MULTI-AGENT DEEP COLLABORATION LEARNING FOR FACE ALIGNMENT UNDER DIFFERENT PERSPECTIVES

Congcong Zhu, Suping Wu, Zhenhua Yu, Xing Wang, and **Hao Liu***

Ningxia University, China

https://haoliuphd.github.io/

Corresponding Author

Hao Liu

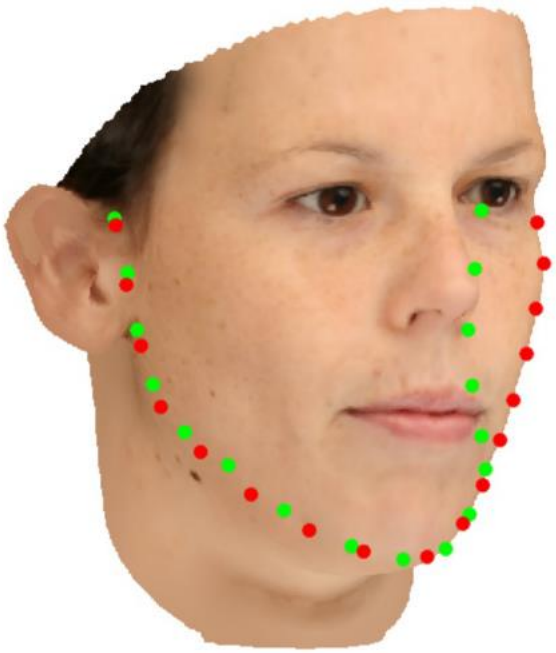
- Associate Professor, Ningxia University, China
- Ph.D supervised by Professor **Jie Zhou** and Professor **Jiwen Lu**
- Researches on facial analysis, particularly face alignment, facial age estimation and deep learning.

Selected Publications

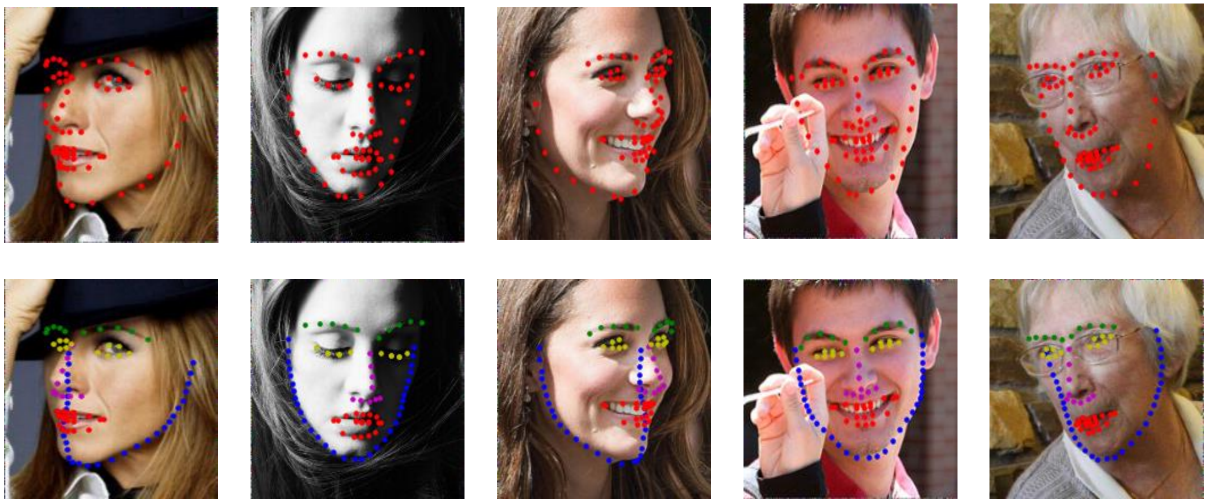
- **Hao Liu**, Jiwen Lu*, Minghao Guo, Suping Wu and Jie Zhou. Learning Reasoning-Decision Networks for Robust Face Alignment, **T-PAMI**, 2019
- **Hao Liu**, Jiwen Lu*, Jianjiang Feng and Jie Zhou. Two-Stream Transformer Networks for Video-based Face Alignment, **T-PAMI**, 2018
- **Hao Liu**, Jiwen Lu*, Jianjiang Feng and Jie Zhou. **T-IP** , 2017
- **Hao Liu**, Jiwen Lu*, Jianjiang Feng and Jie Zhou. Label-Sensitive Deep Metric Learning for Facial Age Estimation, **T-IFS**, 2018
- **Hao Liu**, Jiwen Lu*, Jianjiang Feng and Jie Zhou. Ordinal Deep Learning for Facial Age Estimation, **T-CSVT**, 2018

