Supplementary Materials for: RetinaFace: Single-shot Multi-level Face Localisation in the Wild

Jiankang Deng * [1,2,3] Jia Guo * [2] Evangelos Ververas[1,3]
Irene Kotsia[4] Stefanos Zafeiriou[1,3]
[1]Imperial College [2]InsightFace [3]FaceSoft [4]Middlesex University London
{j.deng16, e.ververas16, s.zafeiriou}@imperial.ac.uk
guojia@gmail.com, i.kotsia@mdx.ac.uk

1. Visualisation Results

We show some challenging results predicted by RetinaFace-ResNet50 on 300VW [4], AFW [7], Pascal [3] and WIDER FACE [6]. The threshold of face score is set as 0.5. We employ the single-scale test with the original resolution. All 3D meshes are rendered by the Vulkan toolkit. RetinaFace shows impressive robustness in the wild.

1.1. 300VW

300VW [4] aims at the evaluation of 2D facial landmark localisation and tracking under variations of pose, expression, illumination, background, occlusion, and image quality. The test set of 300VW includes three scenarios (Scenario 1: 31 videos, Scenario 2: 19 videos, and Scenario 3: 14 videos) with incremental difficulty. In Fig. 2, we show one frame of the whole 300VW dataset. The proposed RetinaFace can not only predict accurate 3D vertices but also estimate precise pose.

1.2. AFW

The AFW dataset [7] contains 205 high-resolution images with 473 faces [3] collected from Flickr. Images in this dataset contain cluttered backgrounds with large variations in viewpoint. In Fig. 3, we show two mesh regression results predicted by the proposed RetinaFace.

1.3. Pascal

The PASCAL face dataset [3] is collected from the PASCAL 2012 person layout subset, includes $1,335$ labelled faces in 851 images with large facial appearance and pose variations (*e.g.* large in-plane rotation). In Fig. 4, we show two mesh regression results predicted by the proposed RetinaFace.

1.4. WIDER FACE

The WIDER FACE dataset [6] consists of $32,203$ images and $393,703$ face bounding boxes with a high degree of variability in scale, pose, expression, occlusion and illumination. The WIDER FACE dataset is split into training (40%), validation (10%) and testing (50%) subsets by randomly sampling from 61 scene categories. For validation and testing, three levels of difficulty (*i.e.* Easy, Medium and Hard) are defined by incrementally incorporating hard samples. In Fig. 5, we show two mesh regression results predicted by the proposed RetinaFace.

