

# Appendix

Kai Zhang<sup>1</sup>  
zhangkai193@mails.ucas.edu.cn

Yifan Sun<sup>2</sup>  
sunnyf15@tsinghua.org.cn

Rui Wang<sup>1</sup>  
wangrui@iscas.ac.cn

Haichang Li<sup>1</sup>  
haichang@iscas.ac.cn

Xiaohui Hu<sup>1</sup>  
hxxh@iscas.ac.cn

<sup>1</sup> Institute of Software,  
Chinese Academy of Sciences, China

<sup>2</sup> Baidu Research, China

## 1 Hyperparameter analysis

$\lambda_{cross}$	0.1	0.3	0.5	0.7
mIoU	55.1	55.3	<b>55.7</b>	55.4

Table 1: Influence of weight parameter  $\lambda_{cross}$ .

**The ratio between online and offline pseudo label supervision.** MFA is novel for its online and offline pseudo label co-supervision. The ratio between the two supervisions is important and is determined by  $\lambda_{cross}$ , given the weighting factor  $\lambda_{self}$  fixed to 1. We vary  $\lambda_{cross}$  from 0.1 to 0.7 in Table 1. It is observed that the achieved mIoU gradually increases, as  $\lambda_{cross}$  increases to 0.5. It indicates that increasing the supervising with online pseudo labels benefits MFA (when  $\lambda_{cross} \leq 0.5$ ). However, further increasing  $\lambda_{cross}$  slightly deteriorates MFA. We infer that it is because the online pseudo labels are prone to instability. Consequentially, when they dominates the supervision of MFA, the learning effect is compromised. With the above two observations, we recommend setting  $\lambda_{cross} = 0.5$  in MFA.

$\rho_{min}$	0.1	0.2	0.3
mIoU	55.1	<b>55.7</b>	55.4

Table 2: Influence of minimum proportion  $\rho_{min}$  of online pseudo labels given  $\rho_{max} = 0.7$

**The minimum and maximum proportion of Online-CBST algorithm.** The minimum proportion  $\rho_{min}$  and maximum proportion  $\rho_{max}$  are the hyperparameters of proposed Online-CBST algorithm. On GTA5-to-Cityscapes, the sensitivities of them are analyzed by fixing one of them and then changing the value of the other. Firstly, We vary  $\rho_{min}$  from 0.1 to 0.3

$\rho_{max}$	0.5	0.7	0.9
mIoU	55.4	<b>55.7</b>	55.6

Table 3: Influence of maximum proportion  $\rho_{max}$  of online pseudo labels given  $\rho_{min} = 0.2$

by setting  $\rho_{max} = 0.7$  (only the results under best setting are reported), as shown in Table 2. The results show that the proposed method achieves the best performance at  $\rho_{min} = 0.2$ . We noticed that the model gets better performance when  $\rho_{min}$  is smaller, which proves our incremental  $\rho(t)$  strategy is effective. Secondly, in order to verify the influence of  $\rho_{max}$  on the model, we change the  $\rho_{max}$  from 0.5 to 0.9, given  $\rho_{min} = 0.2$ . The results indicate that the proposed method is relatively robust to  $\rho_{max}$ , as shown in Table 3. Specifically, the model achieves the best performance at  $\rho_{max} = 0.7$  and the performances are not significantly reduced under other settings.