An Evaluation of 2D Face Alignment methods against Extreme Lighting Conditions and Blur

Avishek Bose
Department of Electrical & Computer Engineering
University of Toronto
40 St George St Rm: BA4158, Toronto, ON M5S 2E4
Phone: 647-832-5165
Email: joey.bose@mail.utoronto.ca

Parham Aarabi
Department of Electrical & Computer Engineering
University of Toronto
40 St George St Rm: BA4134, Toronto, ON M5S 2E4
Phone: 416-946-7893
Email: parham@ecf.utoronto.ca

***Abstract*—The purpose of this work is to empirically evaluate both state of the art and popular models in these extreme conditions. To this end, we first synthetically extend the 300-W IMAVIS test set to incorporate our defined extreme conditions. Then we evaluate the two chosen Face Alignment models on our augmented dataset and answer the question whether current state of the art Face Alignment models are robust to our extreme conditions and how they compare against very popular off the shelf models such as an Ensemble of Cascaded Regressors. We find that state of the art CNN models outperform the Ensemble of Cascaded Regressors model when Brightness is changed However, the Ensemble of Cascaded Regressors model outperforms the state of the art CNN when subject to extreme Blurring. We also observe that under extreme conditions both models are far from optimal.**

***Keywords*—Computer Vision, Face Alignment, Ensemble of Cascaded Regressors, CNN**

I. Introduction

Facial Landmark localization, commonly referred to as Face Alignment, has seen significant recent progress through the proliferation of large annotated datasets [1] [2] and the advent of large deep learning models. Incredibly, specifically designed Convolution Neural Nets [3] such as 2D and 3D Face Alignment Network (FAN) [4] have shown state of the art performance across multiple challenging Facial Landmark datasets to the point of saturation. Furthermore, these deep learning models have been shown to be robust against large head pose variation and occlusion, a key failure mode of many prior 2D Facial Alignment algorithms that do not use a Deep Learning approach.

In many computer vision tasks facial landmarks often provide critical information needed for downstream tasks such as 3D headpose estimation or Gaze Tracking [5]. Consequently, Face Alignment is often an intermediary problem in Computer Vision systems such as the ones required for Facial Recognition where the pipeline first has a Face Detection phase. As such the performance of the Face Alignment models is dependent on the output of a Face Detector i.e. Viola Jones [6]. Furthermore, it has been shown empirically that different initializations to face detection can lead to poorer Facial Landmark locations [7].

In this work we ask the question to what extent is the the Face Alignment problem solved? From our experiments we observe that even state of the art models suffer under extreme conditions of where brightness is increased or decreased. Secondly, we then analyze the failure modes of these models to these extreme conditions that are not well represented in current datasets. We proceed our analysis by first synthetically extending the cropped 300-W test dataset [8] [9] [10] by 10 times the current size and reevaluate the performance on this augmented dataset. Finally we compare the performance of the state of the art Deep Learning model and a popular off the shelf model that uses an Ensemble of Cascaded Regressors against our synthetically generated extreme conditions. Our main contributions can be summarized as follows:

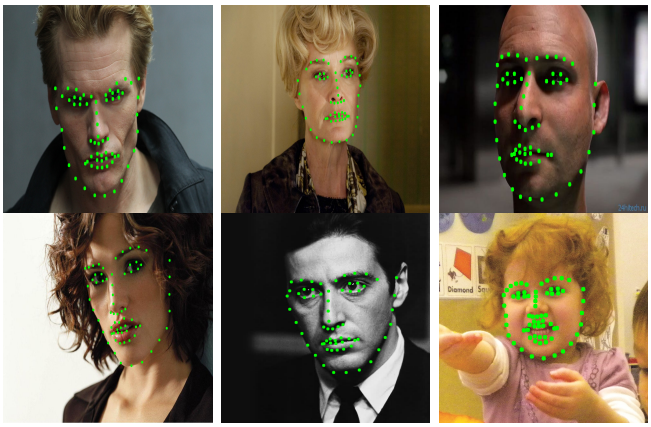

Fig. 1. Selected Performance of 2D FAN on 300W IMAVIS dataset. The green points are ground truth landmark positions while the red points are the predicted landmarks.

