A. Proofs

Proof of Proposition 3.1 The proof is analogous to [30]. The L_1 calibration error for binary classification task is given by $\mathrm{CE}_{cls} := \mathbb{E}[|\mathbb{E}[Y = 1|\hat{S}] - \hat{S}|]$.

Proof.

$$d_{cls} = |\mathbb{E}[Y - \hat{S}]| \tag{16}$$

$$= |\mathbb{E}[Y - \mathbb{E}[[Y = 1] \mid \hat{S}]] + \mathbb{E}[\mathbb{E}[[Y = 1] \mid \hat{S}] - \hat{S}]| \tag{17}$$

$$= \mathbb{E}[Y] - \underbrace{\mathbb{E}[\mathbb{E}[[Y = 1] \mid \hat{S}]]}_{=\mathbb{E}[Y]} + \mathbb{E}[\mathbb{E}[[Y = 1] \mid \hat{S}] - \hat{S}]| \tag{18}$$

$$\leq \underbrace{\mathbb{E}[|\mathbb{E}[[Y = 1] \mid \hat{S}] - \hat{S})|]}_{=\mathrm{CE}_{cls}}. \tag{19}$$

The inequality in (19) is obtained by utilizing the convexity of the absolute value and applying Jensen's inequality. □

Proof of Proposition 3.2

Proof.

$$\mathrm{CE}_{det} = \mathbb{E}[|\mathbb{E}[\mathrm{IoU} \mid \hat{S}] - \hat{S}|] \tag{20}$$

$$= \mathbb{E}[|\mathbb{E}[\mathrm{IoU} \mid \hat{S}] - \mathbb{E}[\hat{S} \mid \hat{S}|] \tag{21}$$

$$\leq \mathbb{E}[\mathbb{E}[|\,\mathrm{IoU} - \hat{S}| \mid \hat{S}]] \tag{22}$$

$$= \mathbb{E}[|\,\mathrm{IoU} - \hat{S}|] = d_{det} \tag{23}$$

□

As before, (22) is obtained by applying Jensen's inequality.

B. Experiments

In this section we will provide additional quantitative and qualitative results, following the structure of the experiments in the main text.

We compare the performance of the binning-based LaECE metric with our estimator $\widehat{\mathrm{CE}}$, for measuring calibration error as obtained by using an identity link in Equation (4). In Figure 4 we can observe that both metrics perform equally well and converge to the ground truth value after 5000 points.

In Tables 7 and 8, we investigate the performance of our estimator in a calibration regularized training framework on Cityscapes and Pascal VOC datasets, and compare it with competing post-hoc and trainable methods. We report two versions of AP, $\widehat{\mathrm{CE}}$, and D-ECE: averaged over 10 thresholds and evaluated at $\mathrm{IoU} = 0.5$. Additionally, we include $\widehat{\mathrm{CE}}$ with an identity link. The models with $\widehat{\mathrm{CE}}$ and TCD

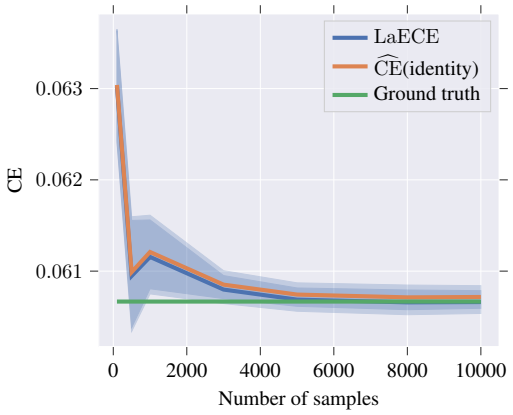

Figure 4. Comparison of LaECE vs. $\widehat{\mathrm{CE}}$ (identity) as a function of the number of points used for the estimation.

auxiliary loss are trained from scratch. We show the results for several values of λ and observe that by increasing the weight of the regularization, the calibration error is reduced, as expected. In all evaluations we use a detection score threshold $\gamma = 0.5$, and the results are averaged across three seeds.

In Tables 9 and 10 we summarize the AP and CE metrics for fine-tuning on Cityscapes and COCO, respectively. The learning rate for fine-tuning is set to the one from the last training iteration, i.e., $1\mathrm{e}{-3}$ for Cityscapes and $1\mathrm{e}{-4}$ for COCO and Pascal VOC.

