



北京邮电大学

BEIJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS

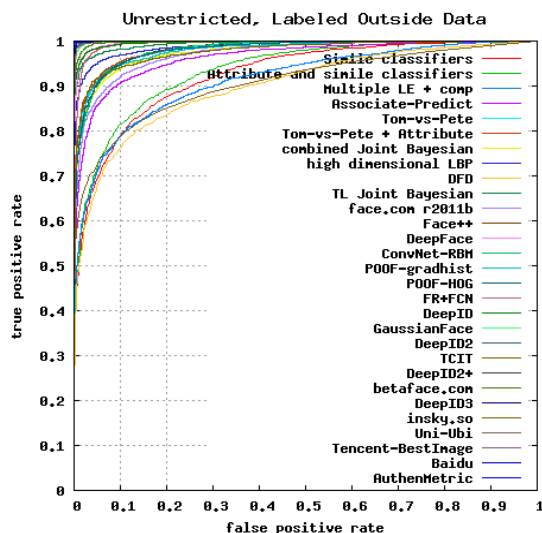
# 人脸识别新问题与数据库

邓伟洪

北京邮电大学

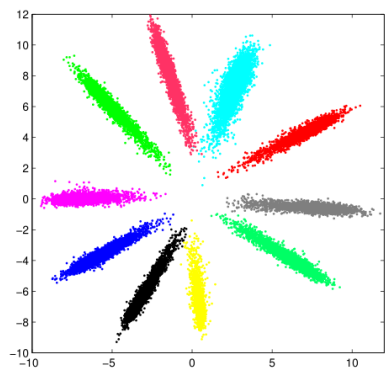
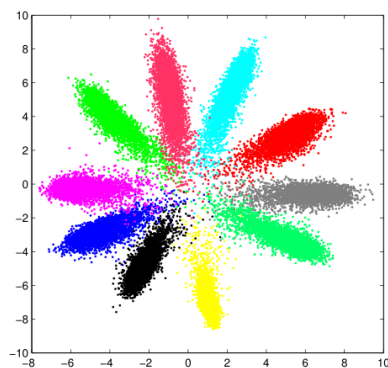
# 人脸识别：研究背景

以LFW为代表的人脸识别  
基准实验精度接近饱和



深度学习提供了LFW的较好的  
解决方案

| Methods                 | #Train      | #Net | LFW Accuracy  |
|-------------------------|-------------|------|---------------|
| Deepface [1]            | 4M          | 3    | 97.00%        |
| DeepID2 [2]             | 0.2M        | 1    | 95.12%        |
| DeepID2+                | 0.2M        | 25   | 99.47%        |
| FaceNet [4]             | 200M        | 1    | 99.63%        |
| VGG-Face [3]            | 2.6M        | 1    | 98.95%        |
| Baidu                   | 1.3M        | 1    | 99.13%        |
| Center-loss [6]         | 0.7M        | 1    | 99.28%        |
| NAN [15]                | 3M          | 1    | -             |
| Baseline A              | 0.5M        | 1    | 97.13%        |
| Baseline B              | 0.5M        | 1    | 98.81%        |
| Baseline C              | 0.5M        | 1    | 99.11%        |
| <b>DCFL</b>             | <b>0.5M</b> | 1    | <b>99.32%</b> |
| <b>DCFL (64-layers)</b> | <b>4.7M</b> | 1    | <b>99.55%</b> |



# 国际权威LFW评测是什么？

Negative pairs in LFW



**LFW: 6000道判断题 (3000是+3000非)**  
**产业界最高准确率99.8%**

# 新的人脸识别评价方法

## 国际学术界的出题思路

- 把验证问题（判断题）变成识别问题（选择题）
- 增加选择题的选项数量（Megaface: 百万级选项）

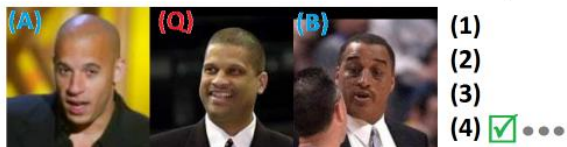


## 我们的出题思路

- 不改变评价规则和数据规模
- 出难题：增加肯定题和否定题的难度

# 人脸识别：众包相似脸选择

采用主流深度学习  
方法选择潜在相似  
度较高的人脸对



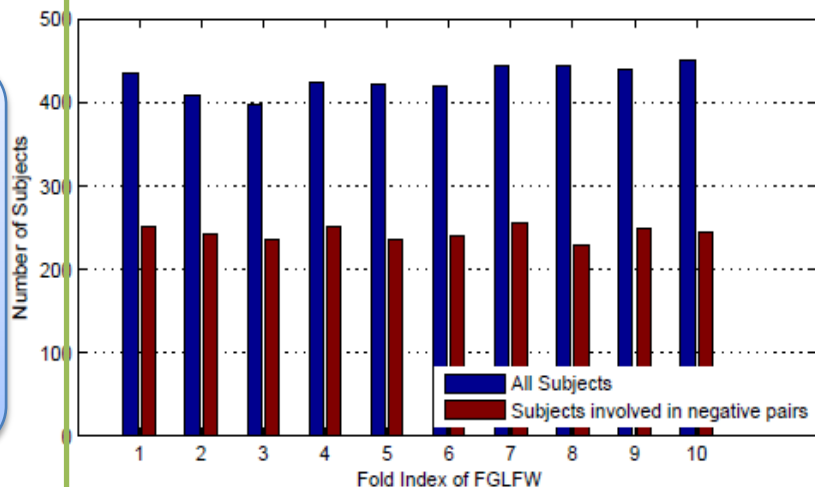
.....



