# FACIAL LANDMARKS DETECTION USING A CASCADE OF RECOMBINATOR NETWORKS

Pedro Diego López, Roberto Valle and Luis Baumela

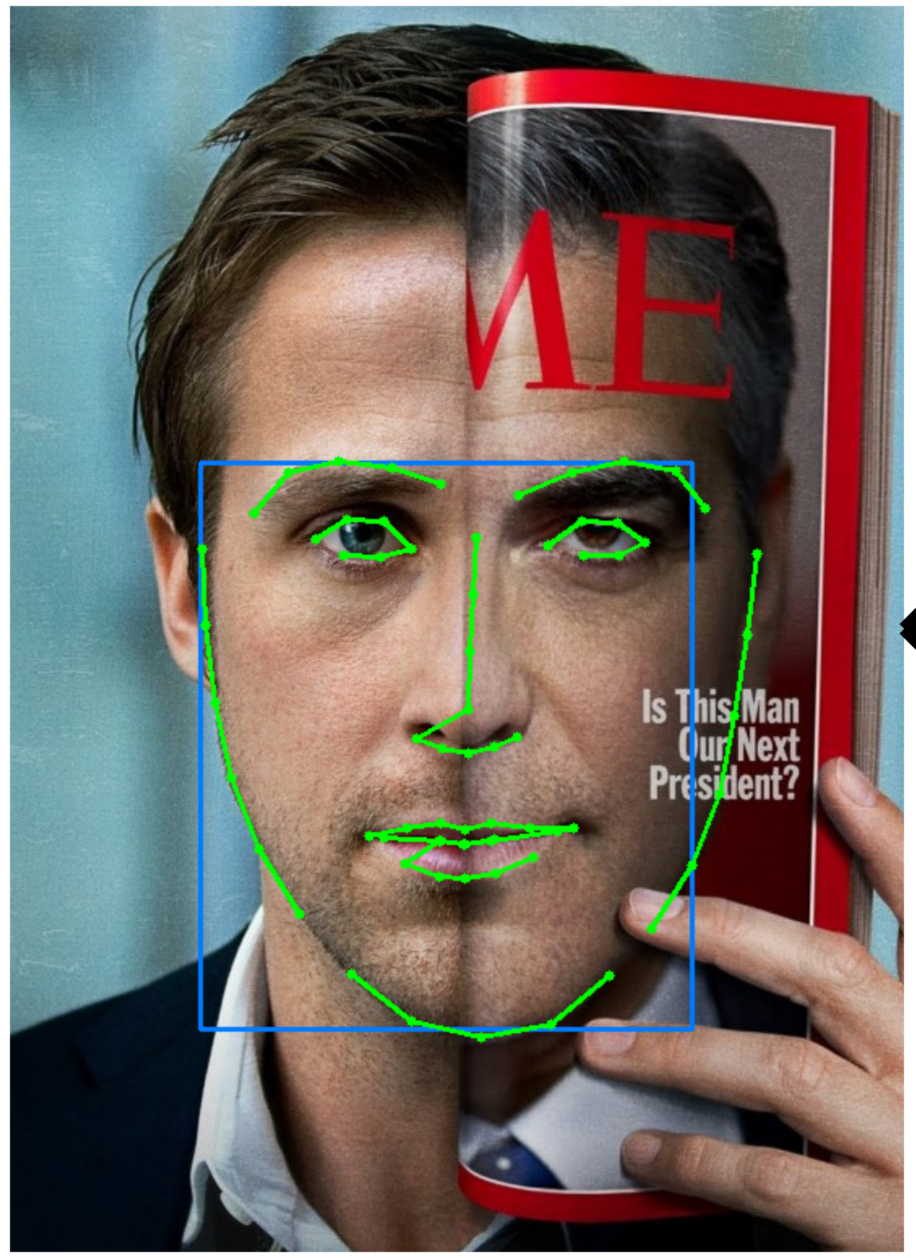
Univ. Politécnica de Madrid

http://www.dia.fi.upm.es/~pcr/research.html PCR

November 21, 2018

# 1. PROBLEM DEFINITION

- **Facial landmarks detection** is a crucial step for many face analysis problems.




