Copula-based conformal prediction for object detection: a more efficient approach

Bruce Cyusa Mukama BRUCE.CYUSA-MUKAMA@HDS.UTC.FR
HEUDIASYC - UMR CNRS 7253, Université de Technologie de Compiègne, 57 avenue de Landshut, 60203 Compiègne Cedex - France

Soundouss Messoudi SOUNDOUSS.MESSOUDI@HDS.UTC.FR
HEUDIASYC - UMR CNRS 7253, Université de Technologie de Compiègne, 57 avenue de Landshut, 60203 Compiègne Cedex - France

Sylvain Rousseau SYLVAIN.ROUSSEAU@HDS.UTC.FR
HEUDIASYC - UMR CNRS 7253, Université de Technologie de Compiègne, 57 avenue de Landshut, 60203 Compiègne Cedex - France

Sébastien Destercke SEBASTIEN.DESTERCKE@HDS.UTC.FR
HEUDIASYC - UMR CNRS 7253, Université de Technologie de Compiègne, 57 avenue de Landshut, 60203 Compiègne Cedex - France

Editor: Simone Vantini, Matteo Fontana, Aldo Solari, Henrik Boström and Lars Carlsson

Abstract

Object detection is an important vision task, and providing statistical guarantees around such detections can be of critical importance. So far, most conformal bounding box regression approaches do not simultaneously account for heteroscedasticity and dependencies between the residuals of each dimension. In this paper, we examine the importance of such dependencies and heteroscedasticity in the context of multi-target conformal regression, we apply copula-based conformal prediction methods to model them and to improve the volume of bounding box prediction regions. We compare these methods to the state-of-the-art conformal object detection approaches (on the KITTI & the BDD100K autonomous driving benchmarks) and the empirical copula-based method shows high-efficiency results that are robust w.r.t. heteroscedasticity and also robust w.r.t. the structure of the dependencies.

Keywords: Uncertainty quantification, object detection, conformal prediction, copulas.

1. Introduction

In recent years, quantifying the uncertainty of predictions made by machine learning algorithms has become increasingly important as these algorithms are deployed in safety-critical applications (Salim and Jayasudha, 2023). In vision-based autonomous driving, object detection algorithms (Gupta et al., 2021) jointly predict the categories and the locations of the objects in a road scene by processing an image or a video frame. Consequently, to ensure that the system is safe, guarantees are not only needed for the classification task but also for the bounding box regression task. In 2D object detection, uncertainty quantification (Hüllermeier and Waegeman, 2021; Abdar et al., 2021) has been done using Bayesian techniques (Harakeh et al., 2020; Sheikh and Shah, 2005), gradient-based techniques (Riedlinger et al., 2023), ensemble techniques (Lyu et al., 2020; Miller et al., 2019), other post-hoc calibration methods (Kuppers et al., 2020; Oksuz et al., 2023) and split conformal prediction (Andéol et al., 2023; De Grancey et al., 2022).

Split conformal prediction (Papadopoulos et al., 2002) provides probabilistic guarantees with a reasonable computational cost and with very few assumptions: the data is only assumed to be exchangeable (Lei et al., 2018).

However, bounding box prediction can be seen as a multi-target regression problem because it takes 4 scalars to represent a box in an image (i.e., the spatial coordinates of the top left corner and the bottom right corner of the rectangle), but the split conformal prediction (SCP) methods that have been proposed by De Grancey et al. (2022) infer conformal predictions without explicitly accounting for dependencies that can exist between the elements of the multivariate regression error, and the method in Andéol et al. (2023) does not account for heteroscedasticity[1] nor dependencies[2]. These limitations are significant because such dependencies can be leveraged to improve the predictions and because the residuals are not guaranteed to be homoscedastic[1]. In fact, Messoudi et al. (2021) proposed a copula-based multivariate conformal regression method that models the dependencies without assuming homoscedasticity and Zhang et al. (2023) improved it by leveraging those dependencies to minimize the volumes of the prediction regions.

In this paper, we examine the impact of dependencies and heteroscedasticity in the context of multi-target conformal regression, we apply copula-based conformal prediction methods to bounding box regression and we compare these approaches to other conformal object detection methods on KITTI (Geiger et al., 2012) & BDD100K (Yu et al., 2020).

The rest of this paper is organized as follows: In section 2, we provide an overview of the previous works that are related to our contributions. In section 3, we formally present our approach to conformal bounding box prediction and the preexisting approaches. In section 4 we present our experiments and discuss the results. Finally, we highlight the key points of this study in section 5.