Figure 5 shows qualitative results on COCO, comparing uncalibrated F-RCNN model with a fine-tuned model using $\widehat{\mathrm{CE}}$ loss. Our estimator effectively adjusts the confidence scores to achieve better alignment with the IoU overlap, thus also reducing the number of wrong detections.

Table 7. Calibration and detection performance of models trained on Cityscapes.

Model	AP	AP_{50}	$\widehat{CE}$	$\widehat{CE}_{50}$	$\widehat{CE}$ (identity)	D-ECE	$D\text{-}ECE_{50}$
F-RCNN	$35.36_{\pm 0.65}$	$54.46_{\pm 0.82}$	$37.45_{\pm 0.46}$	$17.00_{\pm 0.46}$	$28.48_{\pm 0.68}$	$37.83_{\pm 0.76}$	$16.65_{\pm 0.91}$
F-RCNN + TS	$32.29_{\pm 0.25}$	$49.55_{\pm 0.27}$	$26.36_{\pm 0.86}$	$13.59_{\pm 0.33}$	$11.90_{\pm 1.22}$	$26.05_{\pm 0.62}$	$13.71_{\pm 0.48}$
F-RCNN + $\widehat{CE}$ ($\lambda = 1$)	$35.12_{\pm 0.42}$	$54.52_{\pm 0.46}$	$36.62_{\pm 0.90}$	$16.06_{\pm 1.22}$	$27.08_{\pm 1.17}$	$36.41_{\pm 0.66}$	$14.96_{\pm 0.96}$
F-RCNN + $\widehat{CE}$ ($\lambda = 2$)	$35.46_{\pm 0.07}$	$53.58_{\pm 0.32}$	$36.61_{\pm 0.45}$	$16.24_{\pm 0.32}$	$27.62_{\pm 0.74}$	$37.20_{\pm 0.44}$	$16.33_{\pm 0.22}$
F-RCNN + $\widehat{CE}$ ($\lambda = 3$)	$35.04_{\pm 0.26}$	$53.91_{\pm 0.12}$	$36.99_{\pm 0.33}$	$16.61_{\pm 0.27}$	$27.30_{\pm 0.69}$	$36.84_{\pm 0.60}$	$15.64_{\pm 0.39}$
F-RCNN + $\widehat{CE}$ ($\lambda = 4$)	$35.45_{\pm 0.30}$	$53.97_{\pm 0.05}$	$35.93_{\pm 0.75}$	$15.33_{\pm 0.79}$	$26.43_{\pm 1.23}$	$36.34_{\pm 0.89}$	$15.24_{\pm 1.01}$
RetinaNet	$34.60_{\pm 0.23}$	$52.60_{\pm 0.81}$	$32.22_{\pm 0.29}$	$12.91_{\pm 0.95}$	$23.25_{\pm 0.39}$	$32.51_{\pm 0.18}$	$12.54_{\pm 0.66}$
RetinaNet + TS	$34.60_{\pm 0.23}$	$52.60_{\pm 0.81}$	$28.72_{\pm 1.25}$	$11.63_{\pm 1.19}$	$18.67_{\pm 2.07}$	$28.92_{\pm 1.18}$	$11.05_{\pm 1.25}$
