FIVE

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

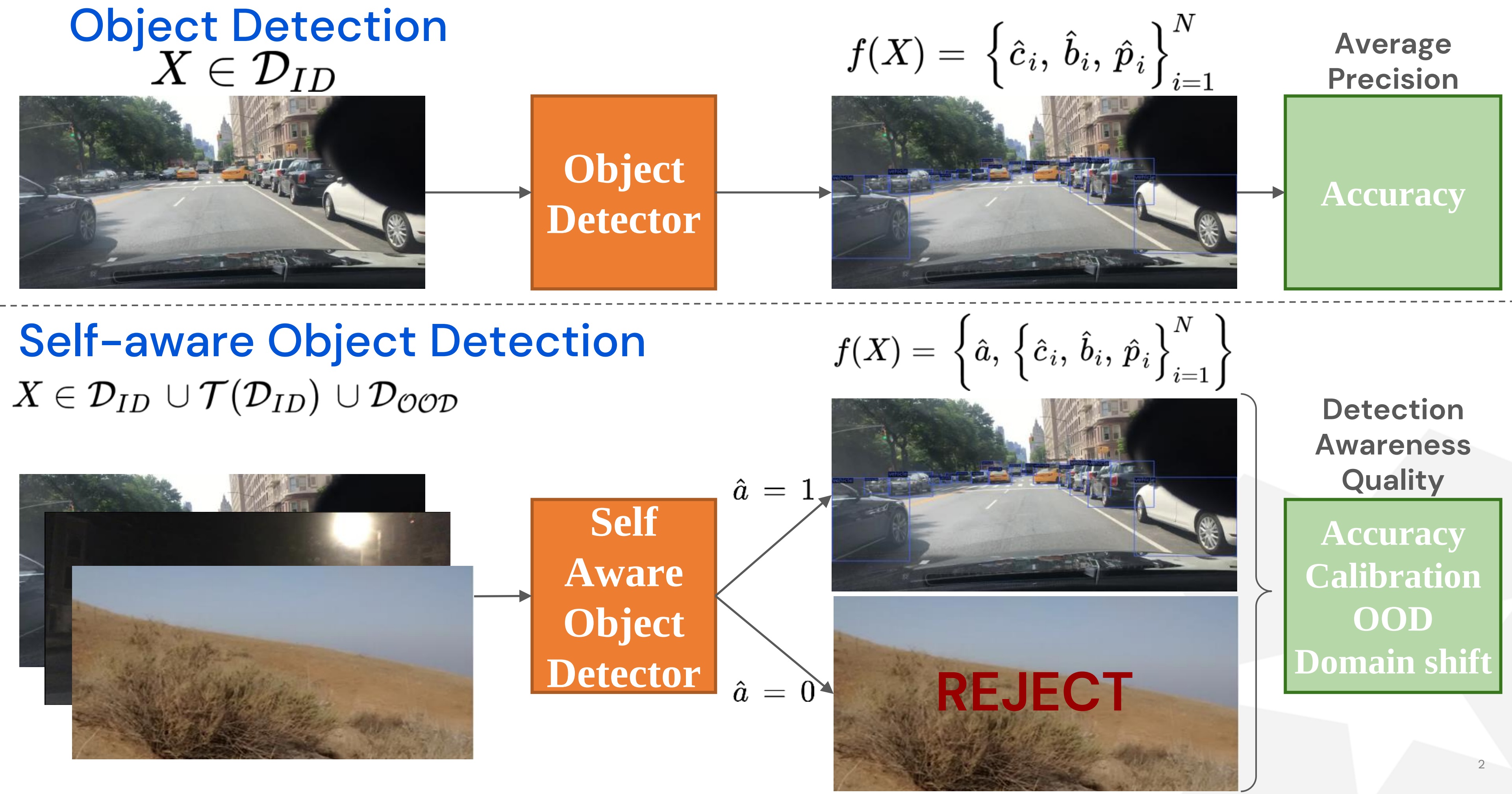
Towards Building Self-Aware Object Detectors via Reliable Uncertainty Quantification and Calibration

Kemal Oksuz
Research Scientist @ Five AI

Tom Joy
Research Scientist @ Five AI

Puneet Dokania
Director of Research @ Five AI
Senior Researcher @ Oxford University