Landmark Ambiguity and Inconsistency

How to describe the semantic position under 2D and 3D views? [Deng *et al.* , TIP 2015]



Major Challenges

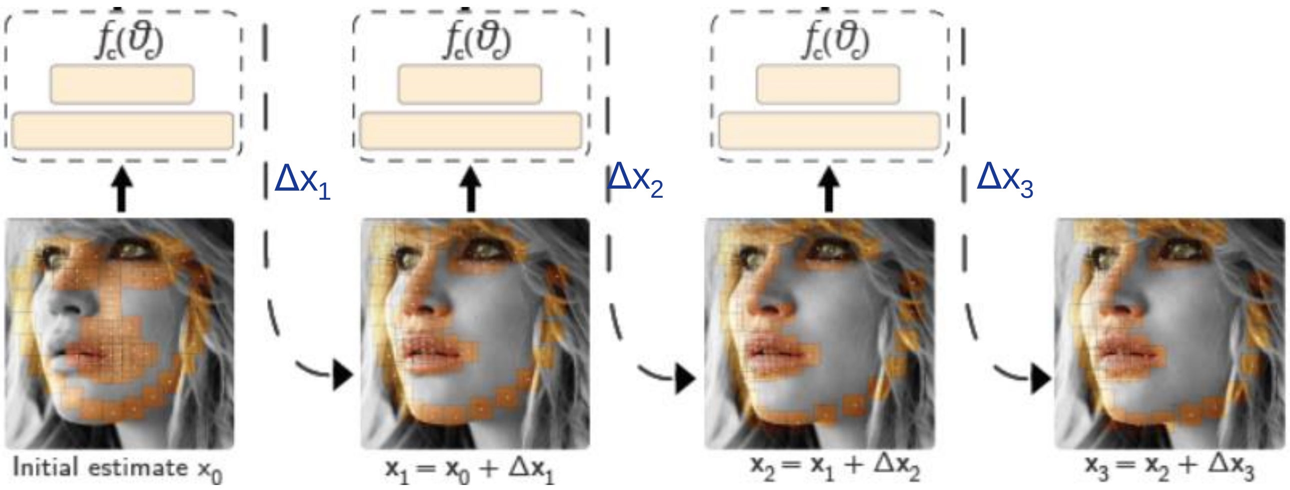


How to describe the facial landmarks more accurately under different perspectives?

- Large variances due to facial expressions and occlusions
- Semantic changes with different views and perspectives
- Ambiguity of different annotations

