JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

THU-AM-309

Intelligent Visual Analytics Lab

Multiclass Confidence and Localization Calibration for Object Detection

Bimsara Pathiraja, Malitha Gunawardhane, Muhammad Haris Khan

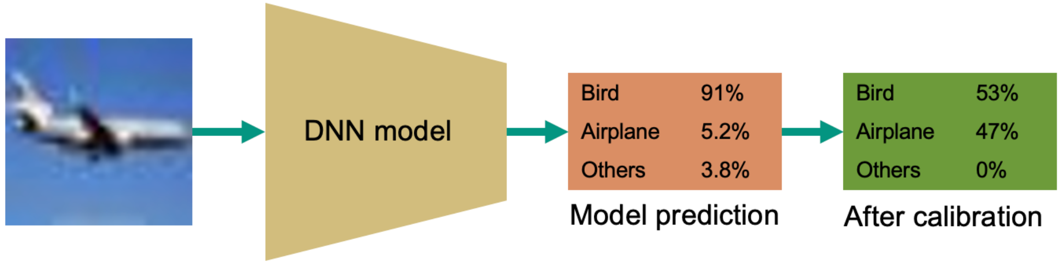
Mohamed Bin Zayed University of Artificial Intelligence, UAE

جامعة محمد بن زايد للذكاء الاصطناعي

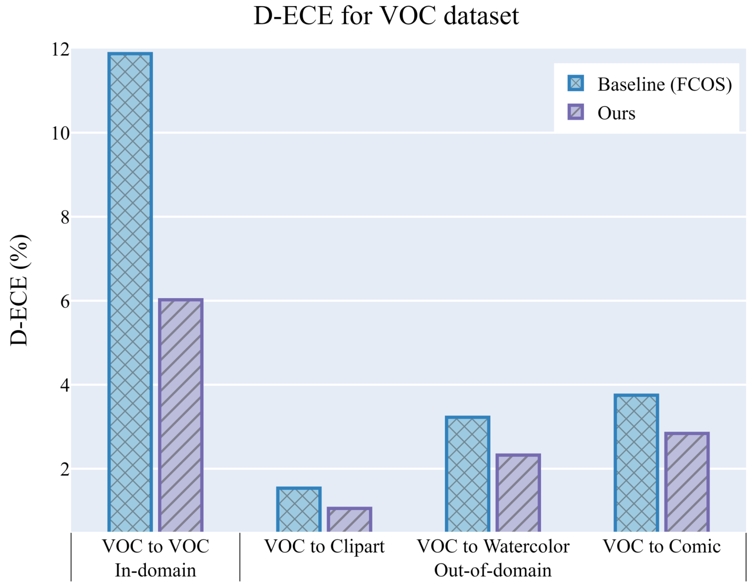
MOHAMED BIN ZAYED UNIVERSITY OF ARTIFICIAL INTELLIGENCE