Paper Tag: WED-AM-099


Object Detection
$X \in \mathcal{D}_{ID}$
Object Detector
$f(X) = \left\{\hat{c}_i,\, \hat{b}_i,\, \hat{p}_i\right\}_{i=1}^{N}$
vehicle
Average Precision
Accuracy
Self-aware Object Detection
$X \in \mathcal{D}_{ID} \cup \mathcal{T}(\mathcal{D}_{ID}) \cup \mathcal{D}_{OOD}$
Self Aware Object Detector
$f(X) = \left\{\hat{a},\, \left\{\hat{c}_i,\, \hat{b}_i,\, \hat{p}_i\right\}_{i=1}^{N}\right\}$
$\hat{a} = 1$
$\hat{a} = 0$
REJECT
Detection Awareness Quality
Accuracy
Calibration
OOD
Domain shift

The Drawbacks of Previous Approaches

- **Main Drawback: No unified approach in evaluation**

Calibration	Domain Shift	OOD Detection	Accuracy
Kuppers et al., Multivariate confidence calibration for object detection, CVPR 2020 Workshop	*Michaelis et al., Benchmarking robustness in object detection: Autonomous driving when winter is coming, NeurIPS 2019 Workshop*	*Harakeh and Waslander, Estimating and evaluating regression predictive uncertainty in deep object detectors, ICLR 2021*	*He et al., Bounding box regression with uncertainty for accurate object detection, CVPR 2019*
Kuppers et al., Parametric and multivariate uncertainty calibration for regression and object detection, ECCV 2022 Workshop	*Wang et al., Robust object detection via instance-level temporal cycle confusion, ICCV 2021*	*Du et al., Towards unknown-aware learning with virtual outlier synthesis, ICLR 2022*	*Cai et al., Learning a unified sample weighting network for object detection, CVPR 2020*
	Wu and Deng, Single-domain generalized object detection in urban scene via cyclic-disentangled self distillation, CVPR 2022		*Harakeh et al., Bayesod: A bayesian approach for uncertainty estimation in deep object detectors, ICRA 2020*
			Choi et al., Active learning for deep object detection via probabilistic modeling, ICCV 2021

- **Other drawbacks in evaluating the robustness aspects:**
  - Unideal data splits
  - Small scale test sets
  - Nontrivial nature of evaluating OOD detection for object detection

Contributions

- Self-aware Object Detection Task
  - Large-scale test sets for two use-cases
  - Performance measures
  - A baseline to convert any detector to the one that is self-aware

- Reliable Uncertainty Quantification in Object Detection
  - Different ways of quantifying uncertainty

- Calibration in Object Detection
  - Localisation-aware Expected Calibration Error
  - Relation between calibration and accuracy
  - Post-hoc calibration methods

Self-aware Object Detection – Task Overview

In-distribution Domain Shift Out-of-distribution

$$X \in \mathcal{D}_{ID} \cup \mathcal{T}(\mathcal{D}_{ID}) \cup \mathcal{D}_{OOD}$$

Self-aware Object Detector

$\hat{a} = 1$

$$f(X) = \left\{ \hat{a}, \left\{ \hat{c}_i, \hat{b}_i, \hat{p}_i \right\}_{i=1}^{N} \right\}$$

$\hat{a} = 0$

- The functionality to reject an image $\hat{a} \in \{0, 1\}$
- Output accurate and calibrated detections
- Be robust to domain shift

Self-aware Object Detection - Dataset Overview

Dataset	$\mathcal{D}_{\text{Train}}$	$\mathcal{D}_{\text{Val}}$	$\mathcal{D}_{\text{Test}}$		
			$\mathcal{D}_{\text{ID}}$	$\mathcal{T}(\mathcal{D}_{\text{ID}})$	$\mathcal{D}_{\text{OOD}}$
SAOD-Gen	COCO$^{(train)}$	COCO$^{(val)}$	Obj45K	Obj45K-C	SiNObj110K-OOD
SAOD-AV	nuImages$^{(train)}$	nuImages$^{(val)}$	BDD45K	BDD45K-C	SiNObj110K-OOD

For each use-case:

- ID : 45K
- Domain Shift : 3 x 45K (ImageNet-C style severities 1, 3, 5)
- OOD : 110K **>> 1-2K images in existing datasets**

Hendrycks and Dietterich, Benchmarking neural network robustness to common corruptions and perturbations, ICLR 2019

Self-aware Object Detection - Overview of Used Detectors

Two-stage:

- Faster R-CNN (F-RCNN)
- Rank & Sort R-CNN (RS-RCNN)

One-stage:

- Adaptive Training Sample Selection (ATSS)

Transformer-based:

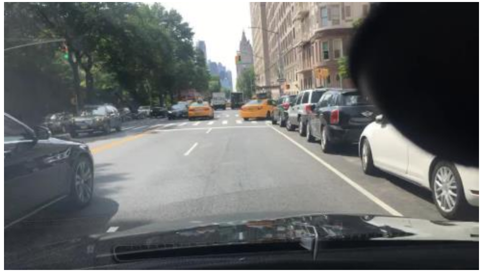
- Deformable DETR (DDETR)

Probabilistic detectors:

- Faster R-CNN minimizing Negative Log-likelihood (NLL-RCNN)
- Faster R-CNN minimizing Energy Score (ES-RCNN)

Image-level Uncertainty – Motivation

<u>Training set</u>



<u>Test set:</u> Can the detector operate reliably in this scene?

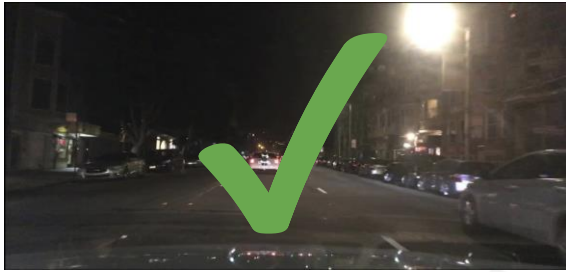


Image-level Uncertainty - How to Obtain?

Step 1: Quantify image-level uncertainty $\mathcal{G} : \mathcal{X} \rightarrow \mathbb{R}$
Step 2: Cross-validate a threshold $\bar{u} \in \mathbb{R}$

- **If $\mathcal{G}(X) < \bar{u}$ ACCEPT; else REJECT.**

Different detection uncertainties

Dataset	Detector	Classification			Localisation		
		$\mathrm{H}(\hat{p}_i^{raw})$	DS	$1-\hat{p}_i$	$\lvert\Sigma\rvert$	$\mathrm{tr}(\Sigma)$	$\mathrm{H}(\Sigma)$
SAOD Gen	F-RCNN	92.6	89.7	**94.1**	N/A	N/A	N/A
	RS-RCNN	93.7	30.0	**94.8**	N/A	N/A	N/A
	ATSS	**94.3**	36.9	94.2	N/A	N/A	N/A
	D-DETR	93.9	73.8	**94.4**	N/A	N/A	N/A
	NLL-RCNN	92.4	89.0	**94.1**	87.6	87.5	87.7
	ES-RCNN	92.8	89.9	**94.1**	85.0	85.2	86.4
SAOD AV	F-RCNN	**97.3**	96.0	**97.3**	N/A	N/A	N/A
	ATSS	97.2	97.1	**97.6**	N/A	N/A	N/A

Aggregate detection uncertainties

Dataset	Detector	sum	mean	top-5	top-3	top-2	min
SAOD-Gen	F-RCNN	20.9	84.1	93.4	<u>94.1</u>	**94.4**	93.8
	RS-RCNN	85.8	85.8	94.3	**94.8**	**94.8**	93.5
	ATSS	66.2	86.3	93.8	**94.2**	<u>94.0</u>	92.6
	D-DETR	85.2	85.2	94.4	**94.7**	<u>94.6</u>	93.3
SAOD-AV	F-RCNN	27.1	84.1	96.4	<u>97.3</u>	**97.4**	96.0
	ATSS	18.8	92.2	**97.7**	<u>97.6</u>	97.3	95.7

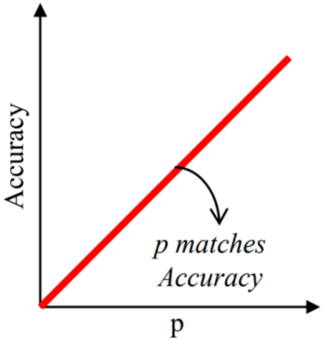
Calibration of Object Detectors – Definition