- (1)  $Q \sim A \ \& \ Q \sim B$
- (2)  $Q \not\sim A \ \& \ Q \not\sim B$
- (3)  $S(Q, A) > S(Q, B)$
- (4)  $S(Q, A) < S(Q, B)$

百万级  
的三元组“相对相似  
度”众包标注

三元组划分为二  
元组



每个fold中相似脸的存在比例  
比预期中高

# 人脸识别：相似脸选择结果

把3000道困难的否定题替换LFW数据库中的否定题



Linda Dano Liza Minnelli



Hartmut Mehdorn Patrick Clawsen



John Swofford Bill McBride



Oxana Fedorova Salma Hayek



Marty Mornhinweg Bob Huggins



Edward Kennedy Dennis Hastert



Donald Keck Roger Cook



Luis Gonzalez Pablo Latras



Kjell Magne Bondevik Cesa Gaviria



Pete Sampras Dominik Hrbaty



Nicolas Massu Fernando Gonzalez



Jason Lezak Brendan Hansen



Patricia Clarkson Jerry Hall



Barbara Brezigar Valerie Harper



John Rowe Richard Armitage



Tim Floyd Michael J Fox



Tom Craddock Stefaan Declerk



Rick Romley Jim O'Brien



Elena Bovina Penelope Taylor



Alessandro Nesta Filippo Inzaghi

# 人机对比实验



Ange lababy

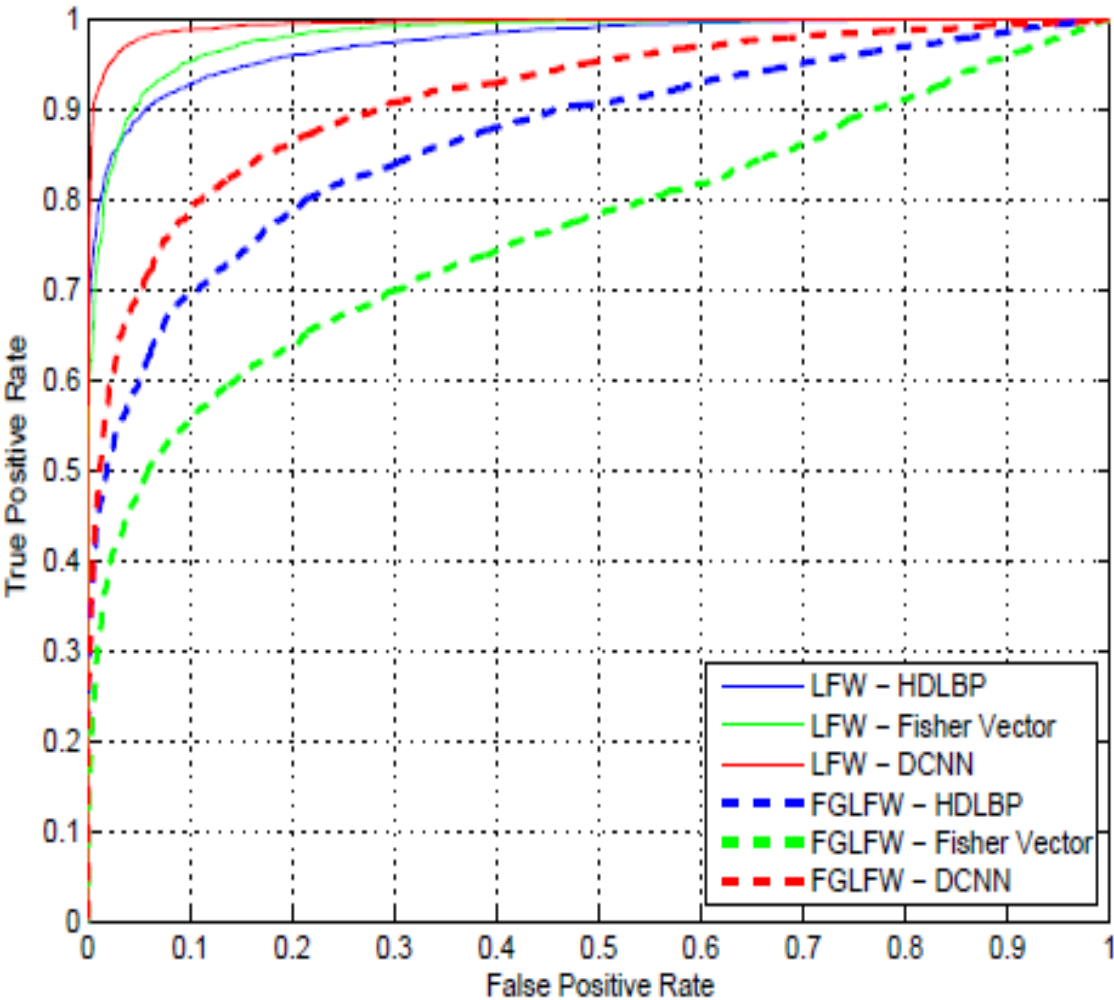


Ange lababy

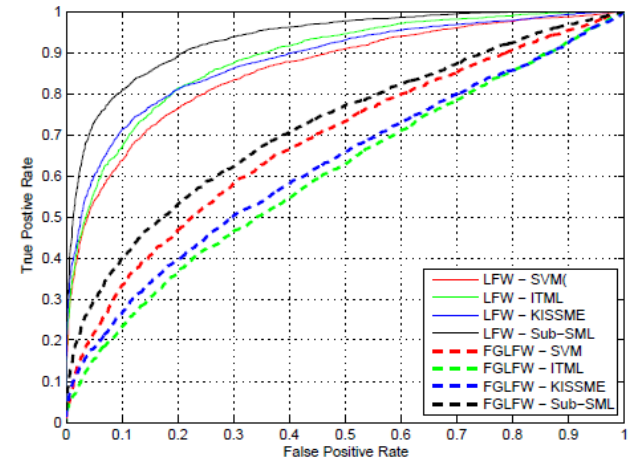
前5题：深度学习正确，人工比对错误  
第6题：人工比对正确，深度学习错误

# 人脸识别：LFW旧题和新题的ROC对比

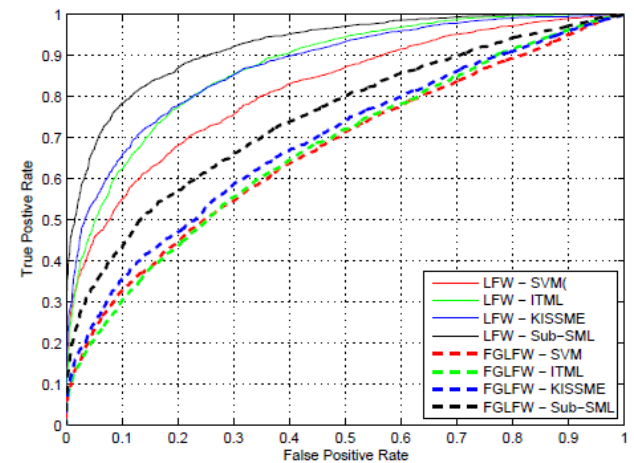
高维特征和深度特征



浅层特征+距离学习



(a) SSIFT



(b) LBP

# 人脸识别：新问题定义

“细粒度”人脸识别问题：如果入侵者选择外貌相似的人进行攻击，人工验证和机器识别都尚不完美，实验中有**约3%的冒名顶替不能被人工检查或者机器算法发现**。

6000道判断题  
(3000是+3000非)  
测试结果



| 深度学习 | 人工比对 | 人机融合 |
|------|------|------|
| 95%  | 92%  | 97%  |



人机高度  
互补性

# 人脸识别：细粒度相似脸甄别

**细粒度相似脸甄别（验证）新问题及数据库**，通过众包方式挑选3000对外貌相似人脸，以更高难度的方式重新定义了国际权威的Labeled Face in-the-Wild (LFW) 人脸识别评测，详情浏览**LFW官方网站的资源链接**。

## Labeled Faces in the Wild



### Menu

- LFW Home
  - Mailing
  - Explore
  - Download
  - Train/Test
  - Results
  - Information
  - Errata
  - Reference
  - **Resources**
  - Contact
  - Support
  - Changes
- Part Labels
- UMass Vision

### Labeled Faces in the Wild Home



### Resources:

Collected resources related to LFW:

- **Fine-grained LFW (FGLFW)**