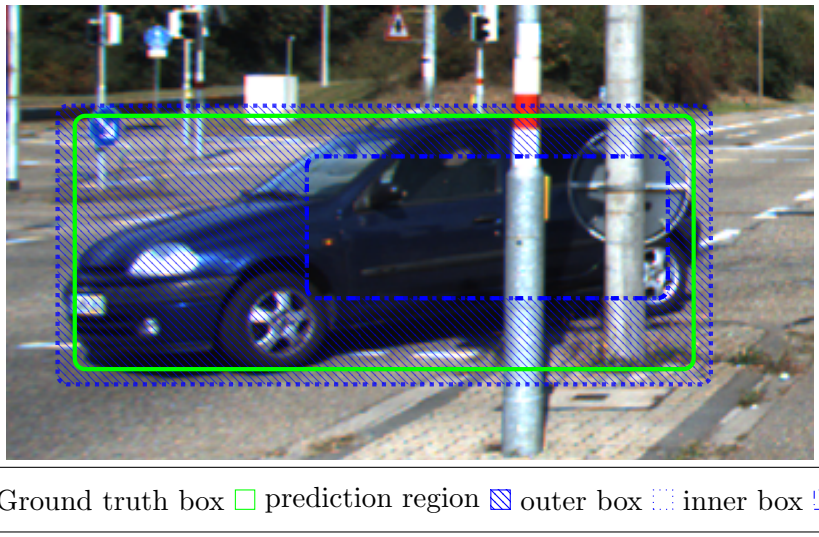

Figure 1: An illustration of a valid bounding box prediction region.

1. In this paper, "heteroscedasticity" will simply stand for heteroscedasticity between residuals of each dimension, and not heteroscedasticity within each dimension, as this is out of the scope of this paper.
2. Similarly, "dependencies" will simply stand for probabilistic dependencies between the dimensions of the regression error.

2. Related Works

In this section, we summarize the previous works that are related to our copula-based SCP approach to bounding box prediction and establish links and distinctions between them.

Object detection (Felzenszwalb et al., 2009) is a multi-task learning problem composed of an object classification task and an object localization task. Object classification predicts the type of an object in an image and object localization predicts the tightest rectangle that encloses all its pixels. In general, a single (end-to-end) neural network is trained for object detection with a loss function that is composed of two parts: a cross-entropy term (for classification) and a penalty for the Jaccard index (the IoU: intersection over union) of the predicted bounding box and the ground truth bounding box. Zhao et al. (2019), Chen et al. (2024), and Su et al. (2024) provide good literature reviews for the various loss functions and neural network architectures that are used in object detection. Although these methods provide a confidence score (Wenkel et al., 2021) for each predicted detection, this score cannot be used for uncertainty quantification (as is) because it is not well calibrated by default and because it is holistic w.r.t. the predicted object's class and its predicted location.

Conformal prediction (Vovk et al., 2022) is an uncertainty quantification (UQ) framework with multiple advantages: it is post-hoc, model agnostic, distribution-free, computationally efficient, non-asymptotic and rigorously calibrated (i.e., with theoretical guarantees). However, most conformal regression methods cannot be directly applied to bounding box regression because it is a multivariate (multi-target) regression problem whilst research in conformal prediction has mainly focused on single-target regression. Thus, De Grancey et al. (2022) used the Bonferroni correction (Bland and Altman, 1995) to compensate for the likelihood of errors that emerge in the multiple comparison problem induced by the 4 dimensions. Later, Andéol et al. (2023) transformed the multidimensional residuals into scalars and applied single-target conformal regression to bounding box prediction. As shown in sections 3.3.1 & 3.3.2, both methods do not explicitly account for the dependencies that can exist between the elements of multivariate dissimilarities and the efficiency of the method in Andéol et al. (2023) depends on the homogeneity of variance (heteroscedasticity).

These are significant limitations because error dependencies can be leveraged to improve the efficiency of multivariate predictions, which is a major objective in conformal prediction. To explicitly account for dependencies, Messoudi et al. (2021) proposed a generalized conformal regression method that applies to multi-target problems, by modeling the dependencies with copulas (Nelsen, 2006; Genest and Favre, 2007), and Zhang et al. (2023) improved that method by minimizing the volumes of the prediction regions that it produces. To the best of our knowledge, copula-based conformal prediction approaches have not been applied to bounding box regression nor compared to other conformal object detection methods.

3. Methodology

In this section, we formally define object detection, box-wise conformal prediction, and the various approaches that are compared in our experiments.

3.1. Object detection

In this study, we only investigate single-class object detection. Given an input image X_i and an object detector f_θ, we denote by $\widehat{Y}_i = f_\theta(X_i)$ the predicted objects and by Y_i the ground truth objects. The j-th object in the i-th image is represented by a rectangle (i.e., its bounding box) with the coordinates of the top left corner $(\underline{x}_{i,j}; \underline{y}_{i,j})$ and the coordinates of the bottom right corner $(\overline{x}_{i,j}; \overline{y}_{i,j})$. Thus, we denote $Y_i = \{B_{i,j}\}_{j=1}^{N_i}$ with $B_{i,j} = \{\underline{x}_{i,j}; \underline{y}_{i,j}; \overline{x}_{i,j}; \overline{y}_{i,j}\}$ and N_i as the number of objects in the i-th image.

