Rethinking Detection Calibration: A Coordinate and Direction Perspective

Juyong Lee[1], Seungjin Jung[1], Jungmin Lee[2], Sunju Lee[1], and Jongwon Choi[1,2,3]*

[1] Dept. of AI, Chung-Ang University, Republic of Korea
[2] Dept. of Advanced Imaging, GSAIM, Chung-Ang University, Republic of Korea
[3] GS. of Virtual Convergence, Chung-Ang University, Republic of Korea
{jylee, sjjung, jngmnlee, lsjoo}@vilab.cau.ac.kr, choijw@cau.ac.kr

Abstract. Deep learning based object detectors require trustworthiness beyond competitive detection performance, but deep neural networks are prone to overconfident predictions, assigning high confidence scores to predictions that are likely to be inaccurate. To improve the alignment between confidence scores and prediction accuracy, existing methods calibrate confidence scores based on box-level localization, such as precision or intersection over union with the ground truth bounding box. However, box-level localization reflects only a measure of agreement between the predicted box and the ground truth, resulting in calibrated confidence scores for box-level accuracy failing to capture the localization accuracy of coordinates of box. To tackle this issue, we propose a novel post-hoc calibration framework, rethinking detection calibration (ReDC), which provides reliable coordinate-level confidence scores, including directional information. The proposed framework defines coordinate-wise alignment and deviation direction between predictions and ground truth. Based on the alignment measure, confidence re-encoding produces reliable coordinate-level confidence scores, while directional displacement estimation predicts coordinate-wise deviation directions. Extensive experiments under in-domain and out-domain scenarios demonstrate that the proposed approach expresses the coordinate-wise localization of detected objects more precisely than existing methods. Furthermore, our method covers the representational scope of prior calibration approaches by aggregating coordinate-level confidence scores into box-level localization.

Keywords: Confidence calibration · Object detection · Explainable computer vision

1 Introduction

Object detection has demonstrated impressive performance through progressive advances [31, 35]. However, deep neural networks are often overconfident [5, 33], which may result in incorrect predictions made with high confidence and limit

* Corresponding author

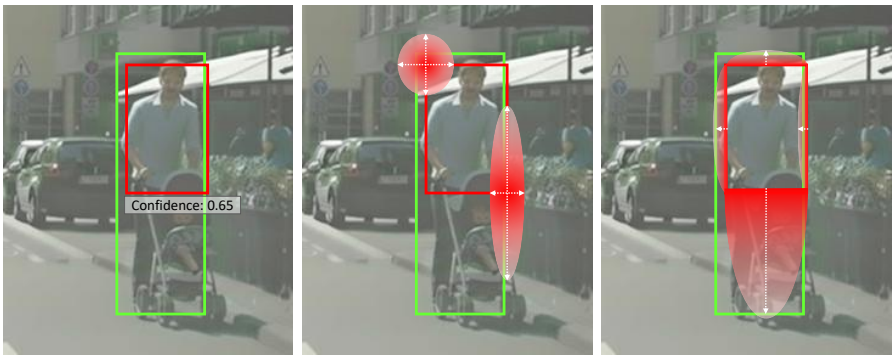

Fig. 1: Conceptual comparison of calibration frameworks in object detection. (a) Deterministic detector: Conventional methods provide a single confidence score for the entire bounding box, failing to identify specific coordinate-wise misalignments. (b) Probabilistic detector: While capturing local confidence, symmetric distributions often fail to represent the specific direction and magnitude of coordinate deviation. (c) Rethinking Detection Calibration (ReDC): Our framework estimates independent coordinate-wise confidence scores aligned with the coordinate-wise alignment ratio (CAR). This approach explicitly captures directional confidence.

the practical adoption of object detection in safety-critical domains such as autonomous driving [4,29], medical applications [10,34], and security systems [17]. To assign lower confidence scores to predictions that are likely to be inaccurate, confidence calibration [5, 11, 28] has been widely studied to align predicted confidence scores with their empirical accuracies. In object detection, confidence calibration further requires properly defining empirical accuracy for bounding box localization [12,18,20], as deterministic detectors do not provide probabilistic estimates for bounding boxes.

Previous studies typically define empirical accuracy at the box-level. One approach measures box-level accuracy using precision with a fixed intersection over union (IoU) threshold [22], while another incorporates the predicted IoU value to estimate box-level accuracy [12]. However, object detection predictions can exhibit different degrees of spatial misalignment and directional shifts relative to the ground-truth bounding box, even when they share identical box-level accuracy. As illustrated in Fig. 1-(a), existing confidence calibration methods capture the degree of alignment only at the bounding-box level because they rely solely on box-level localization accuracy, limiting their ability to reflect fine-grained coordinate-wise misalignment. This limitation can also affect downstream decision-making in real-world applications.

Probabilistic object detection methods [14] estimate coordinate-wise covariance to derive confidence, defining empirical accuracy as whether the ground-truth box falls within the predicted probabilistic region (Fig. 1-(b)). This captures coordinate-wise localization error but, since coordinate-wise estimates are col-

lapsed into box-level confidence and accuracy, still fails to capture the direction of localization error. This matters in practice: in autonomous driving, a box shifted upward and smaller than ground truth can make an object appear farther away, potentially delaying braking, despite identical box-level accuracy. Moreover, these methods require a probabilistic detector, limiting general applicability.

To address the limitations in prior approaches, we propose a new post-hoc calibration framework without restricting its applicability to specific object detectors, rethinking detection calibration (ReDC), consisting of confidence re-encoding (CR), coordinate-wise alignment ratio (CAR), and directional displacement estimation (DDE). CR estimates coordinate-wise confidence by re-scaling the class confidence for each coordinate, while CAR computes coordinate-wise empirical accuracy using the bounding boxes of the ground truth and the instance sample. Furthermore, DDE estimates the deviation direction of each predicted coordinate by utilizing predicted logit vectors and information derived from coordinates. By aligning the resulting confidence and empirical accuracy, the model learns to capture spatial misalignment and directional shifts as illustrated in Fig. 1-(c). Since coordinate-wise alignment alone cannot capture box-level accuracy, we approximate box-level IoU from coordinate-wise confidence and deviation direction, and train a calibration network to align this estimated IoU with actual IoU. This gives ReDC complementary calibration at both the coordinate and box levels. We also introduce two new metrics, as no prior metric jointly captures coordinate-level and directional accuracy: coordinate-wise expected calibration error (C-ECE) for coordinate-level calibration quality, and direction-aware calibration error (Da-CE) for calibration performance incorporating deviation direction.

We compare ReDC against state-of-the-art post-hoc and train-time calibration methods on COCO [16] and Cityscapes [3] under both in-domain and out-of-domain settings. Results show ReDC effectively captures coordinate-wise localization accuracy and deviation direction while maintaining competitive box-level calibration performance.