1. The creation of a synthetically extended version of the 300-W IMAVIS dataset containing 2 dimming factors to simulate extreme low light conditions, 2 whitening factors to simulate lighting conditions with too much light and 5 successive rounds of blurring using a Gaussian Kernel. 2. Analysis of state of the art Deep Learning models on the above conditions. 3. Comparison between the state of the art Deep Learning Model and Ensemble of Cascaded Regressors model on the task of Face Alignment in our defined extreme conditions.

II. Face Alignment Methods

Let I denote a face image whose ith landmark, $x_i \in \mathbb{R}^2$, is the x, y coordinate in the image space. Thus the set of all N landmarks in I, or face shape, is then $S = \{x_i \in \mathbb{R}^2\}_{i=1}^{N}$. The Face Alignment problem is then defined as finding a face shape $\bar{S}$ as close as possible to the ground truth annotation. Formally, the objective is then $min||S - \bar{S}||_2$ where we have chosen the L_2 norm as the distance function between points.

A. 2D FAN

In this section we provide a high level overview of the state of the art CNN specifically designed to tackle the Face Alignment Problem. While a detailed treatment of 2D FAN is beyond the scope of this paper we try to distill the key insights from the original paper. The building block of the 2D FAN model is the Hourglass Network that was originally conceived for human pose estimation [11]. The authors motivate the need for an Hourglass Network due to the need to capture information at every scale. Consequently, the Hourglass Network is composed of Convolutional and max pooling layers that compress the features down to a very low dimension which can be thought of as a low resolution representation of the original input. At each max pooling step, the network branches off and applies more convolutions at the original pre-pooled resolution. After reaching the lowest resolution, the network begins the top-down sequence of upsampling and combination of features across scales. As such at the lowest resolution the Hourglass Network has a bottleneck layer from which the upsampling occurs. In 2D FAN the authors replace the bottleneck layer with a hierarchical, parallel and multi-scale block which has shown greater performance with the same number of parameters. Finally the 2D FAN is then constructed using 4 Hourglass Networks in succession that is used to predict the Facial Landmark locations.

B. Ensemble of Cascaded Regressors

In this section we consider, at a high level, one of the most popular iterations of an Ensemble of Cascaded Regressors [12]. The principal idea in this approach is to iteratively update a predicted face shape $\bar{S}$ based on the output of a regressor $r_t(.,.)$. Specifically, each regressor in the cascade is a function of the original image and the current face shape estimate. Thus each regressor then updates the face shape as follows:

$$\bar{S}(t+1) = \bar{S}^t + r_t(I, \bar{S}^t)$$

Each regressor makes its prediction based on features computed from I and is trained via gradient boosted trees with L_2 distance as the chosen loss function. For a more detailed treatment of the features used in training or the specifics of tree based regression we refer the interested reader to the original paper.

III. Dataset and Evaluation Metrics

The 300-W dataset, was first introduced for Automatic Facial Landmark Detection in-the-Wild Challenge and is widely used as a benchmark for Face Alignment. Landmark annotations are provided following the Multi-PIE 68 points mark-up [13] and the 300-W test set consists of the re-annotated images from LFPW [1], AFW [14], HELEN [2], XM2VTS [15] and FRGC [16] datasets. Moreover, the 300-W test set is split into two categories, indoors and outdoors, of 300 images. In this paper we consider the cropped version of the combined indoor and outdoor splits of the 300-W dataset used for the IMAVIS competition. As we are interested in isolating the effect of Brightness reduction/magnification and blur individually none of our experiments consider the additive effects of Brightness and Blur. Examples for synthetically generated images for each experiment are provided in the appendix.

