Self-adapting confidence estimation for stereo – Supplementary material

Matteo Poggi Filippo Aleotti Fabio Tosi
Giulio Zaccaroni Stefano Mattoccia

University of Bologna, Viale del Risorgimento 2, Bologna, Italy

Abstract. This document contains additional material concerning ECCV 2020 paper "Self-adaptive confidence estimation for stereo". In order, we first illustrate the color encoding used for both disparity and confidence maps (Sec. 1), then reporting additional qualitative results concerning the ablation experiments (Sec. 2), the generalization across datasets (Sec. 3) and the self-adaptation technique enabled by OTB (Sec. 4).

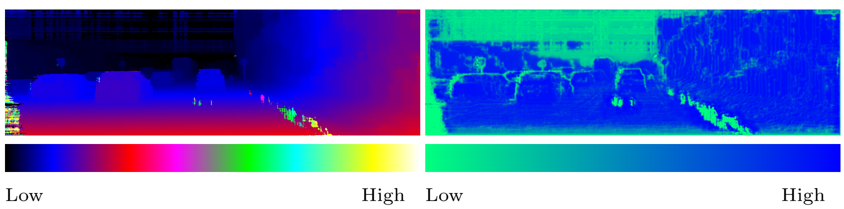

Fig. 1. Colormap encoding for disparity and confidence. We choose the KITTI colormap (on the left) to encode disparity maps and reversed colormap **winter** (on the right) for confidence maps. Best viewed with colors.

1 Color encoding

For visualization purpose, in the submitted paper and this document, we use the KITTI colormap for disparity maps and the reverse **winter** colormap for confidence maps, as shown in Fig. 1.

2 Ablation experiments

To further highlight the impact of different configurations chosen for the MBCE loss, we show qualitative examples of confidence maps obtained by the same network, trained using different $\mathcal{P}, \mathcal{Q}$ sets.

Fig. 2 reports results for the KITTI 2012 stereo pair nr. 000190. We show disparity maps obtained respectively with Census-CBCA (left) and MCCNN-fst-SGM (right) on the first row. Then, we show confidence maps obtained by CCNN for Census-CBCA and ConfNet for MCCNN-fst-SGM, being the two

