Structured Semi-supervised Forest for Facial Landmarks Localization with Face Mask Reasoning

Xuhui Jia[1]
xhjia@cs.hku.hk
Heng Yang[2]
heng.yang@eecs.qmul.ac.uk
Angran Lin[1]
arlin@cs.hku.hk
Kwok-Ping Chan[1]
kpchan@cs.hku.hk
Ioannis Patras[2]
i.patras@eecs.qmul.ac.uk

[1] Department of Computer Science
The Univ. of Hong Kong, HK
[2] School of EECS
Queen Mary Univ. of London, UK

Motivation. Despite the great success of recent facial landmarks localization approaches, the presence of occlusions significantly degrades the performance of the systems [2, 5]. Though occlusion occur frequently in realistic scenarios (e.g. the use of scarf or sunglasses, hands or hair on the face), very few works have addressed this problem explicitly due to the high diversity of occlusion in real world. While [4] tried to model a few synthetic occlusion patterns, the recent method of [1] dealt with the occlusion problem in more realistic sceneries. Both of them only focused on modelling the occlusion in an unstructured way, i.e. treating the visibility of each landmark independently. However in realistic conditions, the occlusion patterns (or called occluders) often occupy a continuous region instead of an individual pixel location, as depicted in Fig 2. Thereby the whole occluded region will consistently affect the landmarks localization.

Contribution. This work attempts to address the face mask reasoning and facial landmarks localization in an unified Structured Decision Forests framework. We first have built a rich face image dataset with face mask annotation. The dataset was built as an extension of the recent datasets: Caltech Occluded Faces in the Wild (COFW), Labeled Face Parts in the Wild (LFPW) and Labeled Face in the Wild (LFW). We manually annotate a portion of images in these datasets with face masks. The face mask indicates whether or not each pixel belongs to the face. Then we incorporate such additional information of dense pixel labelling into training the Structured Classification-Regression Decision Forest. The classification nodes aim at decreasing the variance of the pixel labels of the patches by using our proposed structured criterion while the regression nodes aim at decreasing the variance of the displacements between the patches and the facial landmarks. The proposed framework allows us to predict the face mask and facial landmarks locations jointly. The proposed framework with following properties. First, semi-supervised, it uses training images from the above described augmented dataset, only a portion of which are with face masks. Second, structured, it has a novel structured criterion for split function selection for the pixel labelling (face mask reasoning) problem. Third, joint classification-regression, it predicts face mask label for each pixel (classification) and the landmark locations (regression) at the same time, and more importantly it uses the face mask reasoning results to improve the accuracy of landmark localization . Experiments show our method 1) yields promising results in face mask reasoning; 2) improves the existing Decision Forests approaches in facial landmark localisation, aided by face mask reasoning.

Figure 2: The images on the left side of the two pairs show the results from the standard Random Forests for facial landmarks localization [3], with failure cases under occlusion. The images on the right side of the two pairs show the results of our proposed method. It first explicitly predicts the face mask (the semi-transparent region), then use the face mask information to improve the localization and to predict the occlusion status of the landmarks.

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

[2] X. Cao, Y. Wei, F. Wen, and J. Sun. Face alignment by explicit shape regression. In *CVPR*, 2012.

[3] M. Dantone, J. Gall, G. Fanelli, and L. Van Gool. Real-time facial feature detection using conditional regression forests. In *CVPR*, 2012.

[4] Golnaz Ghiasi and Charless Fowlkes. Occlusion coherence: Localizing occluded faces with a hierarchical deformable part model. In *CVPR*, 2014.

[5] Yi Sun, Xiaogang Wang, and Xiaoou Tang. Deep convolutional network cascade for facial point detection. In *CVPR*, 2013.

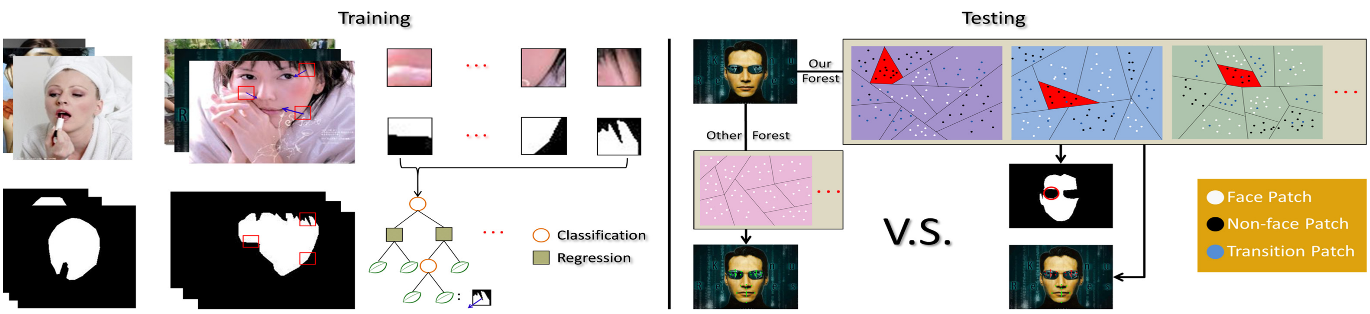

Figure 1: The framework of proposed method. We use face images with annotation of facial landmarks and face masks for training. By randomly switching the information gain function at the internal nodes, the decision trees are optimized with respect to both the offsets to landmarks (regression) and to the local structured label configuration (classification). The forest model is able to predict the face mask and landmark locations jointly. We exploit the face mask prediction to further improve the landmark localization.