A. Fade to Black

To simulate the effect of poor lighting conditions and their effects on Face Alignment algorithms we algorithmically interpolate between the original image and a black image. We follow a simple process to synthetically generate many images at different Brightness levels by first creating a black image (all pixel values are set to 0) of the same size as the original image and where . Finally the output image can be constructed by interpolating between the original image and the degenerate copy using a constant factor, α, as follows:

$$\text{Output} = \text{Black} * (1 - \alpha) + \text{Orig} * \alpha$$

Thus an α value of 0.9 means the Brightness of the output image is reduced by 10% while a value of 1.0 is the original unperturbed image. For our experiments we use α values ranging from 0.9 to 0.1.

B. Fade to White

To increase Brightness levels we follow the exact same procedure outlined in the section above but we instead interpolate with a white image (all pixel values are set to 255).

C. Blur

In our experiments we use a 2D Gaussian kernel defined below to blur the images in the 300-W test set. In our experiments we fix the kernel size and standard deviation at 15 and $\sigma = 5$ respectively, and then apply successive iterations of blurring from the output of the previous round for a total of 5 such iterations.

D. Evaluation Metric

Following the work in [4] we use the Normalized Mean Error (NME), defined below, as our metric of choice.

$$\text{NME} = \frac{1}{N} \sum_{k=1}^{N} \frac{||x_k - y_k||_2}{d}$$

Here x represents the ground truth labels and y represents the predictions and $d = \sqrt{\text{width} * \text{height}}$ of the detected bounding box after face detection. The choice of d as opposed to the inter-ocular distance, which is typical [17] [18], is used to normalize the error as the latter has shown to be biased to frontal faces [14].

IV. Experiments

In this section we present our experimental results on both of the Face Alignment methods outlined in section 2. For the Ensemble of Cascaded Regressors we use the OpenFace Library [19] which makes use of the DLIB [20] model under the hood. The 2D FAN model is publicly available and as such we use it as our reference state of the art model. In order to compare solely the performance on the Face Alignment problem we use the same off the shelf Face Detector for both models. To guide our treatment we first consider the performance of our Face Detector on our augmented 300-W test set. The result of this is presented in Fig 2 and from we conclude all three data augmentations significantly effect face detection performance in their most extreme settings but image whitening is comparatively worse than dimming and about the same as blurring. Finally, we plot $\%$ NME for up to 7% error on only the images where face detection succeeds and as such the results presented in the next section should be viewed in context with the Face Detection results. However, in this work we do not consider the effects of different initializations of a Face Detector or different detectors, interested readers can refer to the work in [7] for a more detailed treatment which is orthogonal to the work done in this paper.

A. Fade to Black Results

In our first experiment we consider the effects of Brightness reduction on the performance of 2D FAN and the Ensemble of Cascaded Regressors model. We evaluate the performance of these models on both indoor and outdoor partitions of our augmented 300-W test set. As can be seen from Fig 3. the 2D FAN model suffers little degradation in performance compared to the Ensemble of Cascaded Regressors and even in the most extreme case where the image is reduced to 10% of its original Brightness the state of the art deep learning model saturates to roughly the same number of images but at a lower NME threshold. Crucially, of the total detected faces in the most extreme setting the Ensemble of Cascaded Regressors model fails to capture $1/3$ of the data below the 7% NME threshold. From this result we draw the conclusion that current state of the art models are comparatively better than Ensemble of Cascaded Regressors at handling small to moderate reductions in lighting conditions and significantly better when lighting is reduced to extremely low levels.

B. Fade to White Results

For our next experiment we now consider the converse question, specifically how do the performance of Face Alignment models vary when Brightness is increased in images? As in the first experiment we evaluate the models on the combined indoor and outdoor partitions of the augmented 300-W dataset. Inspecting the results presented in Fig 4. it is clear that both models suffer noticeable degradation in performance in the most extreme setting where only 10% of the original image is present. However, it is important to note that the 2D FAN model significantly outperforms the Ensemble of Cascaded regressors to the point of saturation in all settings except the most extreme. Through this we conclude that increases in Brightness is a pertinent issue only for the Ensemble of Cascaded Regressors model while the state of the art CNN is quite robust to this except for in the most extreme conditions.

Fig. 2. Face Detection performance on our augmented dataset for Fade to Black, Fade to White and Blur partitions.

