Supplementary material for "Multiclass Confidence and Localization Calibration for Object Detection"

Bimsara Pathiraja, Malitha Gunawardhana, Muhammad Haris Khan
Mohamed bin Zayed University of Artificial Intelligence, UAE

{bimsara.pathiraja,malitha.gunawardhana,muhammad.haris}@mbzuai.ac.ae

In this supplementary material document, we include implementation details on Deformable DeTR experiments (sec. 1), additional quantitative analyses (sec. 2), and qualitative calibration results (sec. 3).

1. Implementation details

We perform all experiments which involve Deformable DeTR as baseline with a multi-gpu setting (8 GPUs), and follow all the training configurations as mentioned in [1]. The weighting/balancing of our MCCL is chosen from $\beta \in \{0.01, 1\}$.

2. Additional Quantitative Analysis

Additional confidence histograms: The confidence histogram bins the samples based on the predicted confidence. We provide additional confidence histograms on more in-domain and out-domain scenarios (see Fig. 1 and Fig. 2). Compared to baselines (FCOS and Deformable DeTR), our method which is based on MCCL better reduces the gap between the overall accuracy and overall confidence.

Reliability diagrams: The reliability diagram reveals the ability of a model in mitigating under/overconfidence. We plot additional reliability diagrams with FCOS and Deformable DeTR baselines in Fig. 3 and Fig. 4, respectively, on more in-domain and out-domain detections. We observe that, compared to baselines, our method is more effective in mitigating the under/overconfidence.

Precision, confidence, and ECE relative to box coordinates: We plot the variation of average precision, confidence and calibration error (ECE) with respect to center-x (c_x) coordinate of the predicted bounding box in more in-domain and out-of-domain scenarios (see Figs. 5 & 6). Compared to baselines, FCOS and Deformable DeTR, our method reduces the calibration error throughout the c_x dimension by aligning the average precision and confidence. It shows that the proposed method is capable of reducing the box-sensitive calibration error throughout the image dimensions.

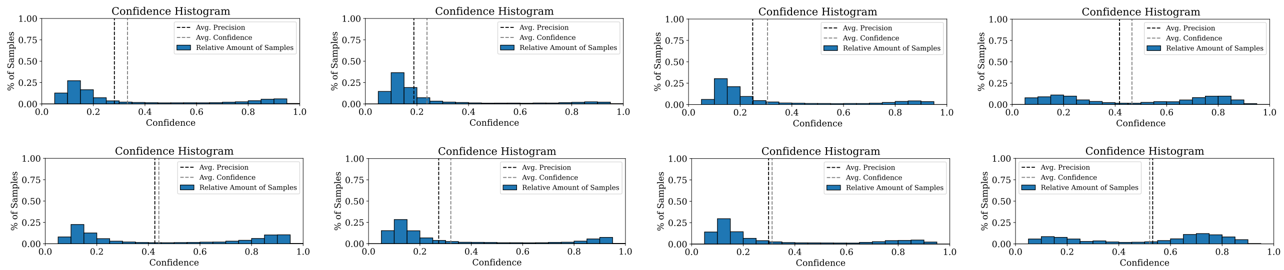
Calibration error heatmaps: We plot precision, confidence and calibration heatmaps relative to center-x c_x and center-y c_y of the predicted bounding boxes in several in-domain and out-domain scenarios (Figs.7,8,9,10). Compared to baselines, FCOS and Deformable DeTR, our method better aligns confidence and precision along both c_x and c_y and so is well-calibrated in both relative c_x and c_y coordinates.

3. Qualitative Calibration Results

We show some qualitative calibration results for baselines and our models in Fig. 11. The confidences are mapped more meaningfully in the predictions from our proposed method (MCCL).

