A Deeply-initialized Coarse-to-fine Ensemble of Regression Trees for Face Alignment

Roberto Valle[1], José M. Buenaposada[2], Antonio Valdés[3], Luis Baumela[1]

[1] Univ. Politécnica de Madrid, [2] Univ. Rey Juan Carlos, [3] Univ. Complutense de Madrid

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

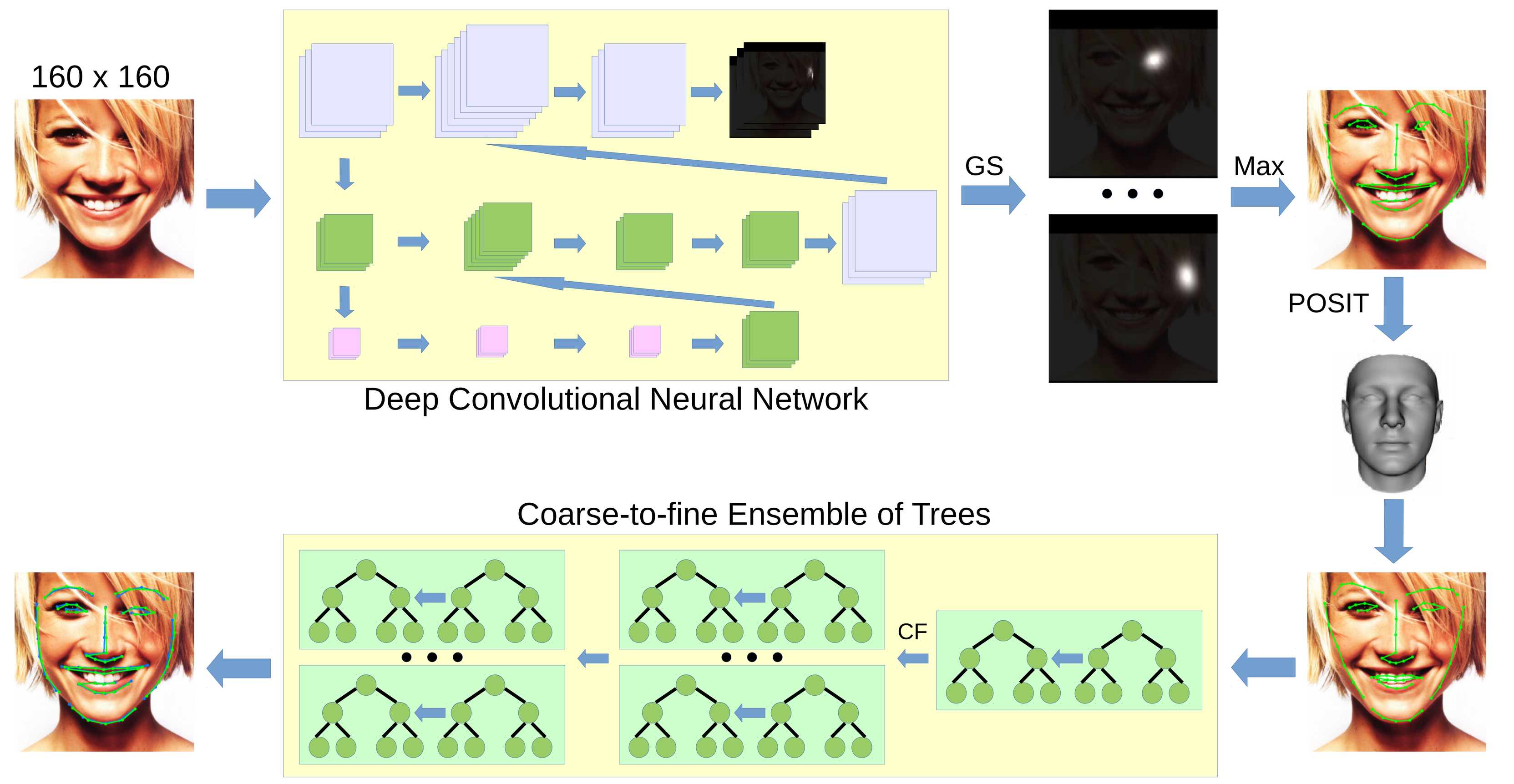
ECCV 2018

Problem Definition and Contribution

Facial landmarks detection is a crucial step for many face analysis problems such as verification, recognition, attributes estimation, etc.

Key contributions: We present DCFE, a robust method that combines the best of existing approaches.

- A CNN to obtain a set of probability maps without face shape enforcement.
- A 3D model to exploit rigid pose information.
- A properly initialized ERT to estimate non-rigid face deformation.