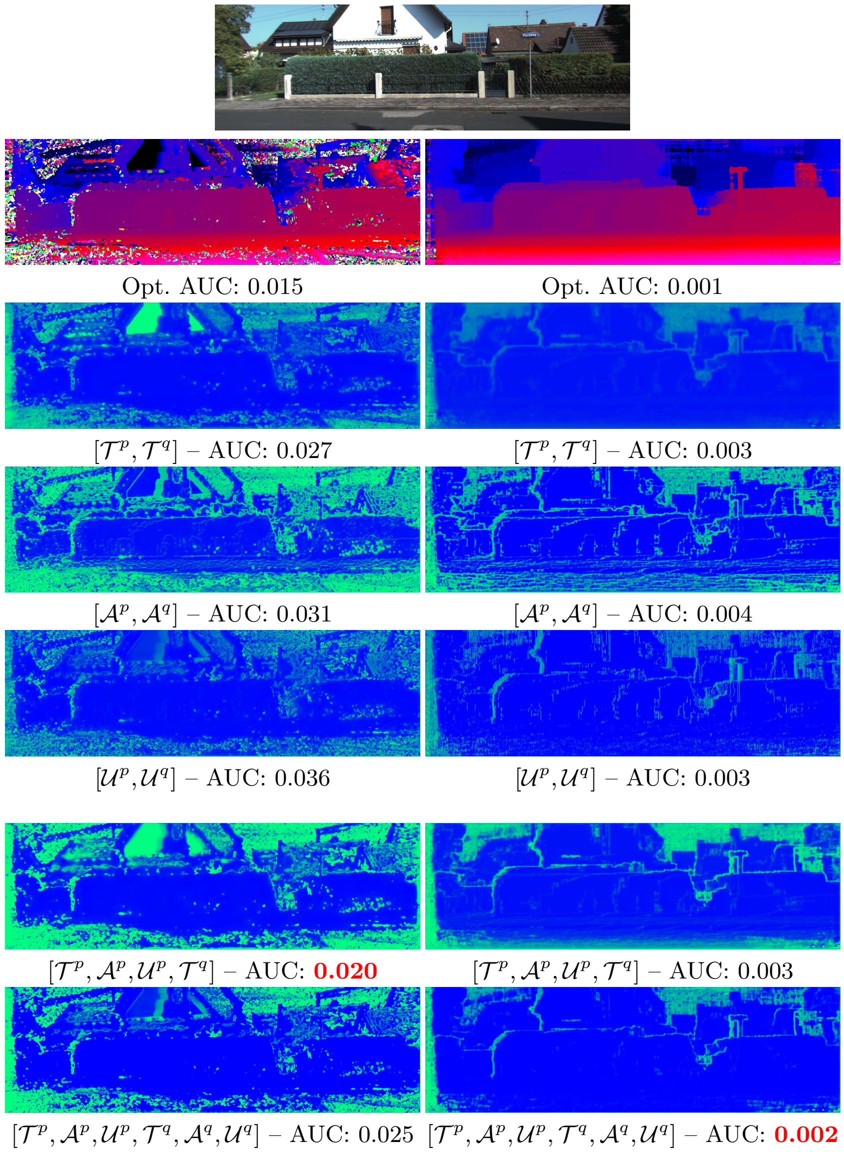

Fig. 2. Ablation study on MBCE loss. We report confidence maps estimated by CCNN with disparity maps from Census-CBCA (left) and by ConfNet with disparity maps from MCCNN-fst-SGM (right), trained with different variants of the MBCE loss.

performing the better according to our experiments in the main paper. We report optimal AUC scores under the disparity map and individual AUC under each corresponding confidence map.

First, we point out the white textureless regions in correspondence of the house in the background. This part leads to a large region of outliers in the case of Census-CBCA (i.e., having disparity equal to zero), while is correctly estimated by MCCNN-fst-SGM. When trained on the single cues (top three confidence maps), CCNN learns to estimate low scores for that portion only thanks to $[\mathcal{T}^p, \mathcal{T}^q]$. When combining the cues (last two rows at the bottom), keeping $\mathcal{T}^q$ only in $\mathcal{Q}$ allows to correctly find the wrong portion in correspondence of the house on Census-CBCA maps, leading to the best AUC score among the other configurations.

On the other hand, considering $\mathcal{T}^q$ only in $\mathcal{Q}$ is less effective when dealing with much more accurate disparity maps, as in the case of MCCNN-fst-SGM algorithm. In this case, using all the cues available enables the network to learn for the most effective confidence estimation.

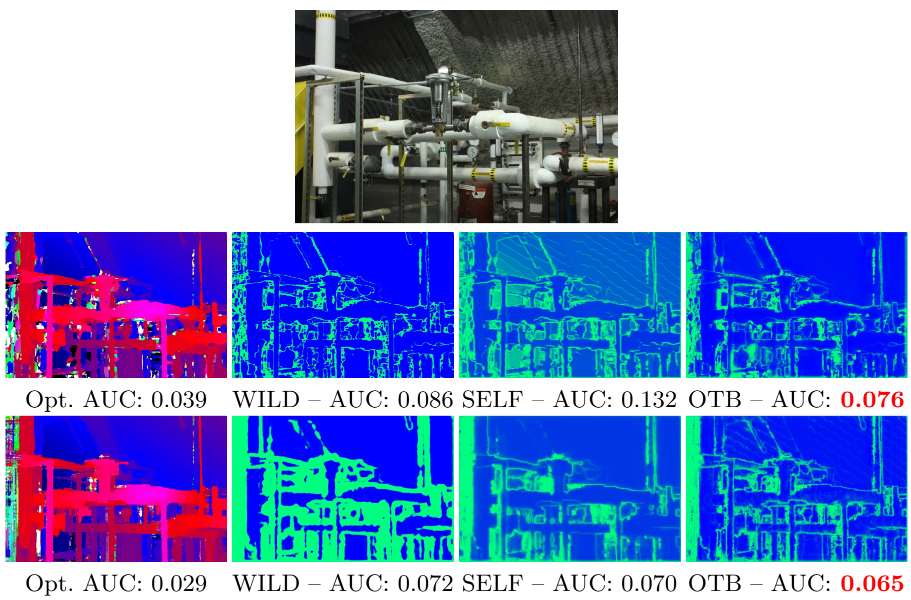

Fig. 3. Qualitative results on Middlebury 2014. We report confidence maps estimated by CCNN on disparity maps from Census-CBCA (top) or by ConfNet on disparity maps from MCCNN-fst-SGM (bottom), trained, from left to right, with WILD [2], SELF [1] and OTB.

