# Supplementary material for How far are we from solving the 2D & 3D Face Alignment problem? (and a dataset of 230,000 3D facial landmarks)

Adrian Bulat and Georgios Tzimiropoulos
Computer Vision Laboratory, The University of Nottingham
Nottingham, United Kingdom
{adrian.bulat, yorgos.tzimiropoulos}@nottingham.ac.uk

## 1. Additional numeric results

Fig. 1 provides additional numerical results for 2D-FAN on the 300-VW training set.

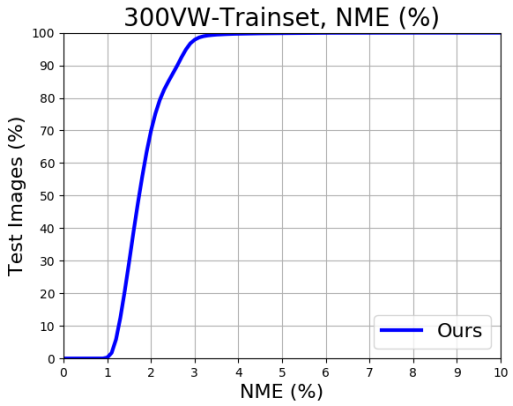


Figure 1: NME (all 68 points used) on 300-VW training set.

## 2. Additional visual results

Figs. 4 and 5 show a series of randomly picked-out samples produced by 2D-FAN and 3D-FAN, respectively, on LS3D-W balanced. Notice that our method copes very well with extreme poses, expressions, and lighting conditions.

## 3. Full 3D face alignment

In this Section, we present an extension of 2D-to-3D-FAN capable of additionally predicting the z coordinate of the facial landmarks.

Similarly to [1], we construct Full-2D-to-3D-FAN by introducing a second subnetwork for estimating the $z$ coordinate (i.e. the depth of each keypoint) on top of 2D-to-3D-FAN. The input to the new subnetwork is the stacked heatmaps produced by 2D-to-3D-FAN alongside the RGB image. The heatmaps play a key role by showing the network where to "look" (i.e at which location should the depth be predicted) incorporating, at the same time, additional pose related information. The proposed subnetwork is based on a ResNet-152 [2] adapted to accept $3 + N$ input channels and to output a vector $N \times 1$ instead of $1000 \times 1$. The network was trained using the L2 loss for 50 epochs and the same learning rates used for the rest of the networks. Fig. 3 reports the numerical error of Full-2D-to-3D-FAN on AFLW2000-3D. For visual results, see Fig. 6.

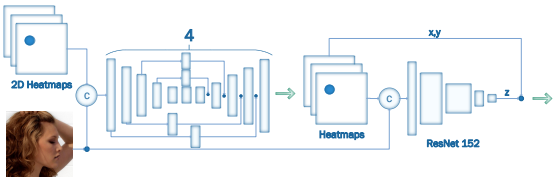


Figure 2: The Full-2D-to-3D-FAN network used for the prediction of the $x, y, z$ coordinates. The network takes as input an RGB image and the 2D landmarks and outputs the corresponding 3D landmarks.

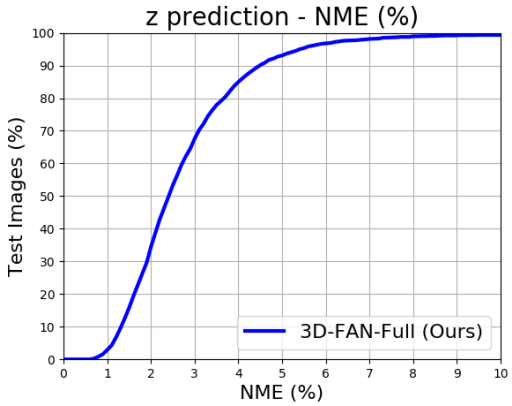


Figure 3: NME on AFLW2000-3D, between the original annotations of [3] and the ones generated by 3D-FAN-Full for depth (z coordinate).

## References


[1] A. Bulat and G. Tzimiropoulos. Two-stage convolutional part heatmap regression for the 1st 3d face alignment in the wild (3dfaw) challenge. In *ECCV*, 2016.

[2] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, 2016.

