Deep Label Distribution Learning for Apparent Age Estimation

Xu Yang, Bin-Bin Gao, Chao Xing, Zeng-Wei Huo, Xiu-Shen Wei, Ying Zhou, Jianxin Wu, and Xin Geng

ChaLearn Looking at People: Workshop and Competitions @ICCV, 2015

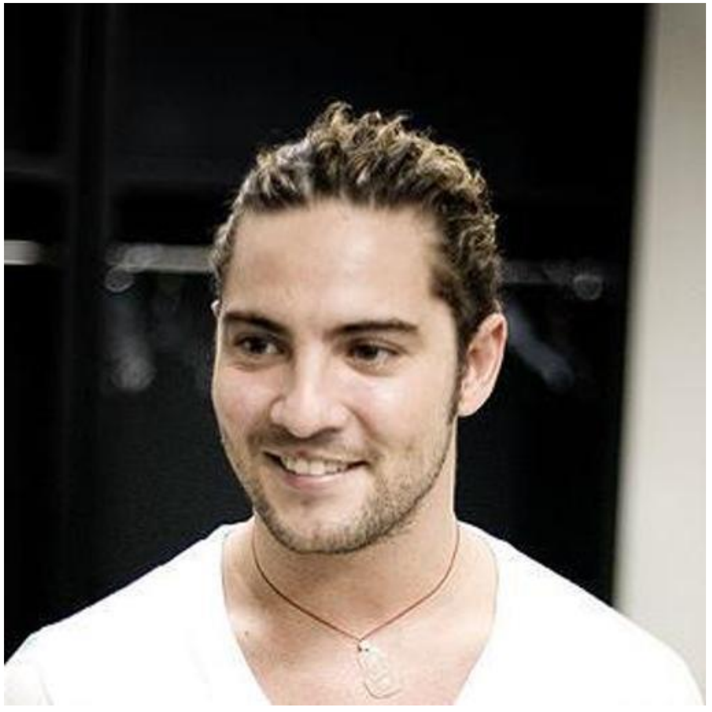
presented by Bin-Bin Gao

Dec. 12, 2015 Santiago de Chile

Why is it difficult?

- It is difficult to provide an exact answer.

How old do these people look like?



A1: 30 **or** 32 years old;
A2: **Around** 31 years old;
......

31 ± 4.24

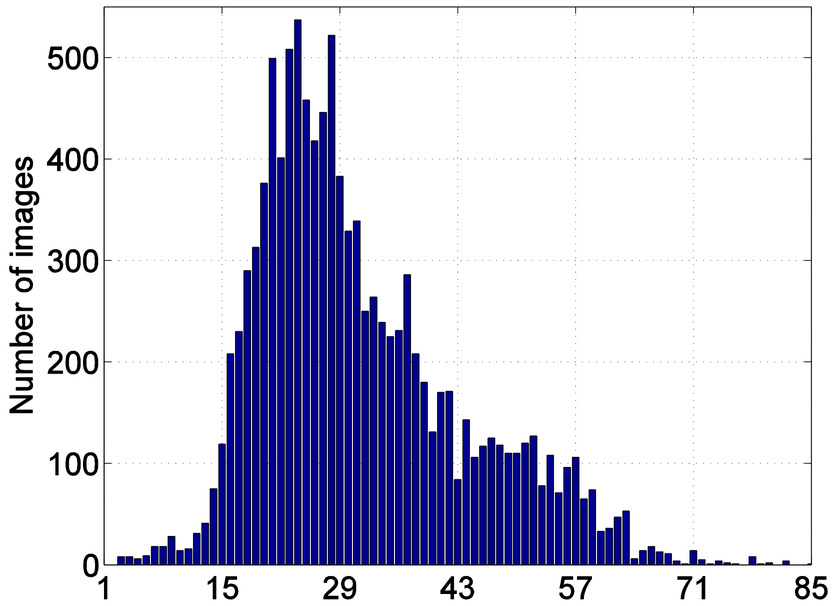
A1: 18 **or** 20 years old;
A2: **May be** 20 years old;
......