RetinaNet + TCD	$33.94_{\pm 0.81}$	$51.83_{\pm 1.44}$	$33.79_{\pm 0.99}$	$13.35_{\pm 0.88}$	$25.01_{\pm 1.13}$	$33.90_{\pm 0.87}$	$13.04_{\pm 0.47}$
RetinaNet + $\widehat{CE}$ ($\lambda = 0.05$)	$32.59_{\pm 0.71}$	$49.22_{\pm 1.38}$	$28.77_{\pm 1.94}$	$11.90_{\pm 0.19}$	$17.33_{\pm 3.15}$	$28.55_{\pm 2.09}$	$11.07_{\pm 0.22}$
RetinaNet + $\widehat{CE}$ ($\lambda = 0.1$)	$33.03_{\pm 0.39}$	$50.95_{\pm 0.75}$	$30.21_{\pm 0.45}$	$10.90_{\pm 0.28}$	$20.15_{\pm 0.11}$	$30.18_{\pm 0.49}$	$10.30_{\pm 0.35}$
FCOS	$34.81_{\pm 0.08}$	$52.31_{\pm 0.29}$	$26.29_{\pm 0.98}$	$13.91_{\pm 1.28}$	$14.43_{\pm 1.31}$	$25.83_{\pm 0.98}$	$13.23_{\pm 1.18}$
FCOS + TS	$33.62_{\pm 0.44}$	$50.23_{\pm 1.17}$	$26.45_{\pm 1.23}$	$13.06_{\pm 0.86}$	$13.33_{\pm 2.16}$	$25.46_{\pm 1.38}$	$11.96_{\pm 0.90}$
FCOS + TCD	$35.57_{\pm 0.23}$	$54.21_{\pm 0.38}$	$28.53_{\pm 0.82}$	$15.09_{\pm 0.83}$	$16.85_{\pm 1.27}$	$27.78_{\pm 0.77}$	$13.68_{\pm 0.86}$
FCOS + $\widehat{CE}$ ($\lambda = 0.05$)	$34.74_{\pm 0.55}$	$52.80_{\pm 0.69}$	$26.39_{\pm 0.48}$	$15.27_{\pm 1.08}$	$15.20_{\pm 0.75}$	$26.35_{\pm 0.54}$	$14.78_{\pm 0.98}$
FCOS + $\widehat{CE}$ ($\lambda = 0.1$)	$34.47_{\pm 0.52}$	$52.34_{\pm 0.71}$	$26.88_{\pm 0.33}$	$16.73_{\pm 1.46}$	$15.63_{\pm 0.47}$	$26.74_{\pm 0.09}$	$15.75_{\pm 1.71}$
FCOS + $\widehat{CE}$ ($\lambda = 0.5$)	$33.73_{\pm 0.57}$	$51.23_{\pm 1.12}$	$25.24_{\pm 0.85}$	$18.15_{\pm 1.07}$	$13.07_{\pm 0.83}$	$24.60_{\pm 0.49}$	$17.45_{\pm 0.61}$
FCOS + $\widehat{CE}$ ($\lambda = 1$)	$32.91_{\pm 0.61}$	$50.09_{\pm 0.66}$	$26.30_{\pm 0.28}$	$19.88_{\pm 0.92}$	$14.73_{\pm 0.63}$	$25.80_{\pm 0.31}$	$18.80_{\pm 0.79}$

