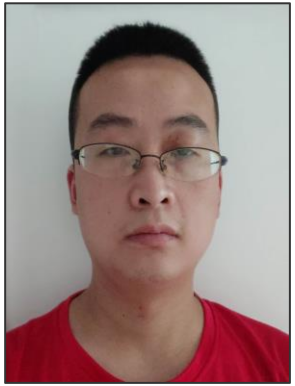
Loss Function Search for Face Recognition

Xiaobo Wang
JD AI Research

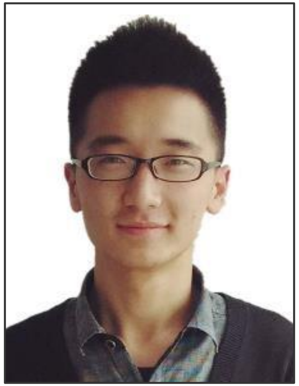
Authors

<u>Xiaobo Wang</u>



Shuo Wang

Cheng Chi

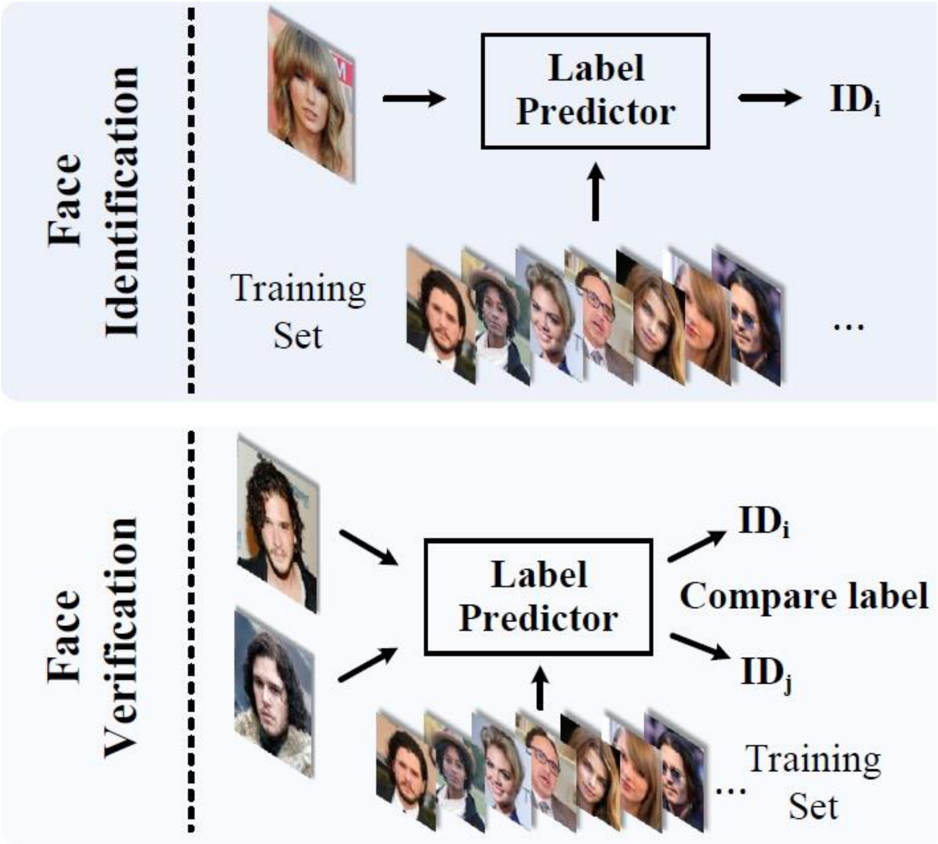


Shifeng Zhang

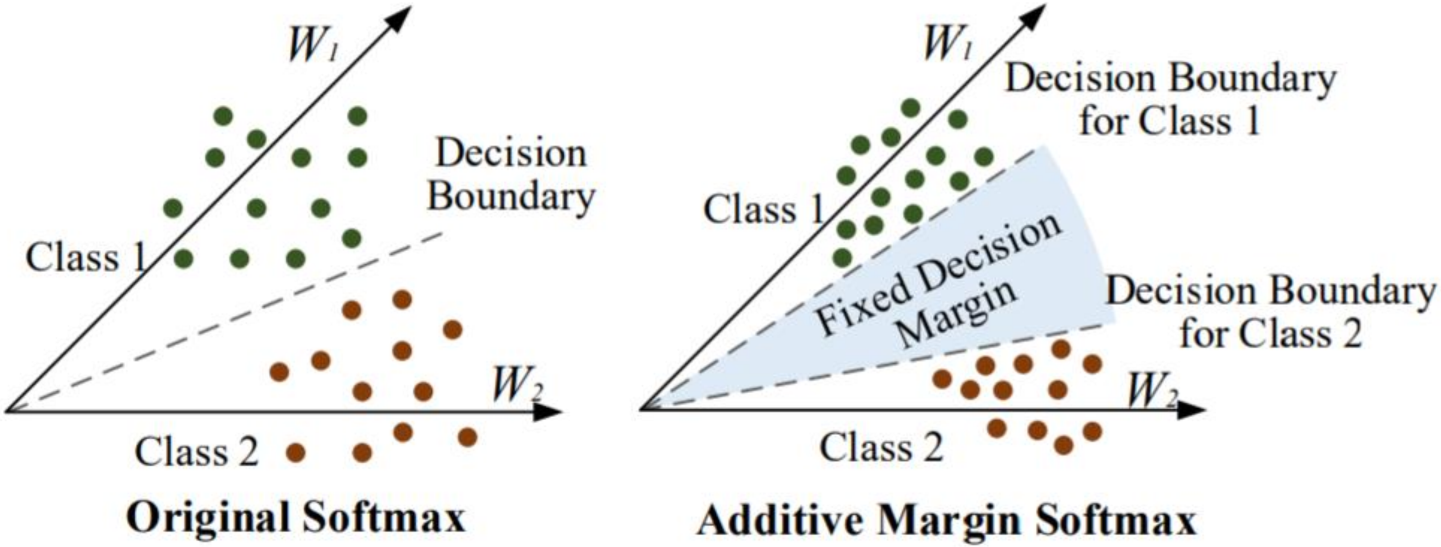
Tao Mei

1. Motivation