Head-pose estimation: *yaw, pitch, roll*

Face recognition/verification: *Ryan Gosling*

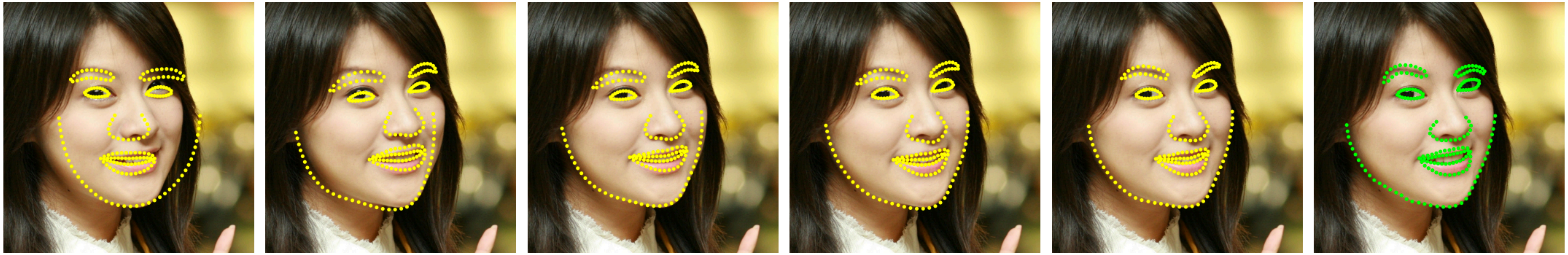
Facial expression recognition: *happy, surprise, …*

Attributes estimation: *gender, age, …*

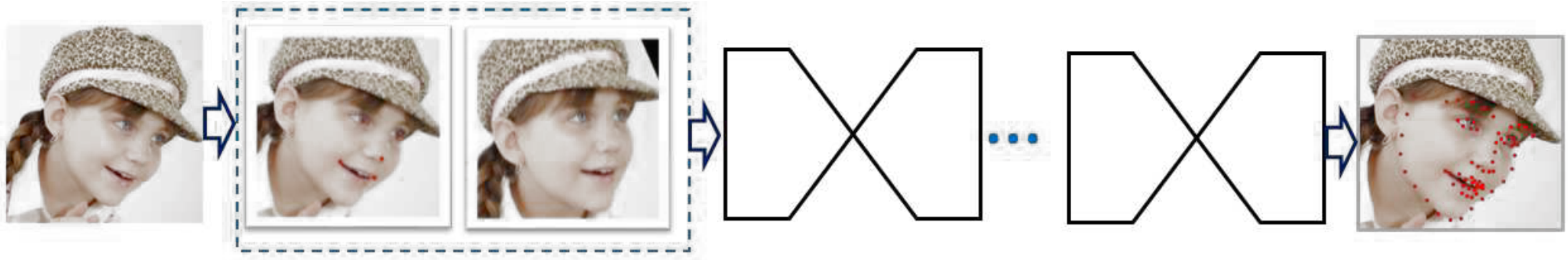
Face alignment/morphing/replacement

# 2. STATE-OF-THE-ART

- **ERT-based models** are easy to parallelize and implicitly impose shape consistency in their estimations (ERT [4], cGPRT [6]). Very sensitive to the starting point of the regression process.



# 2. STATE-OF-THE-ART



- Current state-of-the-art methods are based on **CNN methods**. DAN [5] and SHN [11] are among the top performers. Both use a VGG-based and a Stacked Hourglass network respectively to regress the final shape.