Table 8. Calibration and detection performance of models trained on Pascal VOC.

Model	AP	AP_{50}	$\widehat{CE}$	$\widehat{CE}_{50}$	$\widehat{CE}$ (identity)	D-ECE	$D\text{-}ECE_{50}$
F-RCNN	$52.96_{\pm 0.05}$	$75.88_{\pm 0.08}$	$34.88_{\pm 0.10}$	$17.03_{\pm 0.16}$	$28.79_{\pm 0.12}$	$35.56_{\pm 0.12}$	$16.40_{\pm 0.13}$
F-RCNN + TS	$49.25_{\pm 0.07}$	$70.93_{\pm 0.09}$	$21.43_{\pm 0.03}$	$11.52_{\pm 0.18}$	$8.85_{\pm 0.08}$	$21.43_{\pm 0.03}$	$11.37_{\pm 0.23}$
F-RCNN + $\widehat{CE}$ ($\lambda = 2$)	$52.34_{\pm 0.05}$	$74.90_{\pm 0.03}$	$34.32_{\pm 0.19}$	$15.95_{\pm 0.20}$	$28.00_{\pm 0.22}$	$35.05_{\pm 0.18}$	$15.40_{\pm 0.21}$
F-RCNN + $\widehat{CE}$ ($\lambda = 3$)	$51.95_{\pm 0.01}$	$74.56_{\pm 0.07}$	$34.14_{\pm 0.06}$	$15.59_{\pm 0.08}$	$27.70_{\pm 0.09}$	$34.81_{\pm 0.11}$	$14.99_{\pm 0.12}$
F-RCNN + $\widehat{CE}$ ($\lambda = 4$)	$51.50_{\pm 0.06}$	$74.20_{\pm 0.11}$	$34.23_{\pm 0.17}$	$15.51_{\pm 0.23}$	$27.70_{\pm 0.21}$	$34.95_{\pm 0.16}$	$14.95_{\pm 0.24}$
RetinaNet	$53.20_{\pm 0.02}$	$73.36_{\pm 0.02}$	$24.11_{\pm 0.09}$	$7.95_{\pm 0.19}$	$16.97_{\pm 0.07}$	$23.92_{\pm 0.08}$	$6.92_{\pm 0.15}$
RetinaNet + TS	$53.20_{\pm 0.02}$	$73.36_{\pm 0.02}$	$24.17_{\pm 0.20}$	$7.97_{\pm 0.21}$	$17.05_{\pm 0.27}$	$24.00_{\pm 0.22}$	$6.98_{\pm 0.20}$
RetinaNet + TCD	$53.50_{\pm 0.07}$	$73.74_{\pm 0.10}$	$25.98_{\pm 0.03}$	$9.25_{\pm 0.15}$	$19.58_{\pm 0.02}$	$25.90_{\pm 0.01}$	$8.12_{\pm 0.19}$
RetinaNet + $\widehat{CE}$ ($\lambda = 0.1$)	$52.95_{\pm 0.00}$	$72.89_{\pm 0.08}$	$23.69_{\pm 0.01}$	$7.93_{\pm 0.02}$	$16.49_{\pm 0.03}$	$23.52_{\pm 0.02}$	$7.01_{\pm 0.02}$
RetinaNet + $\widehat{CE}$ ($\lambda = 0.5$)	$52.06_{\pm 0.16}$	$71.93_{\pm 0.28}$	$21.21_{\pm 0.04}$	$7.67_{\pm 0.14}$	$11.97_{\pm 0.05}$	$21.06_{\pm 0.04}$	$7.20_{\pm 0.11}$
RetinaNet + $\widehat{CE}$ ($\lambda = 1$)	$50.61_{\pm 0.14}$	$70.38_{\pm 0.20}$	$19.63_{\pm 0.15}$	$11.01_{\pm 0.06}$	$7.41_{\pm 0.18}$	$19.55_{\pm 0.15}$	$10.77_{\pm 0.04}$
FCOS	$52.13_{\pm 0.02}$	$73.50_{\pm 0.07}$	$22.32_{\pm 0.05}$	$15.77_{\pm 0.10}$	$12.73_{\pm 0.11}$	$22.08_{\pm 0.07}$	$14.46_{\pm 0.04}$
FCOS + TS	$49.71_{\pm 0.08}$	$69.30_{\pm 0.06}$	$19.23_{\pm 0.06}$	$9.91_{\pm 0.03}$	$7.09_{\pm 0.21}$	$19.19_{\pm 0.08}$	$9.82_{\pm 0.05}$
FCOS + TCD	$52.26_{\pm 0.03}$	$73.58_{\pm 0.07}$	$22.55_{\pm 0.16}$	$15.18_{\pm 0.19}$	$12.61_{\pm 0.22}$	$22.21_{\pm 0.18}$	$13.65_{\pm 0.16}$
FCOS + $\widehat{CE}$ ($\lambda = 0.1$)	$52.05_{\pm 0.04}$	$73.15_{\pm 0.17}$	$22.14_{\pm 0.07}$	$16.15_{\pm 0.08}$	$12.56_{\pm 0.09}$	$21.98_{\pm 0.07}$	$15.13_{\pm 0.14}$
FCOS + $\widehat{CE}$ ($\lambda = 0.5$)	$51.46_{\pm 0.05}$	$72.60_{\pm 0.05}$	$22.00_{\pm 0.12}$	$16.89_{\pm 0.17}$	$12.65_{\pm 0.20}$	$21.75_{\pm 0.07}$	$15.75_{\pm 0.06}$
FCOS + $\widehat{CE}$ ($\lambda = 1$)	$51.02_{\pm 0.06}$	$72.10_{\pm 0.08}$	$21.90_{\pm 0.11}$	$18.03_{\pm 0.04}$	$13.00_{\pm 0.23}$	$21.84_{\pm 0.11}$	$17.34_{\pm 0.03}$

Table 9. Comparison of performance before and after finetuning on Cityscapes for 3 epochs and increasing values of λ.