- **1:N matching**
- **1:1 matching**



1. Motivation

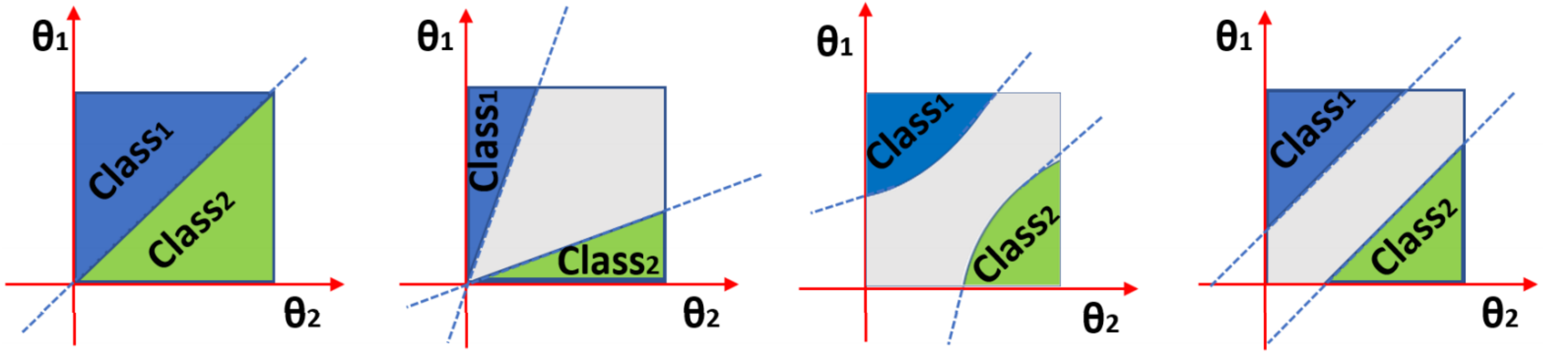


Discriminative

1. Motivation

Hand-Crafted

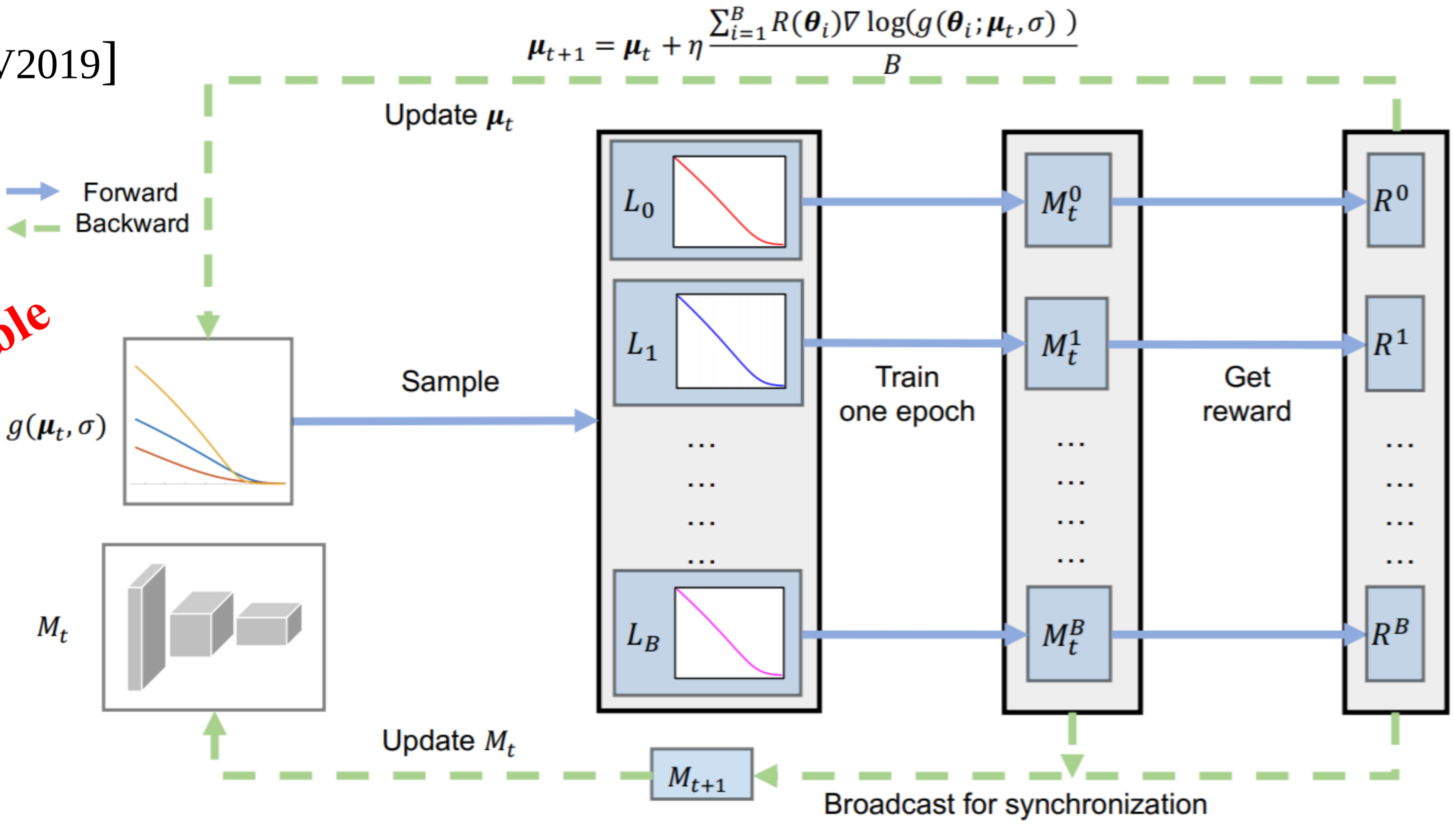
Loss Functions	Decision Boundaries
Softmax	$(W_1 - W_2)x + \ b_1 - b_2 = 0$
SpherFace [CVPR2017]	$\|x\|(\cos m\theta_1 - \cos\theta_2) = 0$
CosFace [CVPR2018]	$s(\cos\theta_1 - m - \cos\theta_2) = 0$
ArcFace [CVPR2019]	$s(\cos(\theta_1 + m) - \cos\theta_2) = 0$



Softmax SphereFace CosFace ArcFace

1. Motivation

AM-LFS [ICCV2019]



Complex & Unstable

1. Motivation

Hand-Crafted:

- Softmax
- SphereFace [CVPR2017]
- CosFace [CVPR2018]
- ArcFace [CVPR2019]

