Conformal prediction for object detection

Sébastien Gerchinovitz
IRT Saint Exupéry & Institut de Mathématiques de Toulouse

Master Class DEEL - Uncertainty Quantification, November 18, 2024

Outline of the talk

This talk is about conformal prediction applied to object detection.

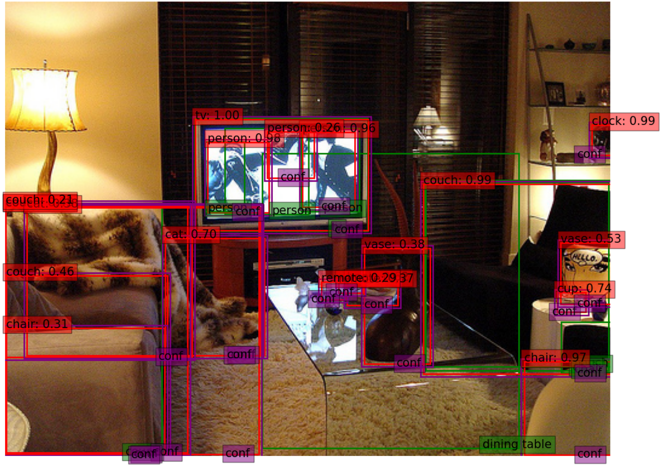
1. Setting
2. Localization task
3. Advanced topics

We present the DEEL works [6, 1] and a forthcoming paper.

Some results are illustrated via the PUNCC library.

Put briefly, given an input image, object detection consists in:

- identifying all objects of interest on an image, and not more;
- localizing them with rectangles (bounding boxes);
- classifying them into object types (classes).



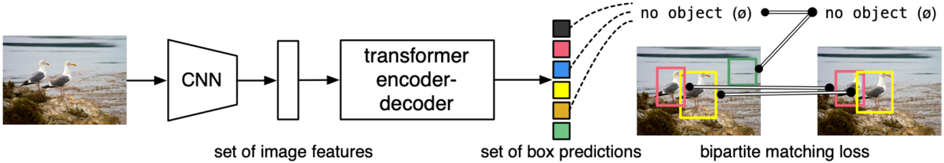
We work with the COCO 2017 dataset [14]. More precisely, we use the COCO validation dataset (5k images), which we split into

- $x\%$ images for calibration
- $(100 - x)\%$ images for test (we cannot use the unlabelled COCO test data)

(a) Iconic object images (b) Iconic scene images (c) Non-iconic images

Figure: Examples from the COCO dataset [14].

Example of model: DETR-ResNet-50, a transformer for object detection with a ResNet-50 backbone [4].



For a given image x, the model f predicts N bounding boxes with softmax scores and confidence scores, i.e., $f(x) = (f_1(x), \ldots, f_N(x)) \in (\mathcal{B} \times \mathcal{C} \times [0,1])^N$, where

- $\mathcal{B} \subset \mathbb{R}^4$ is the set of all admissible bounding boxes (rectangles encoded by the top-left and bottom-right coordinates);
- $\mathcal{C}$ is the set of probability vectors over $\{1, \ldots, 80\}$;
- confidence scores are $[0,1]$-valued.

Need for reliable UQ for autonomous driving, medical diagnosis, anomaly detection, etc. Complex interplay between localization and classification errors.

Earlier works. Until recently, various methods for UQ in OD had been proposed, but with no statistical guarantees on the uncertainty estimates.

- Deep Ensembles [12, 15]
- Monte Carlo-Dropout [17, 16, 3, 19]
- Direct Modeling [13]
- Probabilistic object detectors [10, 7]
- Uncertainty wrappers for safety engineering [9, 8, 5, 18]

Next: conformal prediction methods for OD developed within DEEL since 2022.