Model	AP	AP_{50}	$\widehat{CE}$	$\widehat{CE}_{50}$	$\widehat{CE}$ (identity)	D-ECE	$D\text{-}ECE_{50}$
F-RCNN	$35.36_{\pm 0.65}$	$54.46_{\pm 0.82}$	$37.45_{\pm 0.46}$	$17.00_{\pm 0.46}$	$28.48_{\pm 0.68}$	$37.83_{\pm 0.76}$	$16.65_{\pm 0.91}$
F-RCNN + $\widehat{CE}$ ($\lambda = 1$)	$35.06_{\pm 0.51}$	$53.87_{\pm 0.53}$	$35.90_{\pm 0.81}$	$15.22_{\pm 0.37}$	$25.76_{\pm 0.85}$	$35.29_{\pm 1.00}$	$14.48_{\pm 0.78}$
F-RCNN + $\widehat{CE}$ ($\lambda = 2$)	$34.65_{\pm 0.18}$	$52.84_{\pm 0.13}$	$34.03_{\pm 0.98}$	$14.23_{\pm 0.64}$	$23.69_{\pm 1.06}$	$33.86_{\pm 0.96}$	$13.35_{\pm 0.79}$
F-RCNN + $\widehat{CE}$ ($\lambda = 3$)	$33.90_{\pm 0.30}$	$51.52_{\pm 0.49}$	$33.99_{\pm 0.79}$	$14.80_{\pm 0.32}$	$23.19_{\pm 0.98}$	$33.39_{\pm 0.90}$	$13.33_{\pm 0.62}$
F-RCNN + $\widehat{CE}$ ($\lambda = 4$)	$33.30_{\pm 0.30}$	$50.31_{\pm 0.42}$	$33.10_{\pm 0.64}$	$14.19_{\pm 0.24}$	$22.19_{\pm 0.88}$	$32.91_{\pm 0.90}$	$13.69_{\pm 0.39}$
F-RCNN + $\widehat{CE}$ ($\lambda = 5$)	$33.02_{\pm 0.26}$	$50.05_{\pm 0.76}$	$32.37_{\pm 0.92}$	$14.23_{\pm 0.33}$	$20.62_{\pm 1.46}$	$32.13_{\pm 1.00}$	$13.32_{\pm 0.40}$

Table 10. Finetuning for three epochs on COCO with several values of λ.