- **Require great effort to explore the large design space**

Search: AM-LFS [ICCV2019]

- **Search space is complex and unstable**

2. Formulation

- **Softmax Loss:**

$$\mathcal{L}_1 = -\log \frac{e^{\boldsymbol{w}_y^T \boldsymbol{x}}}{e^{\boldsymbol{w}_y^T \boldsymbol{x}} + \sum_{k \neq y}^{K} e^{\boldsymbol{w}_k^T \boldsymbol{x}}}, \qquad (1)$$

$$\mathcal{L}_2 = -\log \frac{e^{s\cos(\theta_{\boldsymbol{w}_y,\boldsymbol{x}})}}{e^{s\cos(\theta_{\boldsymbol{w}_y,\boldsymbol{x}})} + \sum_{k \neq y}^{K} e^{s\cos(\theta_{\boldsymbol{w}_k,\boldsymbol{x}})}}, \qquad (2)$$

- **Margin-based Softmax Loss:**

$$f(m, \theta_{\boldsymbol{w}_y,\boldsymbol{x}}) \leq \cos(\theta_{\boldsymbol{w}_y,\boldsymbol{x}})$$

$$\mathcal{L}_3 = -\log \frac{e^{sf(m,\theta_{\boldsymbol{w}_y,\boldsymbol{x}})}}{e^{sf(m,\theta_{\boldsymbol{w}_y,\boldsymbol{x}})} + \sum_{k \neq y}^{K} e^{s\cos(\theta_{\boldsymbol{w}_k,\boldsymbol{x}})}}, \qquad (3)$$

2. Formulation

- **Softmax probability**:

$$p = \frac{e^{s\cos(\theta_{\boldsymbol{w}_y,\boldsymbol{x}})}}{e^{s\cos(\theta_{\boldsymbol{w}_y,\boldsymbol{x}})} + \sum_{k\neq y}^{K} e^{s\cos(\theta_{\boldsymbol{w}_k,\boldsymbol{x}})}}$$

- **Margin-based Softmax probability**:

$$p_m = \frac{e^{sf(m,\theta_{\boldsymbol{w}_y,\boldsymbol{x}})}}{e^{sf(m,\theta_{\boldsymbol{w}_y,\boldsymbol{x}})} + \sum_{k\neq y}^{K} e^{s\cos(\theta_{\boldsymbol{w}_k,\boldsymbol{x}})}}$$

$$\boxed{\begin{aligned} p_m &= \frac{e^{sf(m,\theta_{\boldsymbol{w}_y,\boldsymbol{x}})}}{e^{sf(m,\theta_{\boldsymbol{w}_y,\boldsymbol{x}})} + \sum_{k\neq y}^{K} e^{s\cos(\theta_{\boldsymbol{w}_k,\boldsymbol{x}})}} \\ &= \frac{e^{sf(m,\theta_{\boldsymbol{w}_y,\boldsymbol{x}})}}{e^{sf(m,\theta_{\boldsymbol{w}_y,\boldsymbol{x}})} + e^{\cos(\theta_{\boldsymbol{w}_y,\boldsymbol{x}})}(1-p)/p} \\ &= \frac{1}{p + e^{s[\cos(\theta_{\boldsymbol{w}_y,\boldsymbol{x}}) - f(m,\theta_{\boldsymbol{w}_y,\boldsymbol{x}})]}(1-p)} * p \\ &= \frac{1}{ap + (1-a)} * p = h(a,p) * p, \end{aligned}}$$

modulating factor: $a = 1 - e^{s[\cos(\theta_{\boldsymbol{w}_y,\boldsymbol{x}}) - f(m,\theta_{\boldsymbol{w}_y,\boldsymbol{x}})]}$

$(a \leq 0)$

2. Formulation

- The success of margin-based softmax losses is **how to reduce the softmax probability *p***:

$$p_m = h(a,p) * p \qquad (9)$$

where $h(a,p) = \frac{1}{ap+(1-a)} \in (0,1]$ is a modulating function

Method	Modulating Factor a
Softmax	$a = 0$
A-Softmax	$a = 1 - e^{s[\cos(\theta_{\boldsymbol{w}_y,\boldsymbol{x}}) - \cos(m\theta_{\boldsymbol{w}_y,\boldsymbol{x}})]}$
AM-Softmax	$a = 1 - e^{sm}$
Arc-Softmax	$a = 1 - e^{s[\cos(\theta_{\boldsymbol{w}_y,\boldsymbol{x}}) - \cos(\theta_{\boldsymbol{w}_y,\boldsymbol{x}} + m)]}$