[3] X. Zhu, Z. Lei, X. Liu, H. Shi, and S. Z. Li. Face alignment across large poses: A 3d solution. In *CVPR*, 2016.

Figure 4: Fitting examples produced by **2D-FAN** on LS3D-W balanced dataset.

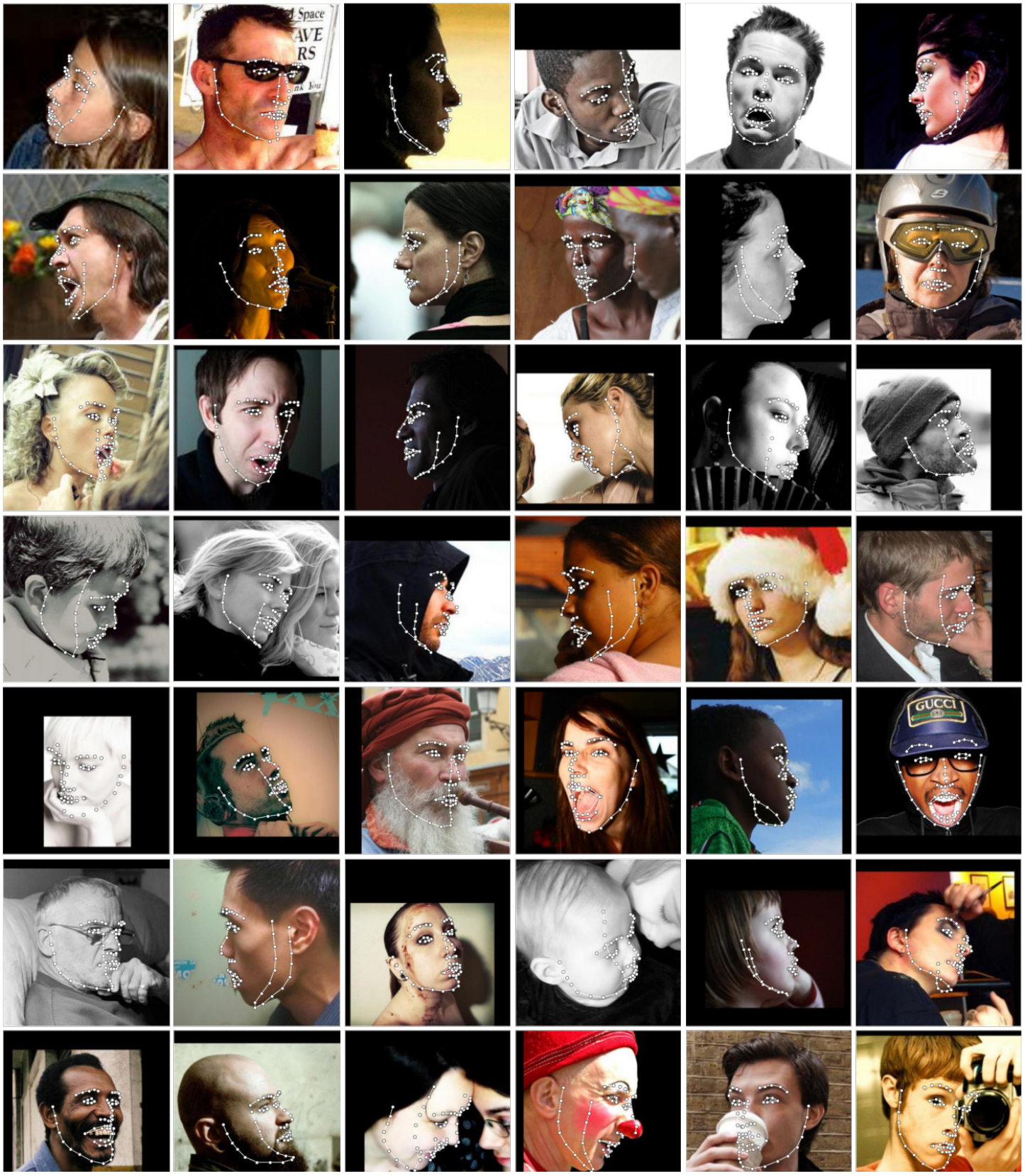


Figure 5: Fitting examples produced by **3D-FAN** on LS3D-W balanced dataset.