In practice, the underlying deep neural network predicts a fixed number of objects that is significantly greater than the number of objects in any image. A confidence score ρ is assigned to each predicted object and a non-maximum suppression algorithm (NMS) is used to remove duplicates and to keep the predictions with high confidence scores. Therefore, the sets Y_i and $\widehat{Y}_i$ are not presented in the same order. The Jaccard index (IoU: intersection over union) is used to match each predicted bounding box $\widehat{B}_{i,j'} = \{\widehat{\underline{x}}_{i,j'}; \widehat{\underline{y}}_{i,j'}; \widehat{\overline{x}}_{i,j'}; \widehat{\overline{y}}_{i,j'}; \rho_{i,j'}\}$ with the most likely ground truth $\widehat{B}_{i,j}$.

$$IoU(B_{i,j}, \widehat{B}_{i,j'}) = \frac{\text{area}(B_{i,j} \cap \widehat{B}_{i,j'})}{\text{area}(B_{i,j} \cup \widehat{B}_{i,j'})}$$

The performance of the detector can be assessed for different levels of thresholds with the mean average precision (mAP), i.e., the area under the precision-recall curve, by considering the predictions with $\rho_{i,j'} \geq \rho_{\text{th}}$ and $IoU(B_{i,j}, \widehat{B}_{i,j'}) \geq IoU_{\text{th}}$ as true-positives. To simplify the notations, we will consider $j' = j$ in the remainder of this paper.

Algorithm 1 Computing bounding box dissimilarity scores

Require: a detection threshold ρ_{th}, an overlap threshold IoU_{th},
a trained object detector f_θ, a calibration dataset D_{cal}.

for $X_i \in D_{\text{cal}}$ **do**
  Predict the bounding boxes: $\widehat{Y}_i = f_\theta(X_i)$,
  for $B_{i,j} \in Y_i, \widehat{B}_{i,j} \in \widehat{Y}_i$ **do**
    if $IoU(B_{i,j}, \widehat{B}_{i,j}) \geq IoU_{\text{th}}$ and $\rho_{i,j} \geq \rho_{\text{th}}$ **then**
      Pair $B_{i,j}$ with $\widehat{B}_{i,j}$
    end if
    $\alpha_{i,j} \leftarrow \{|\widehat{\underline{x}}_{i,j} - \underline{x}_{i,j}|, |\widehat{\underline{y}}_{i,j} - \underline{y}_{i,j}|, |\overline{x}_{i,j} - \widehat{\overline{x}}_{i,j}|, |\overline{y}_{i,j} - \widehat{\overline{y}}_{i,j}|\}$ `// dissimilarity scores`
  end for
end for

3.2. Box-wise conformal prediction

In this study, we only investigate box-wise split conformal prediction approaches: we infer 4-dimensional prediction regions (i.e., bounding box intervals). Box-wise SCP approaches follow the general training and calibration procedure in Algorithm 2. For each bounding box $B_{i,j}$, and given a user-specified (global) significance level $\epsilon^g \in [0, 1]$, we aim to infer a small (efficient) 4-dimensional interval $\mathcal{I}(\widehat{B}_{i,j})$ such that

$$P(B_{i,j} \in \mathcal{I}(\widehat{B}_{i,j})) \geq 1 - \epsilon^g.$$

An inferred interval $\mathcal{I}(\widehat{B}_{i,j})$ is valid if it contains the ground truth $B_{i,j}$ (as in Figure 1) and a UQ method is (marginally) well-calibrated if the proportion of valid inferences matches the desired level $1 - \epsilon^g$.

Thus, box-wise SCP algorithms are identical except for line 4 in Algorithm 2: they only differ by how they process the multidimensional dissimilarities $\alpha_{i,j}$ to determine conformal quantiles $\{\alpha_s^1, \ldots, \alpha_s^1\}$ (the key step for inferring the bounds of the 4-dimensional interval). At test time (or in deployment), only steps 5 & 6 are done to predict a new interval.

Algorithm 2 The generic calibration procedure for box-wise SCP

Require: a global significance level ϵ^g, an object detector f_θ, a dataset D

1: Split the dataset D in two subsets: D_{train} & $D_{\text{cal}} = \{(X_i, Y_i)\}_{i=1}^n$,
2: Fit or fine tune f_θ on D_{train},
3: Follow Algorithm 1 to compute bounding box dissimilarity scores $\{\alpha_{i,j}\}_{i=1}^n$,
4: Compute conformal quantiles $\{\alpha_s^1, \alpha_s^2, \alpha_s^3, \alpha_s^4\}$ from $\{\alpha_{i,j}\}_{i=1}^n$ and ϵ^g,
5: For any new predicted box $\widehat{B}_{n+1,j}$, infer an inner box $\underline{\widehat{B}}_{n+1,j}$ and an outer box $\overline{\widehat{B}}_{n+1,j}$:

$$\underline{\widehat{B}}_{i,j} = \{\underline{\widehat{x}}_{i,j} + \alpha_s^1,\ \underline{\widehat{y}}_{i,j} + \alpha_s^2,\ \overline{\widehat{x}}_{i,j} - \alpha_s^3,\ \overline{\widehat{y}}_{i,j} - \alpha_s^4\} \tag{1}$$

$$\overline{\widehat{B}}_{i,j} = \{\underline{\widehat{x}}_{i,j} - \alpha_s^1,\ \underline{\widehat{y}}_{i,j} - \alpha_s^2,\ \overline{\widehat{x}}_{i,j} + \alpha_s^3,\ \overline{\widehat{y}}_{i,j} + \alpha_s^4\} \tag{2}$$

6: Yield bounding box prediction regions $\mathcal{I}(\widehat{B}_{n+1,j}) \leftarrow [\underline{\widehat{B}}_{n+1,j}, \overline{\widehat{B}}_{n+1,j}]$

3.3. Compared approaches

In this section, we detail how each compared approach processes the multidimensional dissimilarities to determine the conformal quantiles, in order to infer the bounds of bounding box prediction intervals.