17 ± 1.93

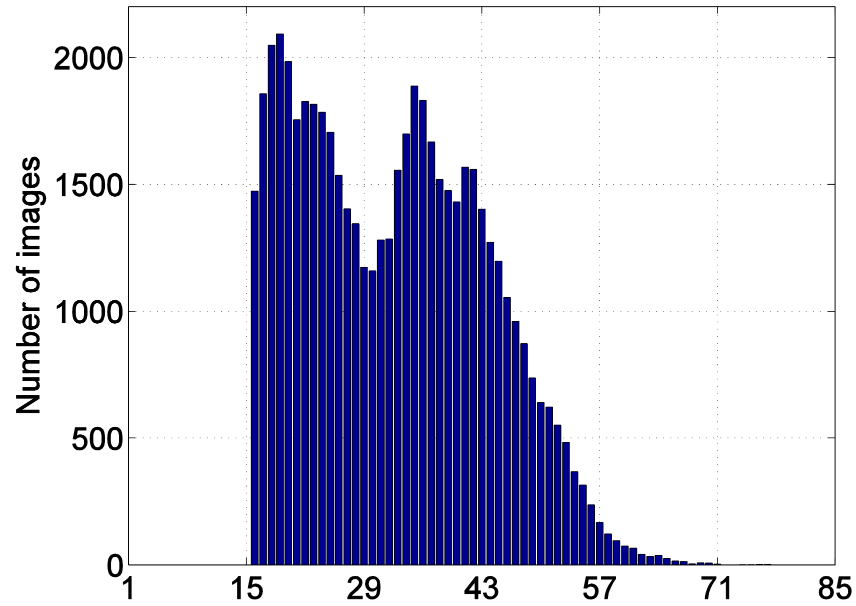
Why is it difficult?

- It is difficult to collect a sufficient and complete training dataset.

✓ ***ChaLearn Competition***
- Ages: from 1 to 85
- Training: 2,476 images
- Validation: 1,036 images



✓ Public age dataset ***Morph Album 2***
- Ages: from 15 to 77
- 55,134 images total



- Training data always has small scale and imbalance.

Some potential applications

- Although the task is very challenging, it has many potential applications.

Cigarette vending machine

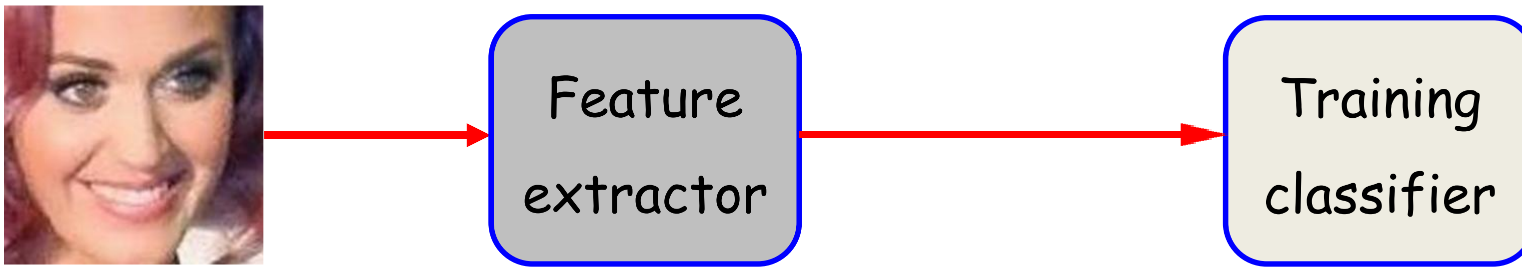
Kraft's vending machine

- Vending machines prevent minors buying cigarettes, alcohol and foods by estimating costumer's apparent age.

http://www.dailymail.co.uk/sciencetech/article-2079048/Kraft-unveils-adults-vending-machine-scans-faces-ensure-children-free-pudding.html

Many existing methods

✓ Hand-crafted feature



- BIF feature [Guo et al., CVPR 2009]
- OHRank [Chang et al., CVPR 2011]
- CCA, rCCA and kCCA [Guo et al. FGR 2013]
- IIS-LLD and CPNN [Gen et al., TPAMI 2013]
- ……

Many existing methods

✓ Deep learning

Low-Level features → Mid-Level features → High-Level features → Training classifier

- Multi-scale CNN [Yi et al., *ACCV* 2014]
- CNN based regression [Huerta *et al.*, PRL 2015]
- CNN for age group classification [Levi & Hassner, CVPR 2015]
- DLA based on CNN features from different layers [Wang *et al.*, WACV 2015]
- ……