C. Blur Results

In our final experiment we evaluate the performance of 2D FAN and the Ensemble of Cascaded Regressors models on images augmented with successive rounds of Gaussian Blur as outlined in the previous section on Dataset and Evaluation Metrics. As can be seen from our results in Fig 5. both models show an immediate decrease in performance even with one round of blurring. Interestingly, the state of the art model is consistently worse than the Ensemble of Cascaded Regressors. Moreover, in the most extreme case Ensemble of Cascaded Regressors is approximately $1.5\times$ better than 2D FANN whose NME error curve saturates. In general, both approaches have poorer performance under blurring than brightness reduction or amplification.

V. Conclusion

In this paper we considered the performance of Face Alignment methods, both state of the art 2D FAN and an

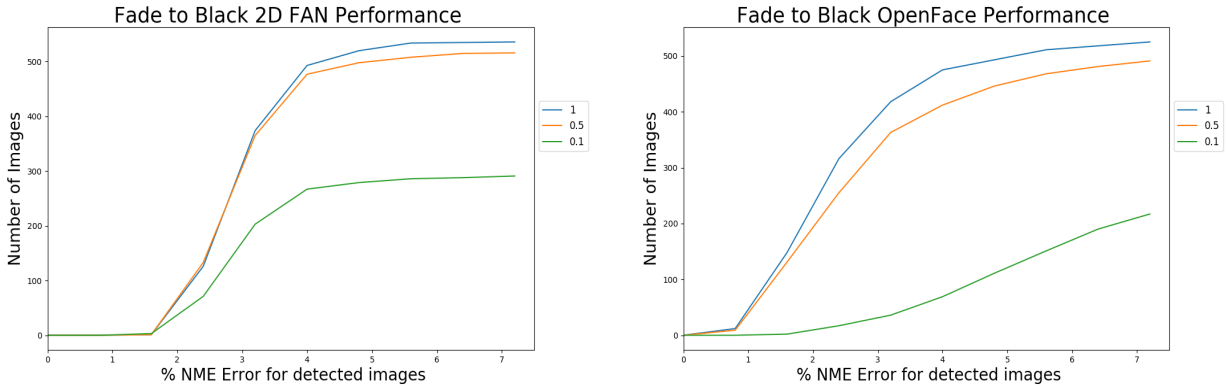

Fig. 3. Performance of 2D FAN and Ensemble of Cascaded Regressors on Fade to Black.

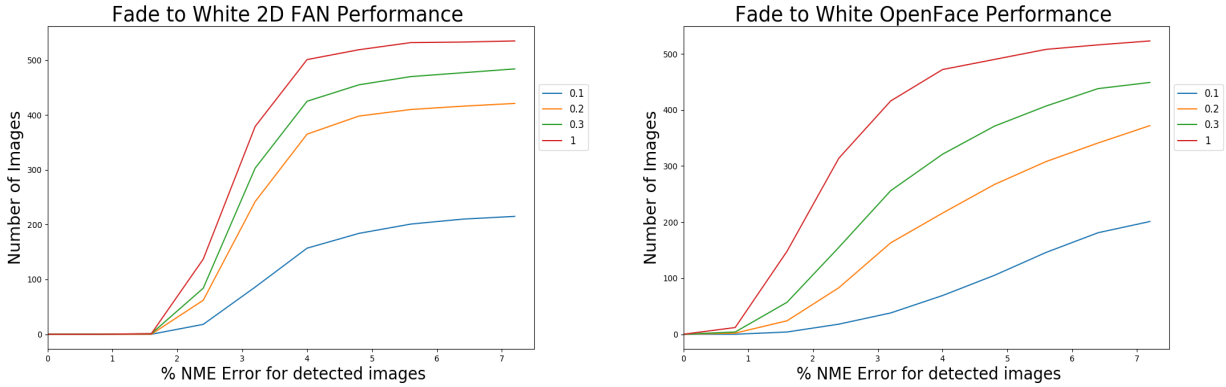

Fig. 4. Performance of 2D FAN and Ensemble of Cascaded Regressors on Fade to White.