3.3.1. The Bonferroni correction

The Bonferroni approach (De Grancey et al., 2022) associates an adjusted significance level ϵ^t to each dimension of the bounding box before applying 4 separate single-target conformal prediction procedures: each conformal quantile α_s^t (associated to the t-th dimension) is computed with the significance level $\epsilon^t = \epsilon^1 = \cdots = \epsilon^4 = \frac{\epsilon^g}{4}$. This adjustment allows to construct multiple confidence intervals while still ensuring that the overall confidence level ϵ^g is maintained. In multiple hypothesis testing, this adjustment is often used to control the family-wise error rate (FWER).

However, this method does not finely nor explicitly model the dependencies because its ϵ^t values do not depend on the 4-dimensional distribution of the $\alpha_{i,j}$ and therefore $\epsilon^t = \frac{\epsilon^g}{4}$, no matter the structure of the dependencies, which is equivalent to making a conservative assumption about the dependencies. Nonetheless, this method accounts for heteroscedasticity because each conformal quantile α_s^t is computed separately.

3.3.2. The max-additive approach

The max-additive approach (Andéol et al., 2023) reduces the multidimensional dissimilarities into a one-dimensional sample by keeping the maxima of residuals.

$$\alpha_{i,j} = \max(|\widehat{\underline{x}}_{i,j} - \underline{x}_{i,j}|, |\widehat{\underline{y}}_{i,j} - \underline{y}_{i,j}|, |\overline{x}_{i,j} - \widehat{\overline{x}}_{i,j}|, |\overline{y}_{i,j} - \widehat{\overline{y}}_{i,j}|). \tag{3}$$

The bounds of $\mathcal{I}(\widehat{B}_{i,j})$ are inferred using equations 1 & 2, with the same conformal quantile for every dimension, i.e., $\alpha_s^1 = \cdots = \alpha_s^4 =$ the $((1-\epsilon^g)(n+1)/n)$-percentile of the dissimilarity scores. Therefore, this method does not explicitely account for any probabilistic dependencies in the dissimilarities. Also, this method assumes the residuals to be homoscedastic, or at least of comparable magnitude: if there is a dimension whose dissimilarity values are always significantly larger than the values on all the other dimensions, the *max* operation will discard the dimensions with small dissimilarity values and thereby infer prediction intervals that are larger than necessary (inefficient).

3.3.3. The copula-based approach

This study applies copula-based conformal prediction (CCP) to object detection for the first time. The copula-based conformal prediction approach (Messoudi et al., 2021) that we devote to object detection finely and explicitly represents the dependencies with copulas.

Copulas $C : [0,1]^t \to [0,1]$ are multivariate cumulative distribution functions with marginals that are uniform on the $[0,1]$ interval. Among other things (Genest and Favre, 2007), they are used to model the dependence structure between the marginals of a joint distribution function. Given two random variables $V^1 \sim F^1$, $V^2 \sim F^2$, their joint cumulative distribution function $(V^1, V^2) \sim F$ and the probability integral transforms $U^1 = F^1(V^1)$ & $U^2 = F^2(V^2)$, we can express F with its copula C as follows:

$$\begin{aligned} F(v^1, v^2) &= P(V^1 \leq v^1,\ V^2 \leq v^2) \\ &= P(F^1(V^1) \leq F^1(v^1),\ F^2(V^2) \leq F^2(v^2)) \\ &= P(U^1 \leq u^1,\ U^2 \leq u^2) \\ &= C(u^1,\ u^2) \end{aligned} \tag{4}$$

The above relations hold for any number of random variables (under mild conditions) and the existence of a copula C is guaranteed by Sklar's theorem (Sklar, 1959). In this study we use the product copula (C_π) to only model independence and we use the Gumbel copula (C_G) and the empirical copula (C_E) to flexibly model independence or interdependency:

$$C_\pi(u^1, \ldots, u^m) = \prod_{t=1}^{m} u^t, \tag{5}$$

$$C_G(u^1, \ldots, u^m) = \exp\left(\sum_{t=1}^{m} (-\ln u^t)^\theta\right)^{\frac{1}{\theta}}, \tag{6}$$

$$C_E(u^1, \ldots, u^m) = \frac{1}{n}\sum_{i=1}^{n}\prod_{t=1}^{m} \mathbb{1}_{u_i^t \leq u^t}, \tag{7}$$

where θ is an estimated parameter and (the indicator function) $\mathbb{1}_{u_i^t \leq u^t}$ takes the value 1 if the inequality is verified and 0 other wise. Messoudi et al. (2021) showed that each significance level ϵ^t in a conformal prediction setting with t outputs can be determined by inverting the copula or by search. From the definition of validity and calibration in conformal prediction, we get that:

$$P(B_{i,j} \in \mathcal{I}(\widehat{B}_{i,j})) = P(|\underline{x}_{i,j} - \widehat{\underline{x}}_{i,j}| \leq \alpha_s^1, \ldots, |\bar{y}_{i,j} - \widehat{\bar{y}}_{i,j}| \leq \alpha_s^4) \geq 1 - \epsilon^g, \tag{8}$$

where for each dimension t, α_s^t denotes the conformal quantile associated with the significance level ϵ^t. By assuming that the method is well-calibrated, we get that:

$$F^g(\alpha_s^1, \ldots, \alpha_s^4) = C(F^1(\alpha_s^1), \ldots, F^4(\alpha_s^4)) = C(1 - \epsilon^1, \ldots, 1 - \epsilon^4) = 1 - \epsilon^g, \tag{9}$$

$$\alpha_s^t = Q^t((1 - \epsilon^t) \times (n + 1)/n) \tag{10}$$

where F^g denotes the joint cumulative distribution, F^t denotes the marginal of the t-th dimension and Q^t denotes its inverse (i.e., the quantile function). There can be many solutions for the quadruplet $(\epsilon^1, \ldots, \epsilon^4)$. In such cases, the efficiency of the method depends on the chosen quadruplet because the bounds depend on the associated conformal quantiles as shown by equations 1 & 2. Instead of simply solving equation 9 for $\epsilon^1 = \cdots = \epsilon^4$ as in Messoudi et al. (2021), the efficiency can be further increased by minimizing the volume of the inferred prediction regions under these specific (copula) constraints:

$$\underset{\epsilon^1, \ldots, \epsilon^4}{\arg\min} \prod_{t=1}^{4} (2 \times \alpha_s^t) \quad \text{s.t.} \quad \begin{cases} C(1 - \epsilon^1, \ldots, 1 - \epsilon^4) \geq 1 - \epsilon^g \\ \epsilon^t \in (0, \epsilon^g] \end{cases} \tag{11}$$

Zhang et al. (2023) solved this optimization problem with a differential evolution algorithm (DE), which is a meta-heuristic optimization technique (Storn and Price, 1997). With this solution, we will use the independent copula to model independence, the (semi-parametric) Gumbel copula to flexibly model correlations, and the (non-parametric) empirical copula to estimate the dependency structure directly from the observations $\alpha_{i,j}$.

Although Zhang et al. (2023) and Messoudi et al. (2021) did not provide formal proofs for the validity guarantees, Sun and Yu (2023) provided a formal proof that relies on the empirical copula (in an application of SCP to multi-step time series forecasting).

4. Experiments

4.1. Experimental setup

In this section, we present the datasets, the protocol, and the evaluation metrics that were used to compare the different box-wise conformal prediction approaches.

The datasets

We use two real autonomous driving datasets (KITTI & BDD100K) as well as two synthetic datasets to compare the different box-wise SCP approaches.

KITTI (Geiger et al., 2012) is a vision benchmark that was produced using a real autonomous driving platform[3] in 2012. To this day, it is still relevant and we use it in this

3. See https://www.cvlibs.net/datasets/kitti/setup.php

study to detect the class with the most instances: car objects. We only use the publicly accessible[4] 2D object detection data, that is, a total of 7841 images and 28742 car instances.

BDD100K (Yu et al., 2020) is a larger autonomous driving dataset that is diverse w.r.t. appearance variation(s). Collected using multiple vehicles, it was published in 2020 and recently used in multi-object tracking (MOT) challenges[5] at the CVPR 2023 Workshop on Autonomous Driving (WAD). In this study, we use it for car bounding box prediction and to emulate settings with abundant data, as opposed to the significantly smaller KITTI dataset. Similarly, we only use the part[6] with publicly accessible annotations, that is, a total of 80K images and 815717 car instances.

The synthetic datasets were used to examine the importance of dependencies and the impact of heteroscedasticity. For the first one, we generated independent residuals from a uniform distribution ($\alpha_{i,j} \sim \mathcal{U}(\Omega)$ with $\Omega = [0, 2.8] \times [0, 2.5] \times [0, 8] \times [0, 2.5]$), and for the second one, we generated correlated residuals from a multivariate normal distribution ($\alpha_{i,j} \sim \mathcal{N}(0, \Sigma)$ with $\Sigma_{11} = 2$, $\Sigma_{22} = 1.5$, $\Sigma_{33} = 15$, $\Sigma_{44} = 1.5$ and $\Sigma_{kl} = 0.8$ for $k \neq l$). We generated about 4000 data points for calibration and 3000 for testing. Although these sizes are arbitrary (and rounded), they are reported because empirical methods are sensitive to calibration sample sizes as shown by Messoudi et al. (2021).

The experimental procedure

As stated above, we focus on predicting cars' bounding boxes. The DE algorithm was initialized with a population of 20 candidates and the number of iterations was set to 200.

To assess the calibration, on each dataset, 12 values of the global significance level ϵ^g were tried (from 0.01 to 0.9) and the variability of our results was assessed in a k-fold cross-validation setting (10-fold for KITTI and 5-fold for BDD100K). For the real datasets, on each cross-validation iteration, we used a different subset for testing (10% of the total) and the training set (i.e., the rest) was further split in two parts: one set for calibration (10% of the total) and one set for fine-tuning (80% of the total). No fine-tuning was done for the synthetic data, the procedure (generation, calibration & test) was repeated 10 times.

For the real datasets, we fine-tuned the YOLO v8 object detector (Jocher et al., 2023) (with the above split ratios) for 100 epochs on KITTI and for 20 epochs on BDD100K. We filtered its predictions with a confidence threshold $\rho_{\text{th}} = 0.3$, and we paired the predicted boxes with their ground truth boxes with a threshold $IoU_{\text{th}} = 0.3$. The code[7] is publicly available and the implementation uses open-source libraries: PyTorch (Paszke et al., 2019), Copulae (Bok et al., 2024) and SciPy (Virtanen et al., 2020).

