Supplementary material for Subpixel Heatmap Regression for Facial Landmark Localization

Adrian Bulat[1]
adrian@adrianbulat.com
Enrique Sanchez[1]
e.lozano@samsung.com
Georgios Tzimiropoulos[1,2]
g.tzimiropoulos@qmul.ac.uk

[1] Samsung AI Center
Cambridge, UK

[2] Queen Mary University London
London, UK

A Datasets

In this paper we conducton experiments on the following datasets:

300W: 300W [22] is a 2D face alignment dataset constructed by concatenating and then manually re-anotating with 68 points the images from LFPW [1], AFW [39], HELEN [17] and iBUG [23]. There are two commonly used train/test splits. Split I: uses 3837 images for training and 600 for testing and Split II that uses 3148 facial images for training and 689 for testing. The later testset comprises of two subsets: common and challenge. Most of the images present in the dataset contain faces found in frontal or near-frontal poses.

300W-LP: 300W-LP [40] is a synthetically generated dataset formed by warping into large poses the images from the 300W dataset. This dataset contains 61,125 pre-warped images and is used for training alone.

Menpo: Menpo [34] is a 2D face alignment dataset that annotates the images using 2 different configurations depending on the pose of the faces. The near frontal facial images are annotated using the same 68 points configuration used for 300W, while the rest using 39 points. In this work, we trained and evaluated on the 68-point configuration.

COFW: The Caltech Occluded Faces in the Wild (COFW) [3] dataset contains 1,345 training and 507 testing facial images captured in real world scenarios and annotated using 29 points. The images were later on re-annotated in [13] using the same 68-point configuration as in 300W.

AFLW: The Annotated Facial Landmarks in the Wild (AFLW) [14] dataset consists of 20,000 training images and 4386 testing images, out of which 1314 are part of the *Frontal* subset. All images are annotated using a 19 point configuration.

WFLW: Wider Facial Landmarks in-the-wild (WFLW) [29] consists of 10,000 images, out of which 7,500 are used for training while the rest are reserved for testing. All images are annotated using a 98 point configuration. In addition to landmarks, the dataset is also annotated with a set of attributes.

300VW: 300VW [24] is a large scale video face alignment dataset consisting of 218,594 frames distributed across 114 videos, out of which 50 are reserved for training while the rest for testing. The test set is further split into 3 different categories (A, B an C) with C being the most challenging one. We note that due to the semi-supervised annotation procedure some images have erroneous labels.

B Metrics

Depending on the dataset, the following metrics were used throughout this work:
Normalized Mean Error (NME). The point-to-point normalized Euclidean distance is the most widely used metric for evaluating the accuracy of a face alignment method and is defined as: $\text{NME}_{type}(\%) = \frac{1}{N}\sum_k^N \mathbf{v}_k \frac{\mathbf{y_k}-\hat{\mathbf{y}}_\mathbf{k}}{d_{type}} \times 100$, where $\mathbf{y}_k$ denotes the ground truth landmarks for the k-th face, $\hat{\mathbf{y}}_\mathbf{k}$ its corresponding predictions and d_{type} is the reference distance by which the points are normalized. $\mathbf{v}_k$ is a visibility binary vector, with values 1 at the landmarks where the ground truth is provided and 0 everywhere else.

Depending on the testing protocol, the NME *type* (*i.e.* how it's computed and defined) will vary. In this paper, we distinguish between the following types: d_{ic} – computed as the inter-occular distance [22], d_{box} – computed as the geometric mean of the ground truth bounding box [2] $d = \sqrt{(w_{bbox} \cdot h_{bbox})}$, and finally d_{diag} – defined as the diagonal of the bounding box.
Area Under the Curve(AUC): The AUC is computed by measuring the area under the curve up to a given user defined cut-off threshold of the cumulative error curve. Compared with NME that simple takes the average, this metric is less prone to outliers.
Failure Rate (FR): The failure rate is defined as the percentage of images the NME of which is bigger than a given (large) threshold.

C Additional comparisons with state-of-the-art

In addition to the comparisons reported in the main paper here in we show how our method performs against an additional set of methods (Tables 1, 2, 3, 3, 4).

