Deep Face Detector Adaptation
without Negative Transfer or Catastrophic Forgetting
- Supplementary Materials -

Muhammad Abdullah Jamal
University of Central Florida
Orlando, FL 32816
a_jamal@Knights.ucf.edu

Haoxiang Li
Adobe Research
San Jose, CA 95110
lhxustcer@gmail.com

Boqing Gong
Tencent AI Lab
Bellevue, WA 98004
boqinggo@outlook.com

Outline

- Training details for the competing methods
- Continuous Scores on FDDB
- Results on COFW
- Qualitative Results

A. Training details of the competing methods

In this section, we give the details of training the competing baselines. The validation sets are used to determine all the free parameters.

- **Fine-tuning [1]:** For Faster-RCNN, we run the experiments for 7,000 iterations in total. The base learning rate is $1e$-4 in the first 4,000 iterations and then reduced to $1e$-5 for the remaining 3,000 iterations. For CascadeCNN, we fine-tune the model with the learning rate of $1e$-4 for 10,000 iterations and another 5,000 iterations with the learning rate of $1e$-5.

- **LWF [2]:** For Faster-RCNN, we train the model for 8,000 iterations. The base learning rate is set to $1e$-4 for the first 6,000 iterations and then reduced to $1e$-5 for the next 2000 iterations. For CascadeCNN, we train the model for 15,000 iterations with the base learning rate of $1e$-4 and reduce it to $1e$-5 at the 10,000th iteration. In both the experiments, the learning rate for the last layer is 10 times the base learning rate of the other layers.

- **GDSDA [3]:** We train the Faster-RCNN, and CascadeCNN with a learning rate of $1e$-4 for 8,000 iterations, and 12,000 iterations respectively.

- **HTL [4]:** We train the CascadeCNN using the learning rate $1e$-4 for 10,000 iterations and another 3,000 iterations using the learning rate $1e$-5.

- *Gradient Reversal [5]:* We train the Faster-RCNN using the base learning rate of $1e$-4 for 20,000 iterations and then reduce it to $1e$-5 for the next 10,000 iterations. Since the training sets of the source domain and the target domain are highly unbalanced, we alternatively take one image from either set to train the *Gradient Reversal*.

B. Continuous scores on FDDB

The FDDB dataset [6] defines both discrete and continuous scores to evaluate the face detection results. We have shown the ROC curves of the discrete scores of different methods in the main text.

Figure 1 and Figure 2 show the ROC curves of the continuous scores on FDDB for the Faster-RCNN and CascadeCNN. The left panels exhibit the curves of the Faster-RCNN and the right panels show the curves for the CascadeCNN. For CasacadeCNN, our approach outperforms all the competing methods in all the settings. For Faster-RCNN, our method can still boost the performance under the supervised setting and under semi-supervised setting when $N = 5$. More importantly, our approach does not incur negative transfer, i.e., the results are either better than or about the same as the source detectors.

It is actually worth pointing out that the annotations are inconsistent between the FDDB dataset and the source where the detectors are trained. As a result, the FDDB under-evaluates the adaptation methods.

C. Results on COFW

In addition to the FDDB dataset used in the main text, we additionally consider COFW [7] as the target domain here. COFW provides 1,345 training faces and 507 testing faces and includes heavy occlusion and large shape variations.

Figure 3 and Figure 4 show the ROC curves of both discrete and continuous scores on COFW for both the Faster-

RCNN and CascadeCNN face detectors under the supervised setting. We can draw the same observation as on the FDDB dataset, that our method shows no negative transfer as compared to other competing methods for both the Faster-RCNN and CascadeCNN detectors. GDSDA is an exception among the competing methods and leads to no negative transfer for CascadeCNN (but not for Faster-RCNN).

We also evaluate the catastrophic forgetting when the detectors are adapted to COFW. Figure 5 shows the performance on the validation set of the WIDER Face for Faster-RCNN. Our approach maintains a good performance in the source domain compared with the original source detector.