Evaluation metrics

We evaluate the performance of box-wise SCP approaches with two metrics: we use the empirical coverage (i.e., the percentage of valid inferences) to assess the calibration and we use the volume of the inferred bounding box intervals to assess the efficiency of the compared box-wise SCP approaches. In the below equations, $\mathbb{1}_{B_{i,j} \in \mathcal{I}(\widehat{B}_{i,j})}$ has value 1 if

4. See https://www.cvlibs.net/datasets/kitti/eval_object.php?obj_benchmark=2d
5. See https://www.vis.xyz/bdd100k/challenges/cvpr2023/
6. See https://doc.bdd100k.com/download.html#k-images
7. See https://gitlab.utc.fr/robust-event-detection/copula-based-conformal-bounding-box-prediction

$B_{i,j} \in \mathcal{I}(\widehat{B}_{i,j})$, and 0 otherwise; N denotes the number of images that were used to test the box-wise SCP approach, N_i denotes the number of objects in the j-th image, and $V_{i,j}$ denotes the volume of the hyper-rectangle $\mathcal{I}(\widehat{B}_{i,j})$. Ideally, the volume should be tiny and the coverage should be equal to the desired coverage level $1 - \epsilon^g$.

$$\text{Coverage} = \frac{1}{N} \times \sum_{i=1}^{N} \left(\frac{1}{N_i} \times \sum_{j=1}^{N_i} \mathbb{1}_{B_{i,j} \in \mathcal{I}(\widehat{B}_{i,j})} \right) \tag{12}$$

$$\text{Average volume} = \frac{1}{N} \times \sum_{i=1}^{N} \left(\frac{1}{N_i} \times \sum_{j=1}^{N_i} V_{i,j} \right) = \prod_{t=1}^{4} \left(2 \times \alpha_s^t \right) \tag{13}$$

4.2. Results

In this section, we report the results of our experiments. We designate by "Bonferroni" the method that is described in section 3.3.1 and we designate by "Max additive" the method that is described in section 3.3.2. For the methods in section 3.3.3, "Independent DE-CCP" is based on the independent copula, "Gumbel DE-CCP" is based on the (semi-parametric) Gumbel copula, and "Empirical DE-CCP" for the (non-parametric) empirical copula. We use graphics to report the general trends and tables to provide finer details for high values of desired coverage levels $(1 - \epsilon^g)$.

4.2.1. Validity

As stated in section 4.1, we used multiple folds and different levels of ϵ^g to assess the calibration of the compared SCP approaches:

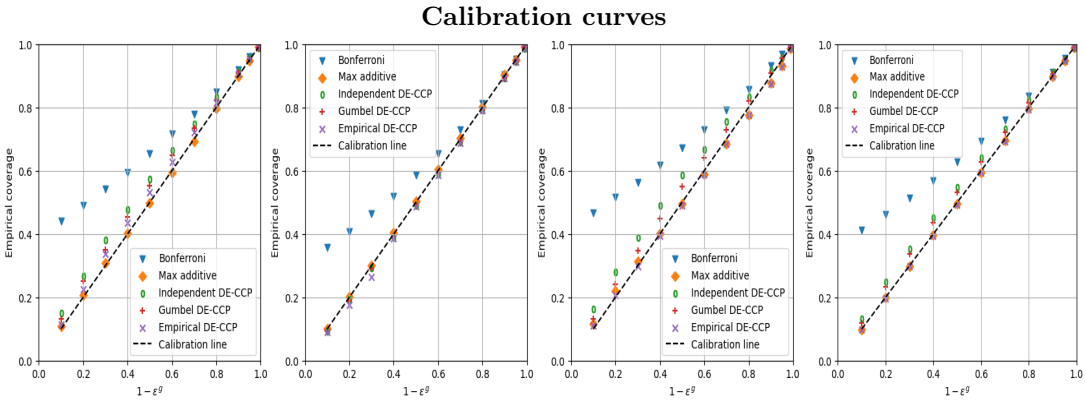

Figure 2: Calibration curves (from left to right) for correlated residuals, for independent residuals, and for YOLO v8's residuals on KITTI & on BDD100K.

As shown in Figure 2, in the case of correlated dissimilarities (on the first subplot, from left to right), we can sort the compared approaches from the closest to the (perfect) calibration line to the furthest as follows: "Max additive", "Empirical DE-CCP", "Gumbel DE-CCP", "Independent DE-CCP" and "Bonferroni". Furthermore, as detailed in Table 1, the results are close but the empirical coverage of the "Max additive" approach is bit lower

than the desired levels. For theses reasons, the results of the "Empirical DE-CCP" approach are put in bold.

In the case of independent residuals (on the second subplot), we can sort the approaches from the closest to the calibration line to the furthest as follows: "Max additive" approach, the copula-based approaches and the "Bonferroni" approach. As shown in Table 2, the validity results of the copula-based methods are very close.