Model	AP	AP_{50}	$\widehat{CE}$	$\widehat{CE}_{50}$	$\widehat{CE}$ (identity)	D-ECE	$D\text{-}ECE_{50}$
F-RCNN	$36.11_{\pm 0.10}$	$53.35_{\pm 0.15}$	$37.33_{\pm 0.09}$	$20.48_{\pm 0.10}$	$31.76_{\pm 0.05}$	$38.87_{\pm 0.08}$	$21.34_{\pm 0.14}$
F-RCNN + TS	$32.86_{\pm 0.03}$	$48.07_{\pm 0.05}$	$24.35_{\pm 0.12}$	$11.97_{\pm 0.21}$	$13.49_{\pm 0.19}$	$25.05_{\pm 0.13}$	$11.85_{\pm 0.11}$
F-RCNN + $\widehat{CE}$ ($\lambda = 0.01$)	$36.11_{\pm 0.13}$	$53.17_{\pm 0.12}$	$37.47_{\pm 0.40}$	$20.65_{\pm 0.52}$	$31.86_{\pm 0.48}$	$38.88_{\pm 0.32}$	$21.36_{\pm 0.44}$
F-RCNN + $\widehat{CE}$ ($\lambda = 0.05$)	$36.09_{\pm 0.15}$	$53.24_{\pm 0.21}$	$37.07_{\pm 0.10}$	$20.29_{\pm 0.09}$	$31.44_{\pm 0.07}$	$38.57_{\pm 0.01}$	$21.15_{\pm 0.07}$
F-RCNN + $\widehat{CE}$ ($\lambda = 0.1$)	$35.98_{\pm 0.12}$	$53.08_{\pm 0.17}$	$36.78_{\pm 0.15}$	$19.99_{\pm 0.17}$	$31.07_{\pm 0.14}$	$38.19_{\pm 0.05}$	$20.84_{\pm 0.10}$
F-RCNN + $\widehat{CE}$ ($\lambda = 0.5$)	$34.72_{\pm 0.09}$	$51.52_{\pm 0.13}$	$33.56_{\pm 0.09}$	$16.63_{\pm 0.08}$	$26.91_{\pm 0.14}$	$34.65_{\pm 0.09}$	$17.10_{\pm 0.14}$
F-RCNN + $\widehat{CE}$ ($\lambda = 1$)	$33.08_{\pm 0.04}$	$49.37_{\pm 0.06}$	$30.69_{\pm 0.17}$	$14.34_{\pm 0.24}$	$23.25_{\pm 0.16}$	$31.77_{\pm 0.02}$	$14.88_{\pm 0.10}$
F-RCNN + $\widehat{CE}$ ($\lambda = 2$)	$29.32_{\pm 0.06}$	$43.98_{\pm 0.07}$	$26.37_{\pm 0.20}$	$13.07_{\pm 0.17}$	$16.96_{\pm 0.24}$	$27.69_{\pm 0.11}$	$13.59_{\pm 0.16}$
F-RCNN + $\widehat{CE}$ ($\lambda = 3$)	$26.16_{\pm 0.12}$	$39.05_{\pm 0.07}$	$25.33_{\pm 0.02}$	$14.57_{\pm 0.12}$	$15.45_{\pm 0.09}$	$26.74_{\pm 0.12}$	$15.04_{\pm 0.16}$
RetinaNet	$30.83_{\pm 0.12}$	$43.33_{\pm 0.22}$	$21.89_{\pm 0.36}$	$12.03_{\pm 0.19}$	$11.03_{\pm 0.56}$	$22.24_{\pm 0.27}$	$11.63_{\pm 0.08}$
RetinaNet + TS	$30.83_{\pm 0.12}$	$43.33_{\pm 0.22}$	$25.76_{\pm 0.38}$	$10.74_{\pm 0.54}$	$17.47_{\pm 0.50}$	$26.34_{\pm 0.29}$	$10.24_{\pm 0.43}$
RetinaNet + $\widehat{CE}$ ($\lambda = 0.01$)	$31.03_{\pm 0.07}$	$43.67_{\pm 0.09}$	$21.98_{\pm 0.15}$	$11.80_{\pm 0.16}$	$11.04_{\pm 0.22}$	$22.06_{\pm 0.22}$	$10.83_{\pm 0.34}$
RetinaNet + $\widehat{CE}$ ($\lambda = 0.05$)	$31.04_{\pm 0.07}$	$43.68_{\pm 0.10}$	$21.98_{\pm 0.15}$	$11.79_{\pm 0.17}$	$11.05_{\pm 0.22}$	$22.07_{\pm 0.20}$	$10.85_{\pm 0.32}$
RetinaNet + $\widehat{CE}$ ($\lambda = 0.1$)	$31.04_{\pm 0.06}$	$43.68_{\pm 0.09}$	$21.95_{\pm 0.15}$	$11.77_{\pm 0.15}$	$11.01_{\pm 0.21}$	$22.05_{\pm 0.19}$	$10.83_{\pm 0.29}$
RetinaNet + $\widehat{CE}$ ($\lambda = 0.5$)	$30.97_{\pm 0.07}$	$43.58_{\pm 0.11}$	$21.87_{\pm 0.12}$	$11.65_{\pm 0.11}$	$10.85_{\pm 0.19}$	$22.07_{\pm 0.16}$	$10.88_{\pm 0.18}$