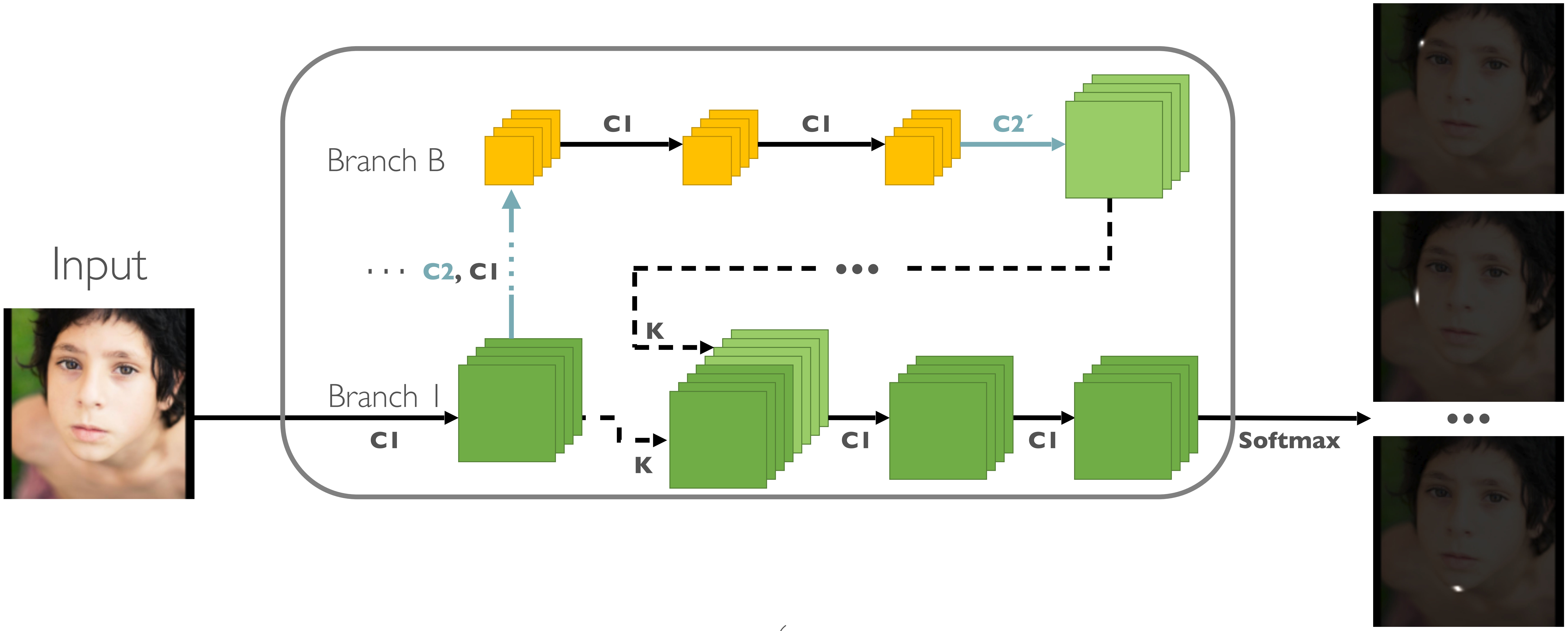
# 2. STATE-OF-THE-ART

- **Advantages**. CNN approaches are robust to face deformations and pose changes due to the large receptive fields of deep nets.

- **Disadvantages**. Loss of feature maps resolution and lack accuracy because of the difficulty in imposing a valid face shape.

# 3. CONTRIBUTIONS

- Maximum of probability maps determine the landmarks positions.



# 3. CONTRIBUTIONS

- CRN is composed of S stages where each stage represents a network that combines features across multiple branches B.

# 3. CONTRIBUTIONS

- We present a **loss function** that is able to handle missing landmarks.

$$\mathcal{L} = \sum_{i=1}^{N} \left( - \frac{1}{||\mathbf{w}_i^g||_1} \sum_{l=1}^{L} (\mathbf{w}_i^g(l) \cdot \mathbf{m}_i^g(l) \cdot \log(\mathbf{m}_i(l))) \right)$$