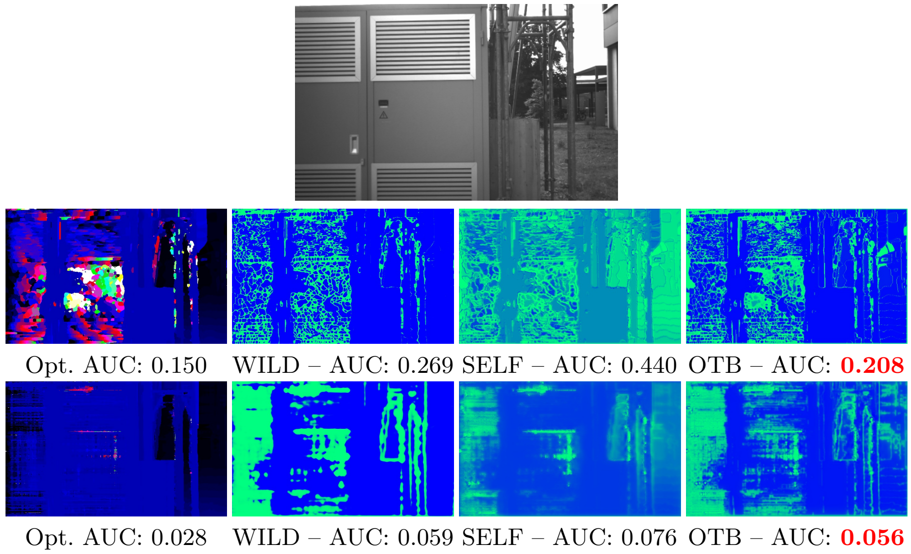

Fig. 4. Qualitative results on ETH3D. We report confidence maps estimated by CCNN on disparity maps from Census-CBCA (top) or by ConfNet on disparity maps from MCCNN-fst-SGM (bottom), trained, from left to right, with WILD [2], SELF [1] and OTB.

3 Generalization

We also report additional qualitative examples concerning the generalization study performed in the main paper. Figures 3 and 4 show two examples respectively from Middlebury 2014 and ETH3D datasets. As for the KITTI example, we report disparity maps by Census-CBCA (left) and MCCNN-fst-SGM (right), together with confidences estimated by CCNN (left) and ConfNet (right) trained with self-supervision strategies WILD, SELF and OTB from top to bottom respectively. We can notice, in both cases, how OTB leads to much sharper confidence maps compared to WILD and SELF supervisions.

4 Self-Adaptation

Moreover, we show some additional examples taken from the DrivingStereo sequence in order to perceive the impact of self-adaptation better. Fig. 5 and 6 collects four frames, starting from the first one and sampling every 2k images. Fig. 5 show disparity maps (second column) by Census-SGM (top) and MADNet (bottom), and the confidence maps estimated by ConfNet trained with OTB, keeping the network frozen (third column) or running self-adaptation (fourth column), while Fig. 6 shows the same for disparity maps by GANet. For this latter, we report four decimals for AUC scores because of the very low amount of outliers.