Conformal prediction for object localization

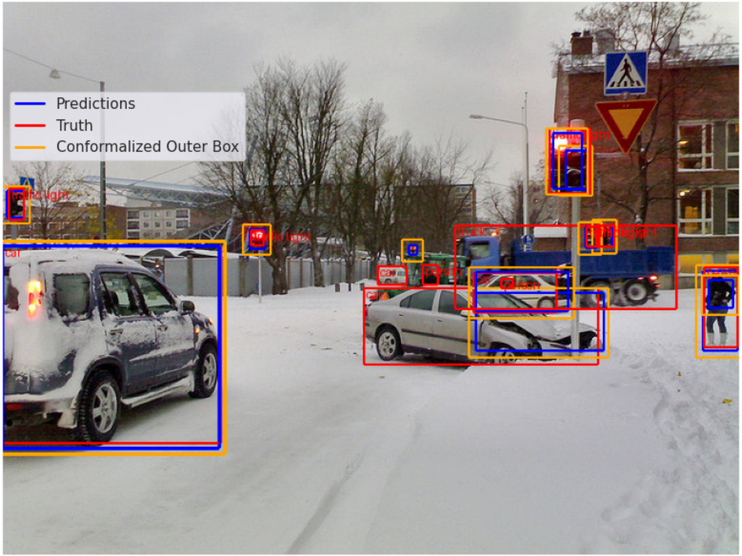
We now present the DEEL work [6], where we apply split conformal prediction to object localization.

Our simplest (box-level) procedure is the following.

1. Box matching: We construct a calibration set made of pairs of boxes $(X_i, Y_i) = (\text{predicted box, true box})$
2. We compute a margin so that most true boxes are covered by predicted+margin boxes

For Step 1 (extraction of pairs of boxes within calibration images), we use the Hungarian matching algorithm [11] with the IoU metric.

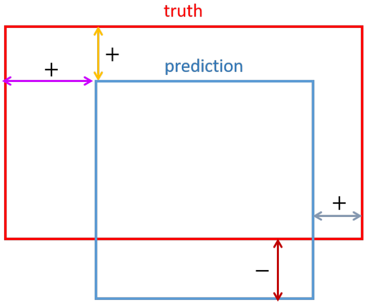
Example



Nonconformity score

For each pair of boxes in our calibration set, we compute the minimum margin $s_i \in \mathbb{R}$ to add around the predicted box so that it contains the true box:

$$s_i = \max\{\leftrightarrow, \updownarrow, \leftrightarrow, \updownarrow\}$$



Errors can be counted positively or negatively.

Conformal prediction: for a risk $\alpha \geqslant 1/(n_{\mathrm{cal}}+1)$, we select the margin s_i at rank $\lceil (n_{\mathrm{cal}}+1)(1-\alpha) \rceil$.

Non-uniform margins. We can also compute 4 margins independently.

For each coordinate $k = 1, \ldots, 4$, the selected margin is the error at rank $\lceil (n_{\mathrm{cal}} + 1)(1 - \alpha/4) \rceil$ in the calibration set (we made a Bonferroni correction).

Multiplicative margins. Instead of applying additive corrections, we can inflate (or deflate) predicted boxes multiplicatively.

To that end, we use normalized errors for the nonconformity score.

PUNCC demo

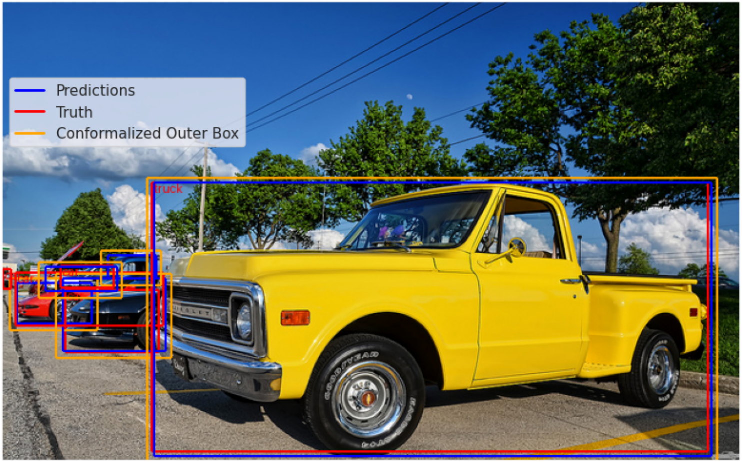
Open the *Conformal Object Detection Tutorial*:

`https://github.com/deel-ai/uq-masterclass`

Tutorial Notebooks

This repository contains the tutorial notebooks for the following topics:

- Part 1: Conformal Regression Tutorial Open in Colab
- Part 2: Conformal Classification Tutorial Open in Colab
- Part 3: Conformal Object Detection Tutorial Open in Colab


Predictions
Truth
Conformalized Outer Box
truck

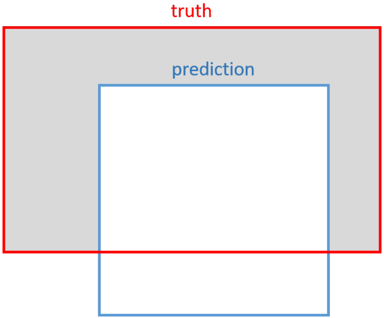

Predictions
Truth
Conformalized Outer Box
TOW - AWAY
NO STOPPING
3 P.M. TO 7 P.M.
traffic light
parking meter
420

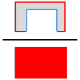
More flexible localization losses

Pixel-recall loss. Previously we penalized incorrect predictions equally. Now consider the fraction of a true box that is not covered by a predicted box [1].

After box matching, we write

- b_i for the i-th true box
- $\widehat{b}_i(\lambda)$ for the i-th predicted box to which we add a margin of λ



Pixel-recall loss: 

More formally:

$$L_i(\lambda) = 1 - \frac{\mathcal{A}(\widehat{b}_i(\lambda) \cap b_i)}{\mathcal{A}(b_i)}$$

The Conformal Risk Control (CRC) algorithm

The CRC method of [2] is an extension of split conformal prediction to bounded losses $L_i(\lambda) \in [0,1]$ that are non-increasing in λ.

Let $\alpha > 1/(n_{\mathrm{cal}}+1)$ be a target risk. The CRC method anticipates a worst-case value $L_{n+1}(\lambda) = 1$ for the loss on the $(n+1)$-th point, and picks

$$\widehat{\lambda}_\alpha = \inf \left\{ \lambda \in \Lambda : \frac{1}{n_{\mathrm{cal}}+1} \sum_{i=1}^{n_{\mathrm{cal}}} L_i(\lambda) + \frac{1}{n_{\mathrm{cal}}+1} \leqslant \alpha \right\} .$$

Theorem (Angelopoulos et al. 2024 [2])

Assume $(X_1, Y_1), \ldots, (X_n, Y_n), (X_{n+1}, Y_{n+1})$ are exchangeable (e.g., i.i.d.). Then, under mild assumptions on Λ and the $L_i(\lambda)$, for any $\alpha > 1/(n_{cal}+1)$, the loss on the new pair (X_{n+1}, Y_{n+1}) satisfies

$$\mathbb{E}\left[L_{n+1}\big(\widehat{\lambda}_\alpha\big) \right] \leqslant \alpha .$$

Calibration at image level

So far we have compared predicted boxes with true boxes. Two limitations:

- false negatives (undetected objects) are ignored
- boxes within the same images might be correlated

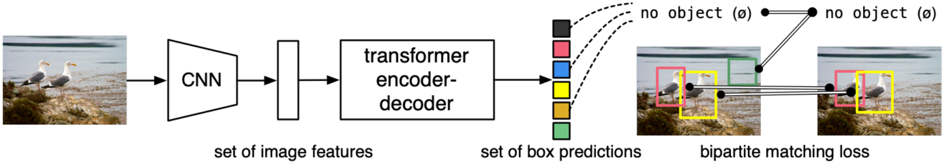
$\rightsquigarrow$ Statistical guarantees can have limited use or be (sometimes) incorrect.