Motivation

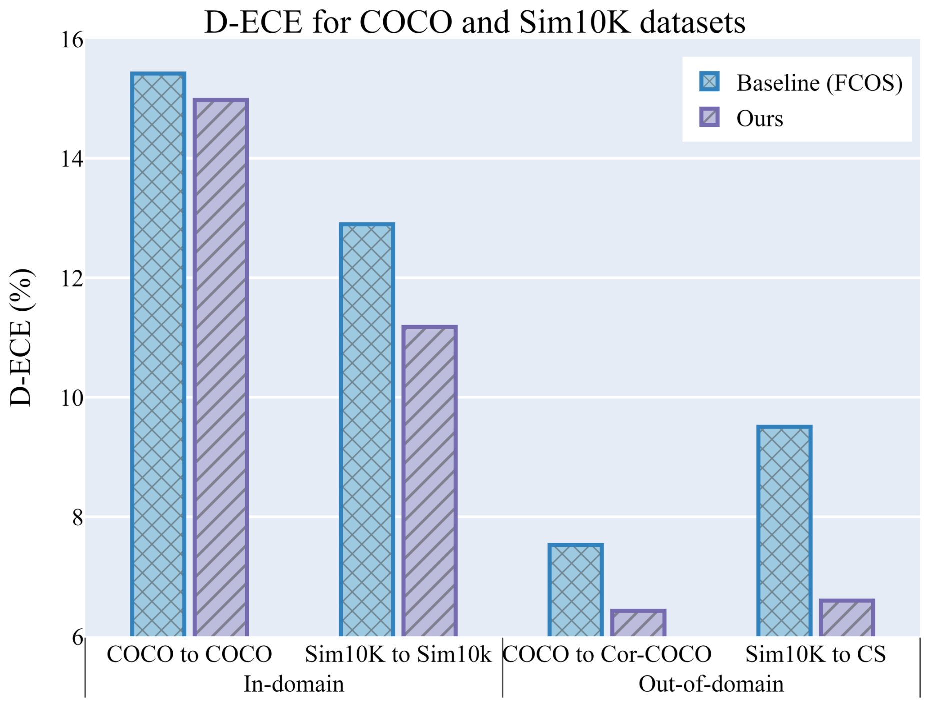
Deep neural networks (DNNs) are **poorly calibrated**

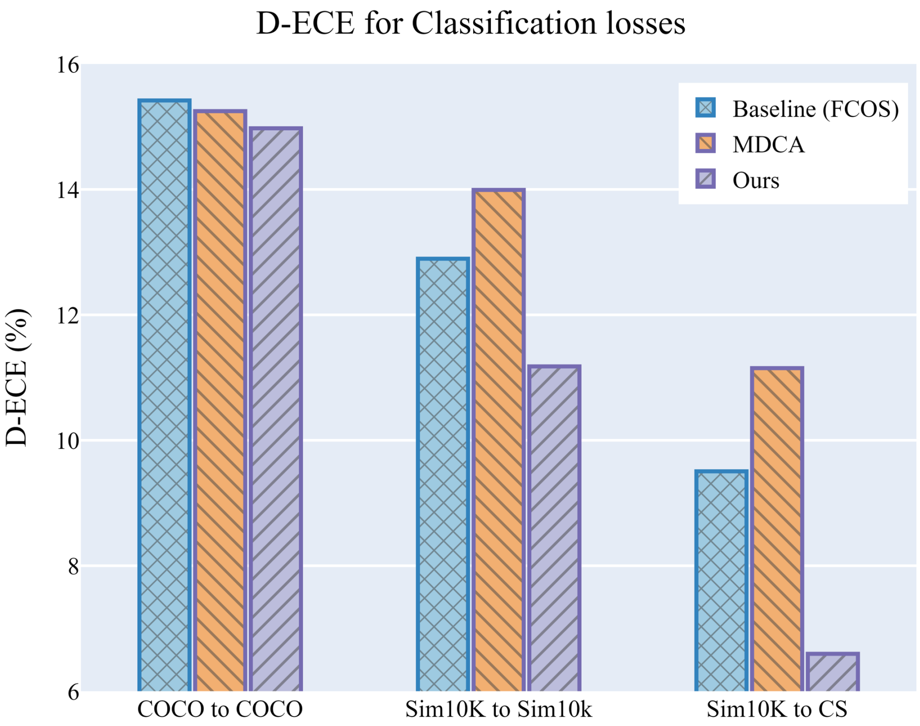


- ❖ Existing attempts for calibration:
  - ❖ are limited to **classification tasks**
  - ❖ restricted to calibrating **in-domain** predictions
- ❖ Little attention toward **calibrating object detectors:**
  - ❖ pivotal space in vision-based security-sensitive and,
  - ❖ safety-critical applications.