Conventional Methods

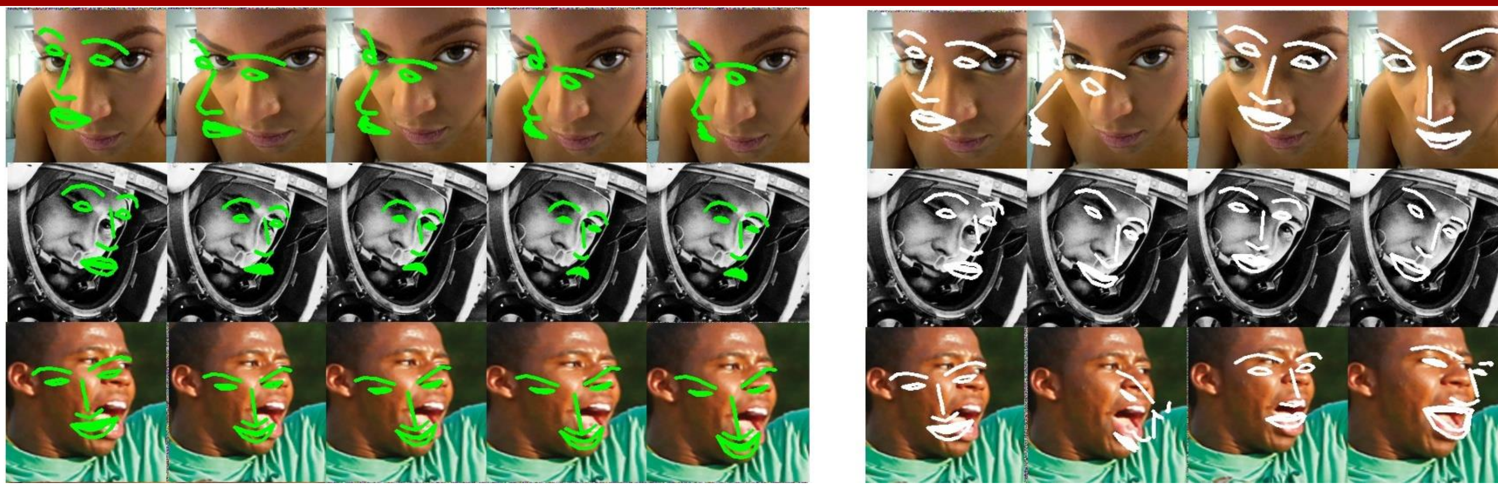
- Deep Regression Framework [CVPR 2016]

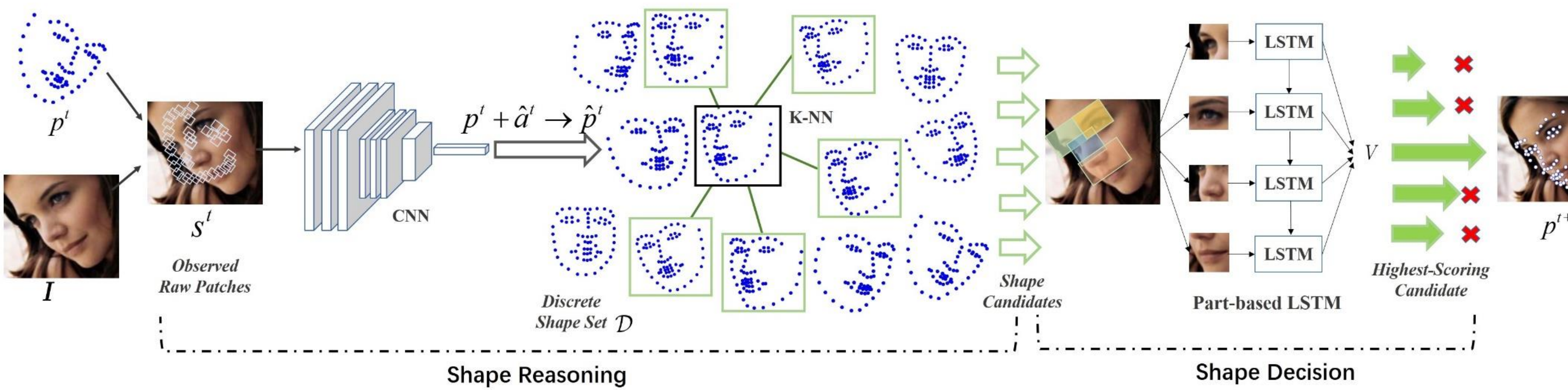


- Limitations
  - Initialization sensitivity
  - local optimization
  - pose sensitivity

