Tsinghua University

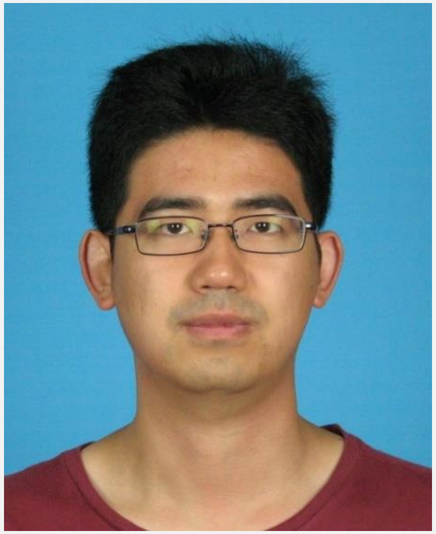
12th IEEE FG 2017
Doctoral Consortium

Deep Feature Learning for Face Alignment and Facial Age Estimation

Hao Liu

Tsinghua University, China

Personal Introduction



Hao Liu

- ✓ Visual Intelligence Group, Department of Automation, Tsinghua University
- ✓ 3rd Year of Ph.D. candidate
- ✓ Supervised by Professor *Jie Zhou* and Associate Professor *Jiwen Lu*
- ✓ Researching Areas: Deep Learning, Facial Age Estimation and Face Alignment

Publications

[1] Hao Liu *et al.*, "Two-Stream Transformer Networks for Video-based Face Alignment," IEEE Transactions on **PAMI**, 2017.

[2] Hao Liu *et al.*, "Learning Deep Sharable and Structural Detectors for Face Alignment," IEEE Transactions on Image Processing (**TIP**), 2017.

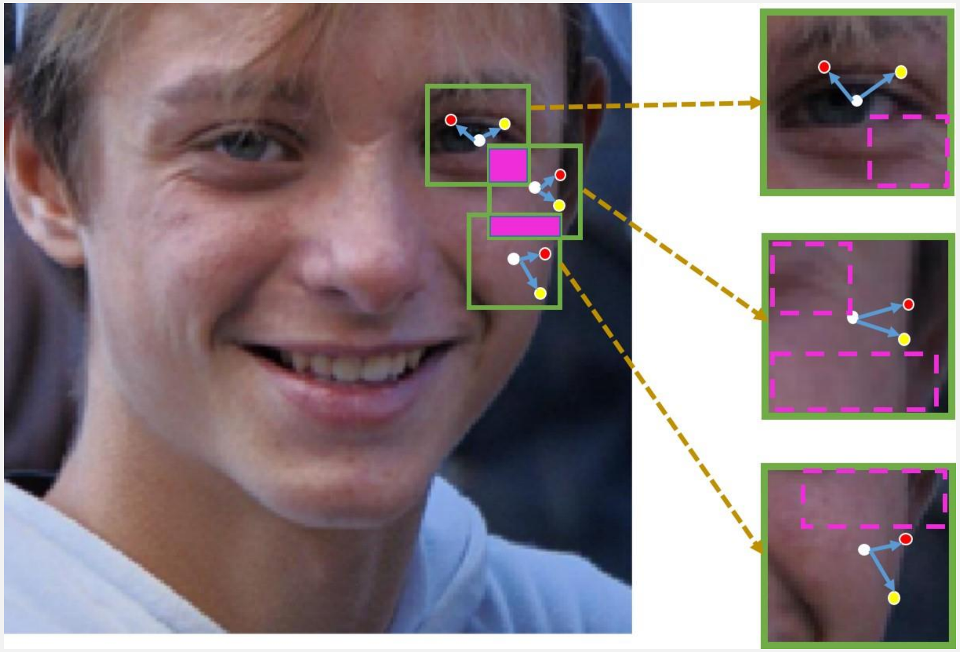
[3] Hao Liu *et al.*, "Label-Sensitive Deep Metric Learning for Facial Age Estimation," IEEE Transactions on **TIFS**, 2017.

[4] Hao Liu *et al.*, "Ordinal Deep Feature Learning for Facial Age Estimation," IEEE International Conference on Automatic Face and Gesture Recognition (**FG**), 2017.