- **AM-LFS**:

$$\mathcal{L}_4 = -\log\left(a_i \frac{e^{s\cos(\theta_{\boldsymbol{w}_y,\boldsymbol{x}})}}{e^{s\cos(\theta_{\boldsymbol{w}_y,\boldsymbol{x}})} + \sum_{k\neq y}^{K} e^{s\cos(\theta_{\boldsymbol{w}_k,\boldsymbol{x}})}} + b_i\right), \quad (4)$$

where a_i and b_i are the parameters of search space. $i \in$

2. Formulation

$$p_m = h(a,p) * p \qquad\qquad p_m = a_i p + b_i$$

- Our search space p_m=h(a,p)*p is **always less than the softmax probability p** while the piece-wise linear functions p_m=a_i*p+b_i are not. The discriminability of AM-LFS is not guaranteed;

- There is **only one parameter a to be searched** in our formulation while the AM-LFS needs search 2M parameters. The search space of AM-LFS is complex and unstable;

- Our method has a reasonable range of the parameter (a<= 0) hence **facilitating the searching procedure**, while the parameters of AM-LFS a_i and b_i are without any constraints.

2. Formulation

$$\mathcal{L}_5 = -\log\left(h(a,p) * p\right), \tag{10}$$

where the modulating function $h(a,p)$ has a bounded range $(0,1]$ and the modulating factor is $a \leq 0$. To validate our

- **Random-Softmax:**
  Randomly set the modulating factor ($a \leq 0$) at each training epoch.

- **Search-Softmax:**
  Update the distribution of a and search the best model from B candidates for the next epoch.

3. Experiments

Table 2. Face datasets for training and test. (P) and (G) refer to the probe and gallery set, respectively.

	Datasets	#Identities	Images
Training	CASIA-WebFace-R	9,879	0.43M
	MS-Celeb-1M-v1c-R	72,690	3.28M
Test	LFW	5,749	13,233
	SLLFW	5,749	13,233
	CALFW	5,749	12,174
	CPLFW	5,749	11,652
	AgeDB	568	16,488
	CFP	500	7,000
	RFW	11,430	40,607
	MegaFace	530 (P)	1M (G)
	Trillion-Pairs	5,749 (P)	1.58M (G)