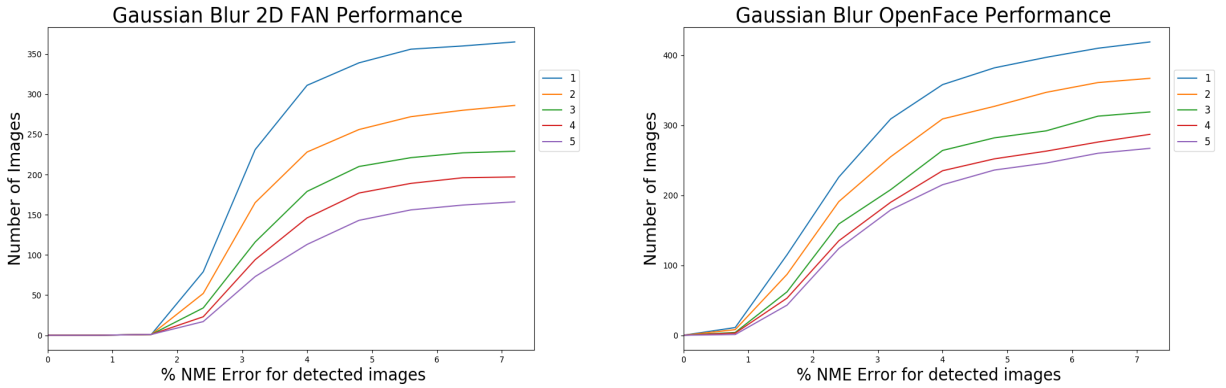

Fig. 5. Performance of 2D FAN and Ensemble of Cascaded Regressors on images smoothed by Gaussian Blur. We compare different blur levels by changing the standard deviation of the Gaussian Kernel between 1-20.

Ensemble of Cascaded Regressors when subject to images of extreme variations in Brightness and Gaussian Blur. First we extended the 300-W IMAVIS test set for each of our transformations then empirically evaluating our chosen models. We find that the state of the art 2D FAN model is less adversely affected due to Brightness reduction than the Ensemble of Cascaded Regressors model and significantly better in the most extreme condition. Furthermore, we also find that in the converse situation where Brightness is increased the 2D FAN model still outperforms the Ensemble of Cascaded Regressors model but both models fail to achieve the 7% NME for a large number of detected faces in the most extreme situation. Lastly, we also evaluate the chosen models on images blurred by a 2D Gaussian kernel and interestingly we note that the 2D FAN model performs worse relative to the Ensemble of Cascaded Regressors model and both model like in the previous experiments still fail on a large number of detected faces. In fact, for solely the Face Alignment problem extreme blurring seems to be the most significant failure mode. However, all of these results must be put into context via the performance of our chosen Face Detector which we argue is a larger bottleneck to performance than the actual Face Alignment algorithm. In general, we find that when a face is properly detected both models do a reasonable job but the 2D FAN model is comparatively better under difficult lighting conditions while the Ensemble of Cascaded Regressors does better under Gaussian Blurring. Through our experiments it is clear that the Facial Alignment Problem is not saturated as datasets and models trained on them cannot handle extreme situations.

REFERENCES

[1] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, and N. Kumar, "Localizing parts of faces using a consensus of exemplars," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 12, pp. 2930–2940, 2013.

[2] V. Le, J. Brandt, Z. Lin, L. Bourdev, and T. S. Huang, "Interactive facial feature localization," in *European Conference on Computer Vision*, pp. 679–692, Springer, 2012.

[3] Y. LeCun, Y. Bengio, *et al.*, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.

[4] A. Bulat and G. Tzimiropoulos, "How far are we from solving the 2d & 3d face alignment problem?(and a dataset of 230,000 3d facial landmarks)," *arXiv preprint arXiv:1703.07332*, 2017.

[5] C.-C. Lai, Y.-T. Chen, K.-W. Chen, S.-C. Chen, S.-W. Shih, and Y.-P. Hung, "Appearance-based gaze tracking with free head movement," in *Pattern Recognition (ICPR), 2014 22nd International Conference on*, pp. 1869–1873, IEEE, 2014.