References

[1] Peter N Belhumeur, David W Jacobs, David J Kriegman, and Neeraj Kumar. Localizing parts of faces using a consensus of exemplars. *TPAMI*, 2013.

[2] Adrian Bulat and Georgios Tzimiropoulos. How far are we from solving the 2d & 3d face alignment problem? (and a dataset of 230,000 3d facial landmarks). In *The IEEE International Conference on Computer Vision (ICCV)*, 2017.

[3] Xavier P Burgos-Artizzu, Pietro Perona, and Piotr Dollár. Robust face landmark estimation under occlusion. In *ICCV*, 2013.

[4] Xudong Cao, Yichen Wei, Fang Wen, and Jian Sun. Face alignment by explicit shape regression. *IJVC*, 2014.

Table 1: Comparison against the state-of-the-art on WFLW in terms of $NME_{inter-ocular}$, AUC_{ic}^{10} and FR_{ic}^{10}.

Method	NME_{ic}(%)	AUC_{ic}^{10}	FR_{ic}^{10} (%)
ESR [4]	11.13	0.277	35.24
SDM [33]	10.29	0.300	29.40
CFSS [37]	9.07	0.366	20.56
DVLN [31]	6.08	0.456	10.84
LAB (w/B) [30]	5.27	0.532	7.56
Wing [12]	5.11	0.554	6.00
MHHN [27]	4.77	-	-
DeCaFa [6]	4.62	0.563	4.84
AVS [20]	4.39	**0.591**	4.08
AWing [28]	4.36	0.572	**2.84**
LUVLi [16]	4.37	0.577	3.12
GCN [18]	**4.21**	0.589	3.04
Ours	**3.72**	**0.631**	**1.55**

Table 2: Comparison against the state-of-the-art on the AFLW-19 dataset.

Method	NME_{diag}		NME_{box}	AUC_{box}^{7}
	Full	Frontal	Full	Full
RCPR [3]	3.73	2.87	-	-
CFSS [37]	3.92	2.68	-	-
CCL [38]	2.72	2.17	-	-
DAC-CSR [11]	2.27	1.81	-	-
LLL [21]	1.97	-	-	-
CPM+SRB [9]	2.14	-	-	-
SAN [8]	1.91	1.85	4.04	54.0
DSNR [19]	1.85	1.62	-	-
LAB (w/o B) [30]	1.85	1.62	-	-
HR-Net [25]	1.57	1.46	-	-
Wing [12]	-	-	3.56	53.5
KDN [5]	-	-	2.80	60.3
LUVLi [16]	1.39	**1.19**	**2.28**	**68.0**
MHHN [27]	**1.38**	**1.19**	-	-
Ours	**1.31**	**1.12**	**2.14**	**70.0**

Table 3: Comparison against state-of-the-art on the 300W Common, Challenge and Full datasets (*i.e.* Split II).

Method	$NME_{inter-occular}$		
	Common	Challenge	Full
ODN [36]	3.56	6.67	4.17
CPM+SRB [9]	3.28	7.58	4.10
SAN [8]	3.34	6.60	3.98
AVS [20]	3.21	6.49	3.86
DAN [15]	3.19	5.24	3.59
LAB (w/B) [30]	2.98	5.19	3.49
Teacher [7]	2.91	5.91	3.49
DU-Net [26]	2.97	5.53	3.47
DeCaFa [6]	2.93	5.26	3.39
HR-Net [25]	2.87	5.15	3.32
HG-HSLE [41]	2.85	5.03	3.28
Awing [28]	**2.72**	**4.52**	**3.07**
LUVLi [16]	2.76	5.16	3.23
Ours	**2.61**	**4.13**	**2.94**

Table 4: Comparison against the state-of-the-art on the COFW-29 dataset.