"Common face verification addresses mainly large intra-class variations, such as pose, illumination, and expression. After inspecting the LFW databases, one can identify a main limiting factor for its unconstrained face verification task: almost all the negative face pairs are quite easy to distinguish. Thus, verification is, by its nature a problem in which many examples are very easy with large inter-class variance, because the collection of LFW database is based on the assumption of random imposter attack. For practical usage, however, it is likely that a desperate impostor may attempt to spoof a genuine user by seeking a similarly-looking people. To simulate this deliberate imposture attack, we construct FGLFW database, which deliberately selects 3000 similarly-looking face pairs within original image folders by human crowdsourcing to replace the random negative pairs in LFW."

## 最新结果

|      | LFW   | FG LFW |
|------|-------|--------|
| 人工比对 | 99.2% | 92%    |
| 深度学习 | 99.3% | 95%    |

Chen & Deng, CVPR17

# 人脸识别：跨年龄的LFW

LFW  
3岁



Angela Lansbury

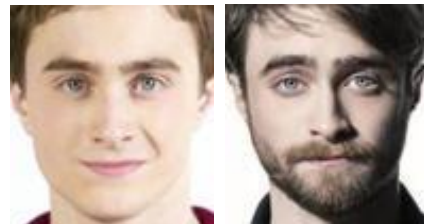
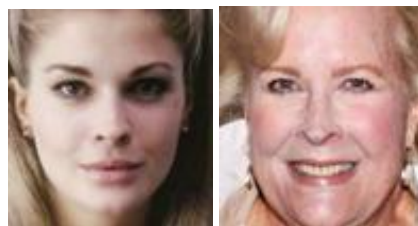
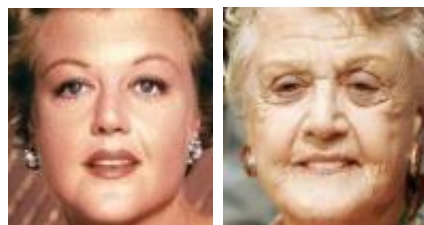


Candice Bergen



Daniel Radcliffe

Cross-Age LFW  
15岁



**跨年龄 Label Face in-the-Wild:** 完全保持LFW的人员名单和规模，逐个人查找跨年龄的图片对。

|      | LFW  | CA LFW |
|------|------|--------|
| 深度学习 | ~99% | ~90%   |

即将公开

# Facial Expression Understanding

---

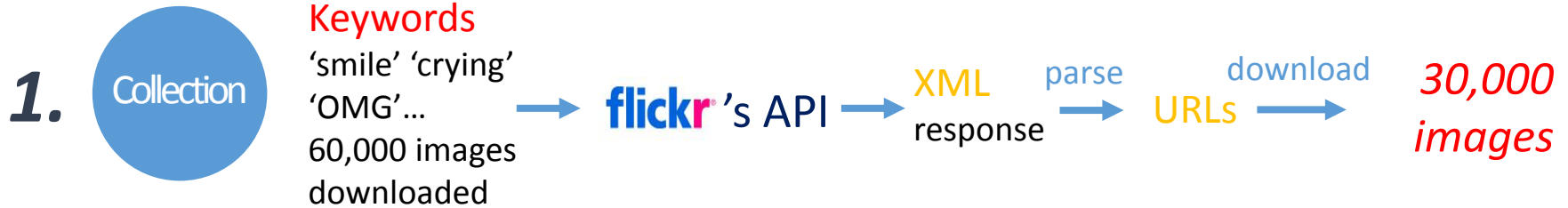
- **Challenges**

- Most widely used facial expression datasets are recorded in “**lab**” controlled environment with **insufficient** data.
- **Annotation** of facial expression categories is difficult and time-consuming.
- Facial emotional images in real world are **various** in subjects’ age, gender and ethnicity, head poses, lighting conditions and occlusions.