T-PAMI 19: Reasoning-Decision Networks

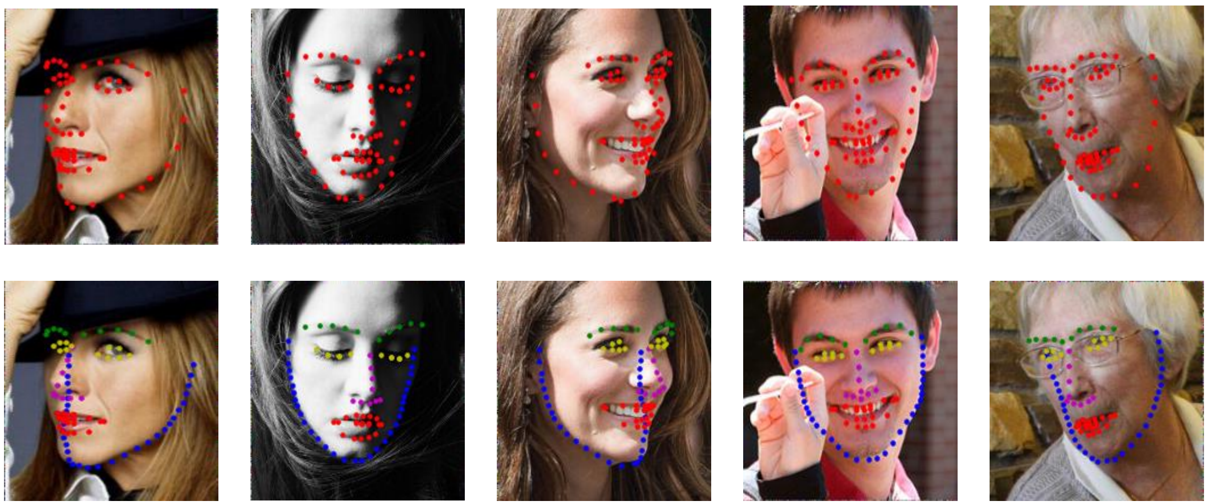
2019/9/24





Our Insight

- Shared information in different annotations.



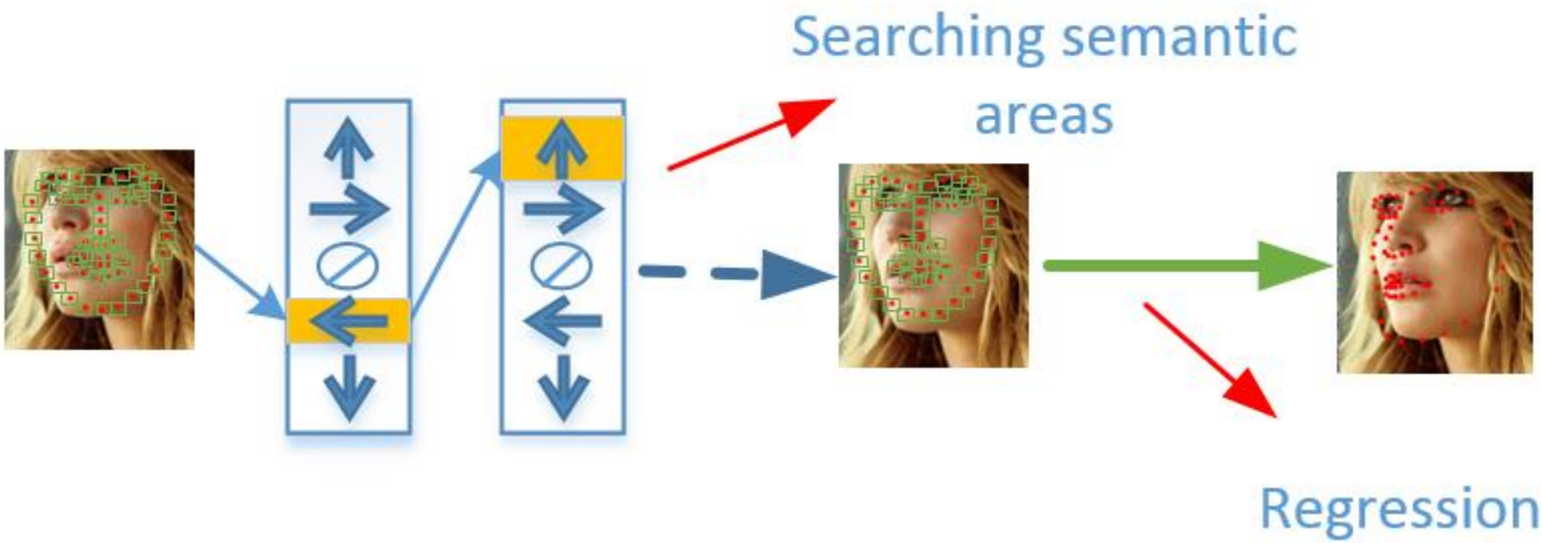
- Our method devolops collaboration learning and initialization adjustment policy to mine more semantic information, so that agents can better reason the undamaged facial shapes and the hidden self-occlusion points.