Calibration at image-level. An alternative approach is to consider images as calibration points, and to work with losses defined on images.

For instance, for the each image i with n_i true boxes (see details in [1]):

$$L_i(\lambda) = 1 - \frac{1}{n_i} \sum_{h=1}^{n_i} \frac{\mathcal{A}(\widehat{b}_h \cap b_h)}{\mathcal{A}(b_h)}$$

Example of model: DETR-ResNet-50, a transformer for object detection with a ResNet-50 backbone [4].



For a given image x, the model f predicts N bounding boxes with softmax scores and confidence scores, i.e., $f(x) = (f_1(x), \ldots, f_N(x)) \in (\mathcal{B} \times \mathcal{C} \times [0,1])^N$, where

- $\mathcal{B} \subset \mathbb{R}^4$ is the set of all admissible bounding boxes (rectangles encoded by the top-left and bottom-right coordinates);
- $\mathcal{C}$ is the set of probability vectors over $\{1, ..., 80\}$;
- confidence scores are $[0,1]$-valued.

Need to tune multiple parameters

In practice the uncertainty may not be summarized by a single parameter.
$\rightsquigarrow$ For localization and classification, at least two parameters are useful.

Recently we extended the previous setting to 3 parameters: $\lambda^{\text{CONF}}, \lambda^{\text{BB}}, \lambda^{\text{CLS}}$.

Inference step: for a given image,

- we obtain N objects from our object detector (bounding boxes with softmax & confidence scores);
- we keep all objects with a confidence score above λ^{CONF};
- we adjust the remaining bounding boxes with an additive margin λ^{BB};
- we label each box with all classes whose softmax score is above λ^{CLS}.

The SeqCRC algorithm (forthcoming paper)

Calibration step: using our calibration data,

- we first tune the confidence threshold λ^{CONF};
- we then tune the additive margin λ^{BB} and the softmax threshold λ^{CLS}.

The resulting "SeqCRC algorithm" will be published soon! (joint work with Léo Andéol, Luca Mossina, and Adrien Mazoyer).

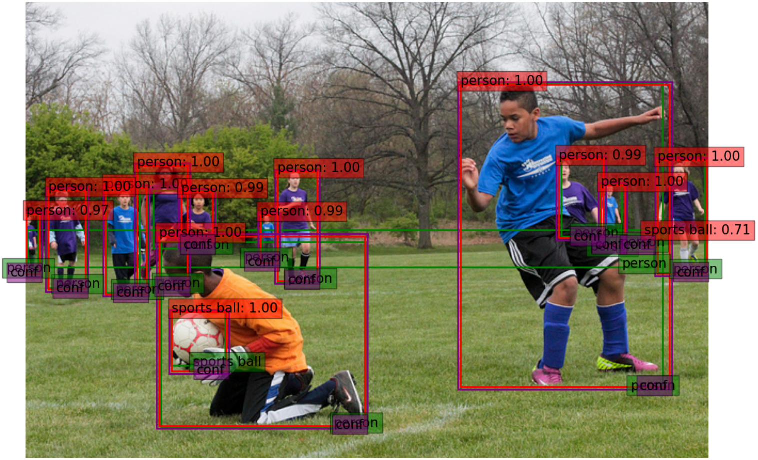
Experiments

We run SeqCRC on COCO, by sequentially tuning $\lambda = (\lambda^{\text{CONF}}, \lambda^{\text{BB}}, \lambda^{\text{CLS}})$.