- **Opportunities**

- **Millions of** images are being uploaded every day by users from different events and social gatherings.
- **Crowdsourcing** is an efficient tool to collect the judgments of annotation results from a large common population.
- The emerging **deep learning** techniques have boosted unconstrained expression recognition to the new state-of-the-art.

# Real-world Affective Face Database (*RAF-DB\**)



## ● Image Collection

### ● Flickr (Image social network)

- [https://api.flickr.com/services/rest/?method=flickr.photos.search&api\\_key={}&text={}&tags={}&per\\_page={}&page={}&sort=relevance](https://api.flickr.com/services/rest/?method=flickr.photos.search&api_key={}&text={}&tags={}&per_page={}&page={}&sort=relevance)
- XML response → Interpreted into URLs of the images → Download

```
:collection PhotosOfOneSearch="Search1">
:photos page="1" pages="127" perpage="500" total="63478">
<photo id="14197338518" owner="74529773@N07" secret="8313e97a1f" server="2909" farm="3" title="null" ispublic="1" isfriend="0" isfamily="0" />
<photo id="8505470995" owner="69642848@N05" secret="375b0c82bc" server="8111" farm="9" title="null" ispublic="1" isfriend="0" isfamily="0" />
<photo id="14744669763" owner="96619214@N04" secret="a818044e97" server="5557" farm="6" title="null" ispublic="1" isfriend="0" isfamily="0" />
<photo id="3568274837" owner="22505098@N04" secret="31f8cd91db" server="3358" farm="4" title="null" ispublic="1" isfriend="0" isfamily="0" />
<photo id="6109626903" owner="45833131@N03" secret="176b96e284" server="6201" farm="7" title="null" ispublic="1" isfriend="0" isfamily="0" />
<photo id="8204874510" owner="35456872@N00" secret="61d0c90451" server="8207" farm="9" title="null" ispublic="1" isfriend="0" isfamily="0" />
<photo id="335131224" owner="85353067@N00" secret="cae5519488" server="151" farm="1" title="null" ispublic="1" isfriend="0" isfamily="0" />
<photo id="5821864725" owner="55919672@N08" secret="81e246c9fe" server="2603" farm="3" title="null" ispublic="1" isfriend="0" isfamily="0" />
<photo id="113989596" owner="64705987@N00" secret="601e305e76" server="56" farm="1" title="null" ispublic="1" isfriend="0" isfamily="0" />
<photo id="858252701" owner="9722602@N05" secret="805210523d" server="1334" farm="2" title="null" ispublic="1" isfriend="0" isfamily="0" />
```

# Real-world Affective Face Database (*RAF-DB\**)



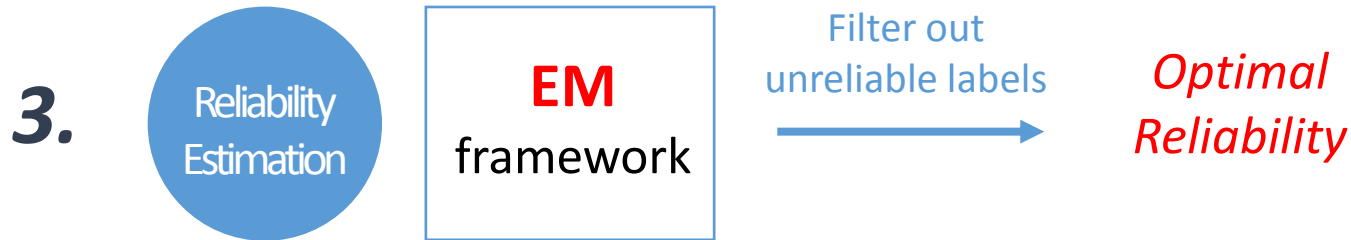
## ● Image Annotation

### ● Crowd-sourcing

- 315 well-trained annotators were asked to label facial images with one of the seven basic categories
- Each image is annotated enough times independently, i.e., around 40 times in our experiment.



# Real-world Affective Face Database (*RAF-DB\**)



## ● Reliability Estimation

### ● Filter noisy labels

- an Expectation Maximization (EM) framework was used to assess each labeler's reliability.

---

#### Algorithm 1 Label reliability estimation algorithm.

---

**Input:** Training set  $D = \{(x_j, t_j^1, t_j^2, \dots, t_j^R)\}_{j=1}^n$

**Output:** Each annotator's reliability  $\alpha_i^*$

**Initialize:**

$\forall j = 1, \dots, n$ , initialize the true label  $y_j$  using majority voting

$$\beta_j := - \sum_{i=1}^R p(t_j^i) \ln p(t_j^i), \alpha_i := 1,$$

The initial value of  $\beta_j$  is image  $j$ 's entropy. The higher the entropy, the more uncertain the image.

**Repeat:**

E-step:

$$Q_j(y_j) := \prod_i p(y_j | t_j, \alpha_i, \beta_j)$$

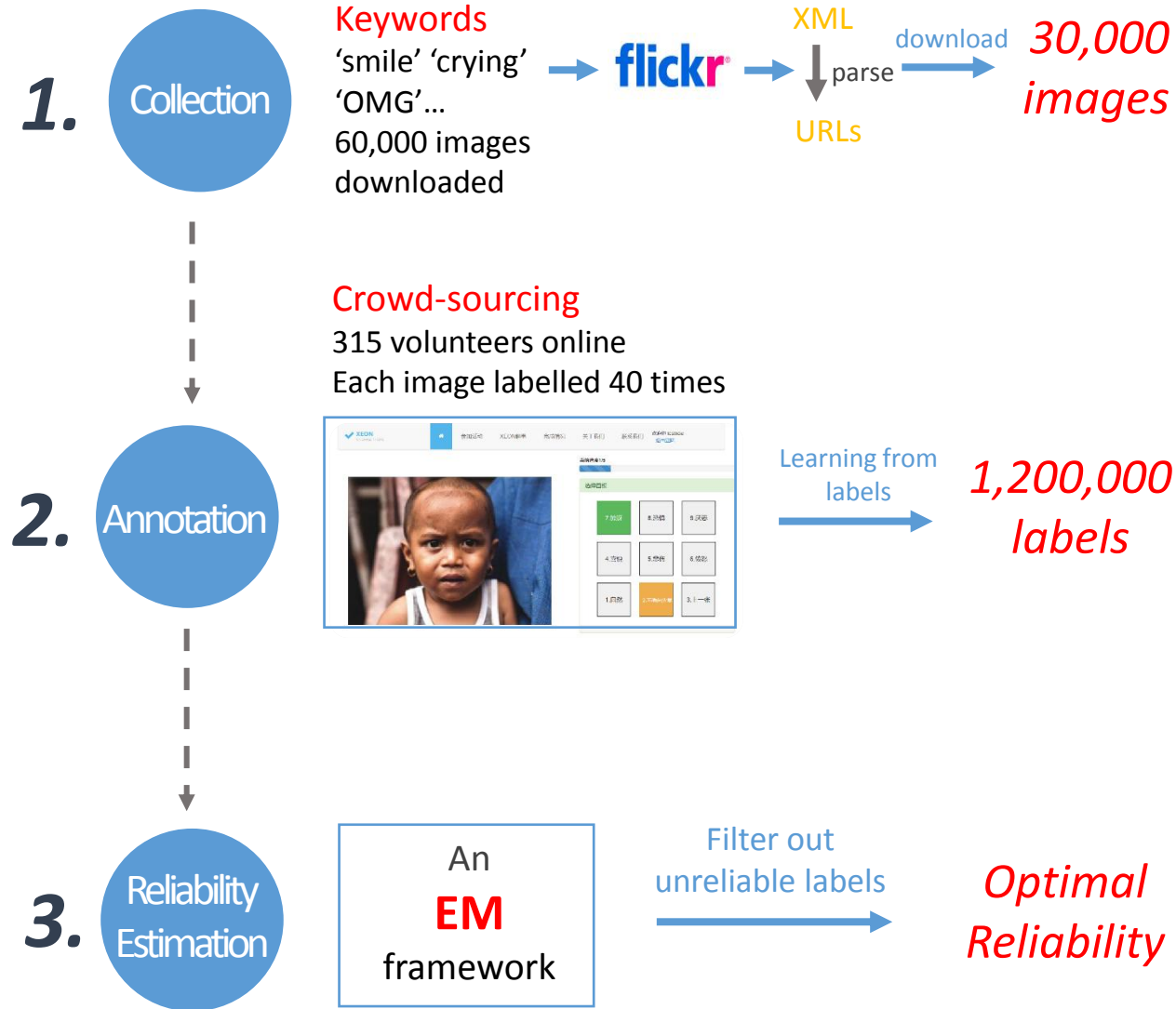
M-step:

$$\alpha_i := \arg \max_{\alpha_i} \sum_j \sum_{y_j} Q_j(y_j) \ln \frac{p(t_j, y_j | \alpha_i, \beta_j)}{Q_j(y_j)}$$

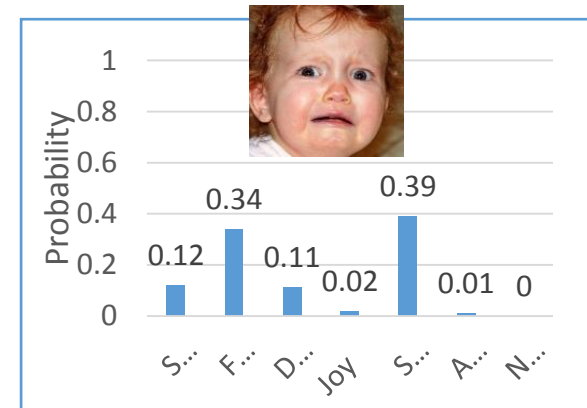
**Until convergence**

---

# Real-world Affective Face Database (*RAF-DB\**)



## Result



**Seven Basic Emotions  
&  
Twelve Compound  
Emotions**

\* Li & Deng, CVPR2017

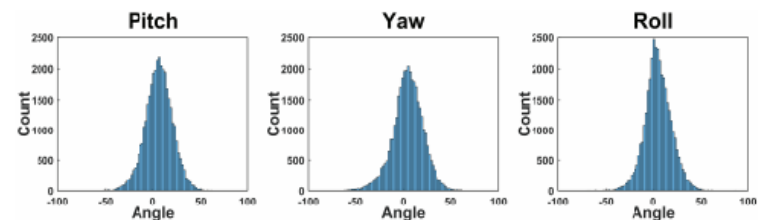
**Multi-label Emotions**

+ Li & Deng, ICCV Submission

# Real-world Affective Face Database (*RAF-DB\**)

- **Database Statistics**

- **29672** number of real-world images,
- a 7-dimensional **expression distribution** vector for each image,
- two different subsets: **single-label subset**, including **7** classes of basic emotions; **two-tab subset**, including **12** classes of compound emotions,
- **5 accurate landmark locations**, **37 automatic landmark locations**, **race**, **age range** and **gender attributes** annotations per image.