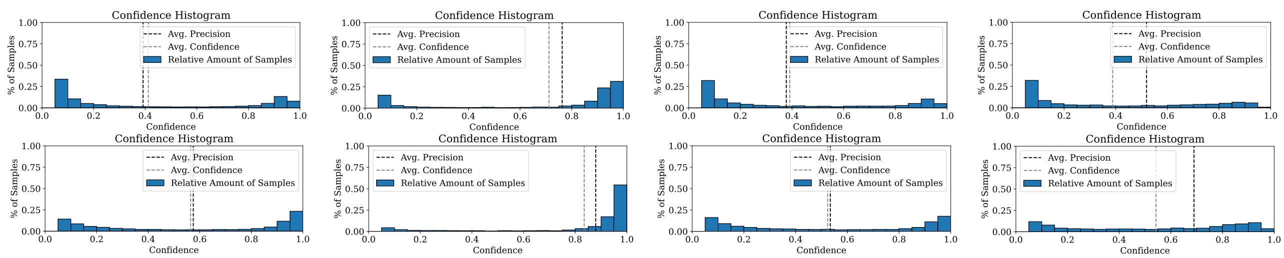
References

[1] Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, and Jifeng Dai. Deformable {detr}: Deformable transformers for end-to-end object detection. In *International Conference on Learning Representations*, 2021. 1



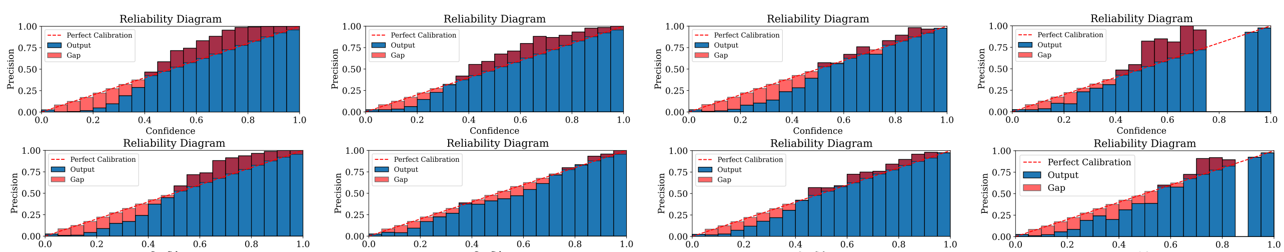
(a) Sim10K (b) Sim10K to CS (c) CS to CS-F (d) VOC

Figure 1. Confidence histograms for models trained with (top row) FCOS and (top row) FCOS+MCCL(ours).



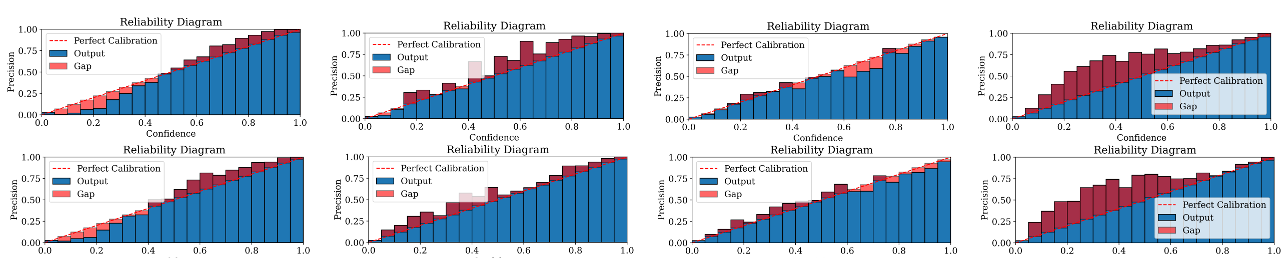
(a) Sim10K (b) KITTI (c) Sim10K to CS (d) KITTI to CS

Figure 2. Confidence histograms for models trained with (top row) Deformable DeTR and (bottom row) Deformable DeTR+MCCL(ours).