Motivation



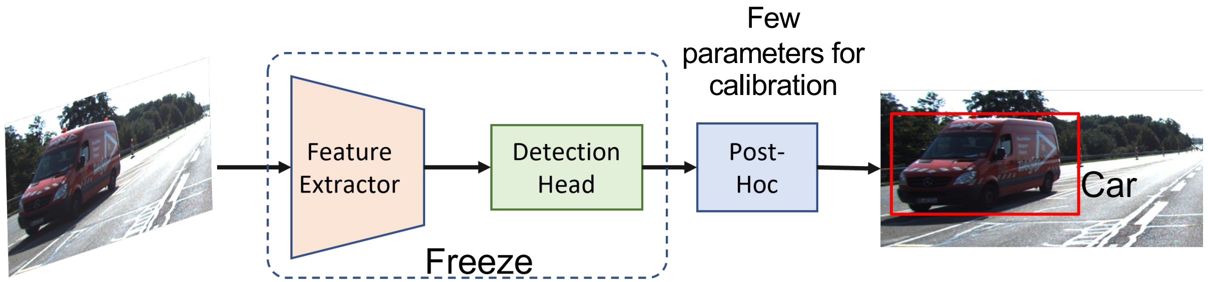


DNN-based object detectors are:

1. inherently **miscalibrated** for in-domain and out-domain predictions
2. image classification-based calibration methods are **sub-optimal** for object detection

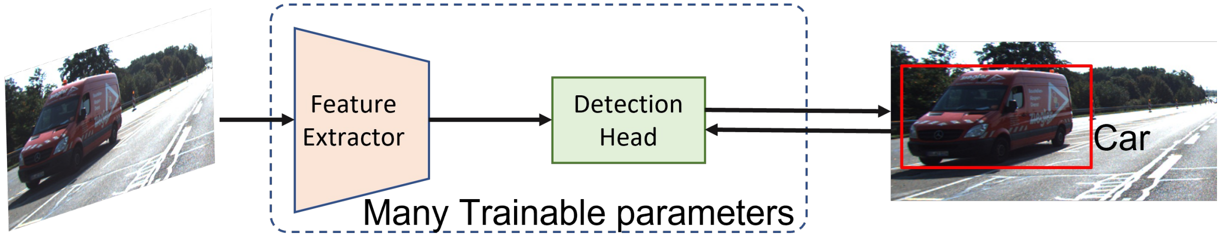
Existing paradigms in calibrating DNNs

Post-hoc



- **Hold-out validation** set to learn calibration parameters
- Calibration is based on **few parameters**

Train-time



- **All model parameters** are engaged
- **No hold-out validation set** is required

Contributions

- Study relatively **unexplored direction** of calibrating object detectors
- Propose a **new train-time calibration** method (MCCL)
- Featuring an **auxiliary loss** term
- Jointly calibrates **multiclass confidences** and **bounding box localization**
- **Differentiable**, operates on mini-batches
- **Extensive experiments** on challenging datasets, featuring several in-domain and out-of-domain scenarios.

Our train-time calibration method - MCCL

Uncertainty quantification

Monte-Carlo Dropout to estimate the following:

$\mathbf{c}_n \in \mathbb{R}^K$	Class-wise certainty
$\bar{\mathbf{s}}_n \in \mathbb{R}^K$	Mean logits-based class-wise confidence
$\bar{\mathbf{b}}_n \in [0,1]^4$	Mean bounding box parameters
g_n	Certainty in bounding box localization

Our train-time calibration method - MCCL

Multi-class confidence calibration

$$\boxed{\mathbf{v}_{l,n}[k] = (\bar{\mathbf{s}}_{l,n}[k] + \mathbf{c}_{l,n}[k])/2.}$$

Multiclass fusion of mean confidence and certainty

$$\boxed{\mathcal{L}_{MCC} = \frac{1}{K}\sum_{k=1}^{K}\left|\frac{1}{M}\sum_{l=1}^{N_b}\sum_{n=1}^{N_{pos}}\mathbf{v}_{l,n}[k] - \frac{1}{M}\sum_{l=1}^{N_b}\sum_{n=1}^{N_{pos}}\mathbf{q}_{l,n}[k]\right|}$$

Absolute difference between the fused vector and the accuracy

$M = N_b \times N_{pos}$. N_b: # samples in the minibatch and N_{pos}: # of positive locations.
$\boldsymbol{q}_{l,n}[k]$ = 1 if k is the ground truth class of bounding box for the nth location in the l^{th} sample.

