

Using Segmentation to Predict the Absence of Occluded Parts

Golnaz Ghiasi
gghiasi@ics.uci.edu

Charless C. Fowlkes
fowlkes@ics.uci.edu

Dept. of Computer Science
University of California
Irvine, CA

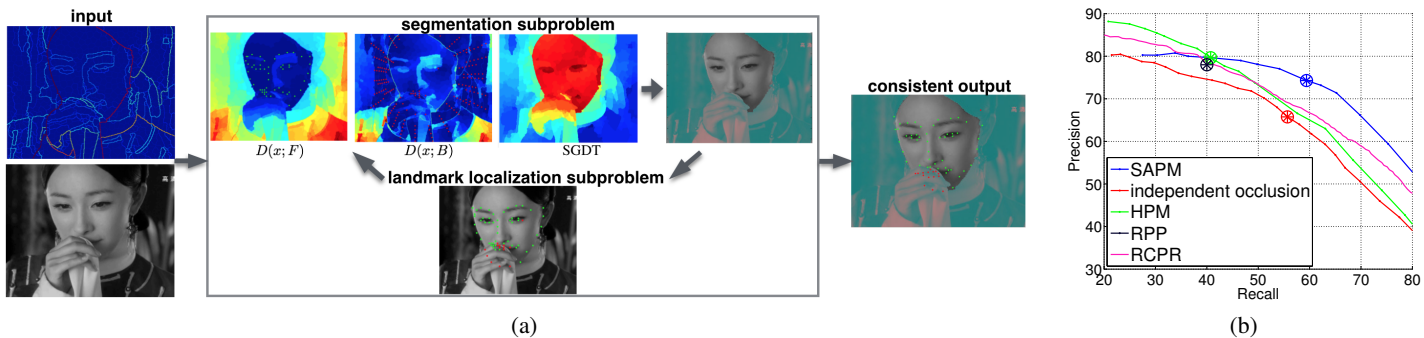


Figure 1: (a) Overall landmark localization and face mask prediction pipeline. Given an input image and superpixel boundary weights, our method alternately optimizes a landmark localization subproblem and segmentation subproblem. (b) Landmark occlusion prediction accuracy for our segmentation-aware part model (SAPM) and baseline methods on the COFW test data. Enforcing segmentation consistency (SAPM) yields improved performance over an equivalent model without segmentation (independent).

Occlusion, arising from interactions between objects in cluttered scenes or between different parts of a single articulated object, creates drastic changes in local visual appearance. This poses a significant barrier to correctly detecting and performing fine-grained analysis of objects in images. We focus on the problem of detecting and localizing faces where the parts in consideration are represented by facial keypoints. Recent work has shown that pose-regression techniques [1] or deformable part models [5] that include explicit states representing part occlusion can yield better detection and localization accuracy.

A key difficulty with part-based occlusion reasoning is that such methods rely heavily on local image appearance in the vicinity of a part to predict whether it is occluded. This is easily confounded by lighting or other appearance variations and ignores long-range dependencies in patterns of occlusion (e.g., the occluding object comprises an extended region of the image demarcated by enclosing contours). We argue that bottom-up segmentation provides a valuable mechanism by which subsets of occluded or visible keypoints can be grouped in a way that is not easily captured by standard pose-regression or deformable part models. We formulate a joint objective that simultaneously attempts to localize parts and determine their occlusion state in a manner that is consistent with image segments suggested by edges in the image.

The contribution of this paper is in combining explicit part occlusion in a detection model with object-specific segmentation using a simple alternating minimization. Unlike previous work that focused primarily on segmenting objects from background, our model solves the problem of identifying occluders with high accuracy. We do not perform inference over a large pool of segmentation proposals (unlike [3, 4]), instead generating a consistent segmentation with only two iterations of segmentation and detection, even when the occluder has similar texture and color to the object.

Figure 1(a) gives an overview of our model which carries out an initial detection followed by alternating segmentation and refinement of detector keypoint locations. These two models are coupled by a unary potential function that enforces agreement between the location and occlusion state of keypoints in the detector and face/non-face assignment of superpixels in the segmentation model.

We use a deformable part model framework to capture the relative layout of facial keypoints. We use the same tree topology used in the hierarchical part model (HPM) described in [5]. Unlike the HPM work we do not restrict the possible occlusion patterns for the parts. Each keypoint can be visible or occluded independent of the other keypoints, allowing us to represent a much larger space of possible occlusion patterns. Instead we make use of bottom-up segmentation to guide the detector towards consistent patterns of occlusion.

To generate a high-quality segmentation, we train a boundary detector to specifically detect those edges relevant for face segmentation. We train a structured random forest [2] on images from the COFW training dataset

in which we manually labeled face foreground masks. The ground-truth boundaries for each training patch were computed based on the segmentation masks. Hence, edges arising from parts of the face such as eyes and lips are labeled as non-boundaries and the boundary detector learns to not return strong boundaries for them. The detected boundaries and the current estimation of the keypoint locations and occlusion state are used to estimate a segmentation for the image. With just two iterations between these two models, we achieve state of the art results for landmark localization, occlusion prediction and face segmentation on the challenging COFW dataset.

Precision/recall of occlusion prediction on the COFW dataset for our model and some baselines are shown in Fig. 1(b). Our model yields substantial gains in the precision of keypoint occlusion prediction. Figure 2 shows example outputs of our model run on several COFW images. Our model produces both a foreground mask and estimates of the keypoint locations and occlusion states.



Figure 2: Examples of landmark localization, occlusion, and segmentation mask estimation for images from the COFW test data produced by our model.

- [1] Xavier P Burgos-Artizzu, Pietro Perona, and Piotr Dollár. Robust face landmark estimation under occlusion. In *ICCV*, 2013.
- [2] Piotr Dollár and C Lawrence Zitnick. Structured forests for fast edge detection. In *ICCV*, pages 1841–1848, 2013.
- [3] Jian Dong, Qiang Chen, Shuicheng Yan, and Alan Yuille. Towards unified object detection and semantic segmentation. In *ECCV*, pages 299–314, 2014.
- [4] Sanja Fidler, Roozbeh Mottaghi, Alan Yuille, and Raquel Urtasun. Bottom-up segmentation for top-down detection. In *CVPR*, pages 3294–3301, 2013.
- [5] Golnaz Ghiasi and Charless C. Fowlkes. Occlusion coherence: Localizing occluded faces with a hierarchical deformable part model. In *CVPR*, pages 1899–1096, 2014.