We choose α^{BB} and α^{CLS} such that $\alpha^{\text{BB}} + \alpha^{\text{CLS}} = 0.1$ (Bonferroni-type correction) and obtain, on the test data:

- a global miscoverage of 0.12 (average value of $\max\{\text{loc loss}, \text{classif loss}\}$)
- an average bounding box "size" of 73.10 pixels;
- an average number of 1.27 predicted classes.

Successful detections.



Examples

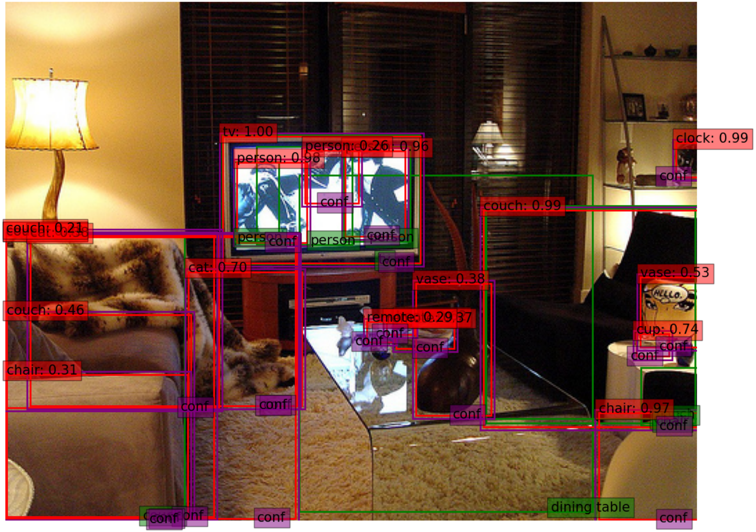
Issues:

- overly confident wrong detection of laptop;
- pizza is detected multiple times.

Typical issues for object detectors. Not penalized by our recall loss (we penalize false negatives), but could be detrimental in some applications.

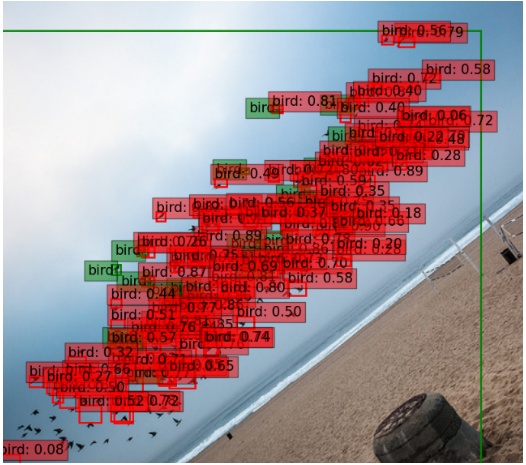
Non-Maximum Suppression could help.

Similar issues.



Examples

The birds are hard to detect individually.
Detection errors, annotation errors, or inappropriate loss functions?



Thank you!

Any questions?

[1] L. Andeol, T. Fel, F. de Grancey, and L. Mossina.
Confident object detection via conformal prediction and conformal risk control: an application to railway signaling.
In H. Papadopoulos, K. A. Nguyen, H. Boström, and L. Carlsson, editors, *Proceedings of the Twelfth Symposium on Conformal and Probabilistic Prediction with Applications*, volume 204 of *Proceedings of Machine Learning Research*, pages 36–55. PMLR, 13–15 Sep 2023.

[2] A. N. Angelopoulos, S. Bates, A. Fisch, L. Lei, and T. Schuster.
Conformal risk control.
In *The Twelfth International Conference on Learning Representations*, 2024.

[3] T. Azevedo, R. de Jong, and P. Maji.
Stochastic-YOLO: Efficient probabilistic object detection under dataset shifts.
In *ML4AD workshop, NeurIPS'20*, 2020.

[4] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko.
End-to-end object detection with transformers.
In A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, editors, *Computer Vision – ECCV 2020*, pages 213–229, Cham, 2020. Springer International Publishing.