Our train-time calibration method - MCCL

Localization calibration

$$\mathcal{L}_{LC} = \frac{1}{N_b}\sum_{l=1}^{N_b}\frac{1}{N_{pos}^{l}}\sum_{n=1}^{N_{pos}^{l}}\left|\left[\mathrm{IoU}(\bar{\mathbf{b}}_{n,l}, \mathbf{b}^{*}_{n,l}) - \mathrm{g}_{n,l}\right]\right|$$

Absolute difference between the bbox overlap (with the ground truth) and the certainty in the bbox prediction

N_{pos}^{l}: # of positive bounding box regions in the l^{th} sample.

Our train-time calibration method - MCCL

Our auxiliary loss

$$\boxed{\mathcal{L}_{MCCL-aux} = \mathcal{L}_{MCC} + \beta \mathcal{L}_{LC}}$$

❖ Also input mean confidence and mean bounding box parameters to task-specific loss

In-domain calibration performance

In-domain performance										
Methods	Sim10K		KITTI		CS		COCO		VOC	
	D-ECE	AP@0.5	D-ECE	AP@0.5	D-ECE	AP@0.5	D-ECE	AP@0.5	D-ECE	mAP
Baseline (FCOS)	12.90	87.45	9.54	94.54	9.40	70.48	15.42	54.91	11.88	59.68
Ours (MCCL)	**11.18**	86.47	**7.79**	93.76	**7.64**	70.22	**14.94**	54.85	**6.02**	59.17

Out-domain calibration performance

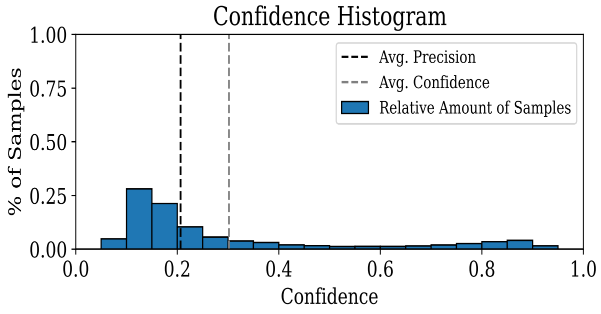
Methods	Sim10K → CS D-ECE	Sim10K → CS AP@0.5	KITTI → CS D-ECE	KITTI → CS AP@0.5	CS → CS-F D-ECE	CS → CS-F AP@0.5	COCO → Cor-COCO D-ECE	COCO → Cor-COCO AP@0.5	CS → BDD100K D-ECE	CS → BDD100K AP@0.5
Baseline (FCOS)	9.51	45.18	7.53	38.11	11.18	19.81	15.90	30.01	18.82	14.18
Ours (MCCL)	**6.60**	44.30	**6.43**	38.73	**8.97**	19.54	**14.45**	29.96	**16.12**	14.20

Out-of-domain performance

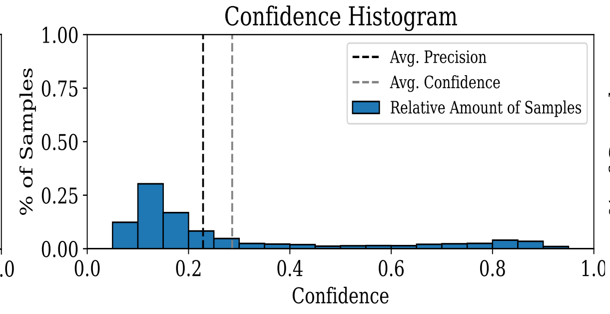
Methods	VOC → clipart D-ECE	VOC → clipart mAP	VOC → watercolor D-ECE	VOC → watercolor mAP	VOC → comic D-ECE	VOC → comic mAP
Baseline (FCOS)	1.54	14.57	3.23	24.23	3.75	9.89
Ours (MCCL)	**1.06**	13.71	**2.33**	28.70	**2.84**	11.50