Contributions

- We model face alignment under different perspectives as a multi-task learning framework. Compared with conventional face alignment methods, we carefully design an initialization strategy based on the MDP.

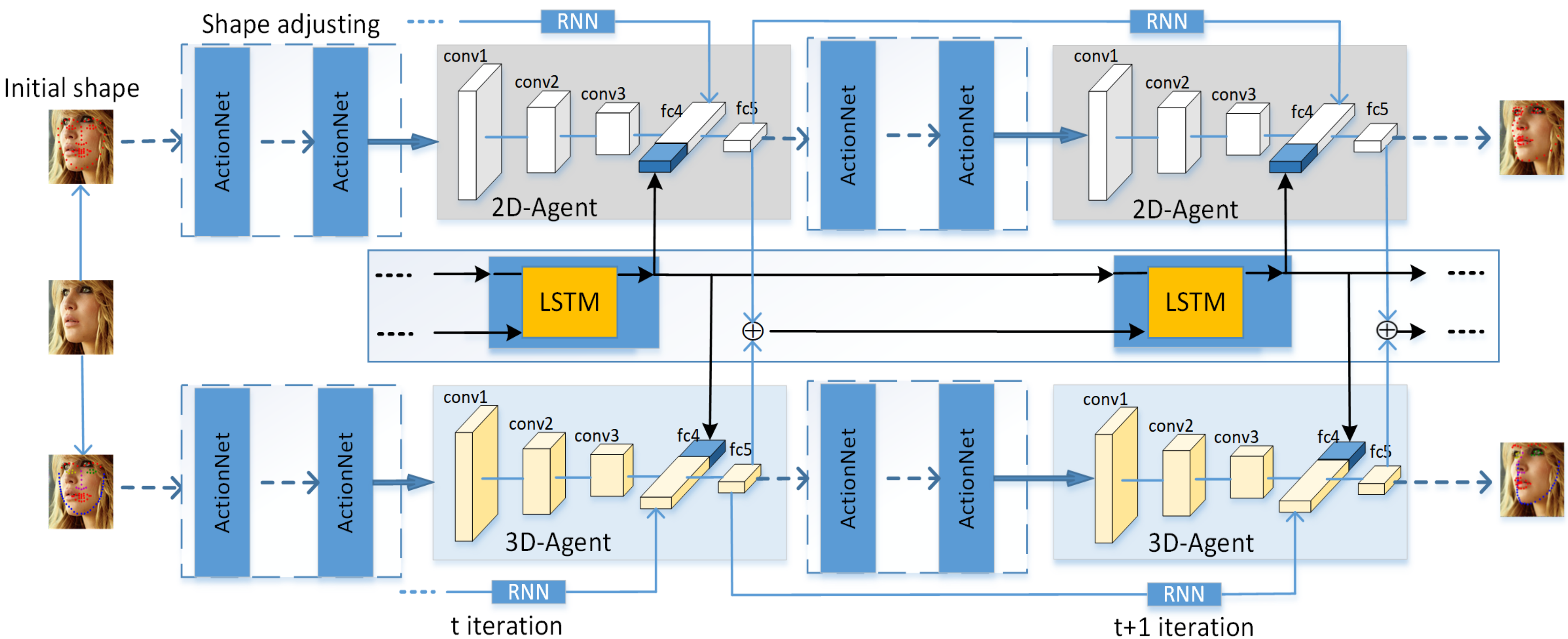
- Following the RDN, our initialization strategy learns a set of actions from the reward function to adjust the initial shape of every iteration to the reasonable location for robust cascade regression process.

Initialization Strategy



- **□** Our initialization strategy.
  - ➢ We defined five actions as the output of ActionNet: up, right, stop, left, down. The initial shape is adjusted to the better initialization position in a limited number of actions before each iteration.