The main contributions of this paper are summarized as follows:

- We analyze the heterogeneity of coordinate-level accuracy among samples sharing similar box-level accuracy and show that existing calibration methods exhibit limitations in independently representing coordinate-level accuracy.
- We propose ReDC, a post-hoc calibration framework that conveys coordinate-wise confidence scores in object detection while incorporating directional information of prediction misalignment with respect to the ground truth.
- Extensive experiments demonstrate that ReDC outperforms existing calibration methods in fine-grained calibration and achieves comparable box-level calibration performance by aggregating fine-grained confidence scores to approximate IoU.

2 Related Work

2.1 Confidence Calibration

Calibration methods aim to align model confidence with empirical accuracy. Temperature scaling (TS) [5], a simple yet effective post-hoc method, has been widely adopted in subsequent works [11, 28, 33]. In object detection, confidence calibration additionally requires properly defining empirical accuracy for bounding box localization [18, 24], as deterministic detectors do not naturally provide probabilistic estimates for bounding boxes. Existing approaches address this problem through train-time calibration objectives integrated into detector training or by aligning detection confidence with localization quality [20]. Post-hoc calibration methods have also been explored, including covariance calibration for bounding box distributions [13], confidence calibration for object detectors [23], and calibration analysis with evaluation baselines for object detection [12, 22]. However, these approaches primarily operate at the box level, calibrating detection confidence without modeling coordinate-level localization information. In contrast, our method calibrates the coordinate-level confidence score and leverages it to estimate the box-level localization score.

2.2 Coordinate-wise Uncertainty Estimation

Recent studies estimate localization uncertainty in object detection by modeling bounding box coordinates as probabilistic distributions. Several works predict Gaussian distributions over bounding box coordinates to estimate coordinate-wise uncertainty [2, 6, 8], while other approaches extend uncertainty estimation to anchor-free detectors [15]. More recent studies focus on uncertainty calibration to align predicted uncertainty with empirical errors [14, 27]. However, existing approaches typically convert localization uncertainty into box-level accuracy and confidence for calibration, limiting their ability to capture directional localization errors. Moreover, since modeling coordinate-wise uncertainty as a probabilistic distribution requires probabilistic detectors, its applicability to commonly used deterministic detectors is limited. In contrast, our method first calibrates coordinate-level confidence scores and then estimates box-level confidence score. Moreover, we learn a lightweight post-hoc module without retraining the detector or relying on specific detectors, improving training efficiency.

3 Preliminary

3.1 Confidence Calibration for Object Detection

Object Detection We define a dataset $\mathcal{D} := \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N}$, where i denotes the sample index and N denotes the total number of samples. Here, $\mathbf{x}_i \in \mathbb{R}^{H \times W \times Z}$ denotes an input image with height H, width W, and Z color channels, while the corresponding label $\mathbf{y}_i = (c_i, b_i)$ consists of a class label c_i and a bounding box

$b_i \in \mathbb{R}^4$. Given an input image $\mathbf{x}_i$, an object detector ϕ_D produces M detections, defined as follows:

$$\phi_D(\mathbf{x}_i) : \mathbf{x}_i \mapsto \{(\hat{c}_j, \hat{b}_j, \hat{p}_j)\}_{j=1}^{M}, \tag{1}$$

where $\hat{c}_j$, $\hat{b}_j$, and $\hat{p}_j$ denote the predicted class label, bounding box, and predicted confidence score for the object, respectively.

Post-hoc Calibration Post-hoc confidence calibration aligns prediction confidence scores $\hat{p}$ with the true probabilities p by training an additional calibration model added to a pretrained model. In classification tasks, perfect calibration [5] is defined as follows:

$$\mathbb{P}(\hat{y} = y \mid \hat{p} = p) = p, \quad \forall p \in [0, 1], \tag{2}$$

where $\hat{y}$ denotes predicted class. Since the true conditional probability is not directly observable, calibration is evaluated in practice using empirical accuracy estimated from samples with similar confidence scores. In object detection, Kuppers et al. [13] proposed training calibrators for the object detection task to satisfy the following equation by setting precision as the accuracy:

$$\mathbb{P}(m = 1 \mid \hat{c} = c, \hat{b} = b, \hat{p} = p) = p, \quad \forall p \in [0, 1], \tag{3}$$

where $\hat{p}$ denotes the confidence score associated with a predicted bounding box inferred by the object detector, and $m \in \{0, 1\}$ indicates whether the detector correctly identifies a ground-truth object. In practice, m is determined based on an intersection over union (IoU) criterion with a threshold τ. To provide information on localization accuracy, Kuzuku et al. [12] proposed setting the IoU threshold τ to 0 and aligning confidence scores with IoU, which yields the following equation:

$$\mathbb{E}_{\hat{b} \in\ B(\hat{p})}[\mathrm{IoU}(\hat{b}, b)] = \hat{p}, \quad \forall \hat{p} \in [0, 1], \tag{4}$$

where $B(\hat{p})$ denotes the set of predicted bounding boxes with confidence score $\hat{p}$, and b denotes the ground-truth bounding box matched to $\hat{b}$.

3.2 Preliminary Analysis

This section describes the limitation that existing object detection calibration methods do not sufficiently explain the localization accuracy of detection results. We analyze the localization alignment characteristics of bounding boxes at the coordinate level.

As shown in the Fig. 2, the coordinate-wise alignment distributions are not uniform within the same class. Some coordinates exhibit higher variance and lower alignment, indicating that localization accuracy differs across bounding box coordinates. This behavior is more evident for objects with complex geometries.

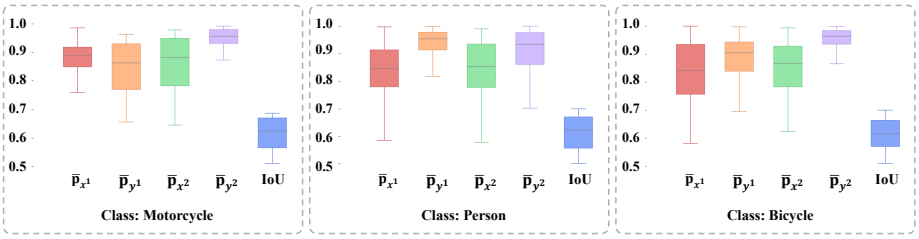

Fig. 2: Box-level IoU fails to capture coordinate-wise alignments. From left to right, the figures correspond to the motorcycle, person, and bicycle classes, respectively. Each box plot shows the distribution of coordinate-wise alignment ratios (CAR) computed over samples belonging to the same class, where CAR measures the degree of correspondence between predicted and ground-truth bounding-box coordinates.