---

## Age distribution

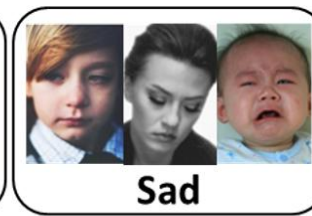
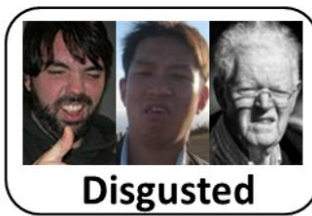
---

| 0~3   | 4~19   | 20~39  | 40~69  | 70+   |
|-------|--------|--------|--------|-------|
| 2696  | 4731   | 16460  | 4696   | 1089  |
| 9.09% | 15.94% | 55.47% | 15.83% | 3.67% |

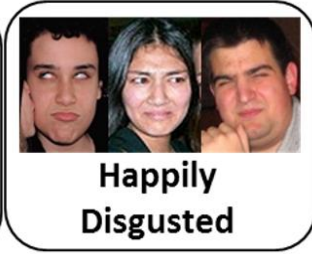
---

# Real-world Affective Face Database (*RAF-DB\**)

## 7 classes Basic Emotions



## 12 classes Compound Emotions



\* Li & Deng, CVPR2017

# Comparison: Real-world & lab-controlled

## LAB-BASED Datasets

Controlled lab conditions



[1]

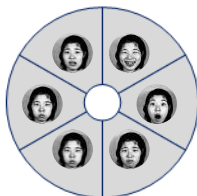
Prototypical emotions



Surprise!

[2]

Balanced distribution



≈1:1:1:1:1:1

1

2

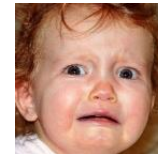
3

## REAL-WORLD RAF

Diverse imaging conditions



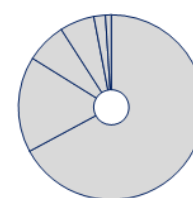
! Compound emotions



Fear?

Sad?

Highly-imbalanced

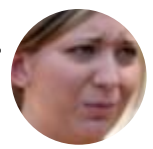


e.g. 'Joy'

'Disgust'



>>



# Action Units: RAF-DB is more diverse

CK+ [4]

RAF



Surprise



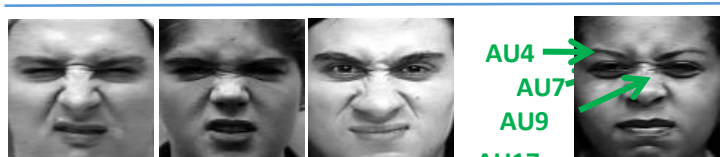
Joy



Fear



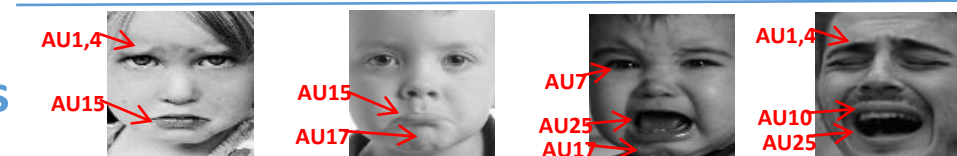
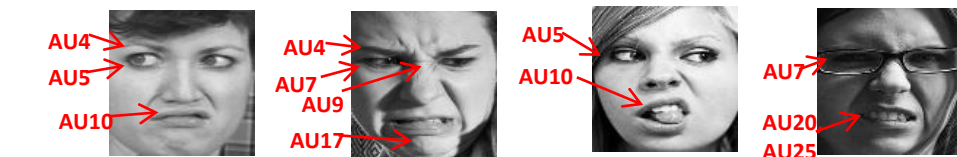
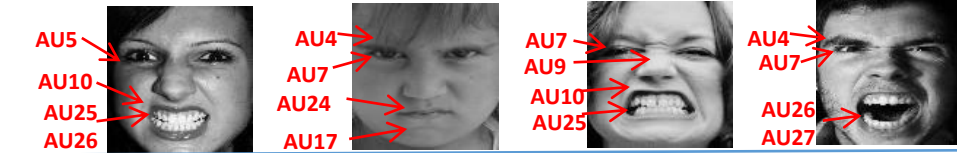
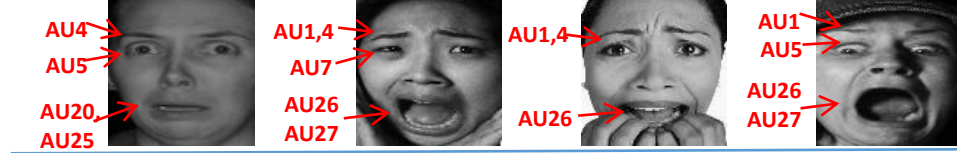
Anger



Disgust



Sadness



# DLP-CNN: Deep Locality-preserving CNN

---

- **Locality Preserving Projection [5] (LPP) –**
  - \* **Preserving the information of its local region**