[6] P. Viola and M. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, pp. 137–154, 2004.

[7] H. Yang, X. Jia, C. C. Loy, and P. Robinson, "An empirical study of recent face alignment methods," *arXiv preprint arXiv:1511.05049*, 2015.

[8] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "A semi-automatic methodology for facial landmark annotation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 896–903, 2013.

[9] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "300 faces in-the-wild challenge: The first facial landmark localization challenge," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 397–403, 2013.

[10] C. Sagonas, E. Antonakos, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "300 faces in-the-wild challenge: Database and results," *Image and Vision Computing*, vol. 47, pp. 3–18, 2016.

[11] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," in *European Conference on Computer Vision*, pp. 483–499, Springer, 2016.

[12] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1867–1874, 2014.

[13] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-pie," *Image and Vision Computing*, vol. 28, no. 5, pp. 807–813, 2010.

[14] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 2879–2886, IEEE, 2012.

[15] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, "Xm2vtsdb: The extended m2vts database," in *Second international conference on audio and video-based biometric person authentication*, vol. 964, pp. 965–966, 1999.

[16] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *Computer vision and pattern recognition, 2005. CVPR 2005. IEEE computer society conference on*, vol. 1, pp. 947–954, IEEE, 2005.

[17] D. Cristinacce and T. F. Cootes, "Feature detection and tracking with constrained local models.," in *BMVC*, vol. 1, p. 3, 2006.

[18] J. Shen, S. Zafeiriou, G. G. Chrysos, J. Kossaifi, G. Tzimiropoulos, and M. Pantic, "The first facial landmark tracking in-the-wild challenge: Benchmark and results," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 50–58, 2015.

[19] B. Amos, B. Ludwiczuk, and M. Satyanarayanan, "Openface: A general-purpose face recognition library with mobile applications," tech. rep., CMU-CS-16-118, CMU School of Computer Science, 2016.

[20] D. E. King, "Dlib-ml: A machine learning toolkit," *Journal of Machine Learning Research*, vol. 10, pp. 1755–1758, 2009.

APPENDIX
SAMPLES FROM THE SYNTHETIC 300-W DATASET

Fig. 6. Fade to Black example on a single image. The top left is the original image and each successive image the brightness is reduced to 50% and 10% of the original.

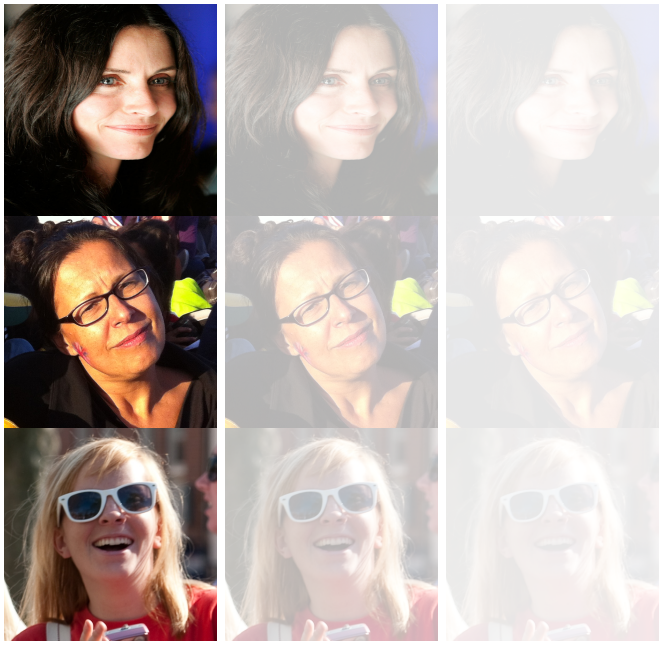

Fig. 7. Fade to White example on a single image. The top left is the original image and the image to its immediate right contains 30% of the original image and finally the rightmost image contains 10% of the original.

Fig. 8. Gaussian Blur applied to a sample from the 300-W test set. The first image on the left is the original unperturbed image followed by the same image under 2 rounds of blurring and finally the last image is after 5 rounds.