RetinaNet + $\widehat{CE}$ ($\lambda = 1$)	$30.77_{\pm 0.05}$	$43.30_{\pm 0.05}$	$21.79_{\pm 0.16}$	$11.94_{\pm 0.03}$	$10.63_{\pm 0.33}$	$22.09_{\pm 0.21}$	$11.32_{\pm 0.07}$
RetinaNet + $\widehat{CE}$ ($\lambda = 2$)	$30.44_{\pm 0.06}$	$42.74_{\pm 0.11}$	$21.63_{\pm 0.18}$	$12.51_{\pm 0.08}$	$10.08_{\pm 0.28}$	$21.69_{\pm 0.22}$	$11.72_{\pm 0.02}$
RetinaNet + $\widehat{CE}$ ($\lambda = 3$)	$30.10_{\pm 0.06}$	$42.23_{\pm 0.10}$	$21.58_{\pm 0.14}$	$13.14_{\pm 0.16}$	$9.72_{\pm 0.18}$	$21.48_{\pm 0.06}$	$12.29_{\pm 0.32}$
RetinaNet + $\widehat{CE}$ ($\lambda = 4$)	$29.70_{\pm 0.08}$	$41.68_{\pm 0.16}$	$21.44_{\pm 0.03}$	$13.69_{\pm 0.28}$	$9.37_{\pm 0.10}$	$21.39_{\pm 0.10}$	$12.98_{\pm 0.20}$
FCOS	$34.02_{\pm 0.06}$	$48.84_{\pm 0.04}$	$24.40_{\pm 0.13}$	$16.55_{\pm 0.21}$	$16.87_{\pm 0.07}$	$24.73_{\pm 0.12}$	$15.57_{\pm 0.20}$
FCOS + TS	$29.51_{\pm 0.05}$	$41.13_{\pm 0.02}$	$20.53_{\pm 0.06}$	$13.66_{\pm 0.14}$	$8.13_{\pm 0.04}$	$20.54_{\pm 0.07}$	$13.00_{\pm 0.23}$
FCOS + $\widehat{CE}$ ($\lambda = 0.01$)	$34.05_{\pm 0.07}$	$48.86_{\pm 0.03}$	$24.43_{\pm 0.15}$	$16.56_{\pm 0.23}$	$16.81_{\pm 0.13}$	$24.73_{\pm 0.11}$	$15.60_{\pm 0.21}$
FCOS + $\widehat{CE}$ ($\lambda = 0.05$)	$34.12_{\pm 0.03}$	$48.89_{\pm 0.01}$	$24.29_{\pm 0.16}$	$16.29_{\pm 0.13}$	$16.68_{\pm 0.14}$	$24.65_{\pm 0.14}$	$15.37_{\pm 0.19}$
FCOS + $\widehat{CE}$ ($\lambda = 0.1$)	$34.03_{\pm 0.08}$	$48.83_{\pm 0.04}$	$24.18_{\pm 0.13}$	$16.34_{\pm 0.12}$	$16.56_{\pm 0.13}$	$24.57_{\pm 0.11}$	$15.40_{\pm 0.14}$
FCOS + $\widehat{CE}$ ($\lambda = 0.5$)	$33.89_{\pm 0.09}$	$48.65_{\pm 0.08}$	$24.09_{\pm 0.12}$	$16.35_{\pm 0.09}$	$16.46_{\pm 0.10}$	$24.45_{\pm 0.09}$	$15.39_{\pm 0.05}$
FCOS + $\widehat{CE}$ ($\lambda = 1$)	$33.73_{\pm 0.12}$	$48.45_{\pm 0.11}$	$23.95_{\pm 0.12}$	$16.35_{\pm 0.10}$	$16.25_{\pm 0.08}$	$24.27_{\pm 0.14}$	$15.36_{\pm 0.06}$
FCOS + $\widehat{CE}$ ($\lambda = 2$)	$33.41_{\pm 0.13}$	$48.00_{\pm 0.15}$	$23.41_{\pm 0.22}$	$16.15_{\pm 0.25}$	$15.56_{\pm 0.16}$	$23.83_{\pm 0.12}$	$15.27_{\pm 0.23}$
FCOS + $\widehat{CE}$ ($\lambda = 3$)	$33.02_{\pm 0.12}$	$47.44_{\pm 0.14}$	$23.18_{\pm 0.19}$	$16.34_{\pm 0.17}$	$15.32_{\pm 0.17}$	$23.72_{\pm 0.17}$	$15.67_{\pm 0.21}$

Figure 5. Qualitative results on COCO using uncalibrated (left) and fine-tuned with $\widehat{\mathrm{CE}}$ (right) F-RCNN model. Detection threshold is set to 0.5. Red boxes denote predictions with their corresponding confidence scores. Ground truth boxes are shown in green dashed lines. The calibration loss is used with identity link, i.e., the scores should be aligned with the IoU overlap with a ground truth box. Best viewed in color and zoom.