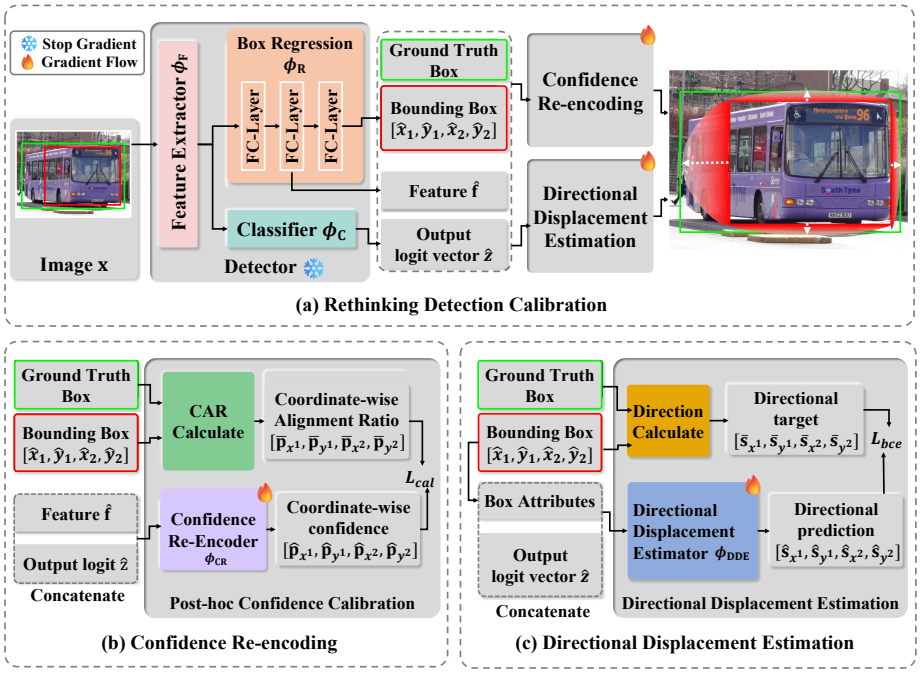

Fig. 3: Overview of ReDC framework. (a) Rethinking detection calibration is a post-hoc calibration method that estimates coordinate-wise confidence scores and deviation directions on top of a pre-trained detector. (b) Confidence re-encoding estimates coordinate-wise confidence scores aligned with coordinate-wise alignment ratio, which represents coordinate-wise accuracy. (c) Directional displacement estimation predicts the direction in which each predicted coordinate deviates from the corresponding ground-truth coordinate.

Such coordinate-specific errors are more pronounced when objects have complex structures or are affected by viewpoint changes.

However, existing confidence calibration methods for the object detection task treat the entire bounding box as a single unit and learn a single confidence score to represent localization accuracy. This approach fails to capture coordinate-wise variations in alignment and does not adequately reflect severe localization errors occurring in specific coordinates. These observations suggest the need for coordinate-wise localization confidence estimation.

4 Method

4.1 Rethinking Detection Calibration

Motivated by the analysis in Sec. 3.2, we introduce a new post-hoc calibration framework, termed rethinking detection calibration (ReDC). ReDC consists of an confidence re-encoder (CR) that re-encodes detection information to calibrate confidence scores for coordinate-wise localization and a directional displacement estimator (DDE) that estimates the direction of coordinate-wise deviation, as illustrated in Fig. 3.

Coordinate-wise Alignment Ratio We propose a coordinate-wise alignment ratio (CAR) to define the localization accuracy for each coordinate. CAR is computed based on coordinate-wise differences and intersection lengths between the predicted and ground-truth bounding boxes, as illustrated in Fig. 4. First, we consider the four bounding box coordinates corresponding to the top-left and bottom-right corners, *i.e.*, (x^1, y^1) and (x^2, y^2), where $x^2 = x^1 + w$ and $y^2 = y^1 + h$, respectively. The predicted coordinates are obtained via a feature extractor ϕ_F and a box regression module ϕ_R of the pretrained object detector ϕ_D, and are defined as $(\hat{x}^1, \hat{y}^1, \hat{w}, \hat{h}) = \phi_R(\phi_F(\mathrm{x}))$. Based on the predicted and ground-truth coordinates, the coordinate-wise differences are formulated as follows:

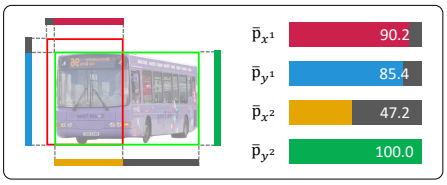

Fig. 4: Coordinate-wise alignment ratio (CAR). CAR measures the coordinate-wise alignment between predicted and ground-truth bounding boxes based on the coordinate-wise absolute difference and axis-wise intersection length.

$$\mathrm{dist}_{x^l} = |\hat{x}^l - x^l|, \quad \mathrm{dist}_{y^l} = |\hat{y}^l - y^l|, \tag{5}$$

where $l \in \{1, 2\}$ indexes the top-left ($l = 1$) and bottom-right ($l = 2$) coordinates of a bounding box. The x- and y-axis intersection lengths are defined as follows:

$$\begin{aligned} \mathrm{inter}_w &= \max\big(0, \min(\hat{x}^2, x^2) - \max(\hat{x}^1, x^1)\big)\,, \\ \mathrm{inter}_h &= \max\big(0, \min(\hat{y}^2, y^2) - \max(\hat{y}^1, y^1)\big)\,. \end{aligned} \tag{6}$$

Using the coordinate-wise differences and intersection lengths defined above, we compute the CAR $\bar{\mathrm{p}} = (\bar{\mathrm{p}}_{x^1}, \bar{\mathrm{p}}_{y^1}, \bar{\mathrm{p}}_{x^2}, \bar{\mathrm{p}}_{y^2})$ as follows:

$$\bar{\mathrm{p}}_{x^l} := \frac{\mathrm{inter}_w}{\mathrm{dist}_{x^l} + \mathrm{inter}_w}, \quad \bar{\mathrm{p}}_{y^l} := \frac{\mathrm{inter}_h}{\mathrm{dist}_{y^l} + \mathrm{inter}_h}. \tag{7}$$

Each CAR value lies in the range $[0, 1]$. For example, $\bar{\mathrm{p}}_{x^1} = 0$ when the intersection width inter_w is zero, indicating no horizontal overlap between the two bounding boxes. In contrast, $\bar{\mathrm{p}}_{x^1} = 1$ when the left x-coordinates of the predicted and ground-truth bounding boxes exactly match. Based on CAR, we require the following definition to be satisfied at the coordinate-wise level, extending the box-level formulation of Kuzucu et al. [12]. Given the coordinate-wise confidence scores $\hat{\mathrm{p}} = (\hat{\mathrm{p}}_{x^1}, \hat{\mathrm{p}}_{y^1}, \hat{\mathrm{p}}_{x^2}, \hat{\mathrm{p}}_{y^2})$, the $\hat{\mathrm{p}}_t$ is perfectly calibrated if satisfying the following equation:

$$\mathop{\mathbb{E}}_{\hat{b} \in B(\hat{\mathrm{p}}_t)} [\bar{\mathrm{p}}_t] = \hat{\mathrm{p}}_t, \quad \forall \hat{\mathrm{p}}_t \in [0, 1], t \in \{x^1.y^1, x^2, y^2\} \tag{8}$$

where $B(\hat{\mathrm{p}}_t)$ denotes the set of predicted bounding boxes $\hat{b}$ whose coordinate-wise confidence scores are equal to $\hat{\mathrm{p}}_t$, and $\bar{\mathrm{p}}_t$ denotes the coordinate-wise true probability associated with $\hat{b}$ as defined in Eq. 7.