TIP 2017：Deep Sharable and Structural Detectors for Face Alignment

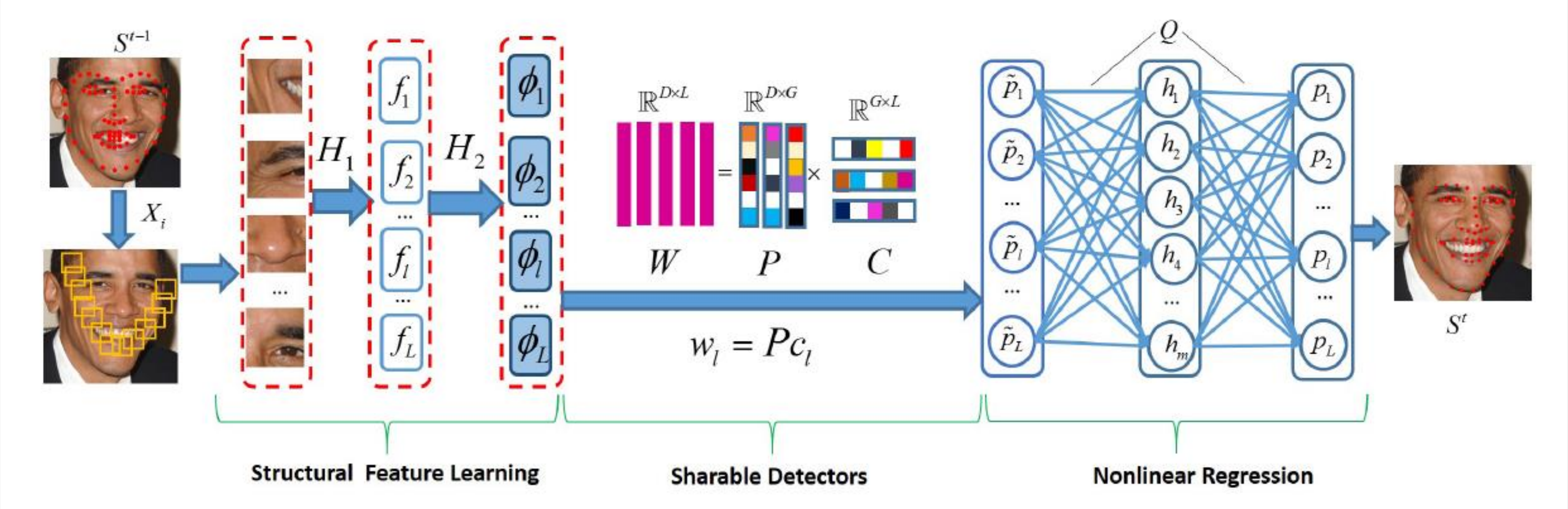
Deep Sharable and Structural Detectors for Face Alignment[1]

- **Motivation**
  - ✓ Facial landmarks are usually spatially correlated.
  - ✓ Conventional approaches utilizing hand-crafted features might lose shape-sensitive details.

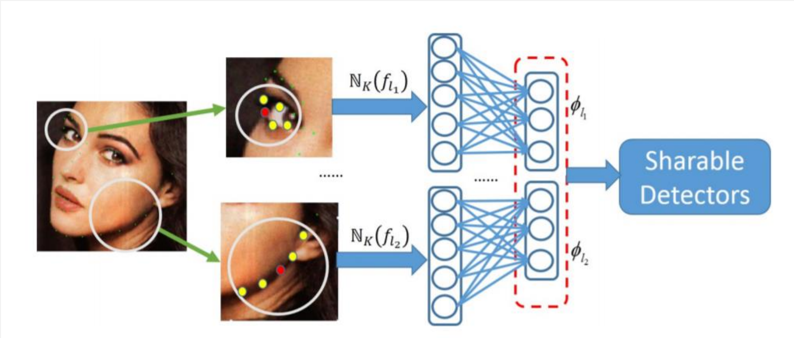


[1] Hao Liu, Jiwen Lu, Jianjiang Feng, Jie Zhou: Learning Deep Sharable and Structural Detectors for Face Alignment. IEEE Transactions on Image Processing (**TIP**) 26(4): 1666-1678 (2017).