Out-of-domain performance

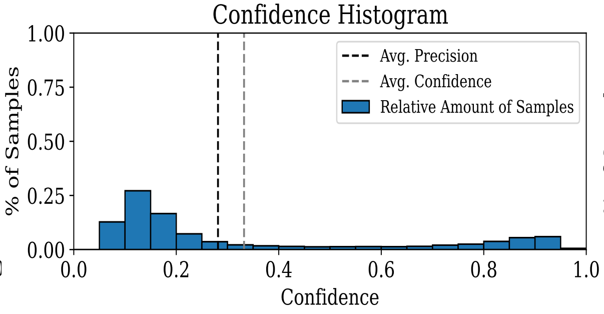
Overcoming under/overconfidence



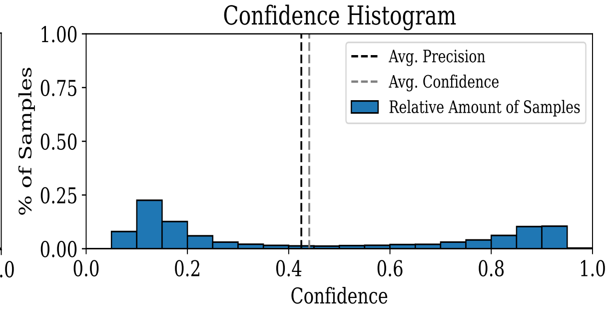
Sim10K – baseline (FCOS)



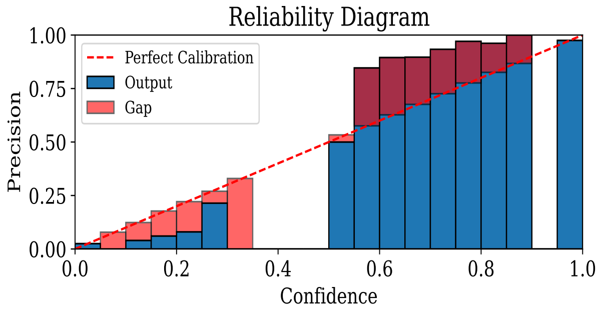
Sim10K – ours



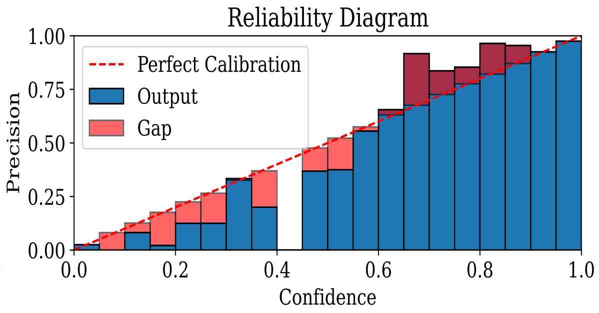
CS to CS-F – baseline (FCOS)



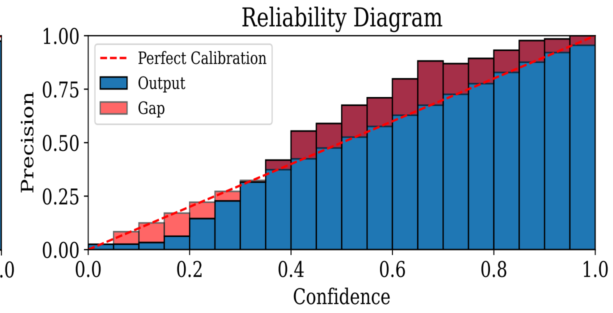
CS to CS-F – ours



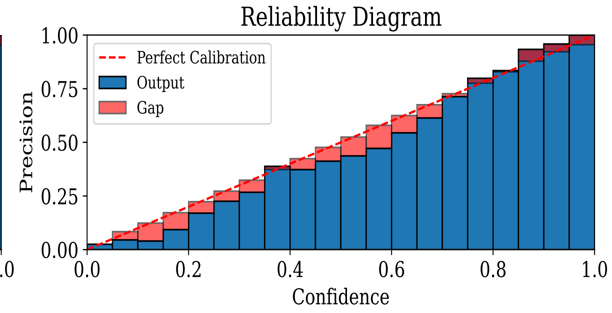
VOC – baseline (FCOS)