Motivations

30±4.17

31±4.24

32±4.23

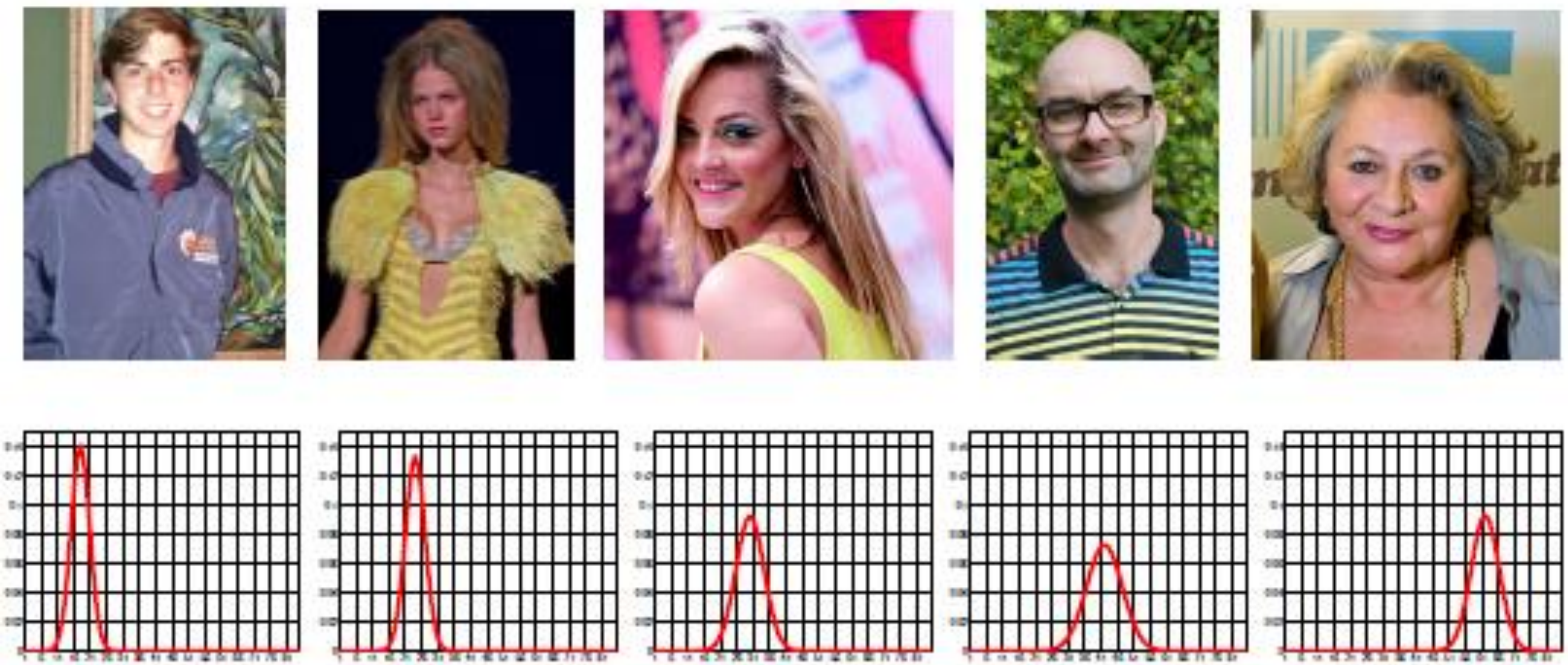
33±2.01

Faces with similar ages look alike in terms of facial details such as wrinkles or skin smoothness. In other words, there is a *correlation among neighboring ages at both* ***image*** and ***feature*** *level*.

How to utilize the correlation?

Motivations

- ✓ How to utilize the correlation?
  - Label Distribution (LD) Learning. But it does not learn the visual representations.



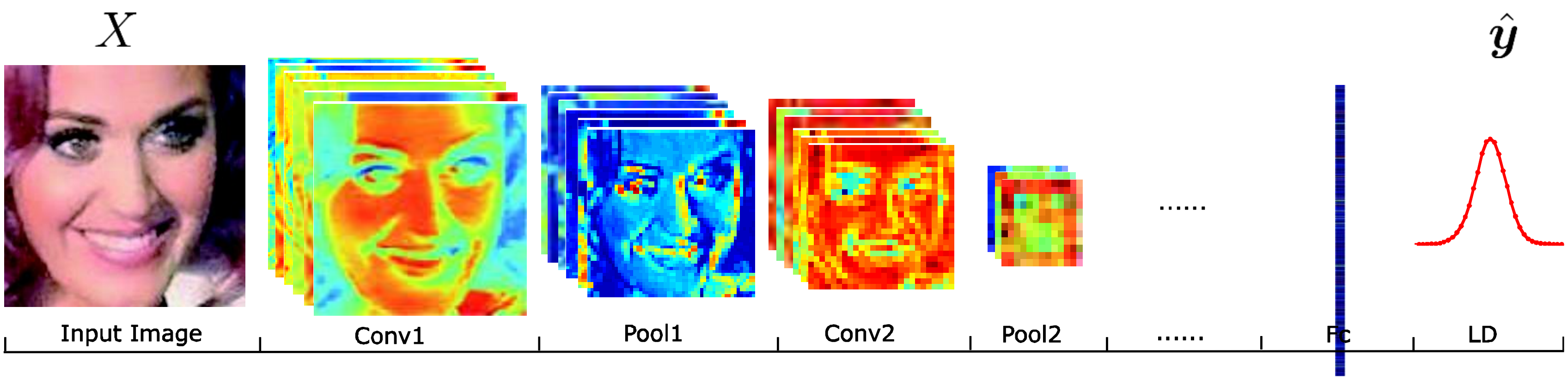
(a) (18,2.82) (b) (24,2.97) (c) (30,4.30) (d) (42,5.45) (e) (62,4.27)