Calibration

$$\mathbb{P}(\hat{c}_i = c_i | \hat{p}_i) = \hat{p}_i, \forall \hat{p}_i \in [0,1]$$

Guo et al., On Calibration of Modern Neural Networks, ICML 2017

Kuppers et al., Multivariate confidence calibration for object detection, CVPR 2020 Workshop (i.e., Kuppers et al. apply it to object detection)



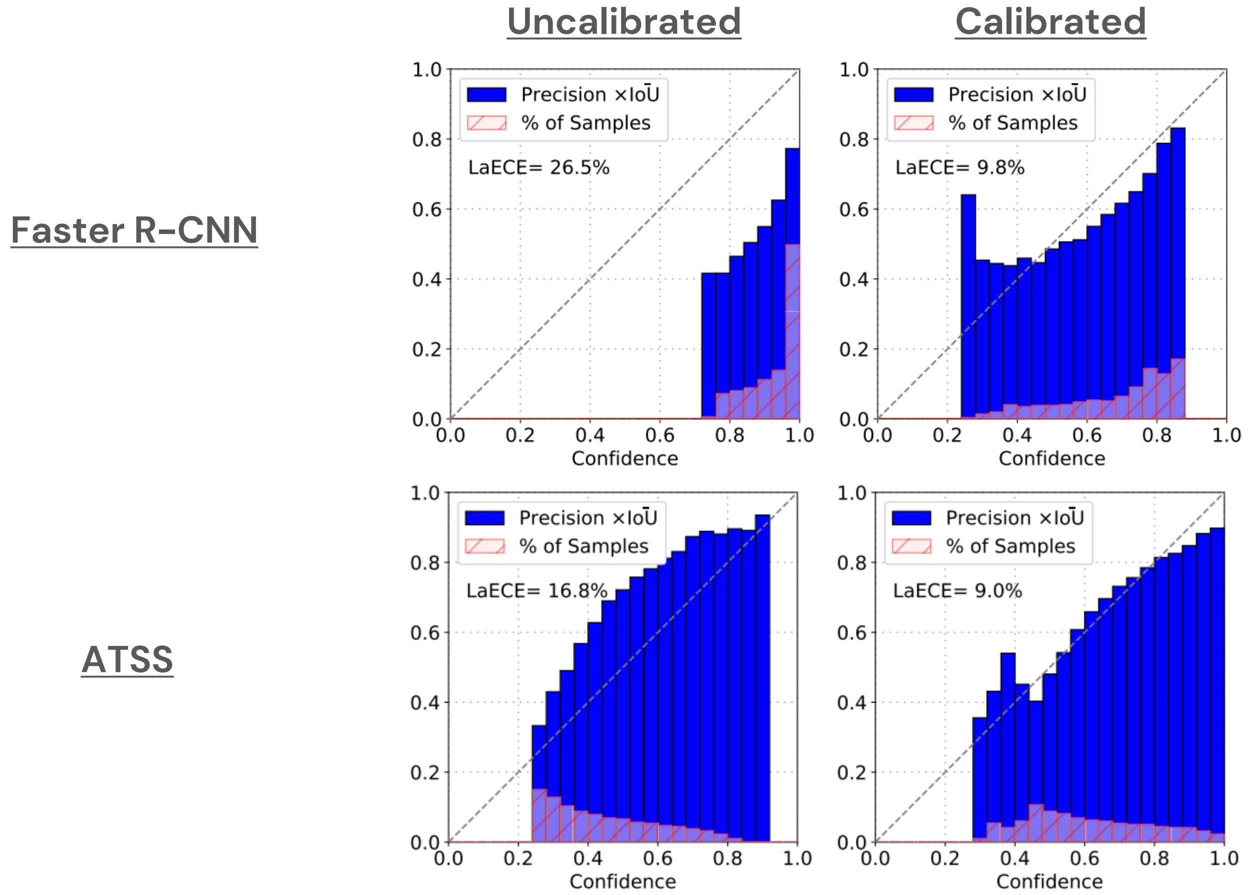
Localisation-aware Calibration of Object Detectors

$$\underbrace{\mathbb{P}(\hat{c}_i = c_i | \hat{p}_i)}_{\text{Classification Performance}} \underbrace{\mathbb{E}_{\hat{b}_i \in B_i(\hat{p}_i)} [\mathrm{IoU}(\hat{b}_i, b_{\psi(i)})]}_{\text{Localisation Performance}} = \hat{p}_i, \forall \hat{p}_i \in [0,1]$$