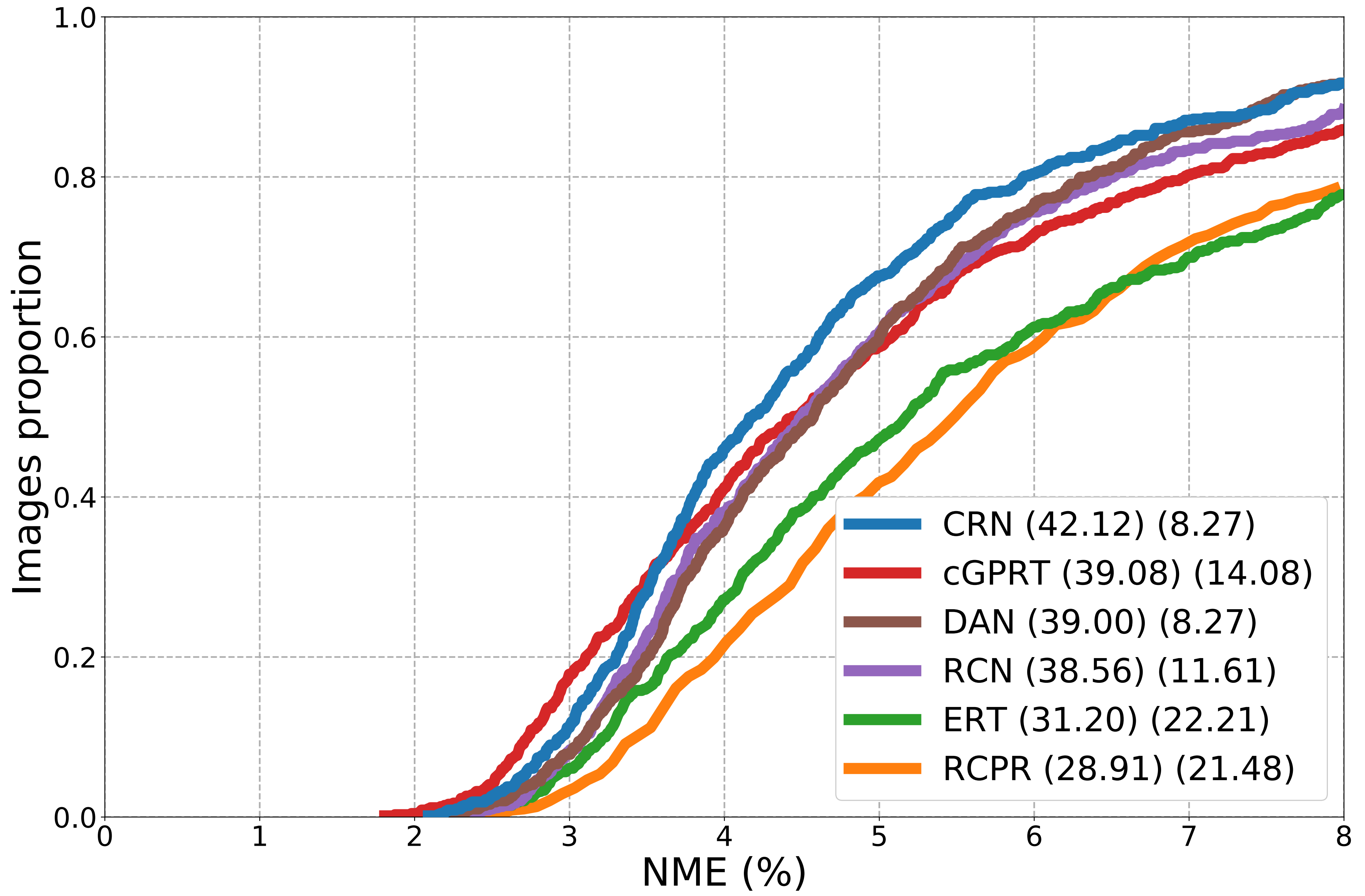
- Aggressive **data augmentation** with large face rotations, translations and scalings, labeling landmarks falling outside of the bounding box as missing.

# 4. EXPERIMENTS

- 300W public

| Method | Common pupils $NME$ | Common corners $NME$ | Challenging pupils $NME$ | Challenging corners $NME$ | Full pupils $NME$ | Full corners $NME$ | Full corners $AUC_8$ | Full corners $FR_8$ |
|---|---|---|---|---|---|---|---|---|
| RCN [3] | 4.70 | - | 9.00 | - | 5.54 | - | - | - |
| RCN+DKM [3] | 4.67 | - | 8.44 | - | 5.41 | - | - | - |
| DAN [5] | 4.42 | 3.19 | 7.57 | 5.24 | 5.03 | 3.59 | 55.33 | 1.16 |
| TSR [7] | 4.36 | - | 7.56 | - | 4.99 | - | - | - |
| RAR [10] | 4.12 | - | 8.35 | - | 4.94 | - | - | - |
| SHN [11] | 4.12 | - | 7.00 | 4.90 | - | - | - | - |
| **CRN ($S$=1)** | 4.26 | 3.07 | 8.69 | 6.01 | 5.09 | 3.62 | 55.62 | 2.75 |
| **CRN ($S$=2)** | 4.12 | 2.97 | 7.90 | 5.47 | 4.83 | 3.44 | 57.44 | 1.88 |

Table 1: Error of face alignment methods on the 300W public test set.