Confidence Re-encoding We designed CR to estimate coordinate-wise confidence scores. To estimate coordinate-wise confidence scores using CR, we compute the following two feature representations: the output logit and the bounding box feature representation. The output logit of i-th prediction is defined as $\hat{\mathrm{z}}_i = \phi_C(\phi_F(\mathbf{x}_i))$. The bounding box feature representation is defined as $\hat{\mathbf{f}}_i = \phi_R[:-2](\phi_F(\mathbf{x}_i))$. To calibrate confidence scores with CAR, we leverage the bounding box feature $\hat{\mathbf{f}}_i$ to encode localization confidence to logit $\hat{\mathrm{z}}_i$. We obtain coordinate-wise confidence scores through CR ϕ_{CR} formulated as follows:

$$\hat{\mathbf{p}}_i = (\hat{\mathrm{p}}_{(x^1,i)}, \hat{\mathrm{p}}_{(y^1,i)}, \hat{\mathrm{p}}_{(x^2,i)}, \hat{\mathrm{p}}_{(y^2,i)}) = \sigma(\frac{\hat{\mathrm{z}}_i}{\phi_{\mathrm{CR}_t}(\hat{\mathbf{f}}_i)} + \beta_t), \quad t \in \{x_1.y_1, x_2, y_2\}. \tag{9}$$

Finally, we train CR ϕ_{CR} to satisfy Eq. 8 by aligning the coordinate-wise confidence scores (Eq. 9) with the corresponding the empirical coordinate-wise accuracies defined by CAR (Eq. 7). The object loss is derived from the negative log-likelihood (NLL) and is formulated as follows:

$$\mathcal{L}_{cal} := \mathbb{E}\left[-(\bar{\mathrm{p}} \log(\hat{\mathrm{p}}) + (1 - \bar{\mathrm{p}}) \log(1 - \hat{\mathrm{p}}))\right]. \tag{10}$$

Directional Displacement Estimation We propose a directional displacement estimation that estimates the relative directional displacement of each predicted bounding box coordinate with respect to the ground truth bounding box. The proposed method determines whether the ground truth coordinate is larger or smaller than the predicted coordinate and thereby identifies the direction of the

coordinate error. The directional displacement estimator ϕ_{DDE} uses an MLP that takes the predicted logit vector $\hat{\mathbf{z}}_i$ and geometric attributes of the bounding box as input, including the center coordinates $\hat{\text{cx}}_i$ and $\hat{\text{cy}}_i$, width $\hat{\text{w}}_i$, height $\hat{\text{h}}_i$, area $\hat{\text{A}}_i$, and width-to-height ratio $\hat{\text{R}}_i$. The directional targets of each t coordinate $\bar{\text{s}}_t$ is defined as follows:

$$\bar{\text{s}}_t = \begin{cases} +1 & \text{if } \hat{t} - t > 0 \\ -1 & \text{otherwise} \end{cases}, t \in \{x^1.y^1, x^2, y^2\}. \tag{11}$$

In addition, the directional displacement estimator ϕ_{DDE} produces the directional prediction $\hat{\mathbf{s}}_i$ as follows:

$$\hat{\mathbf{s}}_i = (\hat{\text{s}}_{(x^1,i)}, \hat{\text{s}}_{(y^1,i)}, \hat{\text{s}}_{(x^2,i)}, \hat{\text{s}}_{(y^2,i)}) = \phi_{\text{DDE}}(\hat{\mathbf{z}}_i, \hat{\text{cx}}_i, \hat{\text{cy}}_i, \hat{\text{w}}_i, \hat{\text{h}}_i, \hat{\text{A}}_i, \hat{\text{R}}_i). \tag{12}$$

The final directional prediction is defined as:

$$\hat{\text{s}}_{(t,i)} = \begin{cases} +1 & \text{if } \sigma(\hat{z}^c_{s_{(t,i)}}) \geq \tau^c_t \\ -1 & \text{otherwise} \end{cases} \quad , t \in \{x^1, y^1, x^2, y^2\}, \tag{13}$$

where $\hat{\text{z}}^c_{s_{(t,i)}}$ denotes the i-th predicted logit for class c obtained through ϕ_{DDE}, $\sigma(\cdot)$ denotes the sigmoid function, and τ^c_t denotes the class-specific threshold for class c, determined based on the direction accuracy during the training process of ϕ_{DDE}. We first train the confidence re-encoder and then train the directional displacement estimator. Since each deviation direction has a binary label of $+1$ or -1, the directional displacement estimator is trained with the binary cross-entropy loss to estimate coordinate-wise deviation directions. During inference, $\sigma(\hat{\text{z}}^c_{s_{(t,i)}})$ is compared with the threshold τ^c_t found during training to obtain the predicted direction $\hat{\text{s}}_{(t,i)}$.

Furthermore, to provide both coordinate-level and box-level calibrated confidence scores, we show that intersection over union (IoU) can be derived from CAR and directional information between predicted and ground-truth bounding boxes. When coordinate-wise expected calibration error (C-ECE) reaches the minimum value, the calibrated confidence score equals CAR. Therefore, IoU approximation uses coordinate-wise calibrated confidence scores with directional information, and the approximated IoU score is further calibrated with ground-truth IoU using isotonic regression [30] and Platt scaling [25]. More details are provided in the supplementary material.