- Generating LD $y_j = \frac{p(l_j|\mu,\sigma)}{\sum_k p(l_k|\mu,\sigma)}$, where $p(l_j|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(l_j-\mu)^2}{2\sigma^2}\right)$

X. Geng, C. Yin, and Z.-H. Zhou. Facial age estimation by learning from label distributions. TPAMI, 35(10):2401–2412, 2013.

Proposed methods

Deep Label Distribution Learning (DLDL)



✓ Formally:

- The goal of DLDL is to directly learn a conditional probability mass function $\hat{\boldsymbol{y}} = p(\boldsymbol{y}|X;\boldsymbol{\theta})$ from D (the training set), where $\boldsymbol{\theta}$ is the parameter of the framework.

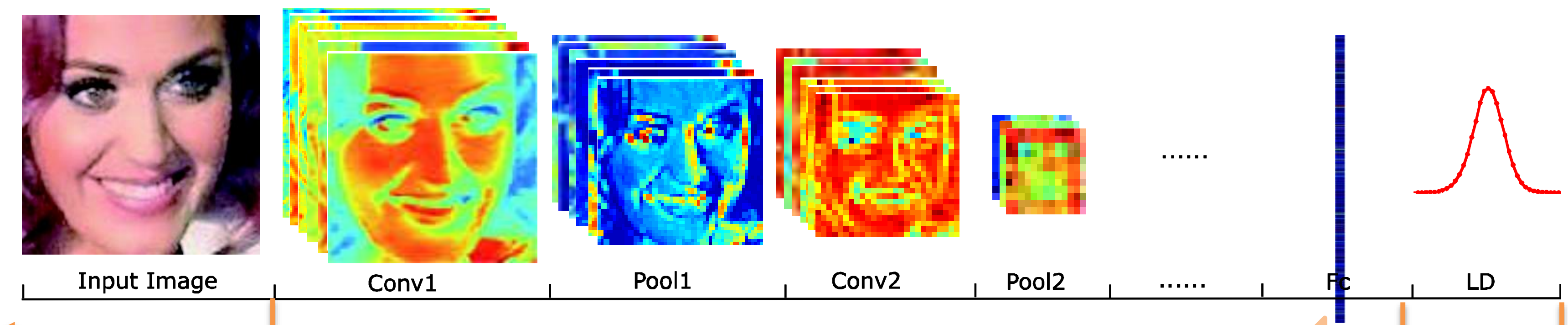
Proposed methods

Learning

$$\hat{y}_j = \frac{\exp(x_j)}{\sum_t \exp(x_t)}$$

✓ Forward:

X $\quad \boldsymbol{x} = \phi(X; \boldsymbol{\theta})$ $\quad \hat{\boldsymbol{y}}$



$$\frac{\partial T}{\boldsymbol{x}}$$

$$\frac{\partial T}{\partial \boldsymbol{\theta}} = \frac{\partial T}{\partial \boldsymbol{x}} \times \frac{\partial \boldsymbol{x}}{\partial \boldsymbol{\theta}}$$

✓ Backward propagation:

- Objective function with K-L divergence:

$$\boldsymbol{\theta}^* = \operatorname{argmin}_{\boldsymbol{\theta}} \sum_k y_k \ln \frac{y_k}{\hat{y}_k} = \operatorname{argmin}_{\boldsymbol{\theta}} \boxed{-\sum_k y_k \ln \hat{y}_k} \; T$$