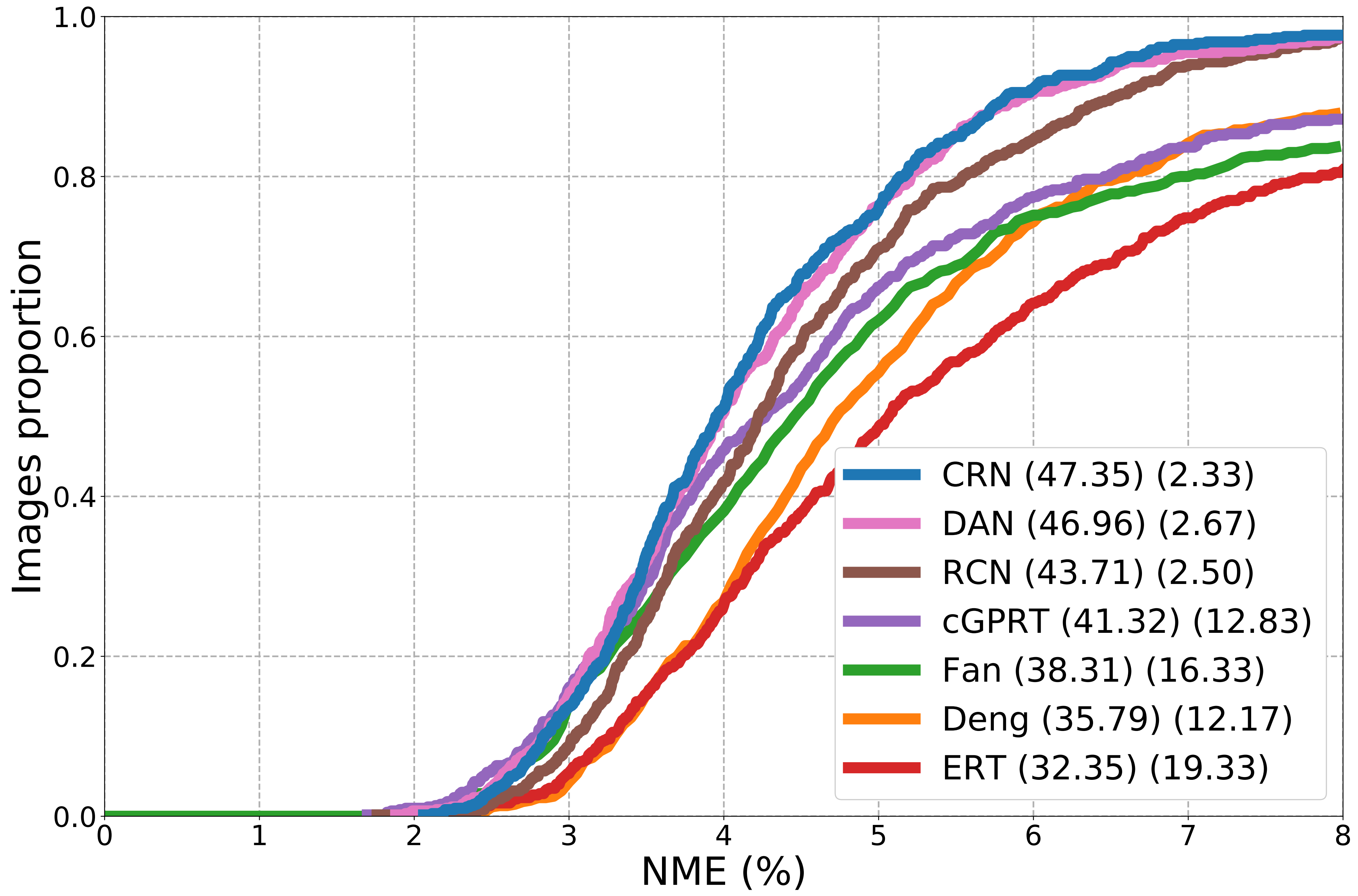


# 4. EXPERIMENTS

- 300W private

| Method | Indoor corners $NME$ | Indoor corners $AUC_8$ | Indoor corners $FR_8$ | Outdoor corners $NME$ | Outdoor corners $AUC_8$ | Outdoor corners $FR_8$ | Full corners $NME$ | Full corners $AUC_8$ | Full corners $FR_8$ |
|---|---|---|---|---|---|---|---|---|---|
| DAN [5] | - | - | - | - | - | - | 4.30 | 47.00 | 2.67 |
| SHN [11] | 4.10 | - | - | 4.00 | - | - | 4.05 | - | - |
| **CRN ($S$=1)** | 4.42 | 45.91 | 1.66 | 4.45 | 45.25 | 2.66 | 4.43 | 45.59 | 2.16 |
| **CRN ($S$=2)** | 4.28 | 47.36 | 2.66 | 4.25 | 47.32 | 2.00 | 4.26 | 47.35 | 2.33 |

Table 2: Error of face alignment methods on the 300W private test set.