D. More qualitative results

We show more qualitative results in Figure 6 and Figure 7. Our method is able to discard some of the false positives from both the source detectors. We can also observe that our method is able to detect the true positives that have not been detected by the source detectors.

References

[1] Ruslan Salakhutdinov and Geoffrey Hinton. Deep boltzmann machines. In *Artificial Intelligence and Statistics*, pages 448–455, 2009. 1

[2] Zhizhong Li and Derek Hoiem. Learning without forgetting. *CoRR*, abs/1606.09282, 2016. 1

[3] Shuang Ao, Xiang Li, and Charles X Ling. Fast generalized distillation for semi-supervised domain adaptation. In *AAAI*, 2017. 1

[4] Ilja Kuzborskij and Francesco Orabona. Stability and hypothesis transfer learning. In *ICML (3)*, pages 942–950, 2013. 1

[5] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *International Conference on Machine Learning*, pages 1180–1189, 2015. 1

[6] Vidit Jain and Erik Learned-Miller. Fddb: A benchmark for face detection in unconstrained settings. Technical report, 2010. 1

[7] Xavier P. Burgos-Artizzu, Pietro Perona, and Piotr Dollár. Robust face landmark estimation under occlusion. In *Proceedings of the 2013 IEEE International Conference on Computer Vision*, ICCV '13, pages 1513–1520, Washington, DC, USA, 2013. IEEE Computer Society. 1