Experiments

300W public	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 [2]	4.67	-	8.44	-	5.41	-	-	-
DAN [3]	4.42	3.19	7.57	5.24	5.03	3.59	55.33	1.16
TSR [4]	4.36	-	7.56	-	4.99	-	-	-
RAR [7]	4.12	-	8.35	-	4.94	-	-	-
SHN [8]	4.12	-	7.00	4.90	-	-	-	-
DCFE	3.83	2.76	7.54	5.22	4.55	3.24	60.13	1.59

300W private	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
MDM [5]	-	-	-	-	-	-	5.05	45.32	6.80
DAN [3]	-	-	-	-	-	-	4.30	47.00	2.67
SHN [8]	4.10	-	-	4.00	-	-	4.05	-	-
DCFE	3.96	52.28	2.33	3.81	52.56	1.33	3.88	52.42	1.83

COFW	pupils NME	pupils AUC_8	pupils FR_8	occlusion precision/recall
DAC-CSR [1]	6.03	-	-	-
Wu *et al.* [6]	5.93	-	-	80/49.11
SHN [8]	5.6	-	-	-
DCFE	5.27	35.86	7.29	81.59/49.57

AFLW	height NME
CCL [9]	2.72
DAC-CSR [1]	2.27
TSR [4]	2.17
DCFE	2.17

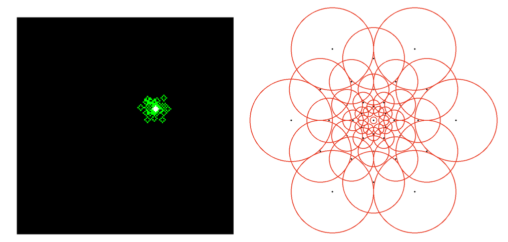
Algorithm

1. CNN to obtain probability maps: Obtain a set of probability maps, $\mathcal{P}(\mathtt{I})$, indicating the position of each landmark in the input image. The maximum of each smoothed probability map determines the 2D landmark locations.

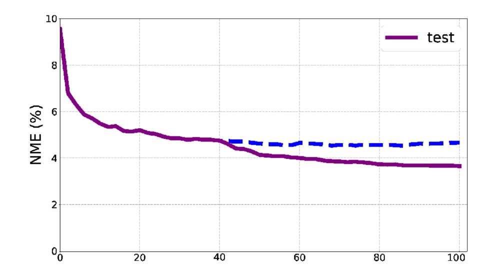
2. 3D face model: Compute the initial shape by fitting a rigid 3D head model to the estimated 2D landmarks locations. We project the 3D model onto the image using the rigid transformation estimated by the POSIT algorithm.

3. ERT for non-rigid shape estimation: ERTs are very efficient and precise when properly initialized. Let $\mathcal{S} = \{s_i\}_{i=1}^N$ be the set of train face shapes, where $s_i = (\mathtt{I}_i, \mathbf{x}_i^g, \mathbf{v}_i^g, \mathbf{w}_i^g, \mathbf{x}_i^0)$: training image, $\mathtt{I}_i$; ground truth shape, $\mathbf{x}_i^g$; ground truth visibility label, $\mathbf{v}_i^g$; annotated landmark label, $\mathbf{w}_i^g$ and initial shape for regression training, $\mathbf{x}_i^0$.

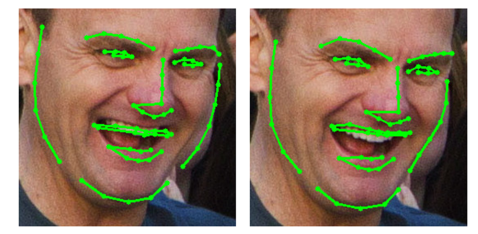
Input: Training data $\mathcal{S}, T$
  Generate augmented training samples set, $\mathcal{S}_A$
  for t=1 **to** T **do**
    Extract features for all samples, $\mathcal{F}_A = \{f_i\}_{i=1}^{N_A} = \{\phi(\mathcal{P}(\mathtt{I}_i), \mathbf{x}_i^{t-1}, \mathbf{w}_i^g)\}_{i=1}^{N_A}$
    Learn coarse-to-fine regressor, $\mathcal{C}_t^{\mathbf{v}}$, from $\mathcal{F}_A$ and $\mathcal{U}_{t-1} = \{(\mathbf{x}_i^{t-1}, \mathbf{v}_i^{t-1})\}_{i=1}^{N_A}$
    Update current shapes and visibilities, $\{(\mathbf{x}_i^t, \mathbf{v}_i^t) = (\mathbf{x}_i^{t-1}, \mathbf{v}_i^{t-1}) + \mathcal{C}_t^{\mathbf{v}}(f_i)\}_{i=1}^{N_A}$
  end for
Output: $\{\mathcal{C}_t^{\mathbf{v}}\}_{t=1}^T$



- **Feature extraction.** The feature is computed as the difference between two pixels values from a FREAK pattern around a random landmark and its associated probability map $\mathcal{P}(\mathtt{I})$.



- **Learn coarse-to-fine regressor.** A key problem is the lack of samples showing all possible combinations of face parts deformations. We introduce the coarse-to-fine ERT architecture to provide local improvements in difficult samples.