Proposed Framework



Problem Formulation

Reward:

$$r^i = \begin{cases} e^i - e^{i+1}, & \text{if} \quad a \in \{up, down, left, right\}, \\ +\eta, & \text{if} \quad a = stop \quad \text{and} \quad e^0 - e^i \geq 0, \\ -\eta, & \text{otherwise} \end{cases}$$

Bellman equation:

$$Q(s, a) = r + \gamma \max Q(s', a'),$$

Objective Function

ActionNet:

$$L = \mathbb{E}\left[Q\left(s^{i}, a^{i}\right) - \left(r^{i} + \gamma \max Q\left(s^{i+1}, a^{i+1}\right)\right)\right]^{2}$$

AgentNet:

$$\min J = \sum_{t=1}^{T} \left\| \Delta P_t - \left(P^{*} - P_{t-1}^{I}\right) \right\|_2^2 ,$$

Experimental Results

Comparisons of averaged errors of our proposed MADCL with the state-of-the-arts on the 2D 300-W (68 landmarks).

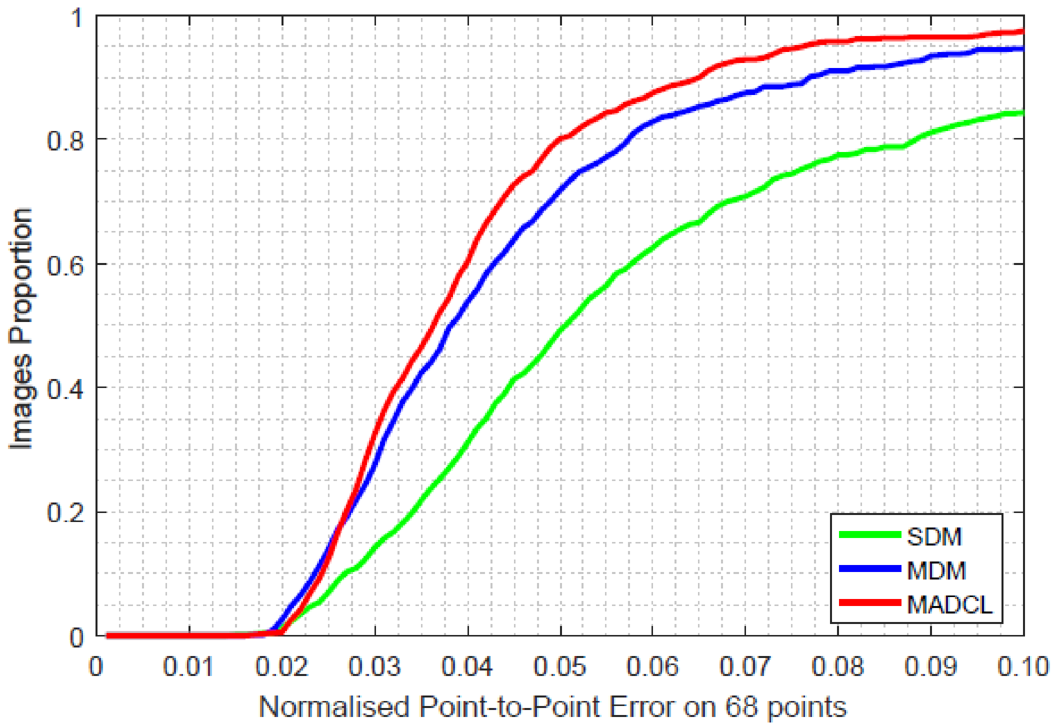
Methods	Challenging	Common	Full
SDM [4]	15.40	5.57	8.35
ESR [5]	17.00	5.28	7.58
LBF [23]	11.98	4.95	6.32
CFSS [6]	9.98	4.73	5.76
PIFA [14]	9.88	5.43	6.30
TCDCN [24]	8.60	4.80	5.54
3DDFA [2]	9.60	4.70	5.98
R-DSSD [25]	8.60	4.80	5.54
MDM [1]	8.87	3.74	4.78
TSR [11]	7.56	4.36	4.99
SBR [26]	8.14	3.39	4.36
MADCL(w/o CM)	6.89	3.53	4.19
MADCL	**6.75**	**3.46**	**4.11**

Experimental Results

Comparisons of averaged errors of our proposed MADCL with the state-of-the-arts on the 3D 300-W (84 landmarks).

