Not all points are created equal - an anisotropic cost function for facial feature landmark location

Farshid Rayhan[1]
f.rayhan@manchester.ac.uk
Aphrodite Galata[1]
A.Galata@manchester.ac.uk
Tim F Cootes[2]
timothy.f.cootes@manchester.ac.uk

[1] Dept of Computer Science
The University of Manchester
Manchester, UK
[2] School of Health Sciences
The University of Manchester
Manchester, UK

Supplementary Materials

1 Implementation details

In our experiments, we used the PyTorch library. The training and testing of our network were conducted on desktop with OS Mint 18, Intel i7 7700 v4 CPU, 8 GB RAM and NVIDIA GeForce RTX 2080ti 11GB card. When training networks we used weight decay=0.0005, momentum=0.9 and batch size of 8.

Each model was trained for 50,000 iterations. We used ResNet50 [13] for all of our experiments and used image resolution 256×256. All the reported results are the average of 5 runs, in some cases we have also reported the standard deviation as well. To perform data augmentation, we followed the same augmentations of [12, 30] where we randomly rotated images $[10, 10]$ degrees. In addition, we also used random horizontal image flipping. Finally, we randomly injected Gaussian blur $(\sigma = 1)$ to 50% of the training images.

Codes: $https://github.com/farshidrayhan-uom/ananisotropic_loss$

2 Datasets

We used the following three datasets in our experiments.

2.1 300W

The 300W [27] is widely used as a 2D face alignment benchmark with 68 annotated landmarks. 300W consists of the following subsets: LFPW [2], HELEN [20], AFW [41], XM2VTS [24] and an additional dataset with 135 images with large pose, occlusion and expressions called iBUG. To compare with other approaches, we adopt the widely used protocol described in [26] to train and evaluate our approach. The full test dataset is split into two subsets, the test dataset of LFPW and HELEN is called the common test dataset, and iBUG is called the challenge test dataset. There is also a 300W private test dataset for the 300W contest, which contains 300 indoor and 300 outdoor faces. We also evaluated our approach on this dataset.

2.2 AFLW

The AFLW [18] dataset contains 24,368 faces with large pose variation. All faces are annotated by up to 21 landmarks per image, while the occluded landmarks were not labelled. For fair comparison with other methods such as wing loss [12], adaptive wing loss [30], we follow the protocol from [40], which provides revised annotations with 19 landmarks. The training set contains 20,000 images, the full test dataset contains 4,368 images.

2.3 WFLW

The WFLW [31] is dataset with 98 manually annotated landmarks that constitutes of 7,500 training images and 2,500 testing images. It also provides attribute annotations including pose, expression, illumination, make-up, occlusion and blur. The WFLW is considered more difficult than commonly used datasets due to its more densely annotated landmarks and difficult faces with occlusion, blur, large pose, makeup, expression and illumination.

Method	NME	AUC8%	FR8%
ESR CVPR 14 [4]	-	32.35	17.00
cGPRT CVPR 15 [21]	-	41.32	12.83
CFSS CVPR 15 [39]	-	39.81	12.30
MDM CVPR 16 [28]	5.05	45.32	6.80
DAN CVPRW 17 [19]	4.30	47.00	2.67
SHN CVPRW17 [35]	4.05	-	-
DCFE ECCV 18 [29]	3.88	52.42	1.83
AWing CVPR 19 [30]	3.56	55.76	0.83
AWing CVPR 19* [30]	3.6	55.7	0.9
Anisotropic loss *	**3.15** $\pm$ **0.2**	**56.87** $\pm$ **0.13**	**0.49** $\pm$ **0.18**
	NME	AUC10%	FR10%
M3-CSR16Õ [5]	-	47.52	5.5
Fan et al. 16Õ [10]	-	48.02	14.83
DR + MDM CVPR 17 [1]	-	52.19	3.67
JMFA17Õ [6]	-	54.85	1.00
LAB CVPR 18 [31]	-	58.85	0.83
AWing CVPR 19 [30]	3.56	64.40	0.33
AWing CVPR 19* [30]	3.6	64.45	0.4
Anisotropic loss *	**3.15** $\pm$ **0.2**	**66.08** $\pm$ **0.09**	**0.26** $\pm$ **0.11**

Table 2: Evaluation on the 300W private dataset. We report mean and variance of Anisotropic loss for 5 runs. '$*$' indicates the experiments we ran.