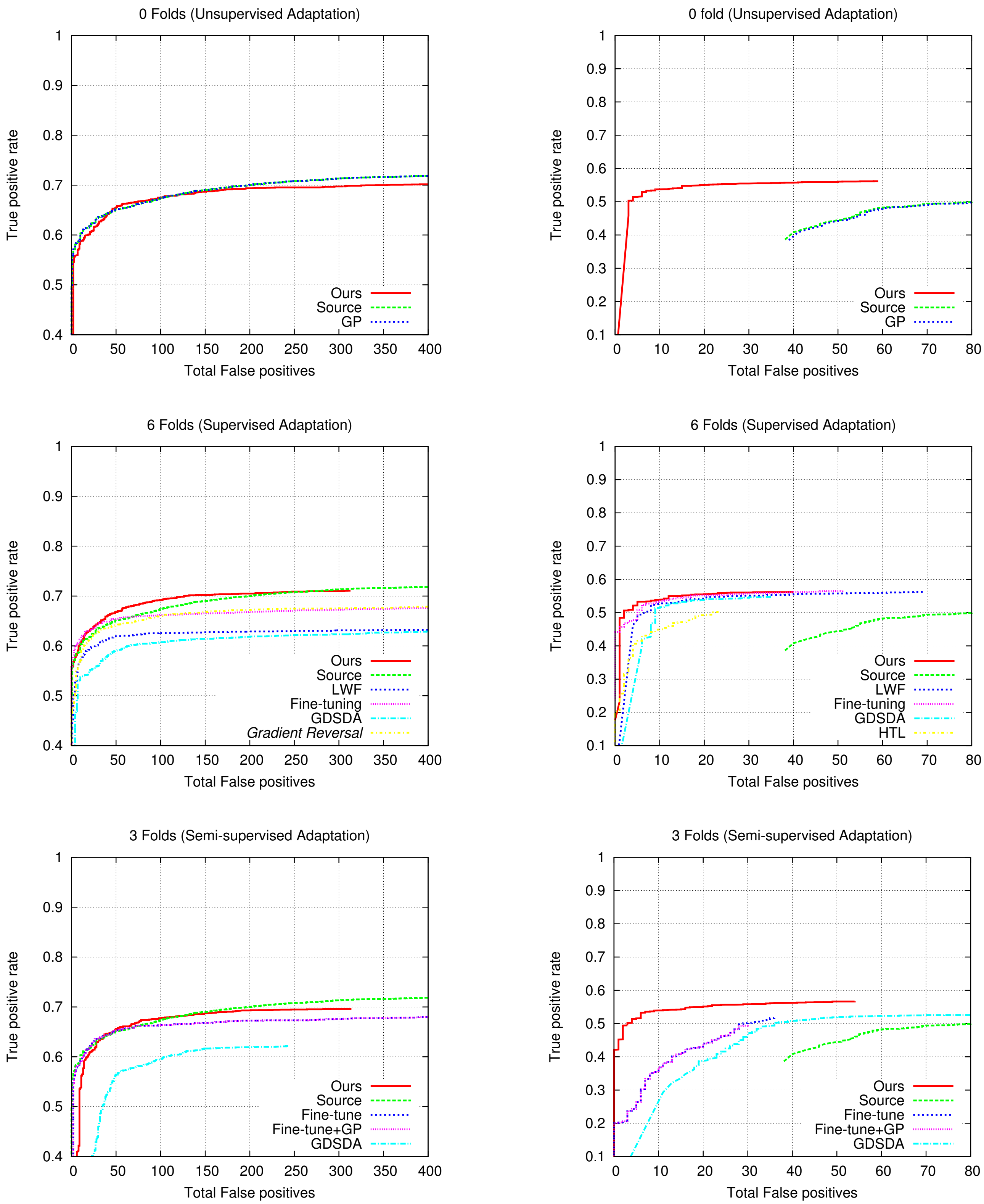


Figure 1. Continuous score results on the FDDB under unsupervised, supervised, and semi-supervised settings (3 out of 6 folds of training images annotated). (Left: Faster-RCNN, Right: CascadeCNN)