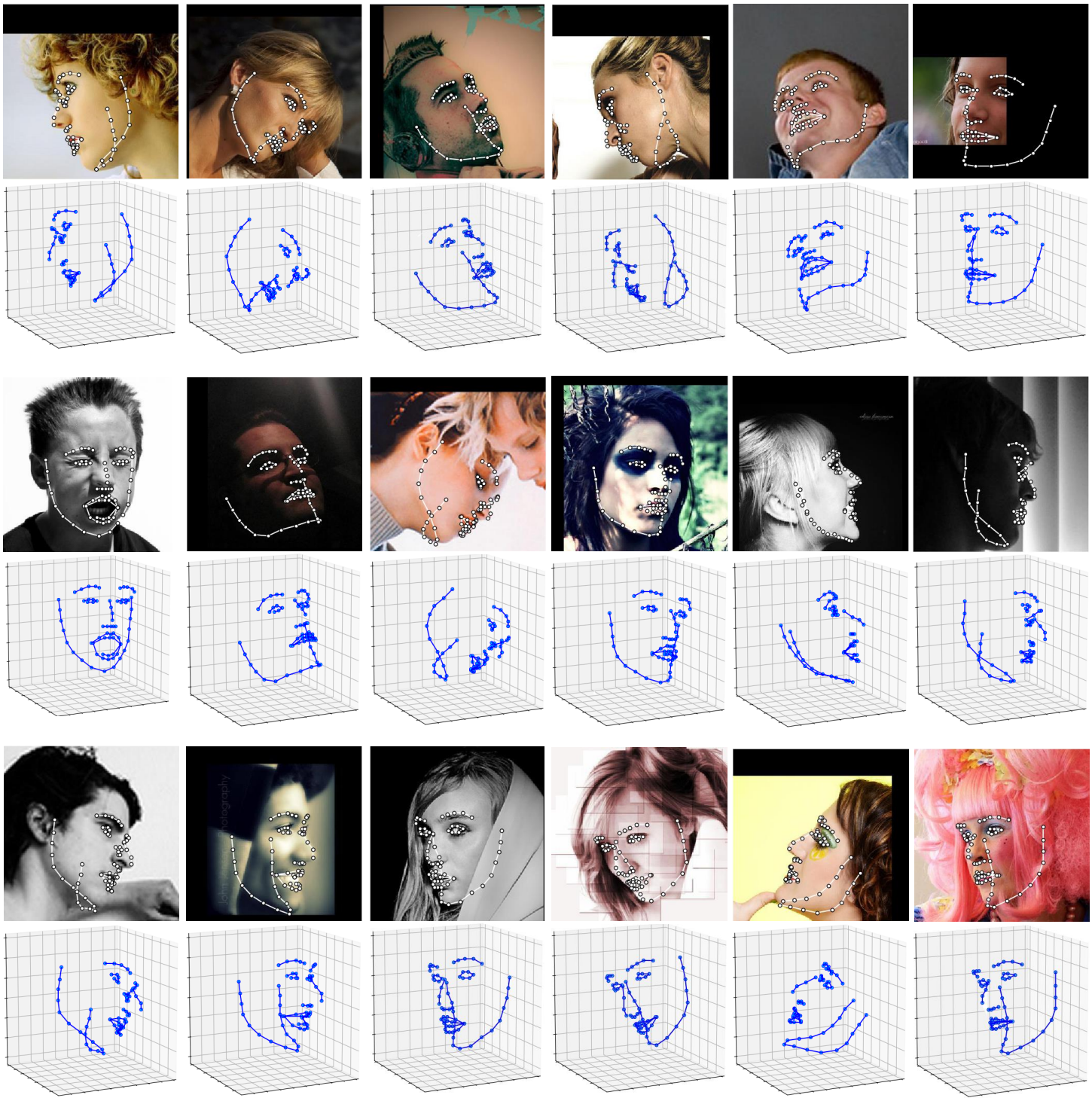

Figure 6: Full 3D fitting examples produced by **Full-2D-to-3D-FAN** on AFLW2000-3D dataset.