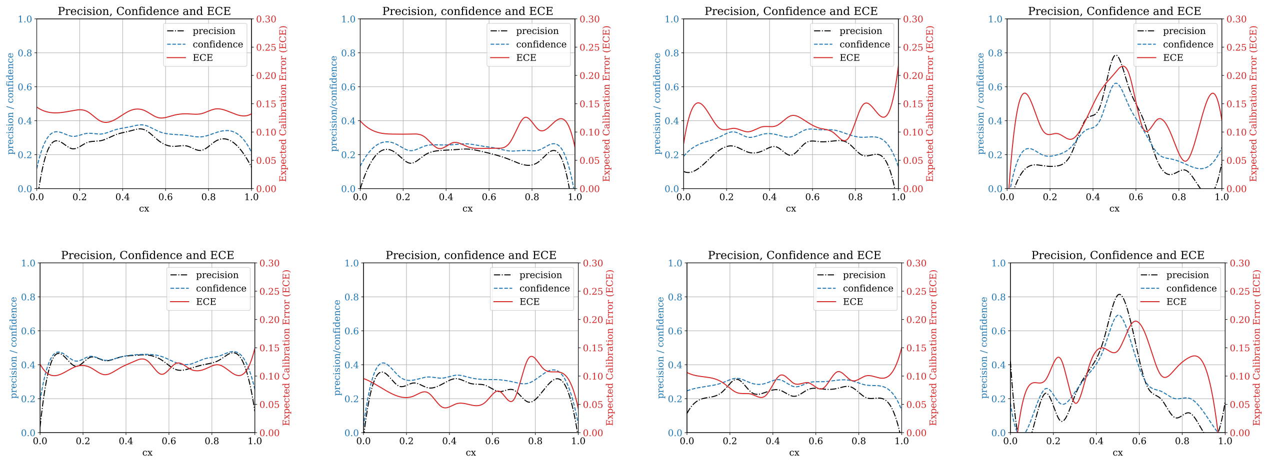
(a) Sim10K (b) Sim10K to CS (c) CS to CS-F (d) VOC

Figure 3. Reliability diagrams for models trained with (top row) FCOS and (bottom row) FCOS+MCCL(ours).



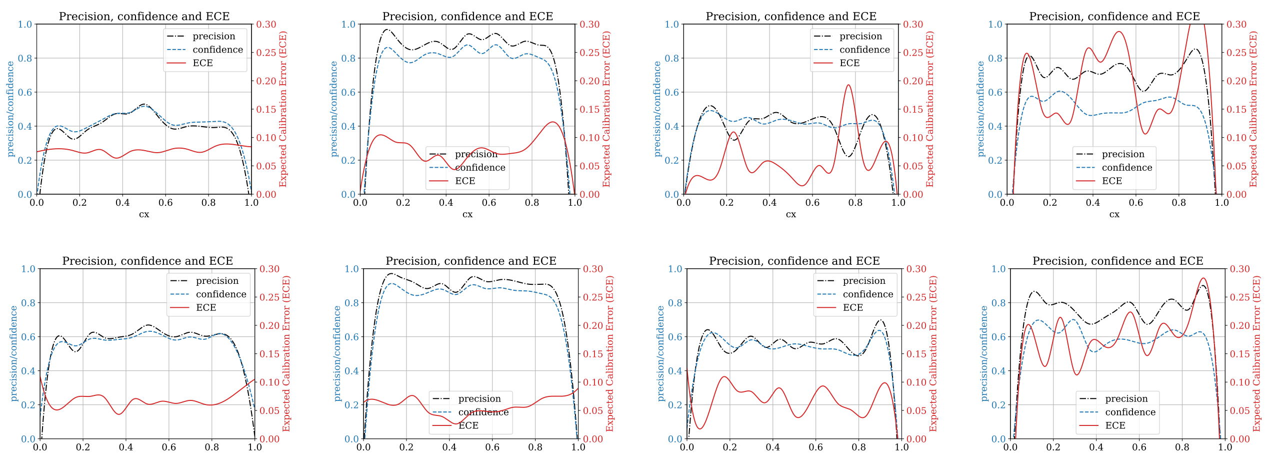
(a) Sim10K (b) KITTI (c) Sim10K to CS (d) KITTI to CS

Figure 4. Reliability diagrams for models trained with (top row) Deformable DeTR and (bottom row) Deformable DeTR+MCCL(ours).