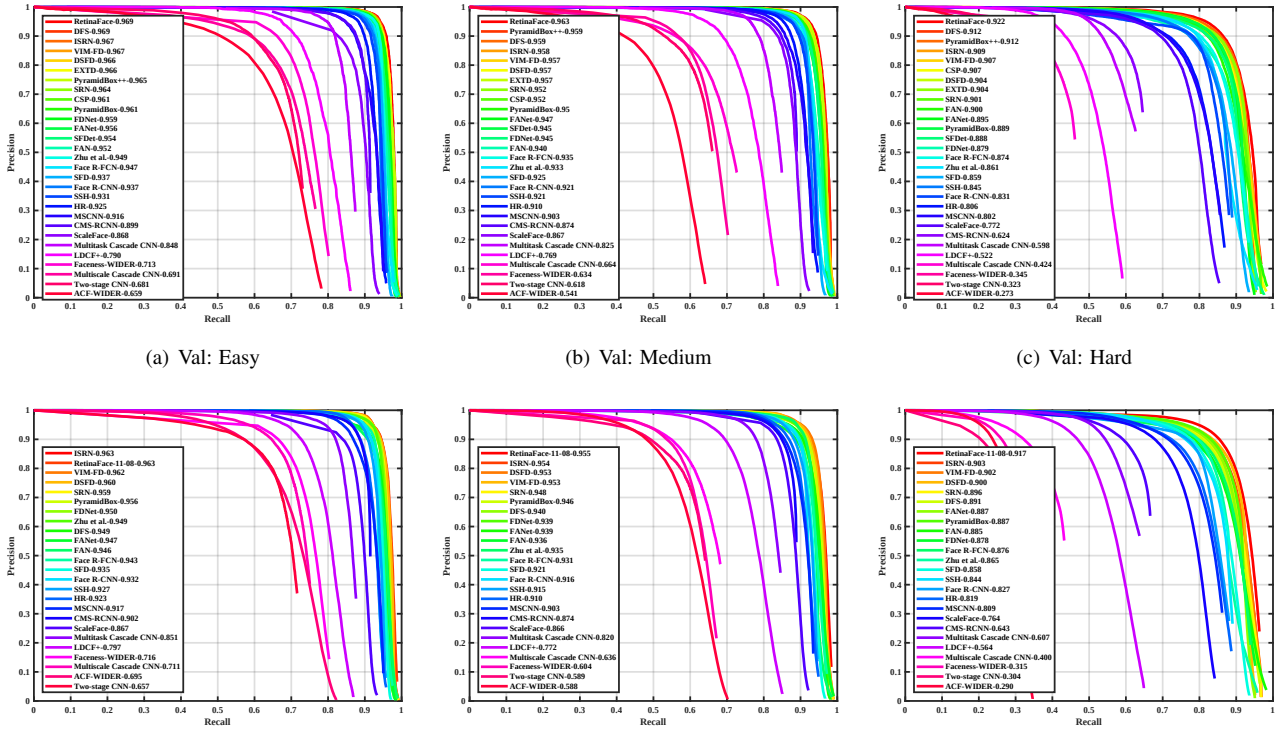
2. Limitations of RetinaFace

As we can see from the above visual results, RetinaFace can not predict facial details (*e.g.* dimple and wrinkle) as our 3D mesh regression branch is designed within the face detector and only predicts 1k vertices considering the efficiency.

In Fig. 6, we show some bad cases from WIDER FACE. Since we set a high threshold (0.5) to avoid false positives, some hard faces (*e.g.* Fig. 6(a)) can be missed by our detector. On WIDER FACE, there are plenty of faces under low-resolution and occlusion, RetinaFace sometimes also makes wrong predictions (*e.g.* Fig. 6(b), 6(c), 6(d)). Nevertheless, the projected face regions in the 3D mesh regression branch still have the effect of attention [5] which can help to improve face detection as confirmed in the section of ablation study.

RetinaFace is robust to exaggerated expression but it has some minimum fitting error around the mouth area (first line, fifth column sample in Fig. 7) as we have not used any specific training data with expression variations. By contrast, MFN [2] employs FacewareHouse [1] as the training data to improve the 3D fitting under exaggerated expressions.

* Equal contributions.
InsightFace is a nonprofit Github project for 2D and 3D face analysis.



(a) Val: Easy (b) Val: Medium (c) Val: Hard

(d) Test: Easy (e) Test: Medium (f) Test: Hard

Figure 1. Precision-recall curves of RetinaFace-ResNet152 on the WIDER FACE validation and test subsets. RetinaFace-ResNet152 achieves state-of-the-art AP on all validation and test subsets.

3. Precision-recall Curves on WIDER FACE

As shown in Fig. 1, RetinaFace-ResNet152 achieves state-of-the-art AP on all validation and test subsets, *i.e.*, 96.9% (Easy), 96.3% (Medium) and 92.2% (Hard) for validation set, and 96.3% (Easy), 95.5% (Medium) and 91.7% (Hard) for test set.

References

[1] Chen Cao, Yanlin Weng, Shun Zhou, Yiying Tong, and Kun Zhou. Facewarehouse: A 3d facial expression database for visual computing. *TVCG*, 2013. 1

[2] Bindita Chaudhuri, Noranart Vesdapunt, and Baoyuan Wang. Joint face detection and facial motion retargeting for multiple faces. In *CVPR*, 2019. 1, 7

[3] Markus Mathias, Rodrigo Benenson, Marco Pedersoli, and Luc Van Gool. Face detection without bells and whistles. In *ECCV*, 2014. 1

[4] 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 *ICCV Workshops*, 2015. 1

[5] Jianfeng Wang, Ye Yuan, and Gang Yu. Face attention network: an effective face detector for the occluded faces. *arXiv:1711.07246*, 2017. 1

[6] Shuo Yang, Ping Luo, Chen-Change Loy, and Xiaoou Tang. Wider face: A face detection benchmark. In *CVPR*, 2016. 1

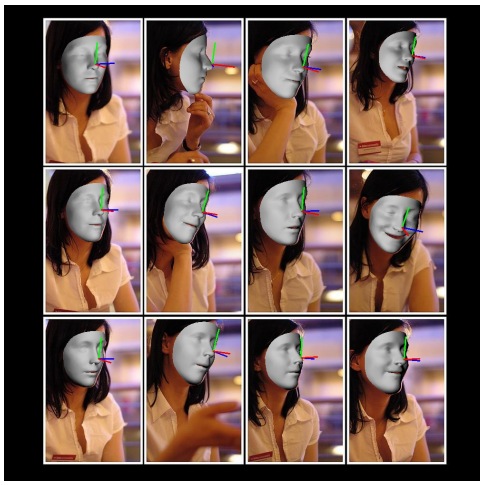
[7] Xiangxin Zhu and Deva Ramanan. Face detection, pose estimation, and landmark localization in the wild. In *CVPR*, 2012. 1