Framework

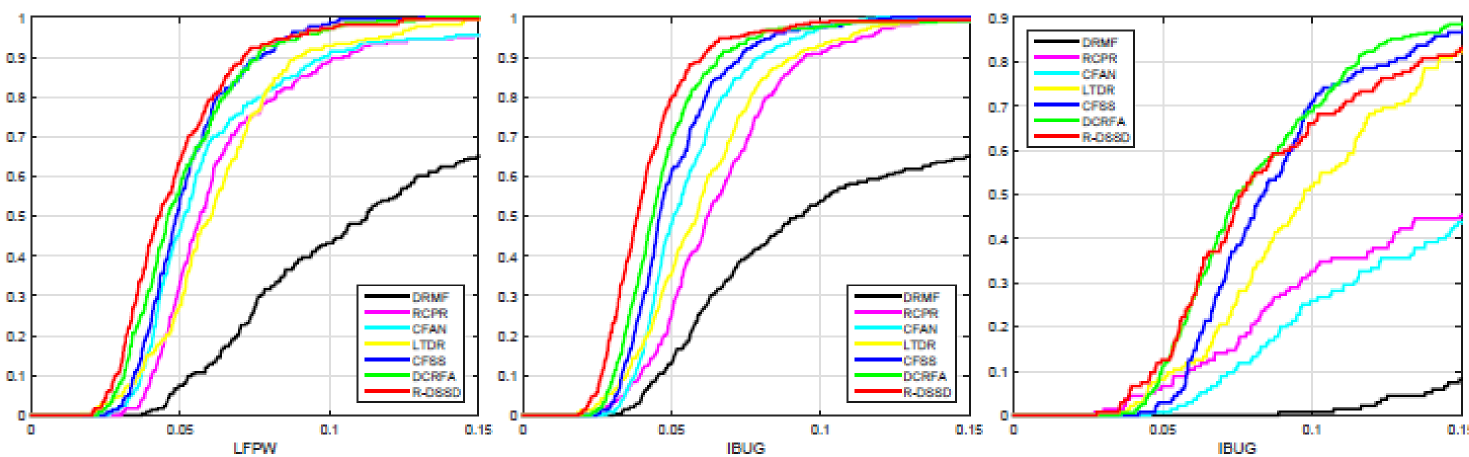


- ✓ **Structural Feature Learning**: model the correlation of neighbouring landmarks to dynamically cover more semantic details.
- ✓ **Sharable Detectors**: remove the noises of spatially overlapped patches.
- ✓ **Nonlinear Regression**: infer occluded part by non-occluded parts.



Experimental Results

Method	LFPW 68-pts	HELEN 68-pts	HELEN 192-pts	Common Set 68-pts	Challenging Set 68-pts	Full Set 68-pts
FPLL [28]	8.29	8.16	-	8.22	18.33	10.20
DRMF [24]	6.57	6.70	-	6.65	19.79	9.22
RCPR [23]	6.56	5.93	6.50	6.18	17.26	8.35
GN-DPM [25]	5.92	5.69	-	5.78	-	-
SDM [8]	5.67	5.50	5.85	5.57	15.40	7.50
CFAN [16]	5.44	5.53	-	5.50	-	-
ERT [11]	-	-	4.90	-	-	6.40
BPCPR [19]	-	-	-	5.24	16.56	7.46
ESR [9]	-	-	5.70	5.28	17.00	7.58
LBF [10]	-	-	5.41	4.95	11.98	6.32
LBF fast [10]	-	-	5.80	5.38	15.50	7.37
Deep Reg [15]	-	-	-	4.51	13.80	6.31
CFSS [12]	4.87	*4.63*	*4.74*	*4.73*	*9.98*	*5.76*
CFSS Practical [12]	4.90	4.72	4.84	4.73	10.92	5.99
TCDCN [47]	*4.57*	4.60	*4.63*	4.80	8.60	*5.54*
DCRFA [39]	*4.57*	*4.25*	-	*4.19*	**8.42**	**5.02**
R-DSSD*	4.77	4.31	4.95	4.57	10.86	5.91
R-DSSD	**4.52**	**4.08**	**4.62**	**4.16**	*9.20*	5.59



Annotated Results

Evaluation on 300-W， where 68 landmarks were employed

Annotated Results

Evaluation regarding with occlusion (68 landmarks):

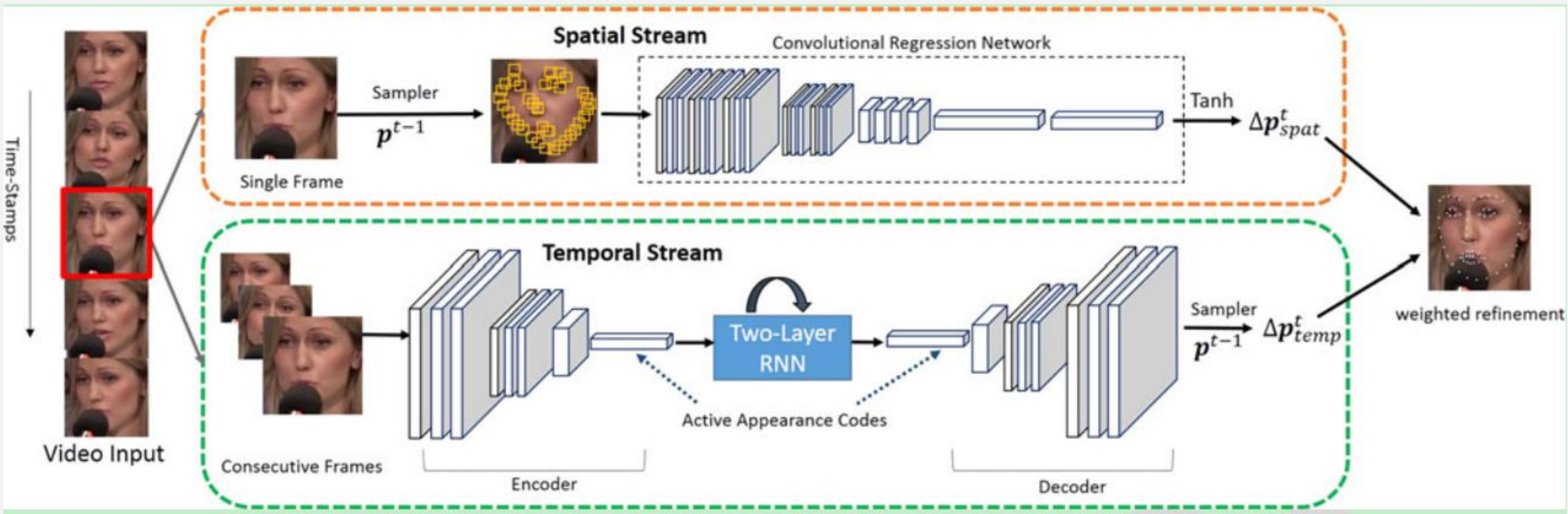
Evaluation regarding with denser landmarks (192 landmarks):