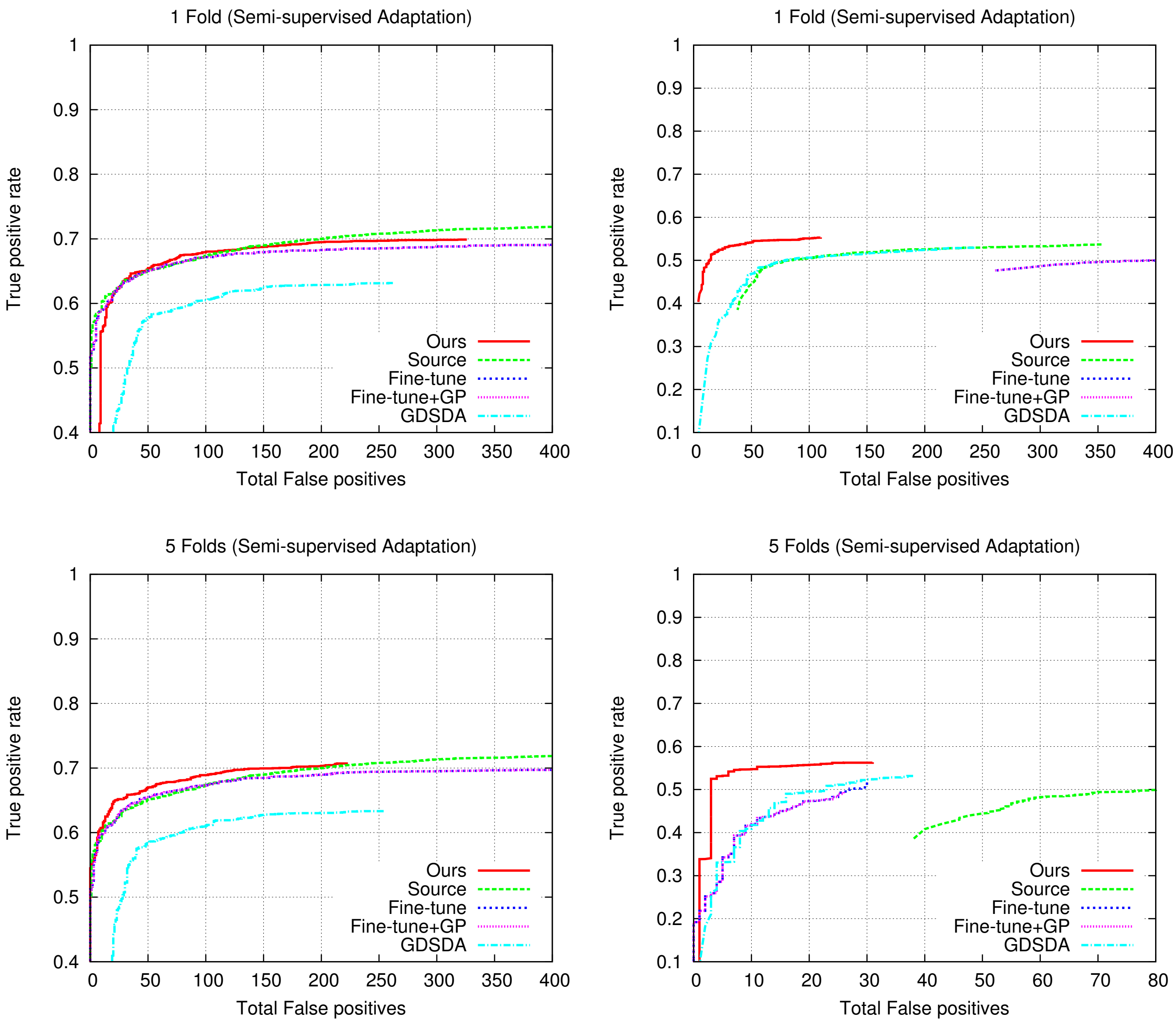


Figure 2. Continuous score results under the semi-supervised settings with $N = \{1, 5\}$ out of 6 folds training images annotated on the FDDB dataset. (Left: Faster-RCNN, Right: CascadeCNN)

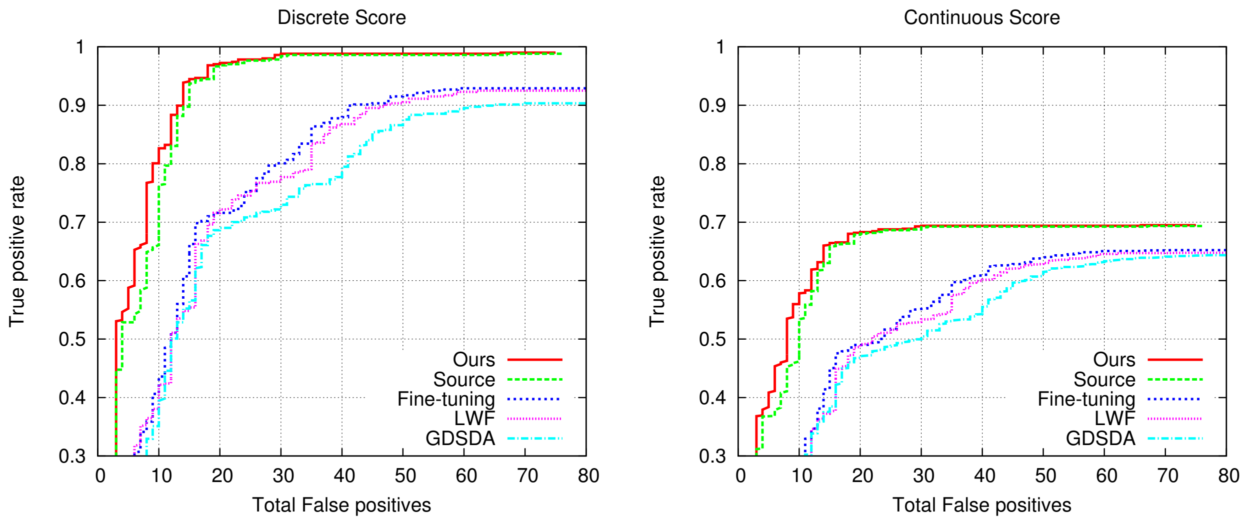


Figure 3. ROC Curves on COFW using **FasterRCNN** (supervised adaptation).