- The derivative of the K-L loss function is given by

$$\frac{\partial T}{\partial \boldsymbol{x}} = \hat{\boldsymbol{y}} - \boldsymbol{y}$$

Our datasets

LAMDA Learning And Mining from DatA

✓ **Internet face images collecting**

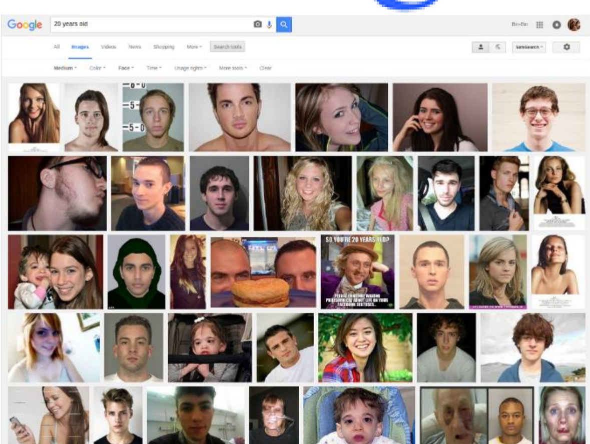
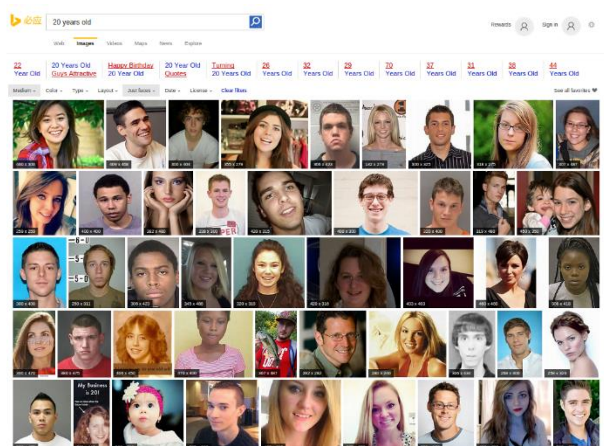
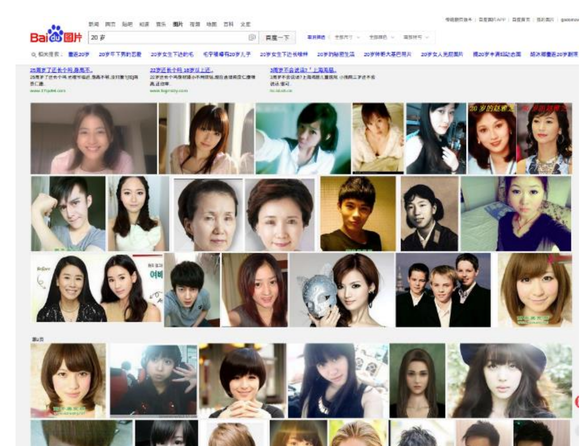
- A set of age related text enquires:

  eg., "20 years old", "20th birthday" and "age-20" for the age of 20 years.

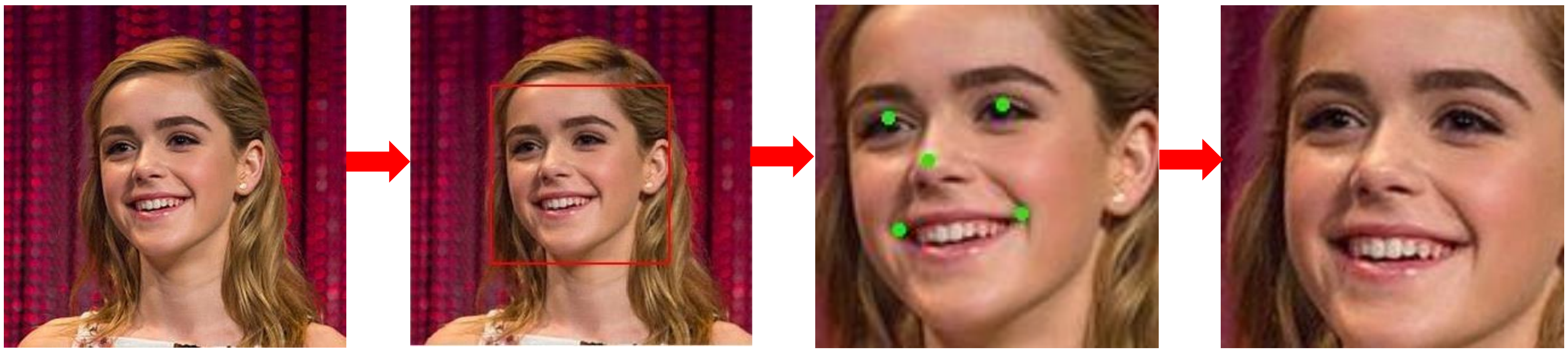
- We use Google, Bing and Baidu image search.