Running Speed： 30FPS@Intel-i5 CPU

TPAMI 2017：Two-Stream Transformer Networks for Video-based Face Alignment

Framework

- Two-Stream Transformer Networks
  - Spatial appearance information
  - Temporal consistency
  - Weighted Fusion



Results

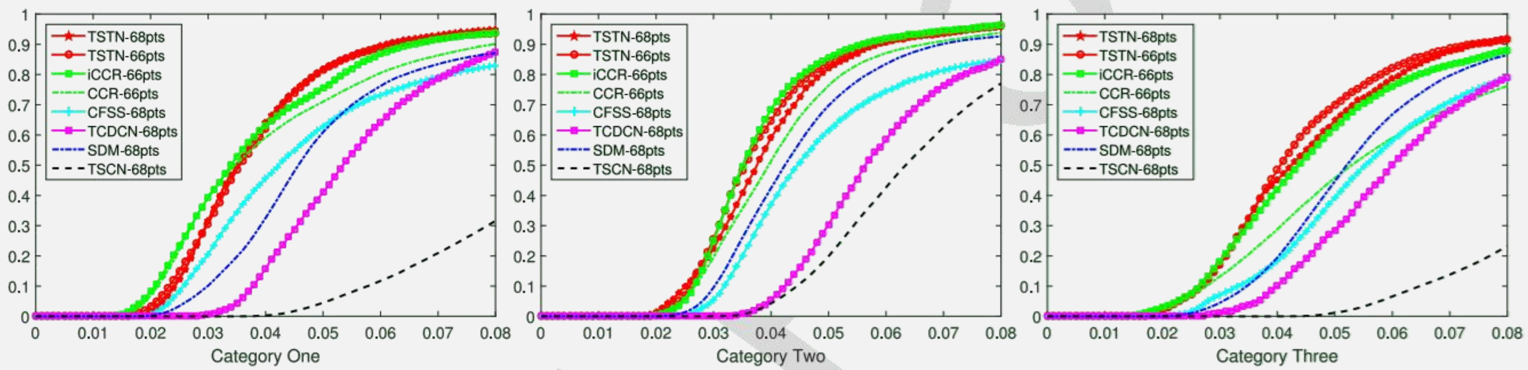


Fig. 3. CED curves of our TSTN compared to the state-of-the-arts on three categories in 300-VW [34] separately. In contrast to the state-of-the-art methods, our TSTN achieves comparable results in category two and superior performance in category one and the most difficult category three.

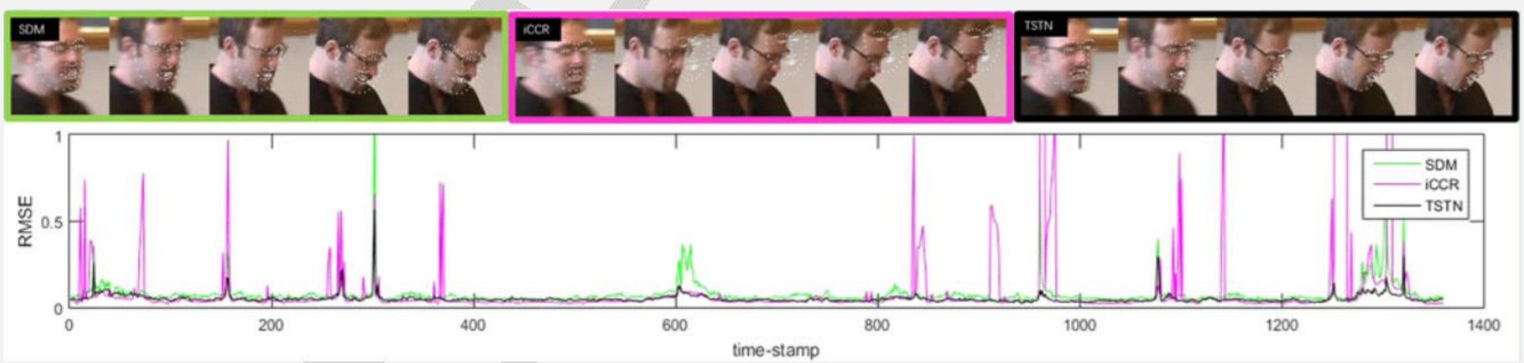


Fig. 4. Resulting examples of our TSTN on the 557th video clip in 300-VW [34] Category Three, where the selected tracked subject undergoes severe poses over time. The bottom subfigure shows that our TSTN exhibits robustness to difficult cases like large variations of facial aspect ratios.