Overlap Removal

3. Experiments

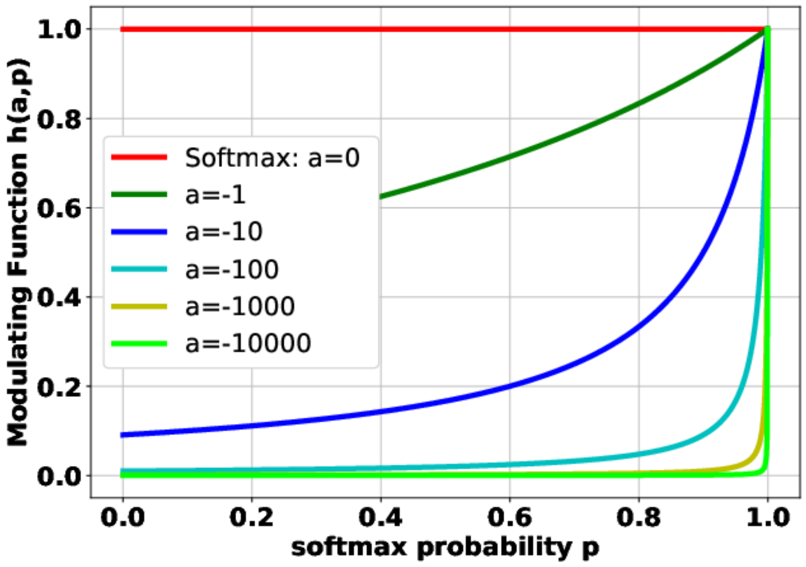
- **Data processing:**

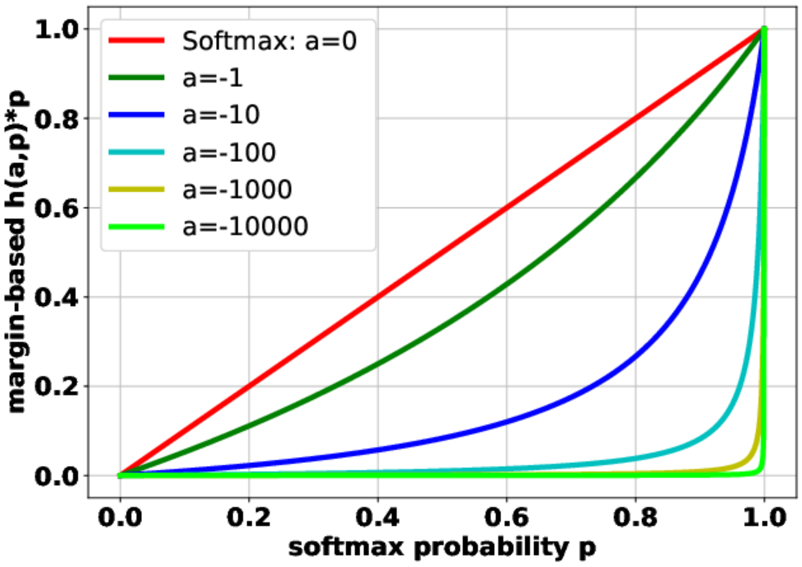
  We detect the faces by adopting the **FaceBoxes detector** and localize five landmarks through a simple 6-layer CNN. The detected faces are cropped and resized to **144×144**.

- **CNN architecture:**

  We use the **SEResNet50-IR** as the backbone, which is also publicly available at the website https://github.com/wujiyang/Face_Pytorch

3. Experiments





3. Experiments

Table 3. Effect of reducing softmax probability by setting the modulating factor $a \leq 0$. The training set is **MS-Celeb-1M-v1c-R**.

	0	-1	-10	-100	-1000	-10000
LFW	99.53	99.56	99.66	99.71	99.61	99.71
SLLFW	98.78	98.91	99.20	99.28	99.36	99.36

Table 4. Effect of the number of sampled models by setting B. The training set is **MS-Celeb-1M-v1c-R**.

	$B = 2$	$B = 4$	$B = 8$	$B = 16$
LFW	99.79	99.78	99.79	99.78
SLLFW	99.31	99.56	99.53	99.58

3. Experiments

Table 9. Performance (%) of different loss functions on the test sets MegaFace and Trillion-Pairs. The training set is **CASIA-WebFace-R**.

Method	MegaFace Id.	MegaFace Veri.	Trillion-Pairs Id.	Trillion-Pairs Veri.
Softmax	65.17	71.29	12.34	11.35
A-Softmax	64.48	71.98	11.83	11.11
V-Softmax	60.09	65.40	9.08	8.65
Arc-Softmax	79.91	84.57	21.32	20.97
AM-Softmax	82.86	87.33	25.26	24.66
AM-LFS	71.30	77.74	16.16	15.06
Random-Softmax	82.51	86.13	27.70	27.28
Search-Softmax	**84.38**	**88.34**	**29.23**	**28.49**

Table 10. Performance (%) of different loss functions on the test sets MegaFace and Trillion-Pairs. The training set is **MS-Celeb-1M-v1c-R**.

Method	MegaFace Id.	MegaFace Veri.	Trillion-Pairs Id.	Trillion-Pairs Veri.
Softmax	91.10	92.30	50.34	46.63
A-Softmax	90.81	93.49	49.99	45.59
V-Softmax	94.45	95.25	63.85	61.17
Arc-Softmax	96.39	96.86	67.60	66.46
AM-Softmax	96.77	97.20	69.02	67.94
AM-LFS	92.51	93.80	54.85	52.76
Random-Softmax	96.15	96.81	68.73	68.03
Search-Softmax	**96.97**	**97.84**	**70.41**	**68.67**

4. Conclusion

- We identify that for margin-based softmax losses, the key to enhance the feature discrimination is actually **how to reduce the softmax probability**.
- We define **a simple but very effective search space**, which involves only one parameter to search. Accordingly, we design a random and a reward-guided method to search the best candidate.
- We conduct **extensive experiments** on a variety of face recognition benchmarks.
- Code and datasets: http://www.cbsr.ia.ac.cn/users/xiaobowang/

Thanks