4.2 Coordinate-wise Calibration Metric

Coordinate-wise Expected Calibration Error To measure coordinate-wise calibration performance for Eq. 8, this work proposes the coordinate-wise expected calibration error (C-ECE), a new metric that quantitatively evaluates the alignment with CAR and coordinate-wise confidence score. The equation is given as follows:

$$\mathrm{C}_t\text{-ECE} = \frac{1}{C}\sum_{c=1}^{C}\sum_{j=1}^{J}\frac{|B_j^c|}{|B^c|}\left|\bar{\mathrm{p}}_t(B_j^c) - \hat{\mathrm{p}}_t(B_j^c)\right| \quad , t \in \{x^1.y^1, x^2, y^2\}, \tag{14}$$

where t denotes the coordinates that represent the top-left and bottom-right, and the estimation follows a class-wise scheme to prevent a specific class from dominating the error. Moreover, the evaluation partitions continuous confidence scores into J bins that are equally spaced over the confidence range. Accordingly, B_j indicates the j-th bin, and $\hat{\mathrm{p}}_t(B_j^c)$ denotes the average coordinate-wise confidence score of the prediction belonging to B_j^c, while $\bar{\mathrm{p}}_t(B_j^c)$ represents the average CAR of the predictions in the corresponding bin.

Direction-aware Calibration Error We introduce direction-aware calibration error (Da-CE) to evaluate coordinate-wise calibration error that incorporates directional information. The Da-CE evaluates the degree of misalignment between predicted bounding box coordinates and ground truth bounding box coordinates while accounting for directional deviation. We first define coordinate-wise mismatch between predicted and ground truth coordinates as:

$$\bar{\mathrm{U}}_{(t,i)} = \bar{\mathrm{s}}_{(t,i)} \cdot (1 - \bar{\mathrm{p}}_{(t,i)}) \quad , t \in \{x^1.y^1, x^2, y^2\}, \tag{15}$$

where $\bar{\mathrm{p}}_t$ denotes the CAR between ground truth bounding box and predicted bounding box and $\bar{\mathbf{s}}_t$ indicates the direction of CAR. Furthermore, predicted mismatch that incorporates the predicted direction is defined as:

$$\hat{\mathrm{U}}_{(t,i)} = \hat{\mathrm{s}}_{(t,i)} \cdot (1 - \hat{\mathrm{p}}_{(t,i)}) \quad , t \in \{x^1.y^1, x^2, y^2\}. \tag{16}$$

With these definitions, Da-CE is defined as:

$$\mathrm{Da}_t\text{-CE} = \frac{1}{C}\sum_{c=1}^{C}\sum_{j=1}^{J}\frac{|B_j^c|}{|B^c|}\Big(\frac{1}{|B_{j_{\mathrm{TP}}}^c|}\sum_{i \in B_{j_{\mathrm{TP}}}^c}|\bar{\mathrm{U}}_{(t,i)} - \hat{\mathrm{U}}_{(t,i)}|\Big), \tag{17}$$

where t denotes a coordinate, i.e., $t \in x^1, y^1, x^2, y^2$, and $B_{j_{\mathrm{TP}}}^c$ denotes the set of true positive samples in the j-th bin for class c. Da-CE partitions predictions into J bins and computes sample-level errors to avoid cancellation among directional mismatches. Since false positive samples lack deviation directions, Da-CE sets false positive errors to zero and uses only true positive samples for error computation.

5 Experiments

Dataset We conduct experiments on COCO [16] and Cityscapes [3]. COCO contains 80 common object classes, while Cityscapes focuses on autonomous driving scenarios with 8 classes. To evaluate robustness under domain shift, we use COCO-C [9], which applies 3 types of corruptions to COCO, and Foggy Cityscapes [26], which simulates foggy conditions on Cityscapes. Further details are provided in the supplementary material.

Table 1: Comparison results with SOTA methods on COCO. For evaluation, we train CR on COCO minival and evaluate CR on COCO minitest. **Bold** and underlined values indicate the best and second-best results, respectively.

	Method	Coordinate-wise				Overall					
		C_{x_1}-ECE↓	C_{y_1}-ECE↓	C_{x_2}-ECE↓	C_{y_2}-ECE↓	D-ECE ↓	LaECE↓	$LaECE_0$ ↓	$LaACE_0$ ↓	AP↑	LRP↓
	Uncalibrated [35]	17.3	17.4	17.4	17.2	14.9	12.2	12.7	27.1	51.3	57.3
Train-time	TCD [19]	18.0	18.1	18.1	17.5	14.4	12.5	13.1	26.8	51.3	57.1
Train-time	BPC [18]	15.5	15.3	15.3	15.2	11.3	12.3	12.8	25.4	50.3	58.4
Train-time	Cal-DETR [20]	14.0	13.9	13.9	13.6	9.8	11.7	11.7	24.6	**52.5**	**56.2**
Post-hoc	IR for $LaECE_0$ [12]	12.0	11.9	11.8	11.9	2.4	8.4	**7.8**	23.1	51.0	57.3
Post-hoc	PS for $LaECE_0$ [12]	13.9	14.0	14.0	13.8	**2.3**	10.0	9.7	23.5	51.3	57.3
Post-hoc	IR for Ours	7.7	10.5	11.4	**10.6**	2.5	**8.1**	7.9	**22.7**	50.4	57.4
Post-hoc	PS for Ours	**7.6**	**10.4**	**10.9**	10.9	2.4	9.9	10.0	23.1	50.3	57.3

Evaluation Metric This work evaluates the calibration performance of fine-grained confidence scores using coordinate-wise expected calibration error (C-ECE) Eq. 14 and direction-aware calibration error (Da-CE) Eq. 17, and adopts established metrics from prior studies to assess the calibration performance of a calibrator on overall accuracy. The evaluation utilizes detection expected calibration error (D-ECE) [13] ($\tau = 0.5$), localization-aware expected calibration error (LaECE) [22]($\tau = 0.5$) and $LaECE_0$ [12]($\tau = 0$). For evaluation, average precision (AP) is computed over the top-100 detections, and localization recall precision (LRP) [21] is computed using the LRP threshold obtained on the validation set. Additional details are described in the supplementary material.

Implementation Details For comparison with existing methods, we leverage deformable-DETR [35] with ResNet-50 [7], which previous methods mainly utilized. The calibration procedure follows Kuzucu et al. [12] and determines the calibration and operating thresholds by cross-validating LRP (IoU $\tau = 0$).

5.1 Comparison with State-of-the-Art Methods

Common Object Scenarios Tab. 1 presents comparison results with SOTA methods on COCO. While prior approaches achieve impressive box-level calibration performance in terms of D-ECE, LaECE, and $LaECE_0$, a single calibrated confidence score fails to capture the diverse distributions of CARs across bounding box coordinates, resulting in limited alignment with coordinate-wise localization accuracy. In contrast, the coordinate-wise confidence scores estimated by CR utilizing PS achieves an average C-ECE of 10.0 over prior methods, while the box-level calibration performance obtained from the approximated IoU remains competitive with existing approaches.