Method	$NME_{ic}(\%)$	FR_{ic}^{10} (%)
Human	5.60	-
ESR [4]	11.20	36.0
RCPR [3]	8.50	20.00
HPM	7.59	13.00
CCR [10]	7.03	10.90
DRDA [35]	6.49	6.00
RAR [32]	6.03	4.14
DAC-CSR [11]	6.03	4.73
LAB (w/o B) [30]	5.58	2.76
Wing [12]	5.07	3.16
MHHN [27]	4.95	1.78
LAB (w/B) [30]	3.92	0.39
HR-Net [25]	**3.45**	**0.19**
Ours	**3.02**	**0.0**

[5] Lisha Chen, Hui Su, and Qiang Ji. Face alignment with kernel density deep neural network. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6992–7002, 2019.

[6] Arnaud Dapogny, Kevin Bailly, and Matthieu Cord. Decafa: deep convolutional cascade for face alignment in the wild. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6893–6901, 2019.

[7] Xuanyi Dong and Yi Yang. Teacher supervises students how to learn from partially labeled images for facial landmark detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 783–792, 2019.

[8] Xuanyi Dong, Yan Yan, Wanli Ouyang, and Yi Yang. Style aggregated network for facial landmark detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 379–388, 2018.

[9] Xuanyi Dong, Shoou-I Yu, Xinshuo Weng, Shih-En Wei, Yi Yang, and Yaser Sheikh. Supervision-by-registration: An unsupervised approach to improve the precision of facial landmark detectors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 360–368, 2018.

[10] Zhen-Hua Feng, Patrik Huber, Josef Kittler, William Christmas, and Xiao-Jun Wu. Random cascaded-regression copse for robust facial landmark detection. *IEEE Signal Processing Letters*, 22(1):76–80, 2014.

[11] Zhen-Hua Feng, Josef Kittler, William Christmas, Patrik Huber, and Xiao-Jun Wu. Dynamic attention-controlled cascaded shape regression exploiting training data augmentation and fuzzy-set sample weighting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2481–2490, 2017.

[12] Zhen-Hua Feng, Josef Kittler, Muhammad Awais, Patrik Huber, and Xiao-Jun Wu. Wing loss for robust facial landmark localisation with convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2235–2245, 2018.

[13] Golnaz Ghiasi and Charless C Fowlkes. Occlusion coherence: Detecting and localizing occluded faces. In *CVPR*, 2014.

[14] Martin Köstinger, Paul Wohlhart, Peter M Roth, and Horst Bischof. Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization. In *ICCV-W*, 2011.

[15] Marek Kowalski, Jacek Naruniec, and Tomasz Trzcinski. Deep alignment network: A convolutional neural network for robust face alignment. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 88–97, 2017.

[16] Abhinav Kumar, Tim K Marks, Wenxuan Mou, Ye Wang, Michael Jones, Anoop Cherian, Toshiaki Koike-Akino, Xiaoming Liu, and Chen Feng. Luvli face alignment: Estimating landmarks' location, uncertainty, and visibility likelihood. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8236–8246, 2020.

[17] Vuong Le, Jonathan Brandt, Zhe Lin, Lubomir Bourdev, and Thomas S Huang. Interactive facial feature localization. In *ECCV*, 2012.

[18] Weijian Li, Yuhang Lu, Kang Zheng, Haofu Liao, Chihung Lin, Jiebo Luo, Chi-Tung Cheng, Jing Xiao, Le Lu, Chang-Fu Kuo, et al. Structured landmark detection via topology-adapting deep graph learning. *arXiv preprint arXiv:2004.08190*, 2020.

[19] Xin Miao, Xiantong Zhen, Xianglong Liu, Cheng Deng, Vassilis Athitsos, and Heng Huang. Direct shape regression networks for end-to-end face alignment. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5040–5049, 2018.

[20] Shengju Qian, Keqiang Sun, Wayne Wu, Chen Qian, and Jiaya Jia. Aggregation via separation: Boosting facial landmark detector with semi-supervised style translation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10153–10163, 2019.

[21] Joseph P Robinson, Yuncheng Li, Ning Zhang, Yun Fu, and Sergey Tulyakov. Laplace landmark localization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10103–10112, 2019.

[22] Christos Sagonas, Georgios Tzimiropoulos, Stefanos Zafeiriou, and Maja Pantic. 300 faces in-the-wild challenge: The first facial landmark localization challenge. In *CVPR*, 2013.