VOC – ours



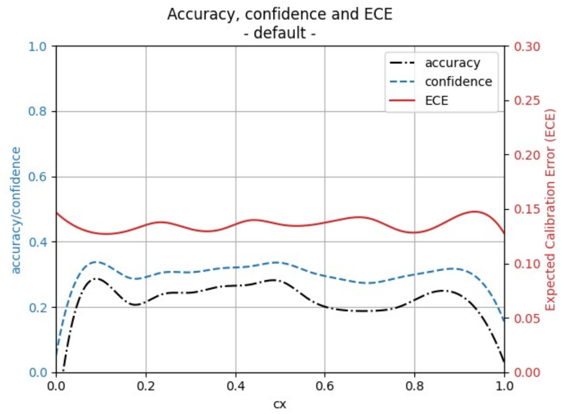
Sim10K to CS – baseline (FCOS)



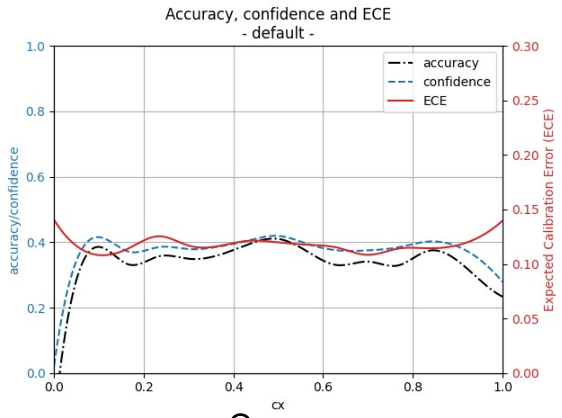
Sim10K to CS – ours

Our MCCL can effectively overcome overconfidence and underconfidence

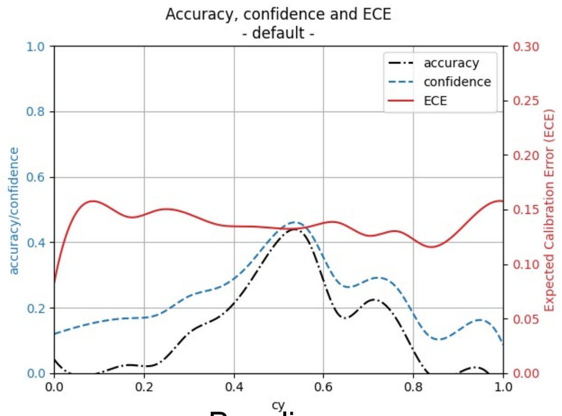
Impact on location-dependent calibration