Autonomous Driving Scenarios Tab. 2 presents comparison results with SOTA methods on Cityscapes. Prior approaches achieve well-calibrated box-level

Table 2: Comparison results with SOTA methods on Cityscapes dataset. For the experiment, we fit the model on Cityscapes minival and evaluate the CR on Cityscapes minitest. **Bold** and underlined values indicate the best and second-best results, respectively.

	Method	Coordinate-wise				Overall					
		C_{x_1}-ECE↓	C_{y_1}-ECE↓	C_{x_2}-ECE↓	C_{y_2}-ECE↓	D-ECE ↓	LaECE↓	$LaECE_0$ ↓	$LaACE_0$ ↓	AP↑	LRP↓
	Uncalibrated [35]	13.7	15.5	14.2	15.5	13.7	11.8	13.4	30.8	44.5	66.4
Train-time	TCD [19]	12.5	13.7	12.4	12.8	15.3	11.7	12.6	29.3	29.1	76.9
Train-time	BPC [18]	13.8	15.4	14.0	15.8	5.4	13.7	14.2	**26.6**	30.5	74.2
Train-time	Cal-DETR [20]	13.4	15.1	13.4	14.5	13.0	11.0	12.7	29.4	38.7	70.6
Post-hoc	IR for $LaECE_0$ [12]	12.2	13.6	11.6	13.9	1.5	8.0	7.5	27.2	43.5	**66.4**
Post-hoc	PS for $LaECE_0$ [12]	12.9	14.7	13.1	15.1	**1.0**	10.0	9.6	27.4	**44.5**	**66.4**
Post-hoc	IR for Ours	6.9	**11.1**	**10.3**	11.3	1.2	**6.2**	**6.5**	29.7	42.5	67.0
Post-hoc	PS for Ours	**6.5**	11.4	10.5	**10.6**	1.1	9.6	9.4	27.6	42.0	66.9

Table 3: Comparison with SOTA methods on corrupted COCO dataset for domain shift scenarios. To simulate a domain shift scenario, we train the confidence re-encoder on COCO minival and report inference results on COCO-C minitest. **Bold** and underlined values indicate the best and second-best results, respectively.

	Method	Coordinate-wise				Overall					
		C_{x_1}-ECE↓	C_{y_1}-ECE↓	C_{x_2}-ECE↓	C_{y_2}-ECE↓	D-ECE ↓	LaECE↓	$LaECE_0$ ↓	$LaACE_0$ ↓	AP↑	LRP↓
	Uncalibrated [35]	20.3	20.3	20.2	20.2	15.9	15.2	15.2	29.4	30.5	74.0
Train-time	TCD [19]	20.6	20.6	20.4	20.3	16.4	14.9	14.8	28.9	29.7	74.4
Train-time	BPC [18]	19.0	19.0	19.0	18.9	13.0	15.0	14.5	27.5	29.6	75.1
Train-time	Cal-DETR [20]	17.8	18.0	17.7	17.7	11.3	15.4	14.6	27.0	**30.5**	**73.9**
Post-hoc	IR for $LaECE_0$ [12]	16.1	16.3	16.0	16.1	3.1	11.7	11.2	26.2	30.2	74.0
Post-hoc	PS for $LaECE_0$ [12]	18.5	18.6	18.4	18.5	**2.6**	14.4	13.7	26.4	**30.5**	74.0
Post-hoc	IR for Ours	11.9	15.1	15.4	**13.9**	2.9	**11.5**	**11.0**	26.2	29.8	74.1
Post-hoc	PS for Ours	**11.7**	**15.0**	**15.3**	14.4	**2.6**	15.3	14.1	**26.1**	29.6	74.3

performance, but coordinate-wise calibration remains inadequate. For example, PS for $LaECE_0$ [12] records 9.6 in $LaECE_0$, yet shows 14.7 in Cy^1-ECE and 15.1 in Cy^2-ECE. This result indicates poor alignment with coordinate-wise localization accuracy. In contrast, CR with PS achieves an average C-ECE of 9.8 across all coordinates, demonstrating consistent coordinate-wise calibration.

Domain Shift Tab. 3 and Tab. 4 present results under domain shift on COCO-C and Foggy Cityscapes, respectively. While existing methods maintain competitive box-level calibration performance under domain shift, they exhibit degraded coordinate-wise calibration, revealing limited ability to represent per-coordinate localization accuracy in unseen domains. In contrast, CR consistently achieves superior coordinate-wise calibration, recording an average C-ECE of 14.1 on COCO-C and 12.0 on Foggy Cityscapes, while maintaining competitive box-level calibration performance using PS, comparable to existing methods.

Table 4: Comparison results with SOTA methods on Foggy Cityscapes. To simulate a domain shift scenario, we train the confidence re-encoder on Cityscapes minival and report the inference results on Foggy Cityscapes minitest. **Bold** and underlined values indicate the best and second-best results, respectively.

	Method	Coordinate-wise				Overall					
		C_{x_1}-ECE↓	C_{y_1}-ECE↓	C_{x_2}-ECE↓	C_{y_2}-ECE↓	D-ECE ↓	LaECE↓	$LaECE_0$ ↓	$LaACE_0$ ↓	AP↑	LRP↓
	Uncalibrated [35]	16.7	18.6	16.9	18.4	13.9	13.5	13.8	28.8	31.0	74.7
Train-time	TCD [19]	16.7	18.2	16.7	18.0	18.0	12.8	13.4	28.3	23.6	81.1
Train-time	BPC [18]	16.9	18.0	17.0	18.8	5.2	17.5	17.2	28.7	23.7	79.9
Train-time	Cal-DETR [20]	16.8	18.6	16.8	18.3	12.4	12.8	13.8	28.4	28.4	76.9
Post-hoc	IR for $LaECE_0$ [12]	13.7	15.8	14.1	15.8	1.3	**7.8**	**7.5**	23.9	30.9	**74.6**
Post-hoc	PS for $LaECE_0$ [12]	17.1	19.4	17.3	19.0	**1.0**	12.0	11.6	24.4	**31.0**	74.7
Post-hoc	IR for Ours	**10.2**	13.5	12.9	11.7	1.2	8.7	8.4	**22.2**	30.3	74.8
Post-hoc	PS for Ours	**10.2**	**13.3**	**12.7**	**11.6**	1.4	11.8	10.6	25.0	30.8	74.9

Table 5: Ablation results of direction influence for C-ECE For the ablation study, we compare results on the COCO dataset over five random seeds with and without prediction deviation direction estimation. **Bold** indicates the best results.