27197 images | 37606 images

Google | bing | Baidu 百度

The face image pre-processing



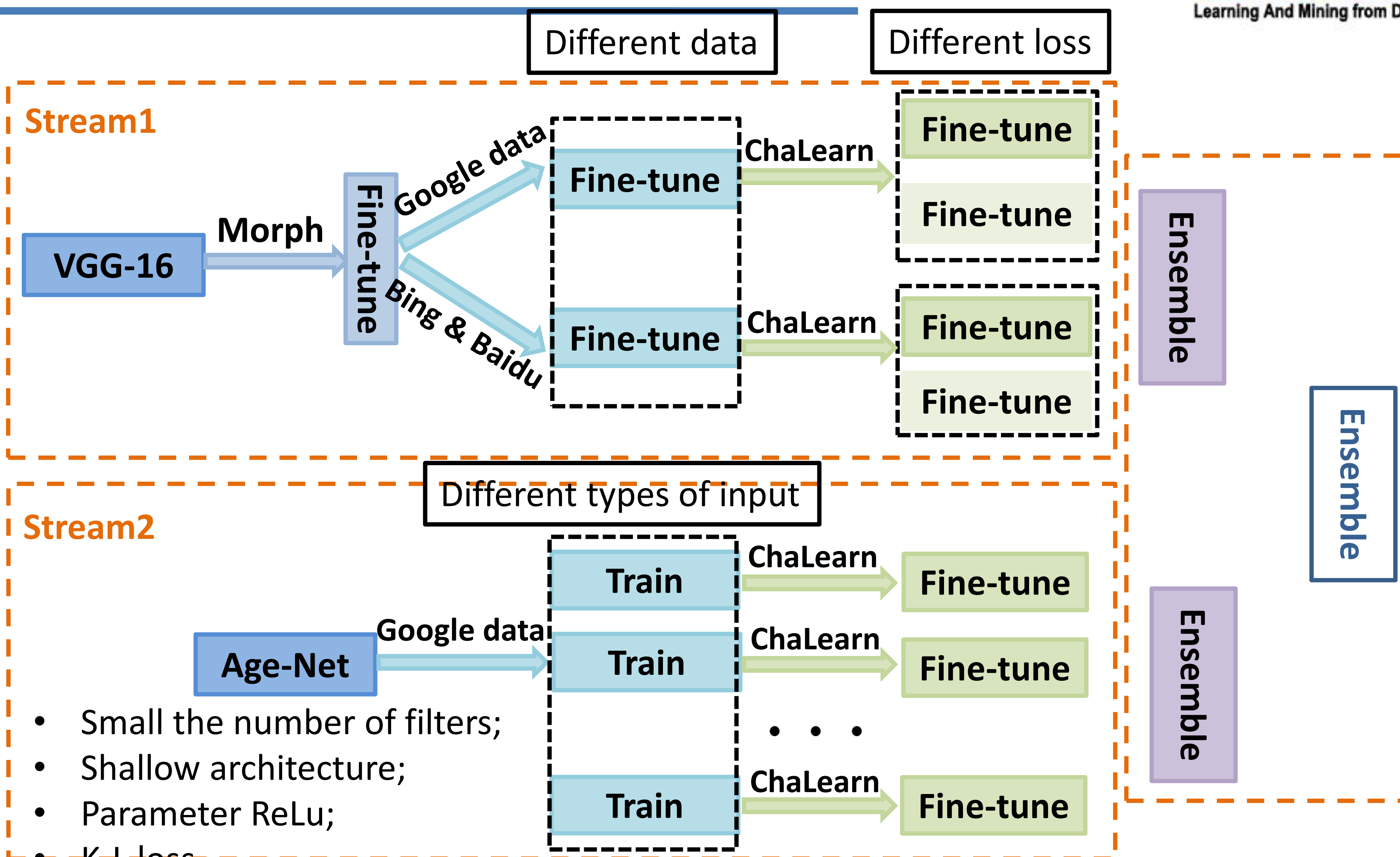
(a) Input (b) Detection (c) Facial points (d) Alignment

Three steps of the images pre-processing

- Face detection
- Facial points detection
- Face alignment

M. Mathias et al., Face detection without bells and whistles. In ECCV, pages 720–735, 2014.
Y. Sun et al., Deep convolutional network cascade for facial point detection. In CVPR, pages 3476–3483, 2013

Model architecture



- Small the number of filters;
- Shallow architecture;
- Parameter ReLu;
- K-L loss.