[5] C.-H. Cheng, T. Schuster, and S. Burton.
Logically sound arguments for the effectiveness of ml safety measures.
In M. Trapp, E. Schoitsch, J. Guiochet, and F. Bitsch, editors, *Computer Safety, Reliability, and Security. SAFECOMP 2022 Workshops*, pages 343–350, Cham, 2022. Springer International Publishing.

[6] F. de Grancey, J.-L. Adam, L. Alecu, S. Gerchinovitz, F. Mamalet, and D. Vigouroux.
Object detection with probabilistic guarantees: A conformal prediction approach.
In M. Trapp, E. Schoitsch, J. Guiochet, and F. Bitsch, editors, *Computer Safety, Reliability, and Security. SAFECOMP 2022 Workshops*, pages 316–329. Springer International Publishing, 2022.

[7] A. Harakeh, M. Smart, and S. L. Waslander.
BayesOD: A Bayesian approach for uncertainty estimation in deep object detectors.
In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 87–93, 2020.

[8] M. Kläs and L. Jöckel.
A framework for building uncertainty wrappers for ai/ml-based data-driven components.
In A. Casimiro, F. Ortmeier, E. Schoitsch, F. Bitsch, and P. Ferreira, editors, *Computer Safety, Reliability, and Security. SAFECOMP 2020 Workshops*, pages 315–327, Cham, 2020. Springer International Publishing.

References III

[9] M. Kläs and L. Sembach.
Uncertainty wrappers for data-driven models.
In A. Romanovsky, E. Troubitsyna, I. Gashi, E. Schoitsch, and F. Bitsch, editors, *Computer Safety, Reliability, and Security*, pages 358–364, Cham, 2019. Springer International Publishing.

[10] F. Kraus and K. Dietmayer.
Uncertainty estimation in one-stage object detection.
In *2019 IEEE Intelligent Transportation Systems Conference (ITSC'19)*, 2019.

[11] H. W. Kuhn.
The hungarian method for the assignment problem.
Naval Research Logistics Quarterly, 2(1-2), 1955.

[12] B. Lakshminarayanan, A. Pritzel, and C. Blundell.
Simple and scalable predictive uncertainty estimation using deep ensembles.
In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS'17)*, 2017.

[13] M. T. Le, F. Diehl, T. Brunner, and A. Knol.
Uncertainty estimation for deep neural object detectors in safety-critical applications.
In *2018 21st International Conference on Intelligent Transportation Systems (ITSC'18)*, 2018.

[14] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick.
Microsoft coco: Common objects in context.
In D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, editors, *Computer Vision – ECCV 2014*, pages 740–755, Cham, 2014. Springer International Publishing.

[15] Z. Lyu, N. Gutierrez, A. Rajguru, and W. J. Beksi.
Probabilistic object detection via deep ensembles.
In *European Conference on Computer Vision (ECCV'20)*, 2020.

[16] D. Miller, F. Dayoub, M. Milford, and N. Sunderhauf.
Evaluating merging strategies for sampling-based uncertainty techniques in object detection.
In *2019 International Conference on Robotics and Automation (ICRA'19)*, 2019.

[17] D. Miller, L. Nicholson, F. Dayoub, and N. Sünderhauf.
Dropout sampling for robust object detection in open-set conditions.
In *2018 International Conference on Robotics and Automation (ICRA'18)*, 2018.

[18] T. Schuster, E. Seferis, S. Burton, and C.-H. Cheng.
Formally compensating performance limitations for imprecise 2d object detection.
In M. Trapp, F. Saglietti, M. Spisländer, and F. Bitsch, editors, *Computer Safety, Reliability, and Security*, pages 269–283, Cham, 2022. Springer International Publishing.

References V

[19] S. H. Yelleni, D. Kumari, S. P.K., and K. M. C.
Monte Carlo DropBlock for modeling uncertainty in object detection.
Pattern Recognition, 146:110003, 2024.