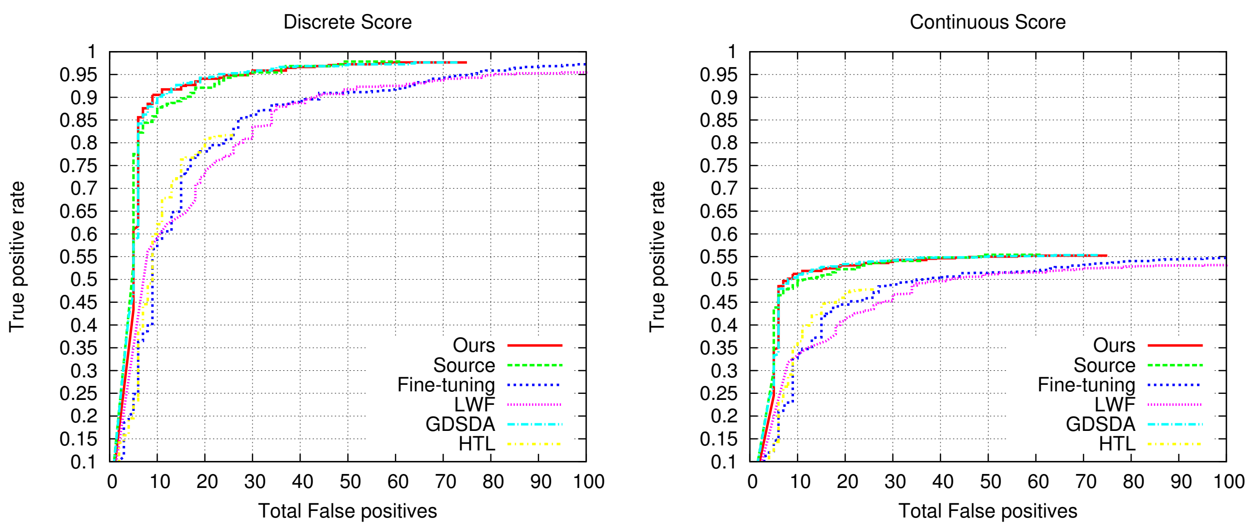
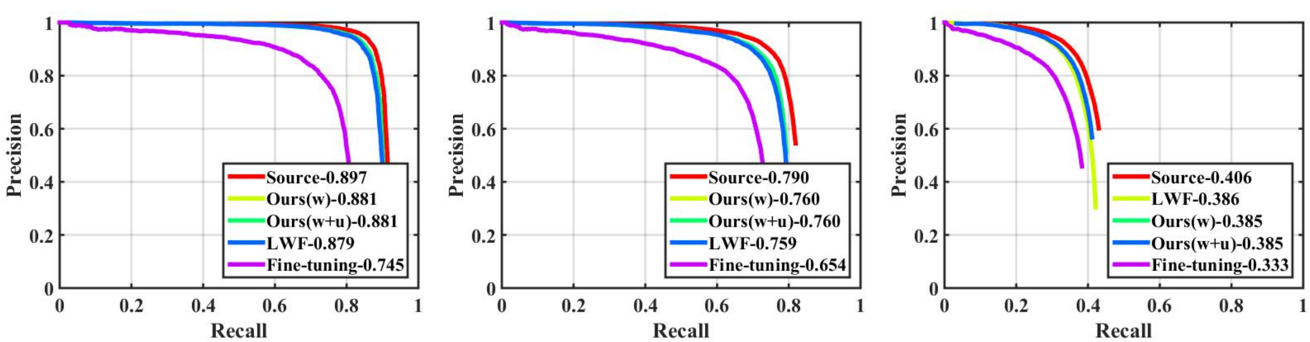