$1-\epsilon^g$	Bonferroni	Max additive	Independent DE-CCP	Gumbel DE-CCP	Empirical DE-CCP
0.99	0.992±0.002	0.991±0.002	**0.990±0.002**	**0.990±0.002**	**0.990±0.002**
0.95	0.960±0.006	0.947±0.007	0.955±0.006	0.953±0.007	**0.952±0.007**
0.90	0.920±0.009	0.898±0.008	0.913±0.009	0.908±0.009	**0.906±0.009**
0.80	0.849±0.011	0.796±0.013	0.834±0.012	0.822±0.012	**0.815±0.012**

Table 1: Empirical coverage (mean and std.) for the correlated residuals dataset.

$1-\epsilon^g$	Bonferroni	Max additive	Independent DE-CCP	Gumbel DE-CCP	Empirical DE-CCP
0.99	**0.990±0.002**	0.991±0.002	0.987±0.002	0.988±0.003	0.987±0.003
0.95	0.952±0.007	**0.951±0.004**	0.945±0.006	0.944±0.004	0.945±0.006
0.90	0.902±0.008	**0.902±0.006**	0.892±0.007	0.892±0.008	0.894±0.009
0.80	0.812±0.011	**0.802±0.009**	0.792±0.010	0.792±0.012	0.791±0.011

Table 2: Empirical coverage (mean and std.) for the independent residuals dataset.

$1-\epsilon^g$	Bonferroni	Max additive	Independent DE-CCP	Gumbel DE-CCP	Empirical DE-CCP
0.99	**0.992±0.005**	0.987±0.006	0.989±0.006	0.989±0.006	0.985±0.006
0.95	0.967±0.010	0.932±0.012	0.958±0.012	**0.956±0.013**	0.930±0.011
0.90	0.932±0.016	0.876±0.012	0.915±0.015	**0.909±0.015**	0.875±0.013
0.80	0.857±0.019	0.775±0.012	0.833±0.016	**0.820±0.017**	0.775±0.011

Table 3: Empirical coverage (mean and std.) for YOLO v8's residuals on KITTI.

In the case of YOLO v8's residuals on the KITTI benchmark (on the third subplot), the "Max additive" approach and the "Empirical DE-CCP" approach are the closest to the calibration line. However, as detailed in Table 3, the calibration results of "Max additive" and "Empirical DE-CCP" get slightly lower than the desired calibration levels when $1-\epsilon^g$ gets very high. Nonetheless, the results of the "Gumbel DE-CCP" approach remain higher but close to the (perfect) calibration line: among the methods that uphold the guarantees that are expressed in Equation 8, this method is the closest to the calibration line.

$1-\epsilon^g$	Bonferroni	Max additive	Independent DE-CCP	Gumbel DE-CCP	Empirical DE-CCP
0.99	**0.990±0.000**	0.989±0.000	0.989±0.000	0.989±0.000	0.989±0.000
0.95	0.954±0.001	0.948±0.001	0.953±0.001	**0.951±0.001**	0.948±0.001
0.90	0.912±0.001	0.898±0.001	0.911±0.001	**0.907±0.001**	0.897±0.001
0.80	0.835±0.001	0.796±0.001	0.823±0.001	**0.816±0.001**	0.795±0.002

Table 4: Empirical coverage (mean and std.) for YOLO v8's residuals on BDD100K.

In the case of YOLO v8's residuals on the BDD100K benchmark (the fourth subplot), we observe the same ranking as on KITTI. However, all the calibration curves are closer to the perfect calibration line. As detailed by Table 4, the "Max additive" approach and the "Empirical DE-CCP" approach are very slightly off and the offsets are significantly smaller than the offsets that we observe on KITTI (Table 3).

4.2.2. Efficiency

In this section we report the average size (volumes) of the inferred bounding box prediction regions for each of the compared approaches, as defined by Equation 12.

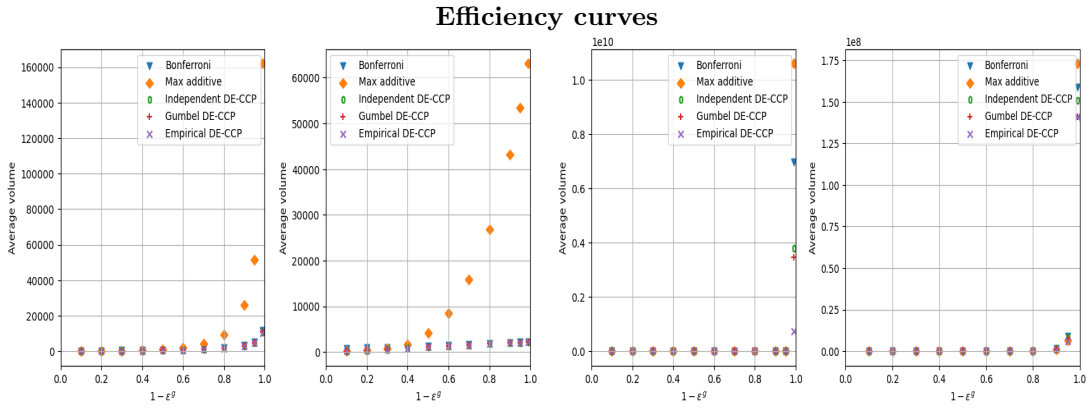

Figure 3: Efficiency curves (from left to right) for correlated residuals, for independent residuals, and for YOLO v8's residuals on KITTI & on BDD100K.