Direction	$\bar{C}_{x_1}$-ECE↓	$\bar{C}_{y_1}$-ECE↓	$\bar{C}_{x_2}$-ECE↓	$\bar{C}_{y_2}$-ECE↓	$LaE\bar{C}E_0$ ↓	$LaA\bar{C}E_0$ ↓
-	11.22 (± 0.058)	11.90 (± 0.106)	10.42 (± 0.054)	10.62 (± 0.134)	9.96 (± 0.022)	**23.14** (± 0.034)
✓	**11.18** (± 0.038)	**10.86** (± 0.042)	**10.26** (± 0.074)	**10.58** (± 0.106)	**9.88** (± 0.022)	**23.14** (± 0.002)

5.2 Analysis of Rethinking Detection Calibration

Ablation Study Tab. 5 presents the ablation results over five random seeds with and without directional displacement estimation (DDE). The model without DDE does not predict deviation directions, and the IoU approximation therefore excludes directional information. The lower $LaECE_0$ and $LaACE_0$ values show that directional information improves box-level IoU approximation and calibration. The improved box-level calibration also enables the estimation of a more appropriate operating threshold for predictions. As a result, DDE improves coordinate-level calibration in terms of C-ECE and reduces variance across random seeds, indicating more robust prediction.

Comparison with GP-Normal GP-Normal [14] performs calibration using probabilistic object detectors that estimate uncertainty through coordinate-wise probability distributions and align the estimated uncertainty with actual localization errors. For a fair comparison with GP-Normal, which estimates uncertainty via coordinate-wise probability distributions, we measure correlation with coordinate-wise error, using normalized inverse standard deviation for GP-Normal and calibrated confidence scores for ReDC. ReDC achieves a higher mean correlation than GP-Normal (0.1771 vs. 0.0263), indicating that ReDC better captures coordinate-wise errors.

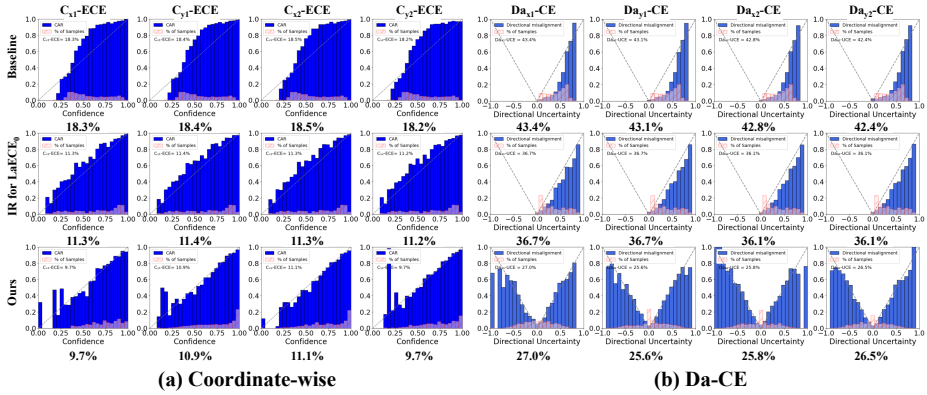

Fig. 5: Reliability Diagram. (a) C-ECE diagrams plot average CAR against predicted confidence per bin. (b) Da-CE diagrams plot average misalignment against predicted directional mismatch. Perfect calibration corresponds to alignment with the diagonal.

Correlation Analysis of y^2 Coordinate Error Fig. 6 shows coordinate-wise confidence distributions for boxes with large localization errors along y^2. As shown in Fig. 6-As shown in (a), prior methods rely on a single confidence score and fail to reflect increasing localization error. In contrast, (b) shows ReDC's confidence score negatively correlates with pixel distance error along y^2, indicating that ReDC quantitatively captures positional confidence and conveys positional information rarely provided by conventional 2D detection frameworks.

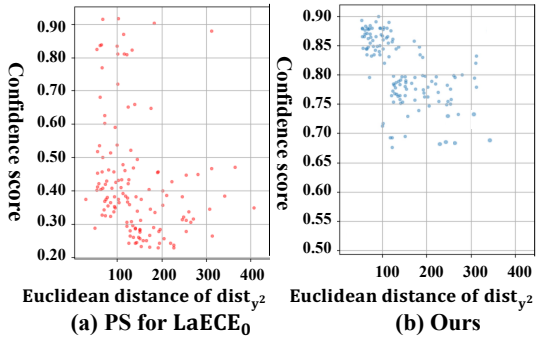

Fig. 6: Analysis of y^2 confidence scores for samples exhibiting large errors in the y^2 coordinate. Samples selected from the top 3% per class exhibiting large errors between ground-truth and predicted bounding boxes along the y^2 coordinate.

Reliable Diagram In Fig. 5, reliability diagrams compare ReDC with IR for LaECE and the baseline (DINO). Since the baseline and IR for LaECE only capture box-level accuracy, they produce under-confident coordinate-wise predictions, as actual coordinate-wise accuracy exceeds the empirical IoU distribution. ReDC, in contrast, achieves well-calibrated coordinate-wise confidence while maintaining competitive box-level calibration. The Da-CE diagrams further reveal systematic directional bias in the baseline and IR for LaECE, whose predicted

Table 6: Model-agnostic post-hoc confidence calibration. Bold and underlined values indicate the best and second-best results, respectively.

	Method	Type	C_{x_1}-ECE↓	C_{y_1}-ECE↓	C_{x_2}-ECE↓	C_{y_2}-ECE↓	Da-CE↓
Base	VFNET [32]	1	22.9	22.9	22.7	22.4	48.1
	Cascade R-CNN [1]	2	16.5	16.0	16.3	16.4	**19.3**
	DINO [31]	ViT	18.3	18.4	18.5	18.2	42.9
Ours	VFNET [32]	1	<u>12.9</u>	12.9	14.2	13.0	<u>25.7</u>
	Cascade R-CNN [1]	2	<u>10.6</u>	**11.0**	<u>12.0</u>	<u>11.8</u>	28.1
	DINO [31]	ViT	**9.4**	**11.0**	**11.1**	**9.6**	26.0

mismatch directions fail to align with actual coordinate-wise error directions. ReDC substantially reduces across all coordinates.

Model-agnostic Coordinate-wise Calibration Tab. 6 reports coordinate-wise calibration results across diverse detector architectures, including one-stage, two-stage, and transformer-based detectors. Since CR leverages bounding box features and logits that any detector already produces, it requires no architectural modifications and achieves consistent coordinate-wise calibration performance regardless of the detector.

6 Conclusion

We propose ReDC, a post-hoc calibration framework that provides coordinate-wise confidence scores for object detection. We introduce CAR to measure localization accuracy at each coordinate and train a lightweight confidence re-encoder to align coordinate-wise confidence scores with CAR. ReDC further captures the directional characteristics of localization errors through DDE, and we introduce C-ECE and Da-CE as new metrics that jointly assess coordinate-level and directional accuracy. Extensive experiments on COCO and Cityscapes, under both in-domain and domain shift scenarios, show that ReDC outperforms existing methods in coordinate-wise calibration while achieving competitive box-level performance across diverse detector architectures.