Training and prediction details

✓ **Training**

- Gaussian random initialization at different layers.

  $1^{st} stream$: The last three layers → The last layer → The last layer.

  $2^{st} stream$: All layers → The last layer.

✓ **Prediction**

- Different fusion strategy

  Early fusion:

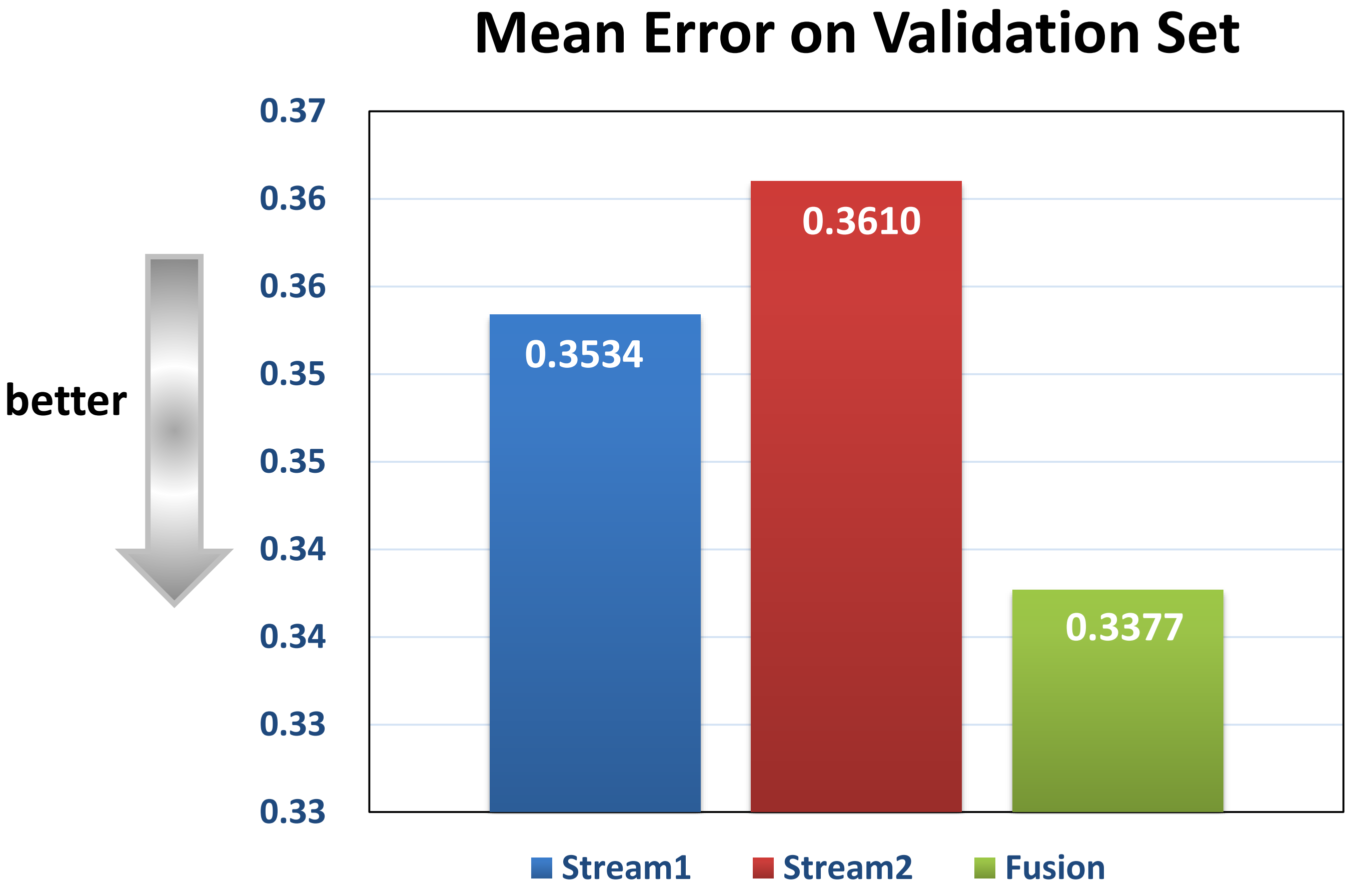
  $1^{st} stream$: Prediction via measuring distance.