Localisation-aware Calibration Error

$$\mathrm{LaECE}^c = \sum_{j=1}^{J} \frac{|\hat{\mathcal{D}}_j^c|}{|\hat{\mathcal{D}}^c|} \left| \bar{p}_j^c - \mathrm{precision}^c(j) \times \bar{\mathrm{IoU}}^c(j) \right|$$

Reliability Diagrams and Post-hoc Calibration



Uncalibrated Faster R-CNN is **overconfident**

Uncalibrated ATSS is **underconfident**

Baseline Self-aware Object Detectors

Requirement 1: The functionality to reject an image

$\mathcal{G}(X)$ ⟶ Average 1 - p of top-3 confident detections in an image

$\bar{u}$ ⟶ Cross validated using pseudo-OOD approach

Requirement 2: Calibrated detections

Linear Regression as a calibrator

Qualitative Example

Input

Standard Faster R-CNN (AV-OD)

[vehicle, p=0.45; vehicle, p=0.11; vehicle, p=0.06]

Self-aware Faster R-CNN

Baseline Self-aware Object Detectors - Evaluation

ID

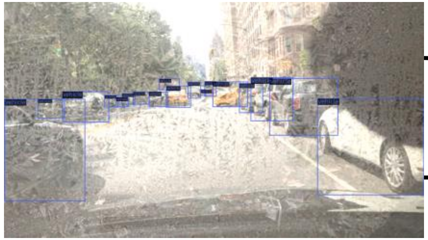
Accuracy

Calibration

In-distribution Quality (IDQ)

HarmonicMean(LRP, LaECE)

Domain Shift



Accuracy

Calibration

In-distribution Quality (IDQ)

HarmonicMean(LRP, LaECE)

OOD

OOD detection

Reliability of Uncertainties

HarmonicMean(TPR, TNR)

Detection Awareness Quality (DAQ)

Baseline Self-aware Object Detectors - Quantitative Results

	Self-aware Detector	DAQ↑	$\mathcal{D}_{\mathrm{OOD}}$	$\mathcal{D}_{\mathrm{ID}}$			$\mathcal{T}(\mathcal{D}_{\mathrm{ID}})$			$\mathcal{D}_{\mathrm{Val}}$	
			BA↑	IDQ↑	LaECE↓	LRP↓	IDQ↑	LaECE↓	LRP↓	LRP↓	AP↑
Gen	SA-F-RCNN	39.7	87.7	38.5	17.3	74.9	26.2	18.1	84.4	59.5	39.9
	SA-RS-RCNN	41.2	**88.9**	39.7	17.1	73.9	27.5	**17.8**	83.5	58.1	42.0
	SA-ATSS	41.4	87.8	39.7	16.6	74.0	27.8	18.2	83.2	58.5	42.8
	SA-D-DETR	**43.5**	**88.9**	**41.7**	**16.4**	**72.3**	**29.6**	17.9	**81.9**	**55.9**	**44.3**
AV	SA-F-RCNN	43.0	**91.0**	41.5	9.5	73.1	28.8	7.2	83.0	54.3	55.0
	SA-ATSS	**44.7**	85.8	**43.5**	**8.8**	**71.5**	**30.8**	**6.8**	**81.5**	**53.2**	**56.9**

Stronger detectors achieve higher DAQ

LaECE is ~10-20% and the performance on val set is significantly better.

Conclusion

- Self-aware Object Detection task that requires the object detectors to consider several robustness aspects
- Datasets and performance measures for evaluation
- A baseline method to convert any detector into one that is self-aware

FIVE

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Kemal Oksuz
Research Scientist @ Five AI

Tom Joy
Research Scientist @ Five AI

Puneet Dokania
Director of Research @ Five AI
Senior Researcher @ Oxford University

Code, datasets and more information: https://github.com/fiveai/saod