3 Experiments

Method		Common Subset	Challenging Subset	Fullset
Inter-pupil Normalisation				
CFAN-ECCV 14	[37]	5.5	16.78	7.69
SDM-CVPR 13	[34]	5.57	15.4	7.5
LBF-CVPR 14	[25]	4.95	11.98	6.32
CFSS-CVPR 15	[39]	4.73	9.98	5.76
TCD-CN 16	[38]	4.8	8.6	5.54
MDM-CVPR 16	[28]	4.83	10.14	5.88
RAR-ECCV 16	[33]	4.12	8.35	4.94
DVLN-CVPR17	[32]	3.94	7.62	4.66
TSR-CVPR17	[22]	4.36	7.56	4.99
DSRN-CVPR18	[23]	4.12	9.68	5.21
DCFE-ECCV18	[29]	3.83	7.54	4.55
LAB-CVPR18	[31]	3.42	6.98	4.12
Wing-CVPR18	[12]	3.27	7.18	4.04
Wing-CVPR18*	[12]	3.31	7.20	4.17
Awing-CVPR 19	[30]	3.77	6.52	4.31
Anisotropic loss *		**3.12** $\pm$ **0.2**	**6.25** $\pm$ **0.47**	**3.94** $\pm$ **0.34**
Inter-ocular Normalisation				
PCD-CNN-CVPR 18	[16]	3.67	7.62	4.44
CPM+SBR-CVPR 18	[7]	3.28	7.58	4.1
SAN-CVPR 18	[7]	3.34	6.6	3.98
LAB-CVPR 18	[31]	2.98	5.19	3.49
Awing-CVPR 19	[30]	2.72	4.52	3.07
Awing-CVPR 19*	[30]	2.8	4.58	3.12
Anisotropic loss *		**2.35** $\pm$ **0.15**	**4.05** $\pm$ **0.5**	**2.91** $\pm$ **0.22**

Table 1: Evaluation on the 300W testset. We report mean and variance of Anisotropic loss for 5 runs. '$*$' indicates the experiments we ran.

b \ a	1	2	3	4	5	6
1	3.3	3.29	3.21	3.15	3.01	3.03
2	3.45	3.38	3.22	3.01	2.95	3.01
3	3.58	3.25	3.14	2.98	**2.91**	2.98
4	3.69	3.33	3.14	2.91	3.07	3.06

Table 3: A comparison of different parameter settings (a and b) for the proposed loss function, measured in terms of the normalised mean error on 300W (Inter ocular normalisation) using ResNet50.

dataset \ CNN	Res18	Res34	ResNet50	ResNet101	DenseNet121	DenseNet 169	Wide Resnet50
300W	4.12*	3.99*	**3.94**	3.97*	4.37	3.96*	3.98*
AFLW	1.8	1.5*	**1.3**	1.7	2.1	1.8	1.7
WFLW	4.31*	4.11*	**4.01**	4.01*	4.27*	4.09*	4.03*

Table 4: A comparison of different CNNs using on 300W, WFLW and AFLW dataset using Anisotropic loss using NME. Values displayed are the mean of 5 runs. '$*$' denotes the results that surpass the state of the art but is not the best score.

Table 1, 5 and 2 displays a comparison among different state of the art models against Anisotropic loss where Anisotropic loss achieves the best result.

In table 3, we observe that for a given value for b, as we increase the value for a the error starts to improve. Since a allows the landmarks to slide along the curve, by giving more freedom on this axis we allow the network to prioritise more on the movement orthogonal to the curve, thus the results improve. However for a given value for a, as we start to increase the value for b, the error usually starts to increase. This occurs because any movement orthogonal to the curve results in predictions that essentially deviates from shape of the face/curve.

We report the mean of 5 runs of state of the art CNNs such as ResNet [13], DenseNet [15] in table 4, Wide ResNet [36] on 300W, AFLW and WFLW dataset using Anisotropic loss . The most evident part of that table is that all the values are very close which means Anisotropic loss works well regardless of the dataset or the CNN architecture. ResNet50 performs overall the best on 2 datasets and provides the same as ResNet101 on the WFLW dataset. For consistency, we used the ResNet 50 for all our experiments. We also denote using '$*$' the results that are state of the art but not the best using those above mentioned network which further demonstrates the effectiveness and robustness of Anisotropic loss .

References

[1] Riza Alp Guler, George Trigeorgis, Epameinondas Antonakos, Patrick Snape, Stefanos Zafeiriou, and Iasonas Kokkinos. Densereg: Fully convolutional dense shape regression in-the-wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6799–6808, 2017.

[2] Peter N Belhumeur, David W Jacobs, David J Kriegman, and Neeraj Kumar. Localizing parts of faces using a consensus of exemplars. *IEEE transactions on pattern analysis and machine intelligence*, 35(12):2930–2940, 2013.

Method	Full(%)	Frontal(%)
RCPRCVPR 13 [3]	3.73	2.87
ERTCVPR 14 [17]	4.35	2.75
LBFCVPR 14 [25]	4.25	2.74
CFSSCVPR 15 [39]	3.92	2.68
CCLCVPR 16 [40]	2.72	2.17
TSRCVPR 17 [22]	2.17	-
DAC-OSRCVPR 17 [11]	2.27	1.81
DCFEECCV 18 [29]	2.17	-
CPM+SBRCVPR 18 [9]	2.14 -	
SANCVPR 18 [8]	1.91	1.85
DSRNCVPR 18 [23]	1.86	-
LABCVPR 18 [31]	1.85	1.62
WingCVPR 18 [12]	1.65	-
RCN+(L+ELT+A)CVPR 18 [14]	1.59	-
AWing [30]	1.53	1.38
Anisotropic loss	**1.3** $\pm$ **0.09**	**1.27** $\pm$ **0.1**

Table 5: Mean error(%) on the AFLW testset

[3] Xavier P Burgos-Artizzu, Pietro Perona, and Piotr Dollár. Robust face landmark estimation under occlusion. In *Proceedings of the IEEE international conference on computer vision*, pages 1513–1520, 2013.