  $2^{st} stream$: Averaging estimation distribution.

  Late fusion :

  Averaging the prediction age of two streams.

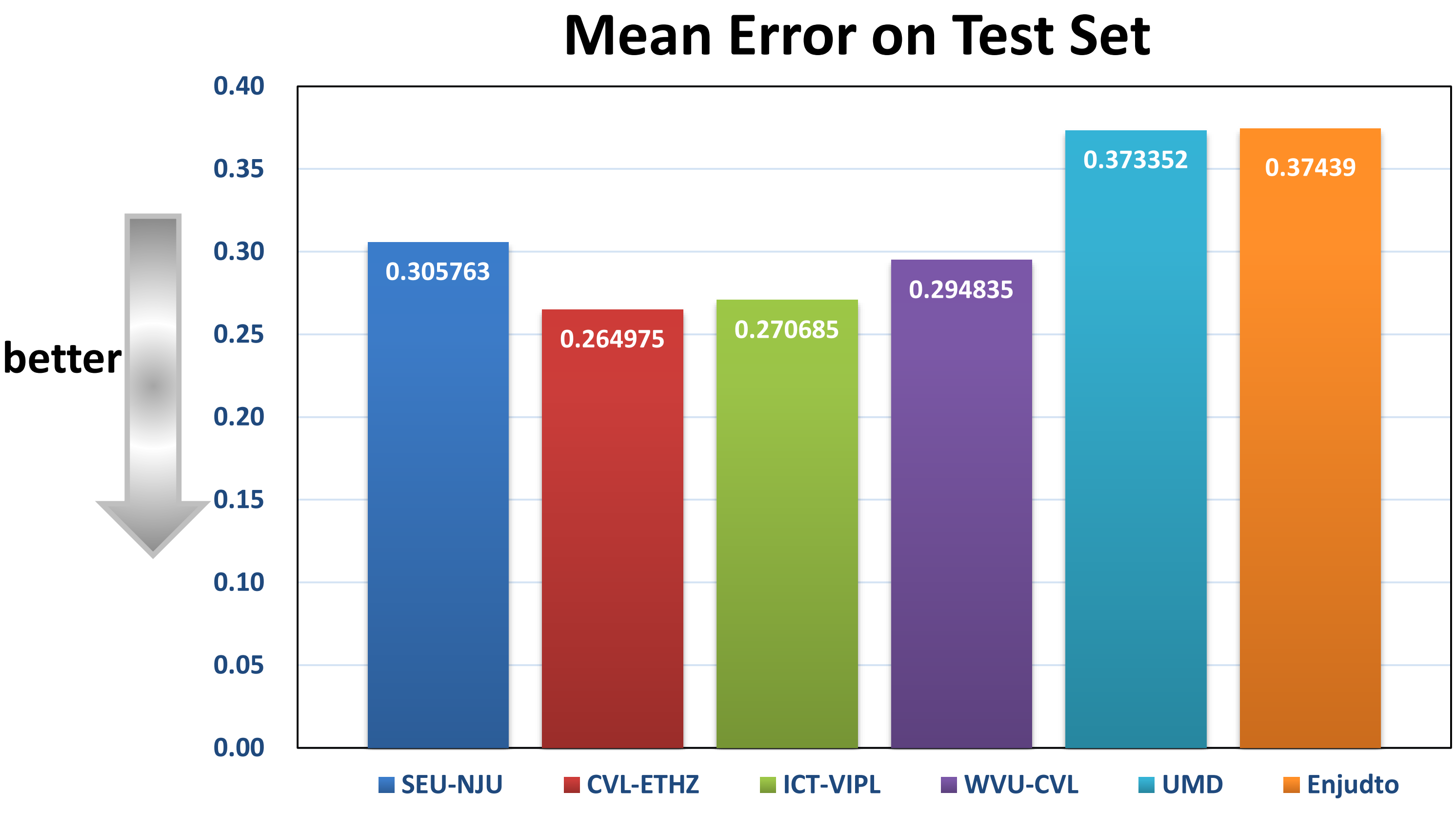
Comparison



The fusion of the two stream is better than single stream.

Final results

The 4nd place with 0.3057 performance.



Conclusions

- ✓ DLDL is an end-to-end learning framework which utilizes the correlation among neighboring labels in both feature learning and classifier learning;
- ✓ DLDL can work when the training set is small.
- ✓ Ensemble strategy: different dataset, different architecture, different initialization and different fusion.

Any questions