References

[1] Feng, Z., Kittler, J., Christmas, W.J., Huber, P., Wu, X.: Dynamic attention-controlled cascaded shape regression exploiting training data augmentation and fuzzy-set sample weighting. CVPR (2017)

[2] Honari, S., Yosinski, J., Vincent, P., Pal, C.J.: Recombinator networks: Learning coarse-to-fine feature aggregation. CVPR (2016)

[3] Kowalski, M., Naruniec, J., Trzcinski, T.: Deep alignment network: A convolutional neural network for robust face alignment. CVPRW (2017)

[4] 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. CVPR (2017)

[5] Trigeorgis, G., Snape, P., Nicolaou, M.A., Antonakos, E., Zafeiriou, S.: Mnemonic descent method: A recurrent process applied for end-to-end face alignment. CVPR (2016)

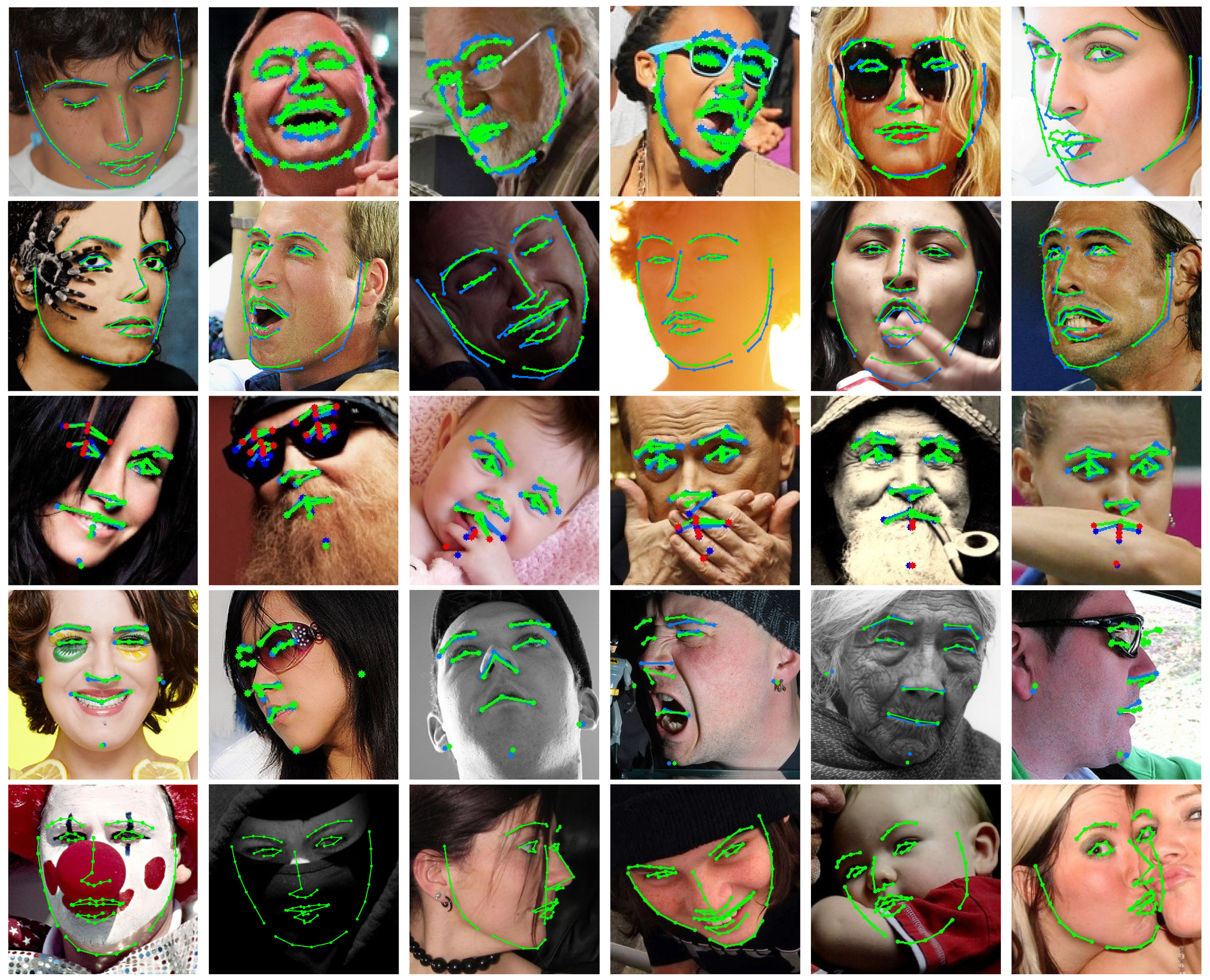
[6] Wu, Y., Ji, Q.: Robust facial landmark detection under significant head poses and occlusion. ICCV (2015)

[7] Xiao, S., Feng, J., Xing, J., Lai, H., Yan, S., Kassim, A.A.: Robust facial landmark detection via recurrent attentive-refinement networks. ECCV (2016)

[8] Yang, J., Liu, Q., Zhang, K.: Stacked hourglass network for robust facial landmark localisation. CVPRW (2017)

[9] Zhu, S., Li, C., Change, C., Tang, X.: Unconstrained face alignment via cascaded compositional learning. CVPR (2016)

Results



Acknowledgements: The authors gratefully acknowledge funding from the Spanish Ministry of Economy and Competitiveness under project TIN2016-75982-C2-2-R.