# 4. EXPERIMENTS

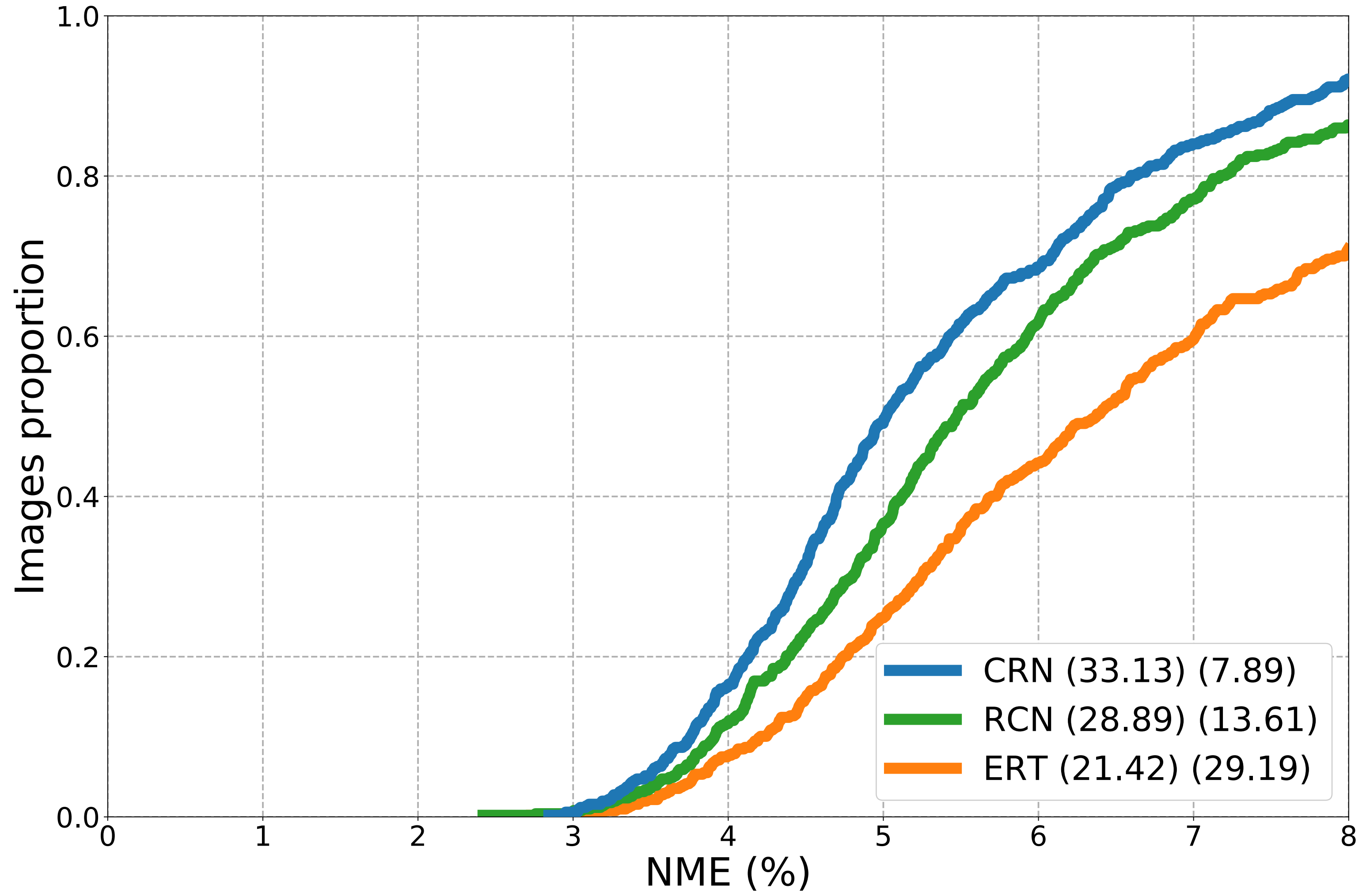
- COFW and AFLW

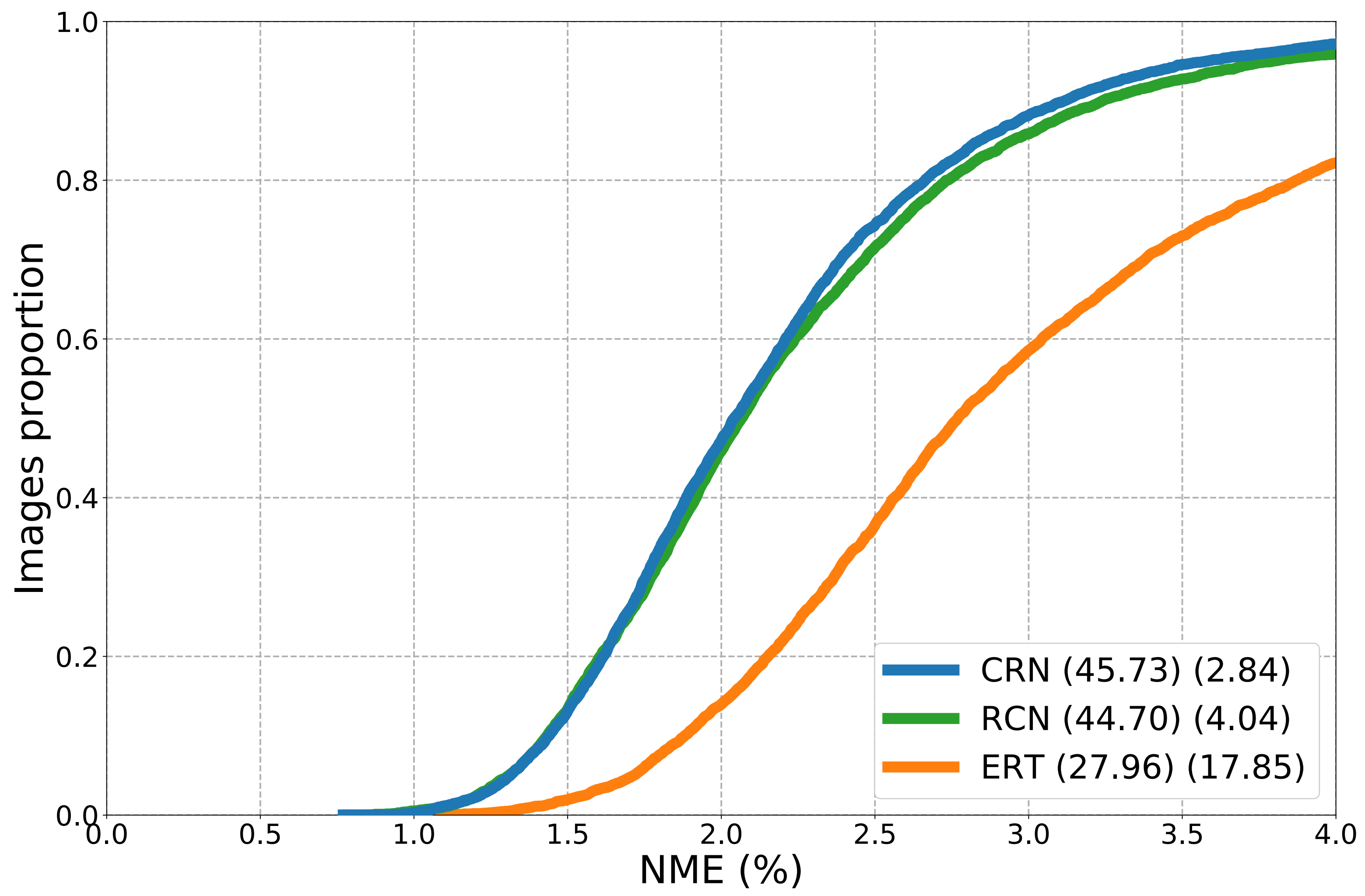
| Method | pupils | | |
|---|---|---|---|
| | $NME$ | $AUC_8$ | $FR_8$ |
| RAR [10] | 6.03 | - | - |
| Wu *et al.* [9] | 5.93 | - | - |
| SHN [11] | 5.6 | - | - |
| **CRN ($S$=1)** | 5.75 | 30.91 | 11.04 |
| **CRN ($S$=2)** | 5.49 | 33.13 | 7.88 |

Table 3: COFW results.