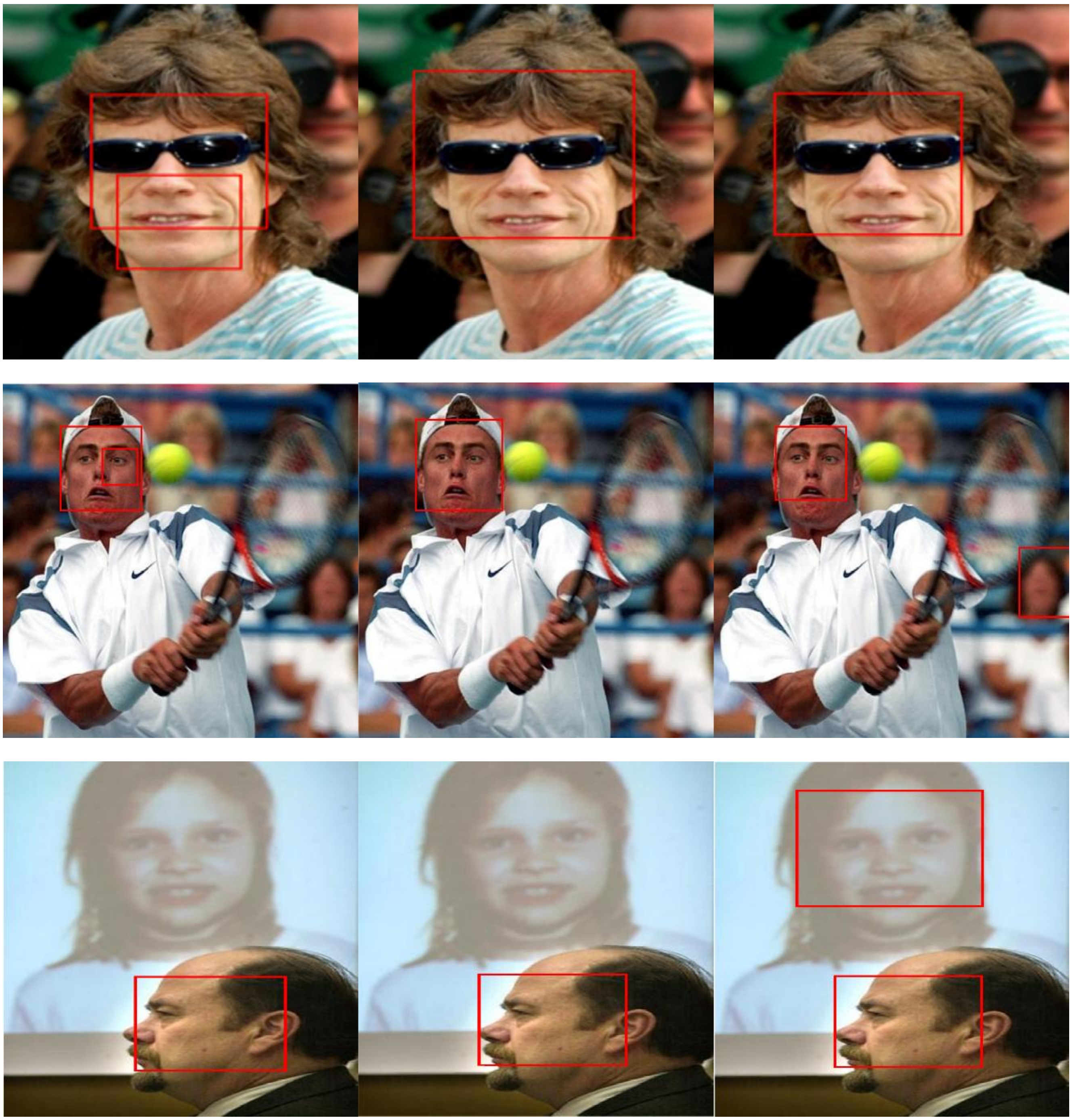
Figure 4. ROC Curves on COFW using **CascadeCNN** (supervised adaptation).



(a) Easy Set (b) Medium Set (c) Hard Set

Figure 5. Evaluation of **catastrophic forgetting** on source domain after supervised adaptation to target domain (COFW): detection results on the validation set of WIDER FACE (Easy, Medium and Hard sets).

Figure 6. Qualitative results of adapting **Faster-RCNN**. The image on the left of each pair shows the detection results by the **source model** and the right image shows **our** method in the supervised adaptation setting.