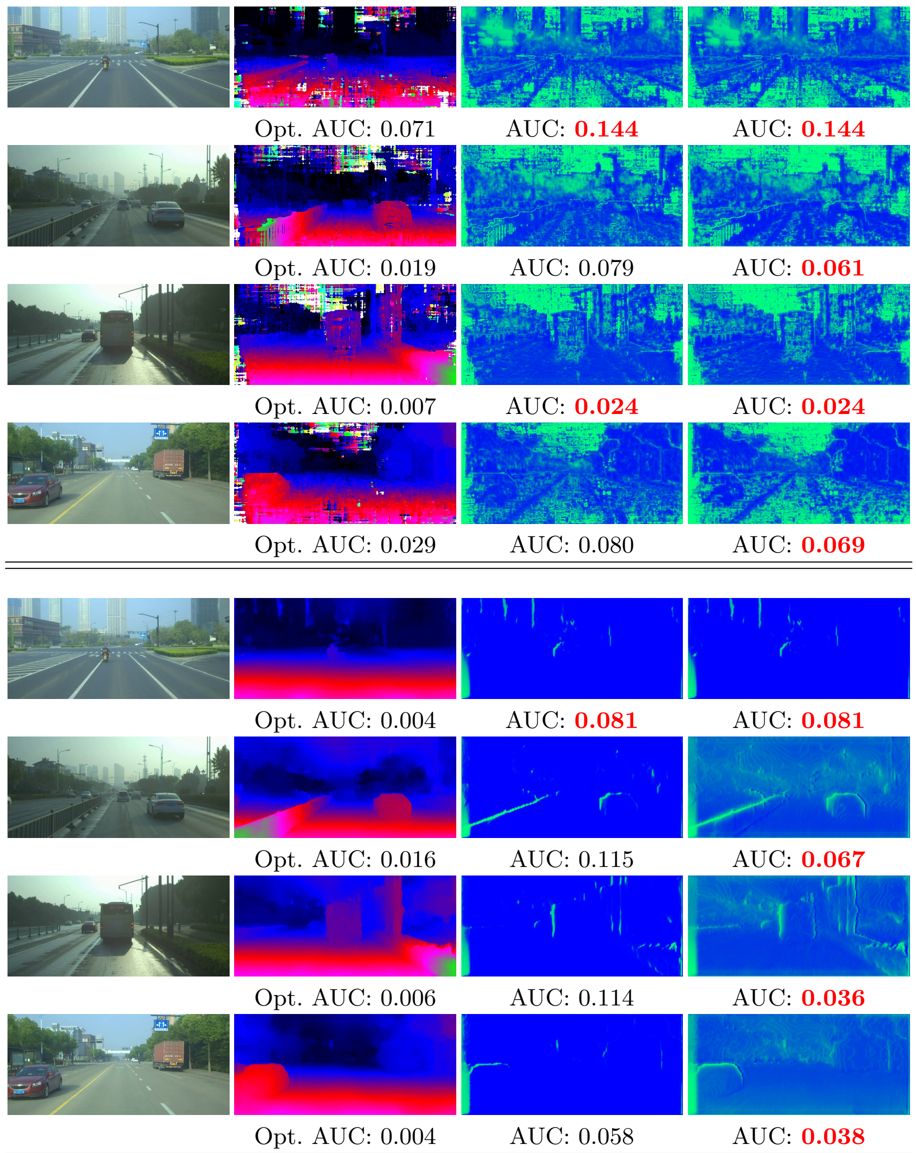

Fig. 5. Online self-adaptation on DrivingStereo. We extract 4 frames from the sequences selected for our experiments, respectively starting from first (top) and sampling every 2000 frames. For each frame, we show from left to right the reference image, disparity map using Census-SGM (top) or MADNet (bottom) and confidence maps estimated by OTB and OTB enabling self-adaptation.

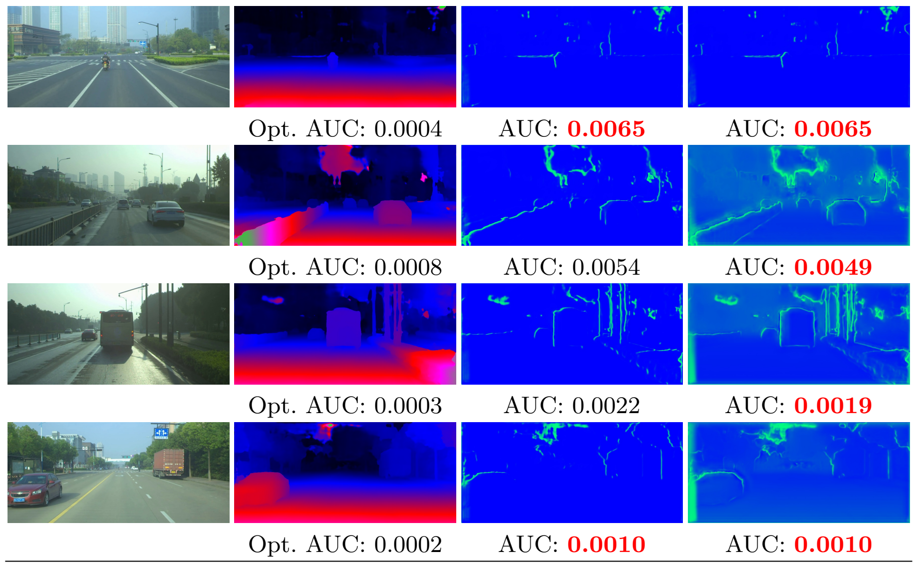

Fig. 6. Online self-adaptation on DrivingStereo, GANet. We extract 4 frames from the sequences selected for our experiments, respectively starting from first (top) and sampling every 2000 frames. For each frame, we show from left to right the reference image, disparity map using GANet and confidence maps estimated by OTB and OTB enabling self-adaptation.

It is evident how, after starting with the same performance on the first frame, self-adaptation rapidly leads to more effective confidence estimation in terms of AUC. This behaviour can also be perceived qualitatively: we can notice on Census-SGM how self-adaptation allows for much sharper confidence maps compare to the frozen confidence network. The same can also be perceived with MADNet and GANet; confidence maps become much more detailed self-adapting after each frame while the frozen confidence network only detects outliers in the left border and occluded regions.

References

1. Mostegel, C., Rumpler, M., Fraundorfer, F., Bischof, H.: Using self-contradiction to learn confidence measures in stereo vision. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (June 2016)
2. Tosi, F., Poggi, M., Tonioni, A., Di Stefano, L., Mattoccia, S.: Learning confidence measures in the wild. In: Proceedings of the British Machine Vision Conference (BMVC). BMVA (Sept 2017)