1. Construct the adjacency graph
2. Choose the weights
3. Computer the eigenvector equation below:

$$XLX^T \mathbf{a} = \lambda XD X^T \mathbf{a}$$

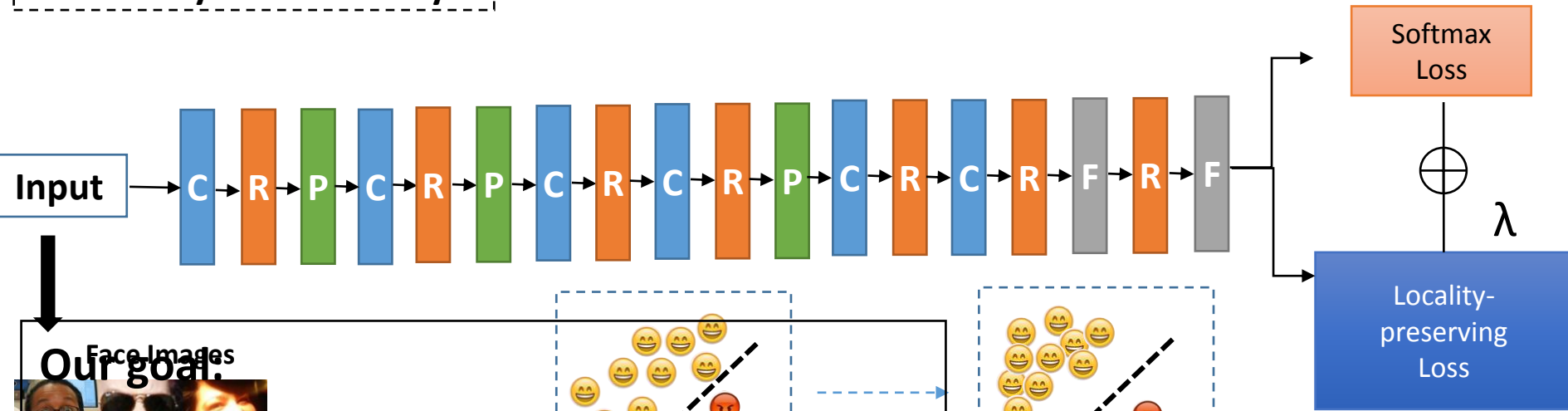
LLP + Deep Learning 

**DLP-CNN**

(Deep Locality-preserving CNN)

# DLP-CNN: Deep Locality-preserving CNN

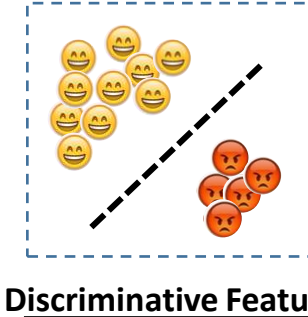
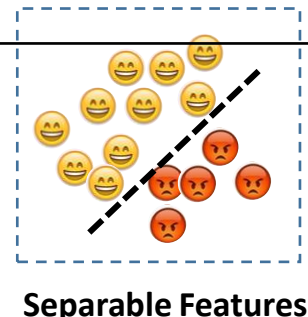
**C:** The convolution layer  
**P:** The max-pooling layer  
**R:** The ReLU layer  
**F:** The fully connected layer



**Our goal:**

Face Images

$$\min_{i,j} \sum_{i,j} S_{ij} \|x_i - x_j\|_2^2$$

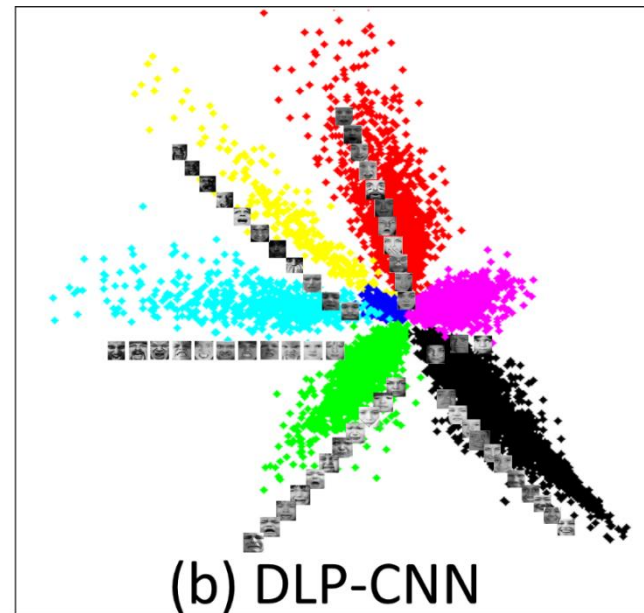
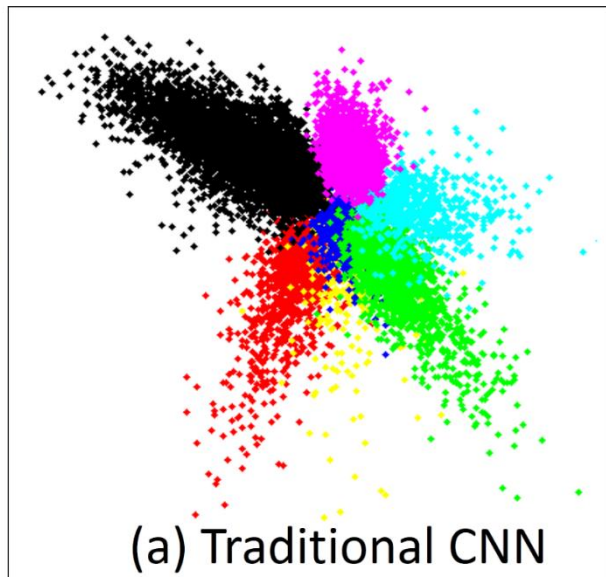


$$S_{ij} = \begin{cases} 1, & x_j \text{ is among } k \text{ nearest neighbors of } x_i \text{ or} \\ & x_i \text{ is among } k \text{ nearest neighbors of } x_j \\ 0, & \text{otherwise} \end{cases}$$