Challenges

- Facial Age Estimation in the wild
  - ✓ Apparent Age Estimation Datasets [ICCVW 15]

5.0 20.07 76.82 55.66 39 21.38 26.23 27.26 20.14 34.76

  - ✓ Challenges
    - Large variances of facial expressions and occlusions
    - Appearance changes with facial makeup
    - Label correlation for human age labels

- Our Focuses
  - ✓ Exploiting *Label Correlation* for Ages
  - ✓ Learning *Robust Features*

State-of-the-arts

- Facial Age Estimation Framework

- Conventional Methods for Facial Age Estimation:
  - ✓ Feature Extraction (Requires strong prior knowledge)
    - Hand-crafted Features: BIF, LBP, SIFT
    - Shallow Feature Learning: CS-LBFL [Lu et al, TIP 2015]
  - ✓ Age Estimator (Should explicitly explore the ordinal relation for ages)
    - LDL [Geng et al,PAMI 2013]
    - OHRANK [Chen et al, CVPR 2012]

Facial Age Estimation x 2

Deep Feature Learning for Facial Age Estimation

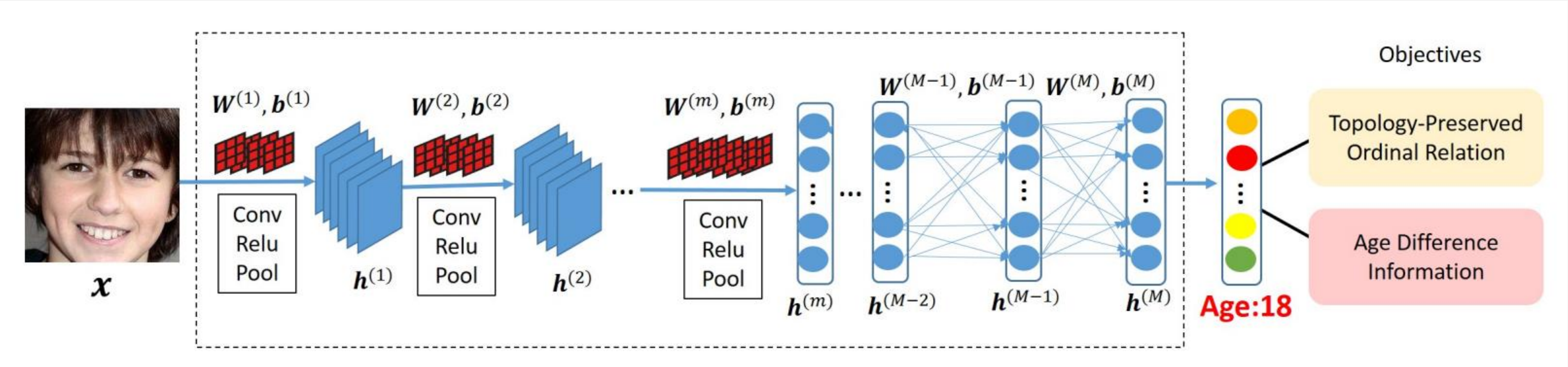
- Why Deep Feature Learning?
  - ✓ **Learning Features** directly from raw pixels
  - ✓ **Modeling Nonlinear Relationship** between Pixels and Labels
  - ✓ Transfer Learning (Fine-tuning)
- Facial Age Estimation
  - ✓ Label Correlation (FG 2017)
  - ✓ Missing Labels (PR 2017)

- Jointly learning feature descriptors for face representation and exploiting the relationship of human age labels

Ordinal Deep Feature Learning for Facial Age Estimation[1][2]

Motivation

- ✓ Deep Convolutional neural networks (CNN) works very well for face recognition.
- ✓ Human age labels are chronologically correlated and age estimation is an ordinal learning computer vision problem.

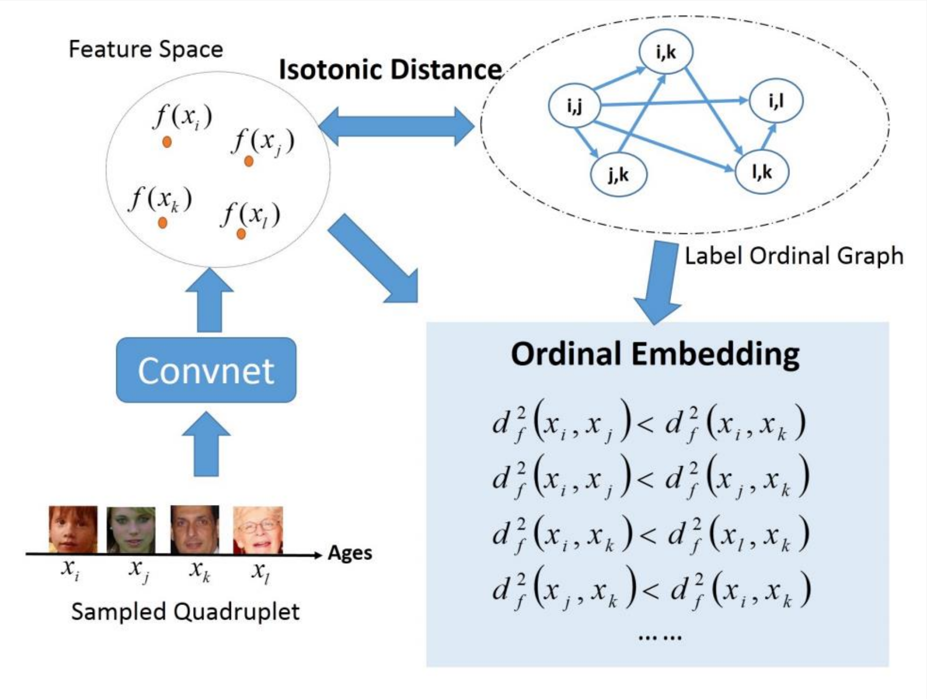


[1] H. Liu, J. Lu, J. Feng, J. Zhou. Ordinal Deep Feature Learning for Facial Age Estimation. In *IEEE Conference on Automatic Face and Gesture* (FG 2017).

