1. Additional Illustrations

In Fig. 2, we show the images from DUTO dataset for which test-time adaptation results in the largest improvement in accuracy. For all of these images, the initial segmentation, before test-time adaptation (column 3, iteration=0) is either empty or contains a much smaller object than the correct segmentation. Since our loss function penalizes objects that are too small, CNN, during test-time adaptation, is forced to come up with a reasonable size segmentation to improve the loss. Note, in particular, that for the top row, the initial segmentation is completely empty. With dense CRF post-processing [1], an empty solution will never improve to a non-empty one, since dense CRF has the smallest possible loss for the empty solution.

In Fig. 3, we show the image from DUTO dataset for which test-time adaptation results in a largest decline in accuracy. We show the original image, the ground truth, and the results after an increasing number of iterations. The initial segmentation (before test-time adaptation) detects mostly wrong areas, the fruits instead of the signs, as salient. Test-time iterations refine the fruit areas, resulting in a clean segmentation of the fruits, but the small areas of the signs that were detected initially as salient, get filtered out, resulting in a decreased $F_\beta$ score.

More failure examples for our method are in Fig. 1.

References

Figure 2. Images from DUTO dataset that improve the most. From left to right: input image, ground truth, and results after increasing number of iterations of test-time adaptation.

Figure 3. Image from DUTO dataset that degrades the most. From left to right: input image, ground truth, and results after increasing number of iterations of test-time adaptation.