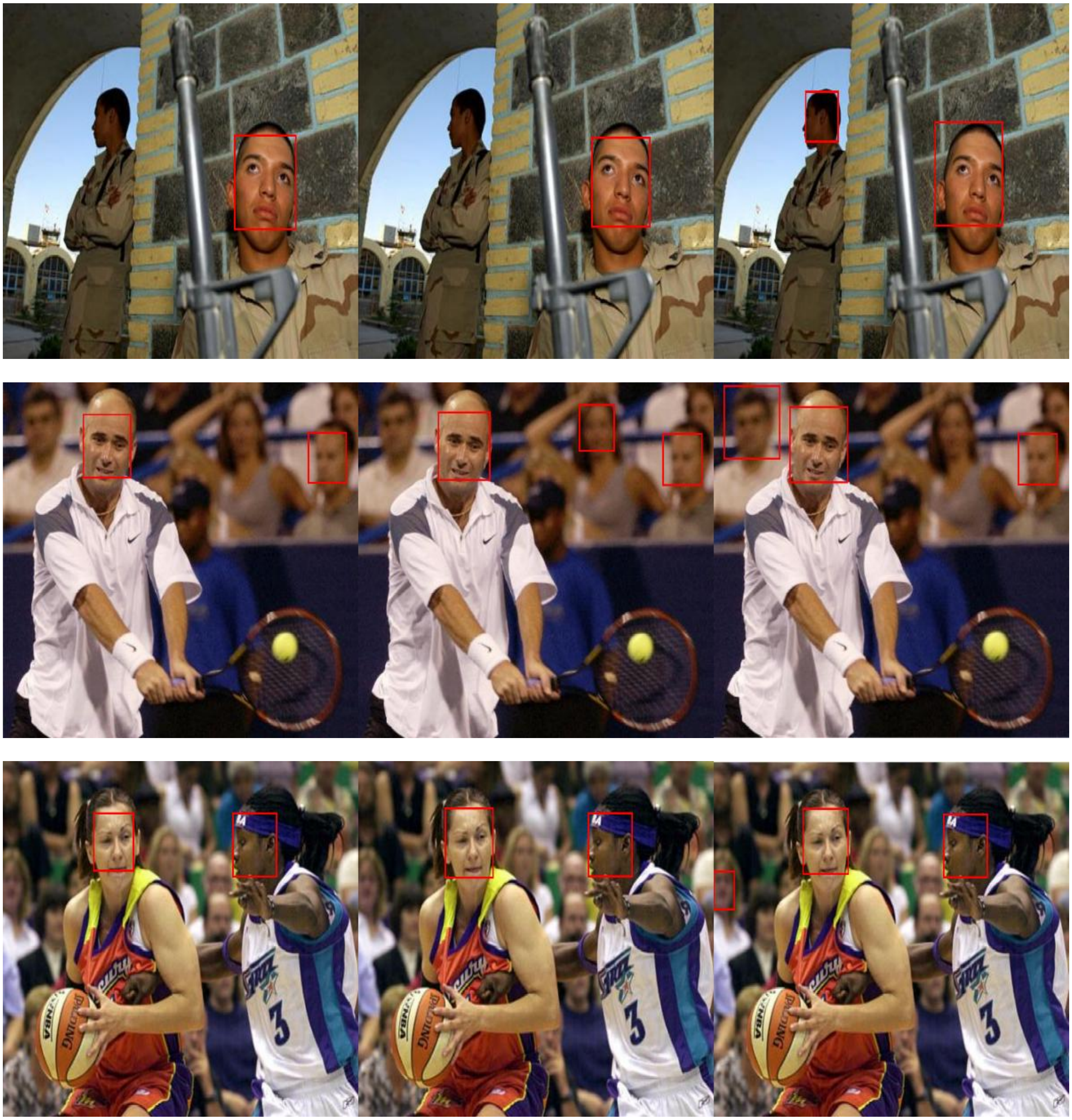


Figure 7. More qualitative results, from left to right are face detection results on FDDB dataset with **a CascadeCNN (1)**, the same detector but adapted by our method to the target domain (FDDB) **with no data annotation (2)**, and **with some data annotation (3)**.