[2] H. Liu, J. Lu, J. Feng, J. Zhou. Ordinal Deep Learning for Facial Age Estimation. Extension of [1] submitted to *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT), under review*.

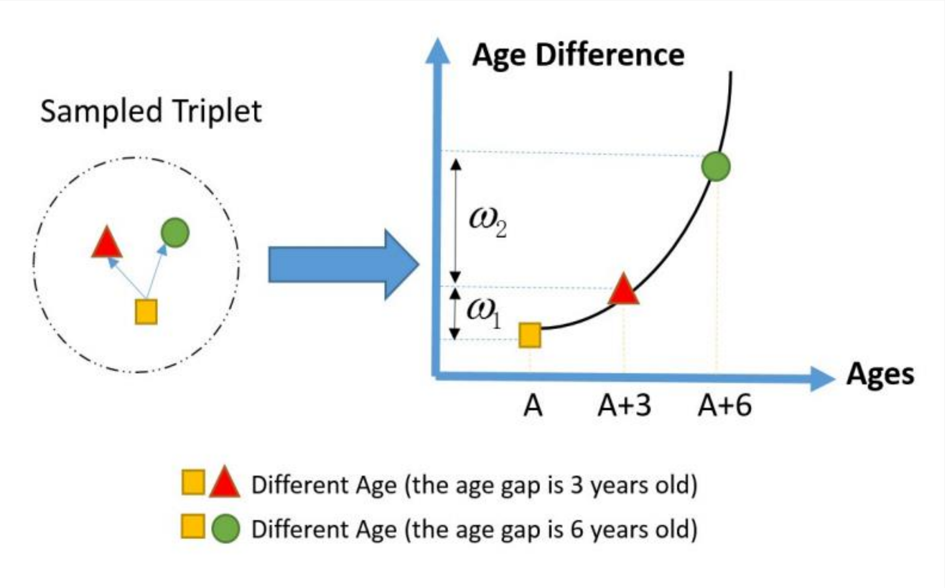
C1:Topology-Preserving Ordinal Relation

- Given a quadruplet of batched data, we construct a label ordinal graph based on the ordinal embeddings [1].
- The dissimilarity of face pairs in the learned feature space should be **isotonic** to that of the ordinal relations within the label ordinal graph.



[1] Kleindessner, M., von Luxburg, U.: Uniqueness of ordinal embedding. In: COLT. (2014) 40–67.

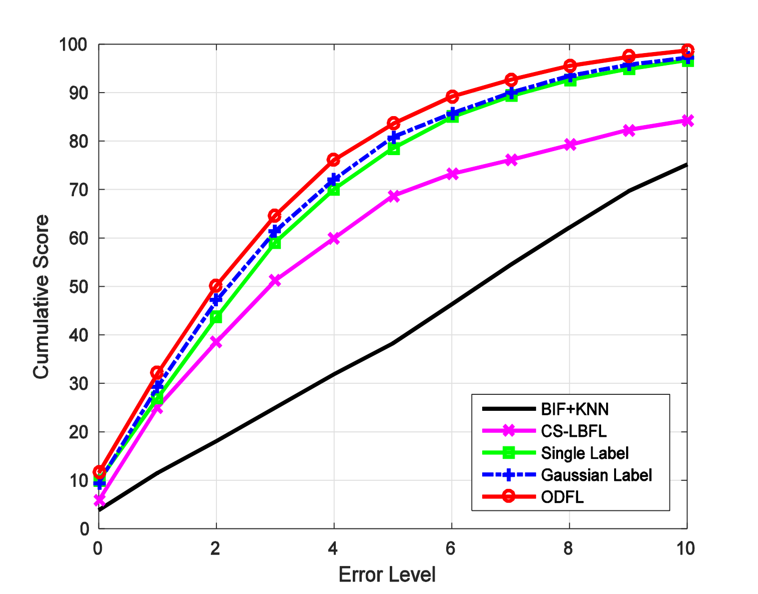
C2: Age Difference Information



- The distance of the face pair with a larger age gap should be smoothly bigger than that of the face pair with a smaller age gap.
- Weighting Function:

$$\omega_{y_{p1},y_{p2}} = \begin{cases} (|y_{p1} - y_{p2}| + 1)^{\eta}, & \text{if} \quad y_{p1} \neq y_{p2} \\ 1, & \text{otherwise} \end{cases}$$