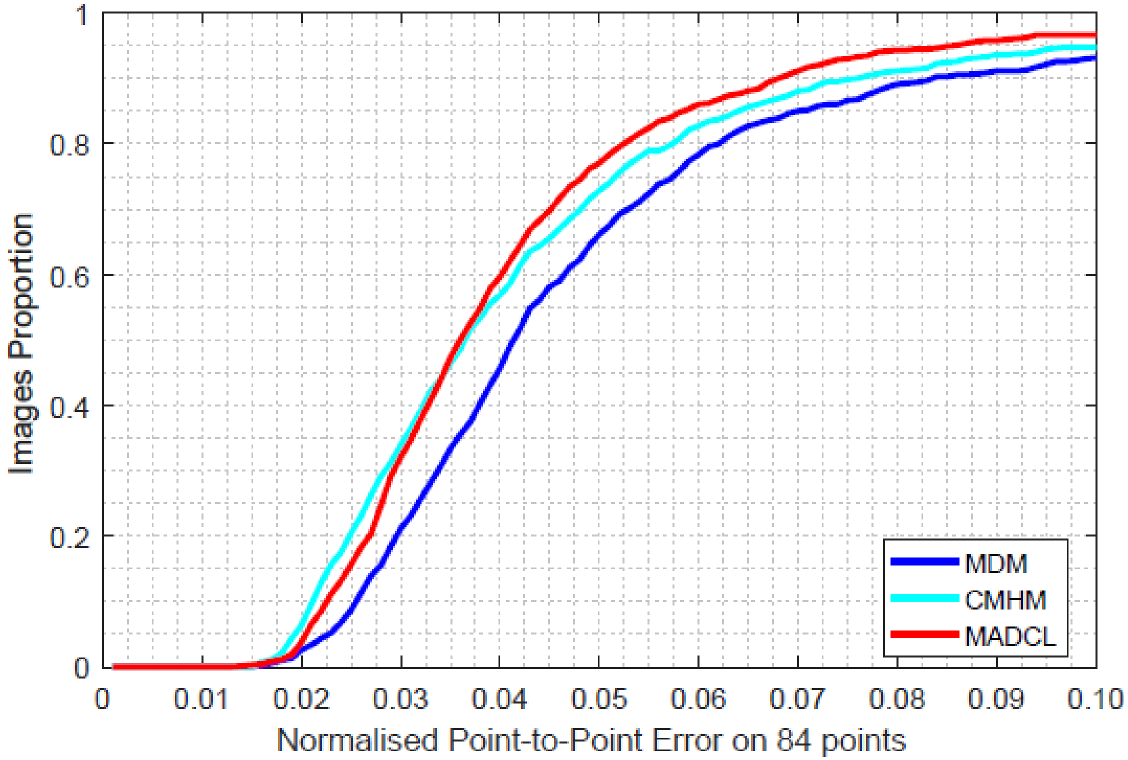
Methods	Challenging	Common	Full
CFSS [6]	11.64	5.61	6.79
PIFA [14]	10.41	5.66	6.59
3DDFA [2]	10.20	4.63	5.72
MDM [1]	9.30	4.02	5.13
MHCH [10]	8.39	3.94	4.81
MADCL(w/o CM)	7.31	3.62	4.34
MADCL	**7.14**	**3.55**	**4.25**

Experimental Results



CED curves of our MADCL compared to the state-ofthe- arts on the 2D 300-W fullset.

Experimental Results



CED curves of our MADCL compared to the state-ofthe-arts on the 3D 300-W fullset.

References

- **Hao Liu**, Jiwen Lu, Minghao Guo, Suping Wu, and Jie Zhou, "Learning reasoning-decision networks for robust face alignment", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018.
- **Hao Liu**, Jiwen Lu, Jianjiang Feng, and Jie Zhou, "Two-stream transformer networks for video-based face alignment", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018.
- **Hao Liu**, Jiwen Lu, Jianjiang Feng, and Jie Zhou, "Learning deep sharable and structural detectors for face alignment", IEEE Transactions on Image Processing, 2017.
- George Trigeorgis, Patrick Snape, Mihalis A Nicolaou, Epameinondas Antonakos, and Stefanos Zafeiriou, "Mnemonic Descent Method: A recurrent process applied for end-to-end face alignment", IEEE International Conference on Computer Vision and Pattern Recognition, 2016.