[23] Christos Sagonas, Georgios Tzimiropoulos, Stefanos Zafeiriou, and Maja Pantic. A semi-automatic methodology for facial landmark annotation. In *CVPR*, 2013.

[24] Jie Shen, Stefanos Zafeiriou, Grigoris G Chrysos, Jean Kossaifi, Georgios Tzimiropoulos, and Maja Pantic. The first facial landmark tracking in-the-wild challenge: Benchmark and results. In *ICCVW*, 2015.

[25] Ke Sun, Yang Zhao, Borui Jiang, Tianheng Cheng, Bin Xiao, Dong Liu, Yadong Mu, Xinggang Wang, Wenyu Liu, and Jingdong Wang. High-resolution representations for labeling pixels and regions. *arXiv preprint arXiv:1904.04514*, 2019.

[26] Zhiqiang Tang, Xi Peng, Kang Li, and Dimitris N Metaxas. Towards efficient u-nets: A coupled and quantized approach. *IEEE transactions on pattern analysis and machine intelligence*, 42(8):2038–2050, 2019.

[27] Jun Wan, Zhihui Lai, Jun Liu, Jie Zhou, and Can Gao. Robust face alignment by multi-order high-precision hourglass network. *IEEE Transactions on Image Processing*, 30: 121–133, 2020.

[28] Xinyao Wang, Liefeng Bo, and Li Fuxin. Adaptive wing loss for robust face alignment via heatmap regression. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6971–6981, 2019.

[29] Wayne Wu, Chen Qian, Shuo Yang, Quan Wang, Yici Cai, and Qiang Zhou. Look at boundary: A boundary-aware face alignment algorithm. In *CVPR*, 2018.

[30] Wayne Wu, Chen Qian, Shuo Yang, Quan Wang, Yici Cai, and Qiang Zhou. Look at boundary: A boundary-aware face alignment algorithm. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2129–2138, 2018.

[31] Wenyan Wu and Shuo Yang. Leveraging intra and inter-dataset variations for robust face alignment. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 150–159, 2017.

[32] Shengtao Xiao, Jiashi Feng, Junliang Xing, Hanjiang Lai, Shuicheng Yan, and Ashraf Kassim. Robust facial landmark detection via recurrent attentive-refinement networks. In *European conference on computer vision*, pages 57–72. Springer, 2016.

[33] Xuehan Xiong and Fernando De la Torre. Supervised descent method and its applications to face alignment. In *CVPR*, 2013.

[34] Stefanos Zafeiriou, George Trigeorgis, Grigorios Chrysos, Jiankang Deng, and Jie Shen. The menpo facial landmark localisation challenge: A step towards the solution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 170–179, 2017.

[35] Jie Zhang, Meina Kan, Shiguang Shan, and Xilin Chen. Occlusion-free face alignment: Deep regression networks coupled with de-corrupt autoencoders. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3428–3437, 2016.

[36] Meilu Zhu, Daming Shi, Mingjie Zheng, and Muhammad Sadiq. Robust facial landmark detection via occlusion-adaptive deep networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3486–3496, 2019.

[37] Shizhan Zhu, Cheng Li, Chen Change Loy, and Xiaoou Tang. Face alignment by coarse-to-fine shape searching. In *CVPR*, 2015.

[38] Shizhan Zhu, Cheng Li, Chen-Change Loy, and Xiaoou Tang. Unconstrained face alignment via cascaded compositional learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3409–3417, 2016.

[39] Xiangxin Zhu and Deva Ramanan. Face detection, pose estimation, and landmark localization in the wild. In *CVPR*. IEEE, 2012.

[40] Xiangyu Zhu, Zhen Lei, Xiaoming Liu, Hailin Shi, and Stan Z Li. Face alignment across large poses: A 3d solution. In *CVPR*, 2016.

[41] Xu Zou, Sheng Zhong, Luxin Yan, Xiangyun Zhao, Jiahuan Zhou, and Ying Wu. Learning robust facial landmark detection via hierarchical structured ensemble. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 141–150, 2019.