Evaluation on Challenge Data

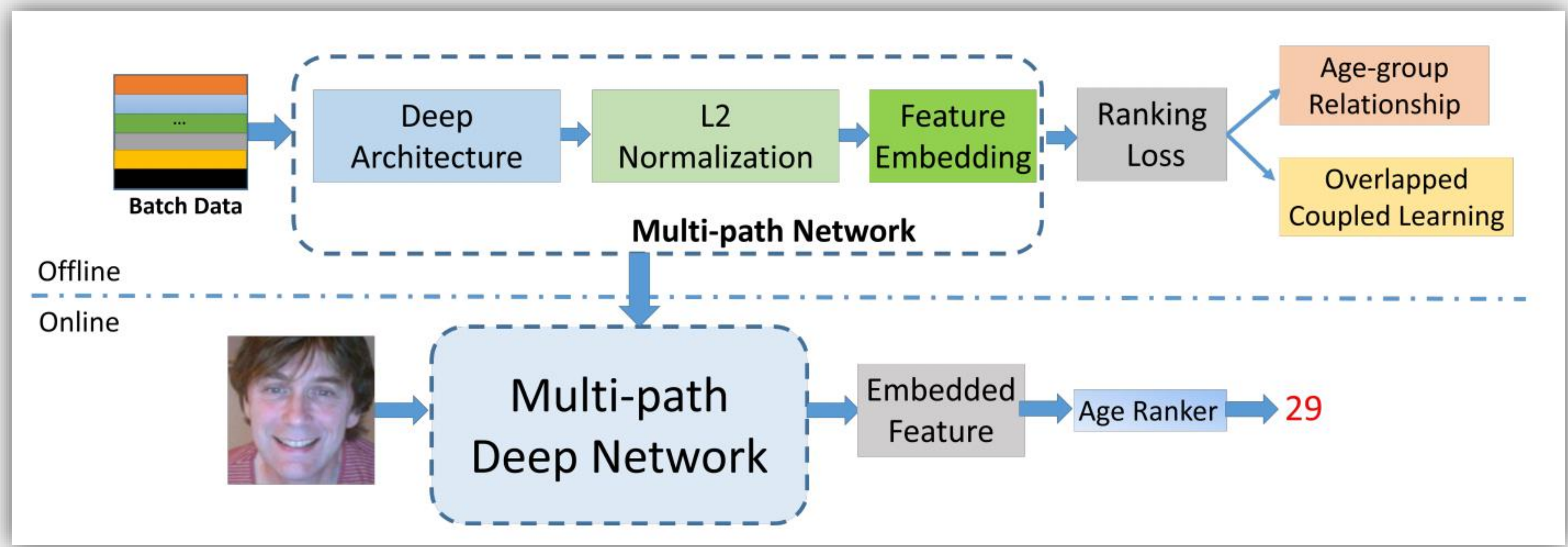


Method	MAE	Gaussian Error
BIF+KNN	7.19	0.620
CS-LBFL	5.12	0.422
Single Label	4.58	0.416
Gaussian Label	4.31	0.363
ODFL	**4.12**	**0.339**

Face resulting samples where the errors are less than one year old.

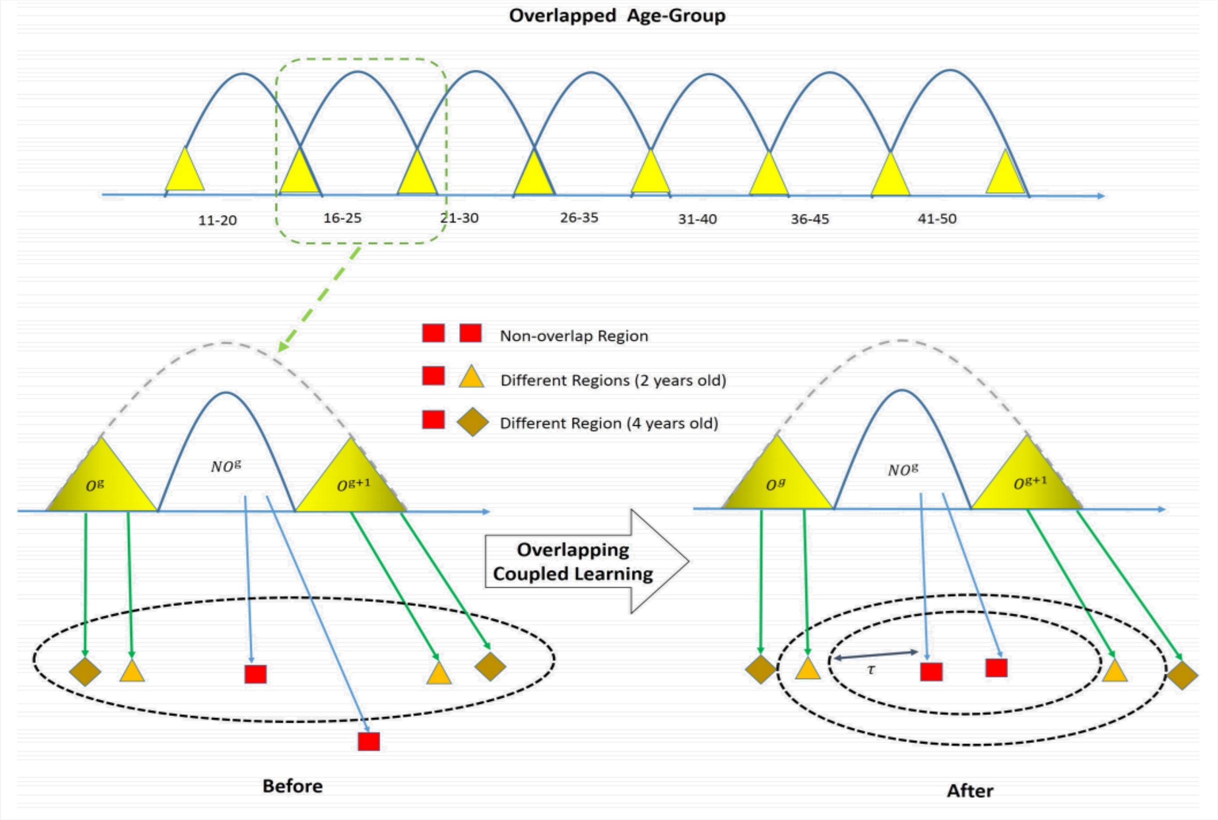
Group-Aware Deep Feature Learning for Facial Age Estimation[1]