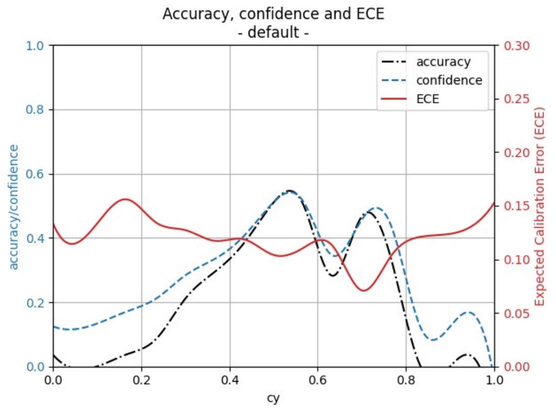
Baseline c_x



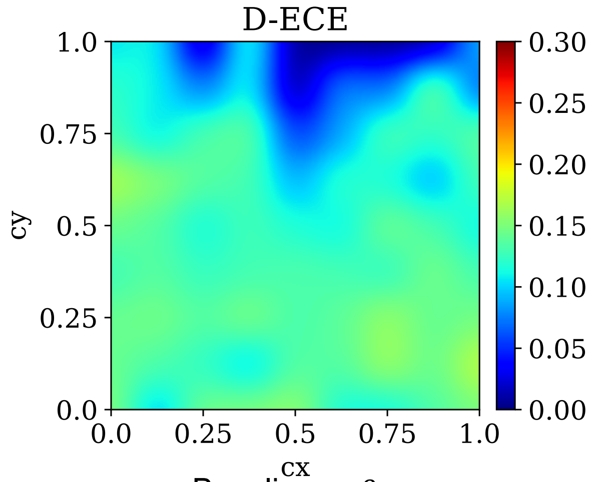
Ours c_x



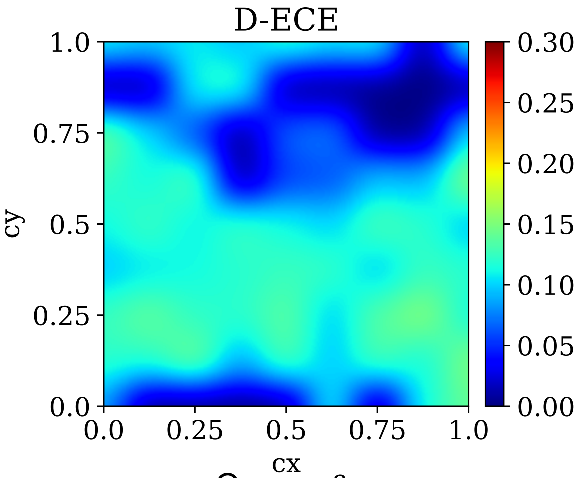
Baseline c_y



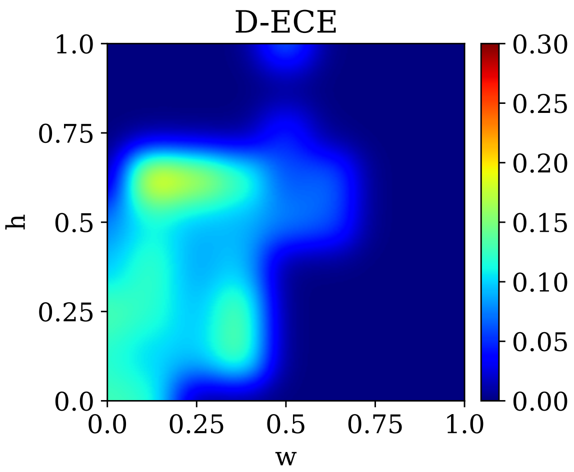
Ours c_y



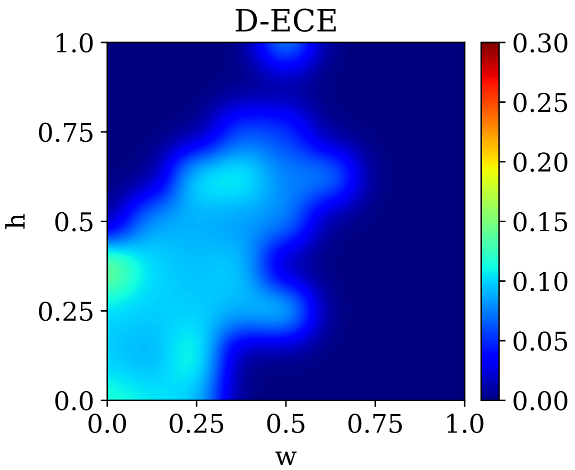
Baseline c_x & c_y



Ours c_x & c_y



Baseline w & h



Ours w & h

Our MCCL can decrease D-ECE at all image locations including image boundaries

Conclusion

Goal: Calibrating object detectors

Approach: A train-time method for calibrating multiclass confidence and box localization.

Promising results:

- Several in-domain and out-domain scenarios
- CNNs and DETR-based detectors
- Effective for location-dependent calibration

Scan me

Thank you for listening!