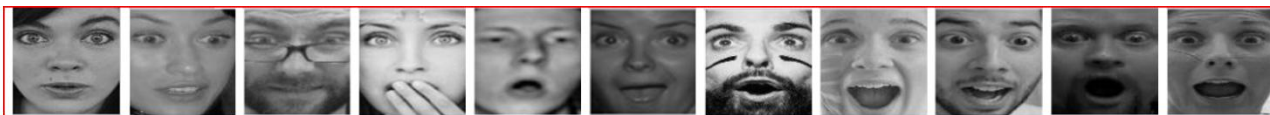
**Locality Preserving Loss:**

$$L_{lp} = \frac{1}{2n} \|x_i - \frac{1}{k} \sum_{x \in N_k\{x_i\}} x\|_2^2$$

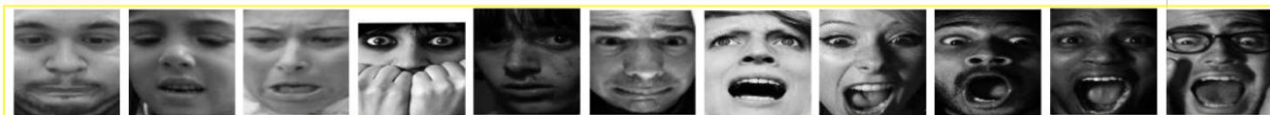
# DLP-CNN: Deep Locality-preserving CNN



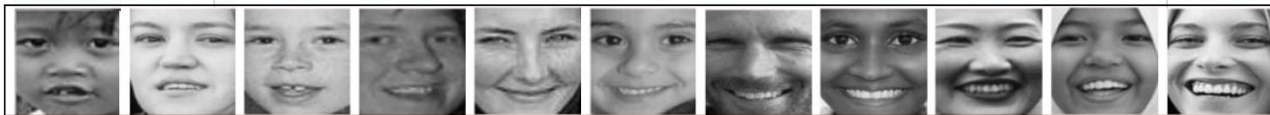
Surprised



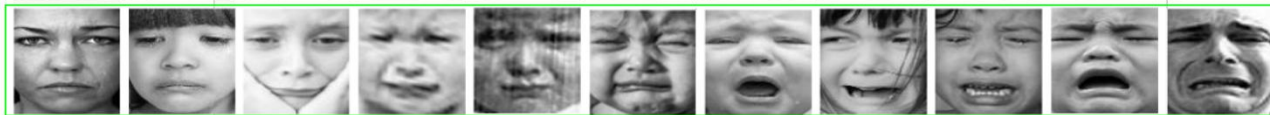
Fearful



Happy



Sad



Angry



# DLP-CNN: Experiment Results

**Table 1.** Expression recognition performance of different DCNNs on RAF. The metric is the mean diagonal value of the confusion matrix.

|      |                 | basic |         |       |           |         |          |         | compound     |              |
|------|-----------------|-------|---------|-------|-----------|---------|----------|---------|--------------|--------------|
|      |                 | Anger | Disgust | Fear  | Happiness | Sadness | Surprise | Neutral | Average      | Average      |
| mSVM | VGG [6]         | 68.52 | 27.50   | 35.13 | 85.32     | 64.85   | 66.32    | 59.88   | <b>58.22</b> | <b>31.63</b> |
|      | AlexNet [7]     | 58.64 | 21.87   | 39.19 | 86.16     | 60.88   | 62.31    | 60.15   | <b>55.60</b> | <b>28.22</b> |
|      | baseDCNN        | 70.99 | 52.50   | 50.00 | 92.91     | 77.82   | 79.64    | 83.09   | <b>72.42</b> | <b>40.17</b> |
|      | center loss [8] | 68.52 | 53.13   | 54.05 | 93.08     | 78.45   | 79.63    | 83.24   | <b>72.87</b> | <b>39.97</b> |
|      | <b>DLP-CNN</b>  | 71.60 | 52.15   | 62.16 | 92.83     | 80.13   | 81.16    | 80.29   | <b>74.20</b> | <b>44.55</b> |
| LDA  | VGG [6]         | 66.05 | 25.00   | 37.84 | 73.08     | 51.46   | 53.49    | 47.21   | <b>50.59</b> | <b>16.27</b> |
|      | AlexNet [7]     | 43.83 | 27.50   | 37.84 | 75.78     | 39.33   | 61.70    | 48.53   | <b>47.79</b> | <b>15.56</b> |
|      | baseDCNN        | 66.05 | 47.50   | 51.35 | 89.45     | 74.27   | 76.90    | 77.50   | <b>69.00</b> | <b>28.23</b> |
|      | center loss [8] | 64.81 | 49.38   | 54.05 | 92.41     | 74.90   | 76.29    | 77.21   | <b>69.86</b> | <b>27.33</b> |
|      | <b>DLP-CNN</b>  | 77.51 | 55.41   | 52.50 | 90.21     | 73.64   | 74.07    | 73.53   | <b>70.98</b> | <b>32.29</b> |

# DLP-CNN: Experiment Results

**Table 2.** Comparison results of DLP-CNN and other state-of-the-art methods on CK+, SFEW and MMI databases. To validate the generalization of our model, the well-trained DLP-CNN has been employed as a feature extraction tool without finetune.

| (a) CK+            |               | (b) SFEW 2.0       |               | (c) MMI            |               |
|--------------------|---------------|--------------------|---------------|--------------------|---------------|
| Methods            | Accuracy      | Methods            | Accuracy      | Methods            | Accuracy      |
| CSPL [9]           | 88.89%        | DL-GPLVM [16]      | 24.70%        | 3DCNN-DAP [12]     | 63.4%         |
| FP+SAE [10]        | 91.11%        | AUDN [11]          | 26.14%        | DTAGN [21]         | 70.24%        |
| AUDN [11]          | 92.05 %       | STM-ExpLet [17]    | 31.73%        | CSPL [9]           | 73.53%        |
| AURF [11]          | 92.22 %       | Inception [13]     | 47.7%         | AUDN [11]          | 74.76%        |
| 3DCNN-DAP [12]     | 92.4 %        | SFEW third [18]    | 48.5%         | STM-ExpLet [22]    | 75.12%        |
| Inception [13]     | 93.2%         | SFEW second [19]   | 52.29%        | F-Bases [23]       | 75.12%        |
| Dis-ExpLet [14]    | 95.1%         | SFEW best [20]     | 52.5%         | Inception [13]     | 77.6%         |
| ESL [15]           | 95.33%        | <b>DLP-CNN</b>     | <b>51.05%</b> | Dis-ExpLet [14]    | 77.6%         |
| <b>DLP-CNN</b>     | <b>95.78%</b> | (without finetune) |               | <b>DLP-CNN</b>     | <b>78.46%</b> |
| (without finetune) |               |                    |               | (without finetune) |               |

# 感谢各位老师和同学！

**FGLFW** (人脸识别, PR 17, CVPR 17a)

<http://vis-www.cs.umass.edu/lfw/#resources>

**RAF** (表情识别, CVPR 17b)

<http://www.whdeng.cn/RAF/model1.html>