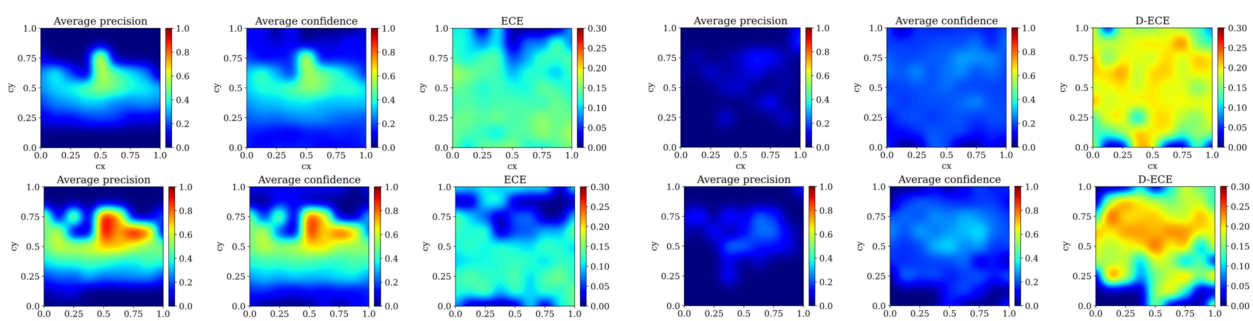
(a) Sim10K (b) Sim10K to CS (c) CS to CS-F (d) VOC

Figure 5. Calibration precision, confidence, and ECE relative to c_x for (top row) a modal trained with FCOS and (bottom row) a model trained with FCOS+MCCL (ours).



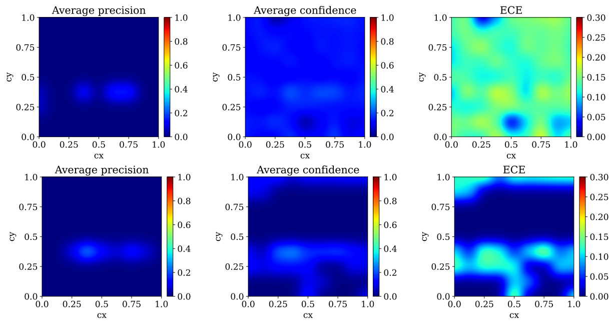
(a) Sim10K (DETR) (b) KITTI (c) Sim10K to CS (d) KITTI to CS

Figure 6. Calibration precision, confidence, and ECE relative to c_x for (top row) a modal trained with Deformable DeTR and, (bottom row) a model trained with Deformable DeTR+MCCL(ours).

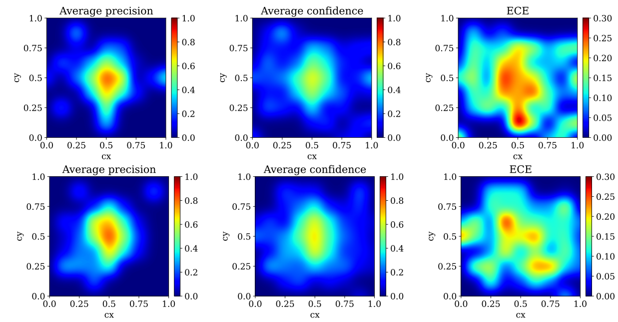


(a) Sim10K (b) COCO

Figure 7. Precision, confidence and calibration heatmaps w.r.t c_x and c_y for (top row) a model trained with FCOS and (bottom row) a model trained with FCOS+MCCL in in-domain scenarios.