(a) 3D Vertices on Face

(b) Mesh and Pose on Face

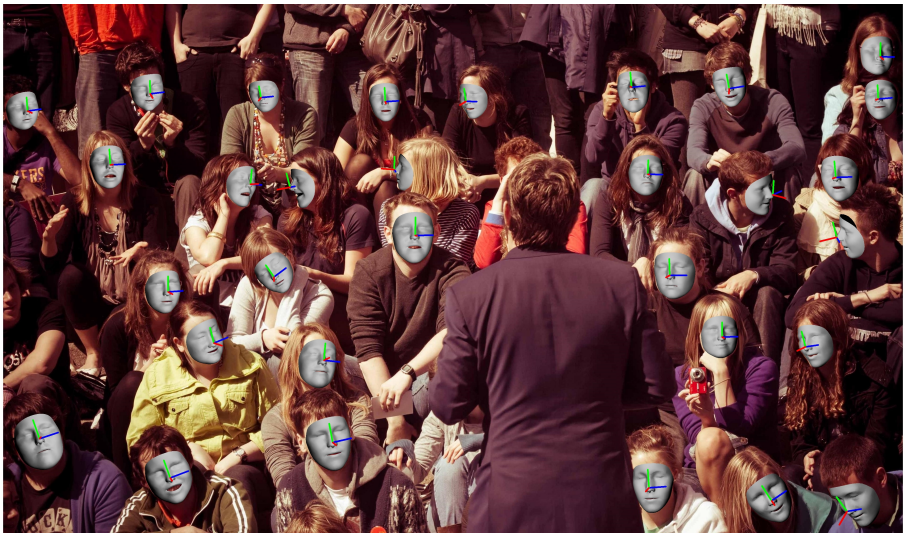
Figure 2. Exemplar 3D mesh regression results on 300VW.

(a) 3D Vertices on Face



(b) Mesh and Pose on Face

(c) 3D Vertices on Face



(d) Mesh and Pose on Face

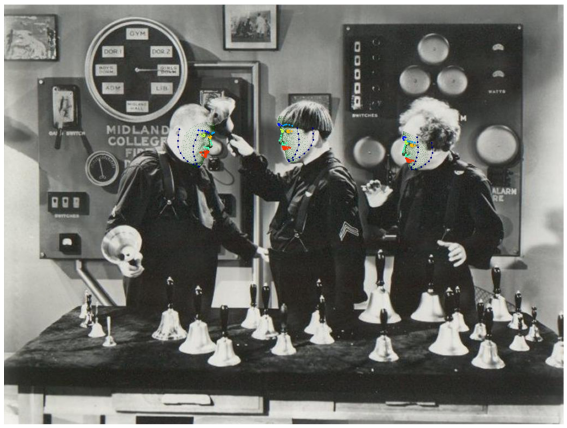
Figure 3. Exemplar 3D mesh regression results on AFW.

(a) 3D Vertices on Face (b) Mesh and Pose on Face

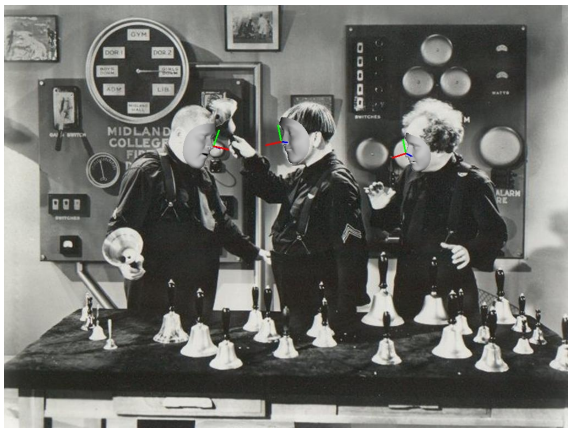
(c) 3D Vertices on Face

(d) Mesh and Pose on Face

Figure 4. Exemplar 3D mesh regression results on PASCAL FACE.



(a) 3D Vertices on Face



(b) Mesh and Pose on Face

(c) 3D Vertices on Face

(d) Mesh and Pose on Face

Figure 5. Exemplar 3D mesh regression results on WIDER FACE.

(a) Missed Face

(b) Wrong Shape

(c) Wrong Pose

(d) Crashed Shape

Figure 6. Bad cases on WIDER FACE. (a) the missed face is annotated by the blue box. (b) (c) and (d) wrong mesh regression results are annotated by the red boxes.

Figure 7. Testing results of RetinaFace (ResNet-50) compared to MFN [2] (First row). We show both the predicted 1k 3D vertices (Second row) and the 3D meshes rendered by the Vulkan toolkit (Third row).