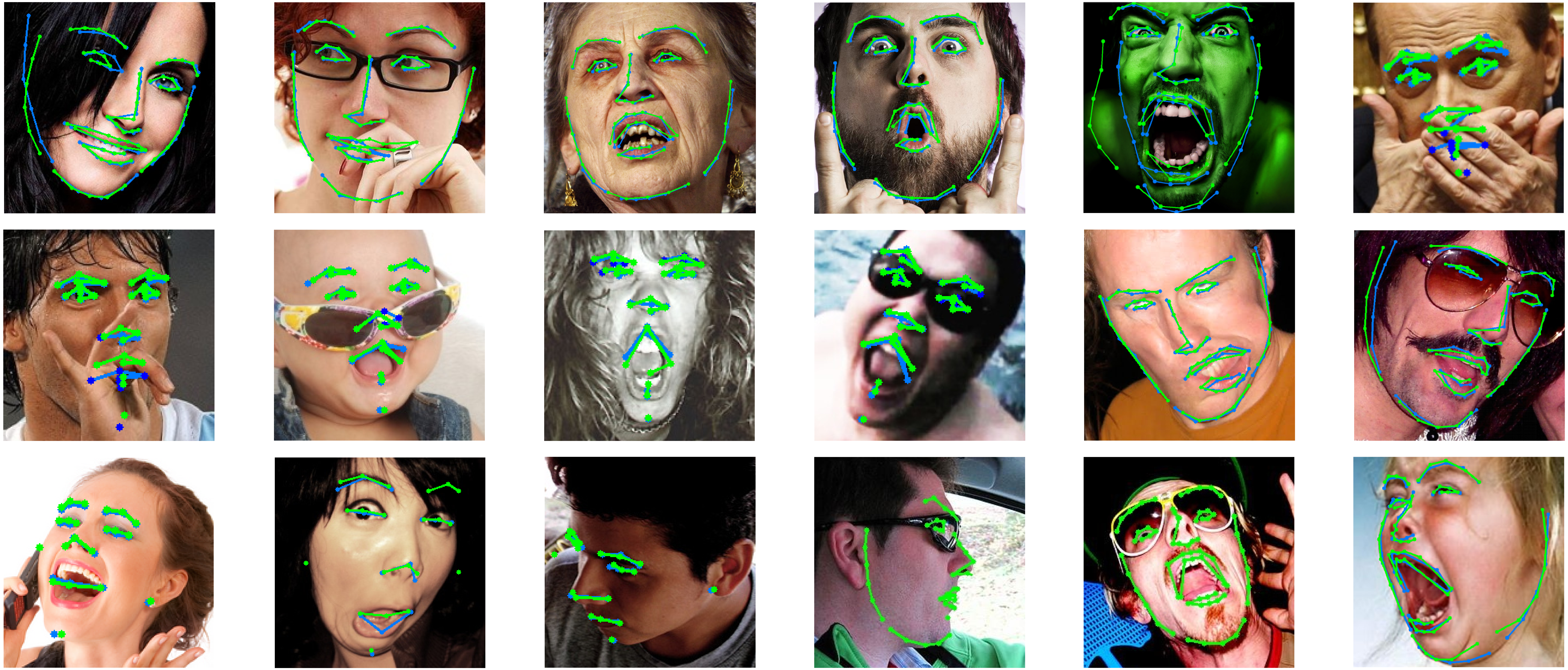
| Method | height $NME$ |
|---|---|
| Bulat *et al.* [2] | 2.85 |
| CCL [13] | 2.72 |
| TSR [7] | 2.17 |
| **CRN ($S$=1)** | 2.29 |
| **CRN ($S$=2)** | 2.21 |

Table 4: AFLW results.



# 5. RESULTS



# 6. CONCLUSIONS

- Our improvements to the RCN baseline together with the cascade approach and the data augmentation achieve state-of-the-art results in 300W, COFW and AFLW in-the-wild data sets.

- Is the facial landmarks detection problem solved? No.

- **Future work**. Multitask learning for estimate landmark visibilities and head-pose simultaneously.

# 7. REFERENCES

1. Amador, E., Valle, R., Buenaposada, J.M., Baumela, L.: Benchmarking head pose estimation in-the-wild. In: Proc. Iberoamerican Congress on Pattern Recognition (CIARP) (2017)
2. Bulat, A., Tzimiropoulos, G.: Binarized convolutional landmark localizers for human pose estimation and face alignment with limited resources. In: Proc. International Conference on Computer Vision (ICCV) (2017)
3. Honari, S., Yosinski, J., Vincent, P., Pal, C.J.: Recombinator networks: Learning coarse-to-fine feature aggregation. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
4. Kazemi, V., Sullivan, J.: One millisecond face alignment with an ensemble of regression trees. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2014)
5. Kowalski, M., Naruniec, J., Trzcinski, T.: Deep alignment network: A convolutional neural network for robust face alignment. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2017)
6. Lee, D., Park, H., Yoo, C.D.: Face alignment using cascade gaussian process regression trees. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015)
7. Lv, J., Shao, X., Xing, J., Cheng, C., Zhou, X.: A deep regression architecture with two-stage re-initialization for high performance facial landmark detection. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
8. Sun, Y., Wang, X., Tang, X.: Hybrid deep learning for face verification. IEEE Trans. Pattern Analysis and Machine Intelligence (TPAMI) 38, 1997–2009 (2016)
9. Wu, Y., Ji, Q.: Robust facial landmark detection under significant head poses and occlusion. In: Proc. International Conference on Computer Vision (ICCV) (2015)
10. Xiao, S., Feng, J., Xing, J., Lai, H., Yan, S., Kassim, A.A.: Robust facial landmark detection via recurrent attentive-refinement networks. In: Proc. European Conference on Computer Vision (ECCV) (2016)
11. Yang, J., Liu, Q., Zhang, K.: Stacked hourglass network for robust facial landmark localisation. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2017)
12. Zafeiriou, S., Trigeorgis, G., Chrysos, G., Deng, J., Shen, J.: The menpo facial landmark localisation challenge: A step towards the solution. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2017)
13. Zhu, S., Li, C., Change, C., Tang, X.: Unconstrained face alignment via cascaded compositional learning. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)

# 8. QUESTIONS