Acknowledgements

This work was partly supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) [RS-2021-II211341, Artificial Intelligence Graduate School Program (Chung-Ang University); No.RS-2024-00437576, Development of autonomous performance improvement technology for video surveillance system based on edge-analysis server connection] and a grant (22193MFDS471) from the Ministry of Food and Drug Safety in 2024.

References

1. Cai, Z., Vasconcelos, N.: Cascade r-cnn: Delving into high quality object detection. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2018)
2. Choi, J., Chun, D., Kim, H., Lee, H.J.: Gaussian yolov3: An accurate and fast object detector using localization uncertainty for autonomous driving. In: The IEEE International Conference on Computer Vision (ICCV) (2019)
3. Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B.: The cityscapes dataset for semantic urban scene understanding. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
4. Feng, D., Harakeh, A., Waslander, S.L., Dietmayer, K.: A review and comparative study on probabilistic object detection in autonomous driving. IEEE Transactions on Intelligent Transportation Systems **23**(8), 9961–9980 (2021)
5. Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q.: On calibration of modern neural networks. In: International Conference on Machine Learning (ICML) (2017)
6. Harakeh, A., Smart, M., Waslander, S.L.: Bayesod: A bayesian approach for uncertainty estimation in deep object detectors. In: 2020 IEEE International Conference on Robotics and Automation (ICRA). pp. 87–93. IEEE (2020)
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
8. He, Y., Zhu, C., Wang, J., Savvides, M., Zhang, X.: Bounding box regression with uncertainty for accurate object detection. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
9. Hendrycks, D., Dietterich, T.: Benchmarking neural network robustness to common corruptions and perturbations. In: International Conference on Learning Representations (ICLR) (2019)
10. Jaeger, P.F., Kohl, S.A., Bickelhaupt, S., Isensee, F., Kuder, T.A., Schlemmer, H.P., Maier-Hein, K.H.: Retina u-net: Embarrassingly simple exploitation of segmentation supervision for medical object detection. In: Machine learning for health workshop. pp. 171–183. PMLR (2020)
11. Jung, S., Seo, S., Jeong, Y., Choi, J.: Scaling of class-wise training losses for post-hoc calibration. In: International Conference on Machine Learning (ICML) (2023)
12. Kuzucu, S., Oksuz, K., Sadeghi, J., Dokania, P.K.: On calibration of object detectors: Pitfalls, evaluation and baselines. In: The European Conference on Computer Vision (ECCV) (2024)
13. Küppers, F., Kronenberger, J., Shantia, A., Haselhoff, A.: Multivariate confidence calibration for object detection. In: Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2020)
14. Küppers, F., Schneider, J., Haselhoff, A.: Parametric and multivariate uncertainty calibration for regression and object detection. In: European Conference on Computer Vision Workshops (ECCVW) (2022)
15. Lee, Y., Hwang, J.W., Kim, H.I., Yun, K., Kown, Y., Hwang, S.J.: Localization uncertainty estimation for anchor-free object detection. In: European Conference on Computer Vision Workshops (ECCVW) (2022)
16. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: European conference on computer vision (ECCV) (2014)
17. Liu, Y., Jourabloo, A., Liu, X.: Learning deep models for face anti-spoofing: Binary or auxiliary supervision. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2018)

18. Munir, M.A., Khan, M.H., Khan, S., Khan, F.: Bridging precision and confidence: A train-time loss for calibrating object detection. Conference on Computer Vision and Pattern Recognition (CVPR) (2023)
19. Munir, M.A., Khan, M.H., Sarfraz, M., Ali, M.: Towards improving calibration in object detection under domain shift. Advances in Neural Information Processing Systems (NeurIPS) (2022)
20. Munir, M.A., Khan, S., Khan, M.H., Ali, M., Khan, F.: Cal-detr: Calibrated detection transformer. Advances in Neural Information Processing Systems (NeurIPS (2023)
21. Oksuz, K., Cam, B., Akbas, E., Kalkan, S.: Localization recall precision (lrp): A new performance metric for object detection. In: European Conference on Computer Vision (ECCV) (2018)
22. Oksuz, K., Joy, T., Dokania, P.K.: Towards building self-aware object detectors via reliable uncertainty quantification and calibration. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2023)
23. Pan, T.Y., Zhang, C., Li, Y., Hu, H., Xuan, D., Changpinyo, S., Gong, B., Chao, W.L.: On model calibration for long-tailed object detection and instance segmentation. In: Advances in Neural Information Processing Systems (NeurIPS) (2021)
24. Pathiraja, B., Gunawardhana, M., Khan, M.H.: Multiclass confidence and localization calibration for object detection. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2023)
25. Platt, J., et al.: Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. Advances in large margin classifiers **10**(3), 61–74 (1999)
26. Sakaridis, C., Dai, D., Van Gool, L.: Semantic foggy scene understanding with synthetic data. International Journal of Computer Vision **126**(9), 973–992 (2018)
27. Song, H., Diethe, T., Kull, M., Flach, P.: Distribution calibration for regression. In: International Conference on Machine Learning (ICML) (2019)
28. Tomani, C., Cremers, D., Buettner, F.: Parameterized temperature scaling for boosting the expressive power in post-hoc uncertainty calibration. In: In European Conference on Computer Vision (ECCV) (2022)
29. Wu, B., Iandola, F., Jin, P.H., Keutzer, K.: Squeezedet: Unified, small, low power fully convolutional neural networks for real-time object detection for autonomous driving. In: Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2017)
30. Zadrozny, B., Elkan, C.: Transforming classifier scores into accurate multiclass probability estimates. In: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 694–699 (2002)
31. Zhang, H., Li, F., Liu, S., Zhang, L., Su, H., Zhu, J., Ni, L., Shum, H.Y.: DINO: DETR with improved denoising anchor boxes for end-to-end object detection. In: International Conference on Learning Representations (ICLR) (2023)
32. Zhang, H., Wang, Y., Dayoub, F., Sünderhauf, N.: Varifocalnet: An iou-aware dense object detector. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2021)
33. Zhang, J., Kailkhura, B., Han, T.: Mix-n-match: Ensemble and compositional methods for uncertainty calibration in deep learning. In: International Conference on Machine Learning (ICML) (2020)
34. Zhang, L., Wang, X., Yang, D., Sanford, T., Harmon, S., Turkbey, B., Wood, B.J., Roth, H., Myronenko, A., Xu, D., et al.: Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation. IEEE transactions on medical imaging **39**(7), 2531–2540 (2020)

35. Zhu, X., Su, W., Lu, L., Li, B., Wang, X., Dai, J.: Deformable detr: Deformable transformers for end-to-end object detection. In: International Conference on Learning Representations (ICLR) (2021)