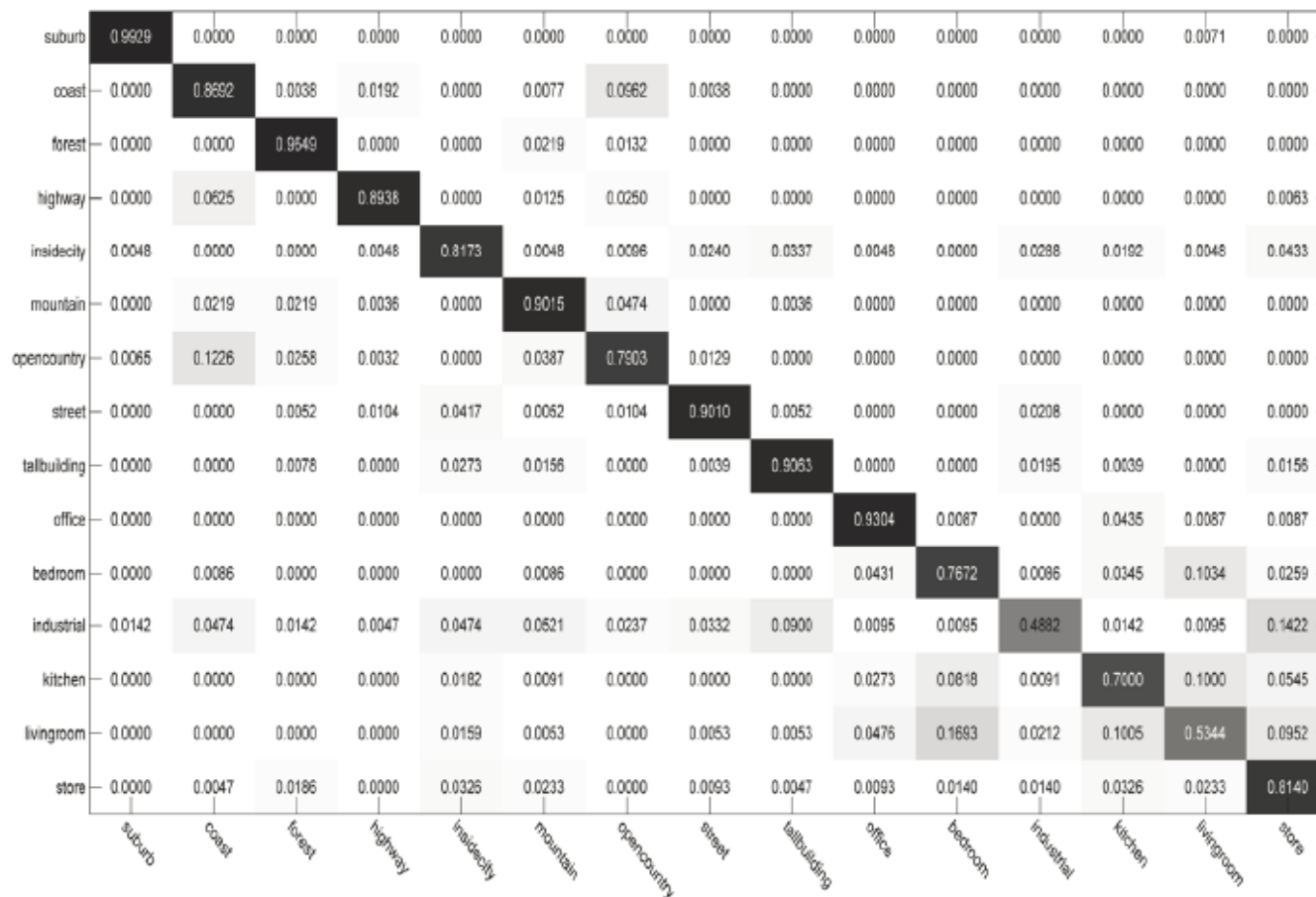
	Scene-15
SPM	79.95±0.27
LLC	79.81±0.35
LLCDC	80.30±0.62
LLCDCSIFT	<b>82.40±0.35</b>
Ours with $\Theta_1$	82.23±0.49
Ours with $\Theta_2$	81.51±0.54





# Our work

41



多任务学习及其在图像分类中的应用



# Our work

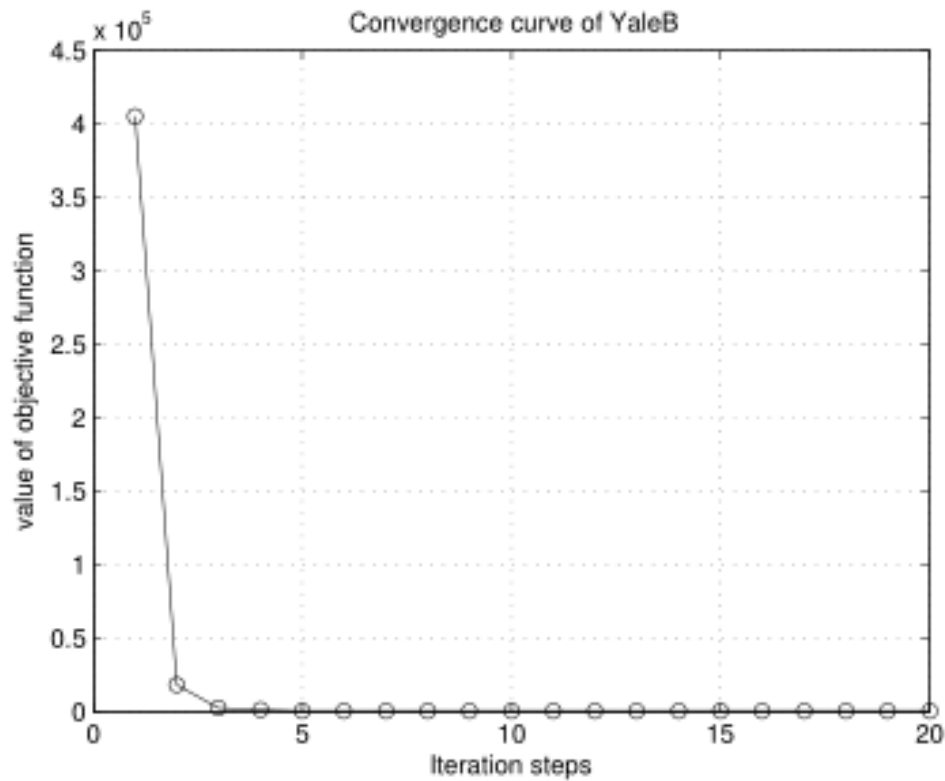
42

## □ Visualization Result





## □ Convergence Analysis





- 几个常见的多任务学习模型
- 基于多任务学习的自回归分类模型
- 未来可以挖掘的潜力



□ 未来可以尝试的工作:

(1) 降维算子学习;

(2) 非线性映射方式: 使用核技术

(3) 考虑引入属性特征应用于其它视觉任务;



*Thank you for your  
attentions!*