[4] Xudong Cao, Yichen Wei, Fang Wen, and Jian Sun. Face alignment by explicit shape regression. *International Journal of Computer Vision*, 107(2):177–190, 2014.

[5] Jiankang Deng, Qingshan Liu, Jing Yang, and Dacheng Tao. M3 csr: Multi-view, multiscale and multi-component cascade shape regression. *Image and Vision Computing*, 47: 19–26, 2016.

[6] Jiankang Deng, George Trigeorgis, Yuxiang Zhou, and Stefanos Zafeiriou. Joint multiview face alignment in the wild. *IEEE Transactions on Image Processing*, 28(7):3636–3648, 2019.

[7] Piotr Dollár, Peter Welinder, and Pietro Perona. Cascaded pose regression. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1078–1085. IEEE, 2010.

[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] Haoqiang Fan and Erjin Zhou. Approaching human level facial landmark localization by deep learning. *Image and Vision Computing*, 47:27–35, 2016.

[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] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[14] Sina Honari, Pavlo Molchanov, Stephen Tyree, Pascal Vincent, Christopher Pal, and Jan Kautz. Improving landmark localization with semi-supervised learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1546–1555, 2018.

[15] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[16] Amin Jourabloo and Xiaoming Liu. Large-pose face alignment via cnn-based dense 3d model fitting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4188–4196, 2016.

[17] Vahid Kazemi and Josephine Sullivan. One millisecond face alignment with an ensemble of regression trees. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1867–1874, 2014.

[18] Martin Koestinger, 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 *2011 IEEE international conference on computer vision workshops (ICCV workshops)*, pages 2144–2151. IEEE, 2011.

[19] 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.

[20] Vuong Le, Jonathan Brandt, Zhe Lin, Lubomir Bourdev, and Thomas S Huang. Interactive facial feature localization. In *European conference on computer vision*, pages 679–692. Springer, 2012.

[21] Donghoon Lee, Hyunsin Park, and Chang D Yoo. Face alignment using cascade gaussian process regression trees. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4204–4212, 2015.

[22] Jiangjing Lv, Xiaohu Shao, Junliang Xing, Cheng Cheng, and Xi Zhou. A deep regression architecture with two-stage re-initialization for high performance facial landmark detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3317–3326, 2017.

[23] 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.

[24] Alejandro Newell, Kaiyu Yang, and Jia Deng. Stacked hourglass networks for human pose estimation. In *European conference on computer vision*, pages 483–499. Springer, 2016.

[25] Shaoqing Ren, Xudong Cao, Yichen Wei, and Jian Sun. Face alignment at 3000 fps via regressing local binary features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1685–1692, 2014.

[26] Shaoqing Ren, Xudong Cao, Yichen Wei, and Jian Sun. Face alignment via regressing local binary features. *IEEE Transactions on Image Processing*, 25(3):1233–1245, 2016.

[27] Christos Sagonas, Georgios Tzimiropoulos, Stefanos Zafeiriou, and Maja Pantic. 300 faces in-the-wild challenge: The first facial landmark localization challenge. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 397–403, 2013.

[28] 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. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4177–4187, 2016.

[29] Roberto Valle, Jose M Buenaposada, Antonio Valdes, and Luis Baumela. A deeply-initialized coarse-to-fine ensemble of regression trees for face alignment. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 585–601, 2018.

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

[31] 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.

[32] 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.

[33] 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.

[34] Xuehan Xiong and Fernando De la Torre. Supervised descent method and its applications to face alignment. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 532–539, 2013.

[35] Jing Yang, Qingshan Liu, and Kaihua Zhang. Stacked hourglass network for robust facial landmark localisation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 79–87, 2017.

[36] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. *arXiv preprint arXiv:1605.07146*, 2016.

[37] Jie Zhang, Shiguang Shan, Meina Kan, and Xilin Chen. Coarse-to-fine auto-encoder networks (cfan) for real-time face alignment. In *European conference on computer vision*, pages 1–16. Springer, 2014.

[38] Zhanpeng Zhang, Ping Luo, Chen Change Loy, and Xiaoou Tang. Learning deep representation for face alignment with auxiliary attributes. *IEEE transactions on pattern analysis and machine intelligence*, 38(5):918–930, 2015.

[39] Shizhan Zhu, Cheng Li, Chen Change Loy, and Xiaoou Tang. Face alignment by coarse-to-fine shape searching. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4998–5006, 2015.

[40] 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.

[41] Xiangxin Zhu and Deva Ramanan. Face detection, pose estimation, and landmark localization in the wild. In *2012 IEEE conference on computer vision and pattern recognition*, pages 2879–2886. IEEE, 2012.