Mirror, mirror on the wall, tell me, is the error small?

Heng Yang and Ioannis Patras
{heng.yang, i.patras}@qmul.ac.uk, School of EECS, Queen Mary University of London

It is a common practice to flip the training images for data augmentation in the model training process of object detection or object object part localization. Do object part localization methods trained in this way produce bilaterally symmetric results on mirror images? In order to answer this question we first introduce the concept of **mirrorability**, i.e., the ability of an algorithm to give on a mirror image bilaterally symmetric results, and a quantitative measure called the **mirror error**. The latter is defined as the difference between the detection result on an image and the mirror of detection result on its mirror image, that is

$$e_{\mathrm{m}} = \frac{1}{K}\sum_{k=1}^{K} ||{}^{\mathrm{q}}\mathrm{x}_k - {}^{\mathrm{p}\rightarrow\mathrm{q}}\mathrm{x}_k||. \tag{1}$$

${}^{\mathrm{q}}\mathrm{x}_k$ denotes the result of the k-th part on the original image. ${}^{\mathrm{p}\rightarrow\mathrm{q}}\mathrm{x}_k$ denotes the mirror result of the k-th part from the mirror image. We evaluate the mirrorability of several state of the art algorithms in two representative problems (face alignment and human pose estimation) on several datasets. One would expect that a model that has been trained on a dataset augmented with mirror images to give similar results on an image and its mirrored version. However, as can be seen in Fig. 2 first column, several state of the art methods in their corresponding problems sometimes struggle to give symmetric results in the mirror images. And for some samples the mirror error is quite large. By looking at the mirrorability of different approaches in human pose estimation and face alignment, we arrive at three interesting findings.

- Most of the models struggle to preserve the mirrorability - the mirror error is present and sometimes significant, as shown in Fig. 2.
- The low mirrorability is not caused by training or testing sample bias - all algorithms are trained on both the original images and their mirrored versions.
- The mirror error of the samples is highly correlated with the corresponding ground truth error, with correlation coefficients $\rho(e_{\mathrm{m}}, e_{\mathrm{a}}) \approx 0.7$. An example is shown in Fig. 1.

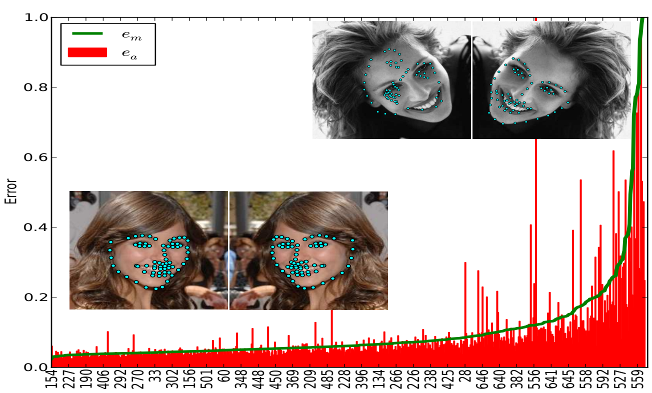
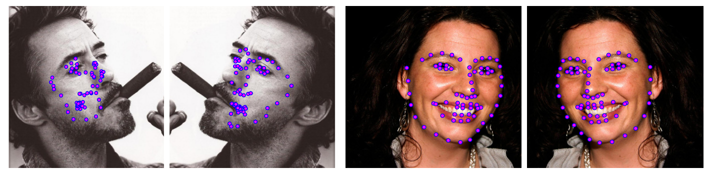


Figure 1: Mirror error and alignment error of RCPR [2] on 300W test images. Results are calculated over 68 facial points.

Since the mirror error is calculated without knowledge of the ground truth, we show two interesting applications - in the first it is used to guide the selection of difficult samples and in the second to give feedback in a popular Cascaded Pose Regression method for face alignment.

Difficult samples selection We apply several state of the art face alignment methods (IFA [1], SDM [4], GN-DPM [3], RCPR [2]) on the test images of the 300W and get the detection results. Then we sort the normalized mirror error e_m in descending order and select the first M samples as being the most difficult ones. We demonstrate two findings in the experiments: 1). the samples that we have selected in this way are truly 'difficult'; 2) the difficult samples selected in this way show high consistency across different approaches.

(a) Mirror error 0.2. (b) Mirror error 0.02.



(c) Mirror error 0.6. (d) Mirror error 0.02.

Figure 2: Example pairs of localization results on original (left) and mirror (right) images. First row: Human Pose Estimation [5], second row: Face Alignment by RCPR [2]. The first column (a and c) shows large mirror error and the second (b and d) small mirror error. Can we evaluate the performance without knowing the ground truth?

Feedback on cascaded face alignment We propose to use the mirror error as a feedback to close the *open* cascaded face alignment system. We apply the baseline face alignment model (RCPR) on the original test image and the mirror image and calculate the mirror error. If the mirror error is above a threshold we restart the process using different initializations, otherwise we keep the detection results. We keep the one that has the smallest mirror error as the final output. The proposed scheme 1) is very effective in selecting test samples to re-start with high precision/recall; 2) improves the performance of state of the art face alignment model, as shown in Table 1.

Methods	*RCPR-F2*	*RCPR-F1*	RCPR-S	RCPR-O	SDM	IFA	GN-DPM	CFAN[6]
49P	**5.35**	6.07	6.59	7.14	7.12	8.31	12.42	7.24
68P	**6.25**	7.11	7.42	7.73	-	-	-	7.72

Table 1: 49/68 facial landmark mean error comparison on 300-W dataset.

[1] Akshay Asthana, Stefanos Zafeiriou, Shiyang Cheng, and Maja Pantic. Incremental face alignment in the wild. In *CVPR*, 2014.

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

[3] Georgios Tzimiropoulos and Maja Pantic. Gauss-newton deformable part models for face alignment in-the-wild. In *CVPR*, 2014.

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

[5] Yi Yang and Deva Ramanan. Articulated human detection with flexible mixtures of parts. *T-PAMI*, 35(12):2878–2890, 2013.

[6] Jie Zhang, Shiguang Shan, Meina Kan, and Xilin Chen. Coarse-to-fine auto-encoder networks (cfan) for real-time face alignment. In *ECCV*, 2014.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.