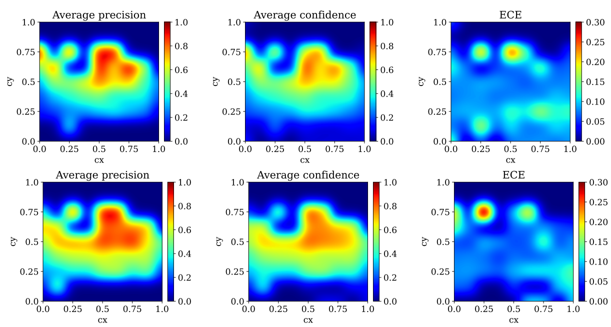


(a) CS to CS-F

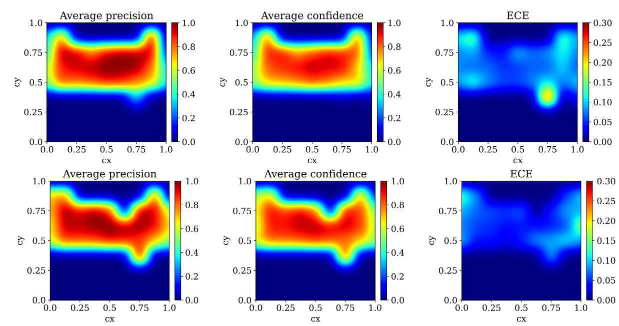


(b) VOC

Figure 8. Precision, confidence and calibration heatmaps w.r.t c_x and c_y for (top row) a model trained with FCOS and (bottom row) a model trained with FCOS+MCCL in both in-domain and out-of-domain scenarios.

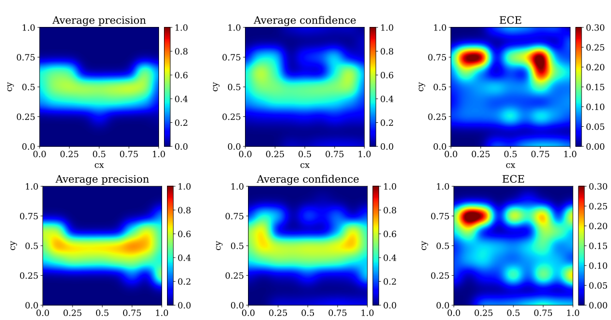


(a) Sim10K

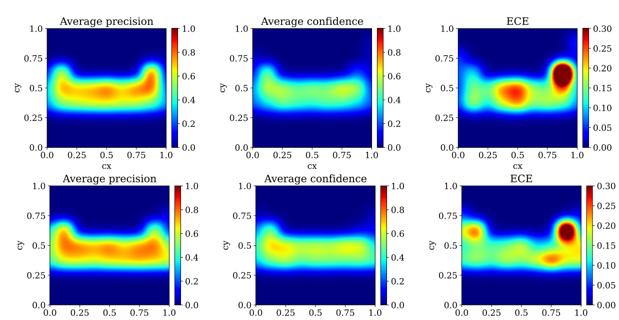


(b) KITTI

Figure 9. Precision, confidence and calibration heatmaps w.r.t c_x and c_y for (top row) a model trained with Deformable DeTR and (bottom row) a model trained with Deformable DeTR+MCCL in in-domain scenarios.

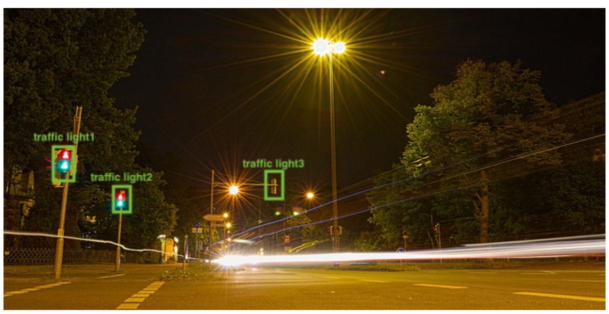


(a) Sim10K to CS



(b) KITTI to CS

Figure 10. Precision, confidence and calibration heatmaps w.r.t c_x and c_y for (top row) a model trained with Deformable DeTR and (bottom row) a model trained with Deformable DeTR+MCCL in out-of-domain scenarios.



(a) **Calibration mappings**: traffic light1 74% $\rightarrow$ 66%, traffic light2 64% $\rightarrow$ 58%, traffic light3 57% $\rightarrow$ 55%

(b) **Calibration mappings**: person1 48% $\rightarrow$ 50%, person2 84% $\rightarrow$ 83%

Figure 11. Predicted confidences from a model trained with FCOS (left of arrow) and a model trained with FCOS+MCCL (right of arrow). We see that FCOS+MCCL reduces its predicted confidence (traffic lights) or increases its predicted confidence (person1) based on how uncertain it is in terms of confidences and box localization.