- Motivation
  - ✓ Densely collecting face samples across a large range of age labels is difficult (unbalanced training samples across age classes)



[1] Hao Liu, Jiwen Lu, Jianjiang Feng, Jie Zhou: Group-aware deep feature learning for facial age estimation. Pattern Recognition 66: 82-94 (2017)

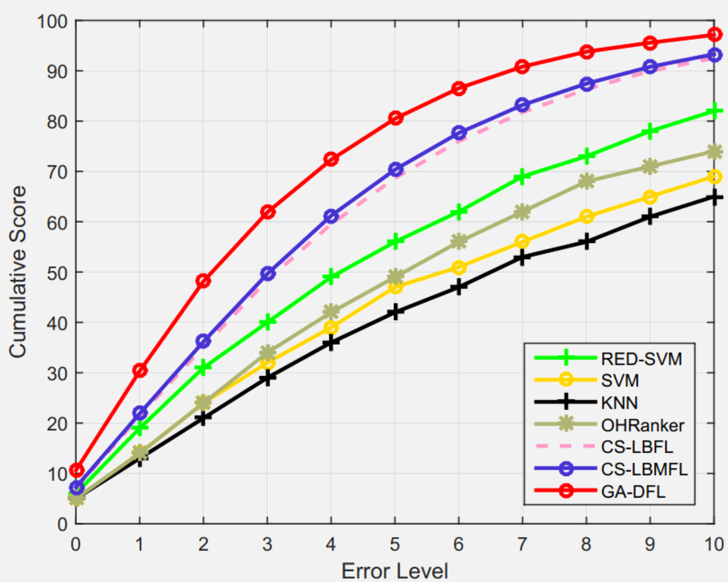
Basic Idea



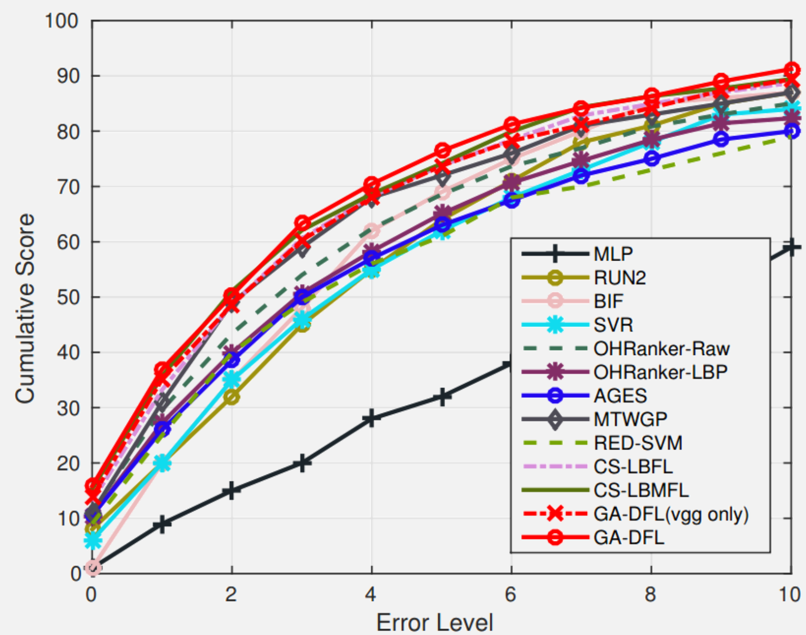
- Group-Aware Relationship: 1) inter-group variances are maximized; 2) intra-group variances are minimized.
- Smoothness of Overlaps: face samples within overlaps should be smoothly weighted according to the age differences.

Evaluation on MORPH and FG-NET

- MORPH:



- FG-NET:



THANK YOU!