$1-\epsilon^g$	Bonferroni	Max additive	Independent DE-CCP	Gumbel DE-CCP	Empirical DE-CCP
0.99	1.15e+04±8.89e+02	1.62e+05±8.09e+03	1.09e+04±7.70e+02	1.06e+04±6.77e+02	**1.05e+04±6.73e+02**
0.95	5.16e+03±1.88e+02	5.16e+04±2.93e+03	4.95e+03±1.91e+02	4.79e+03±1.68e+02	**4.72e+03±1.69e+02**
0.90	3.29e+03±1.70e+02	2.59e+04±1.66e+03	3.14e+03±1.43e+02	3.01e+03±1.35e+02	**2.96e+03±1.38e+02**
0.80	1.92e+03±8.06e+01	9.48e+03±6.74e+02	1.76e+03±6.94e+01	1.67e+03±6.08e+01	**1.59e+03±5.80e+01**

Table 5: Empirical efficiency (mean and std.) for the correlated residuals dataset.

As shown on Figure 3, in the case of correlated residuals (on the first subplot, from left to right), all the efficiency curves are close but the curve of the "Max additive" approach

$1-\epsilon^g$	Bonferroni	Max additive	Independent DE-CCP	Gumbel DE-CCP	Empirical DE-CCP
0.99	2.22e+03±3.86e+00	6.31e+04±3.43e+02	2.21e+03±4.39e+00	2.21e+03±4.07e+00	**2.21e+03±3.45e+00**
0.95	2.13e+03±1.12e+01	5.34e+04±5.71e+02	2.12e+03±7.99e+00	**2.12e+03±7.66e+00**	2.12e+03±9.99e+00
0.90	2.02e+03±7.93e+00	4.32e+04±4.71e+02	2.00e+03±1.07e+01	**2.00e+03±1.03e+01**	2.00e+03±1.04e+01
0.80	1.82e+03±8.97e+00	2.69e+04±4.58e+02	1.78e+03±1.31e+01	1.77e+03±1.37e+01	**1.77e+03±1.16e+01**

Table 6: Empirical efficiency (mean and std.) for the independent residuals dataset.

is significantly above the other curves when $1-\epsilon^g \geq 0.7$ and it reaches 1.62×10^5 (when $1-\epsilon^g = 0.99$) whereas the other curves stay lower than 5.2×10^4, as detailed in Table 5. In the case of independent residuals (the second subplot), the efficiency curve of the "Max additive" approach is also significantly above the other curves when $1-\epsilon^g \geq 0.5$ and it reaches 6.31×10^4 (when $1-\epsilon^g = 0.99$) while the other curves stay lower than 2.3×10^3, as detailed in Table 6.

$1-\epsilon^g$	Bonferroni	Max additive	Independent DE-CCP	Gumbel DE-CCP	Empirical DE-CCP
0.99	6.96e+09±5.47e+09	1.06e+10±1.34e+10	3.78e+09±3.56e+09	3.45e+09±3.76e+09	**7.33e+08±1.02e+09**
0.95	2.10e+07±2.10e+07	4.33e+05±2.88e+05	6.97e+06±9.28e+06	3.44e+06±4.07e+06	**2.87e+05±1.85e+05**
0.90	4.42e+05±4.99e+05	4.86e+04±1.64e+04	1.84e+05±1.24e+05	1.21e+05±6.99e+04	**3.48e+04±1.11e+04**
0.80	2.31e+04±1.01e+04	6.95e+03±1.31e+03	1.56e+04±5.45e+03	1.14e+04±3.30e+03	**5.42e+03±7.63e+02**
0.70	6.93e+03±1.80e+03	2.32e+03±3.17e+02	4.41e+03±9.40e+02	3.31e+03±5.90e+02	**1.87e+03±2.32e+02**
0.60	3.08e+03±6.06e+02	9.40e+02±1.17e+02	1.79e+03±3.21e+02	1.35e+03±1.91e+02	**8.07e+02±1.19e+02**
0.50	1.69e+03±3.01e+02	4.37e+02±6.24e+01	8.25e+02±1.25e+02	6.23e+02±7.18e+01	**3.90e+02±5.27e+01**
0.40	1.03e+03±1.76e+02	2.17e+02±3.21e+01	4.04e+02±5.52e+01	3.00e+02±3.47e+01	**1.95e+02±2.42e+01**
0.30	6.60e+02±1.11e+02	1.07e+02±1.29e+01	1.93e+02±2.61e+01	1.43e+02±1.58e+01	**9.55e+01±1.25e+01**
0.20	4.53e+02±5.68e+01	4.58e+01±6.19e+00	8.39e+01±1.19e+01	5.95e+01±6.80e+00	**3.98e+01±4.89e+00**
0.10	3.16e+02±4.22e+01	1.36e+01±1.48e+00	2.65e+01±3.24e+00	1.77e+01±1.71e+00	**1.16e+01±1.36e+00**

Table 7: Empirical efficiency (mean and std.) for YOLO v8's residuals on KITTI.

$1-\epsilon^g$	Bonferroni	Max additive	Independent DE-CCP	Gumbel DE-CCP	Empirical DE-CCP
0.99	1.59e+08±9.23e+06	1.73e+08±9.97e+06	1.51e+08±7.40e+06	**1.41e+08±7.79e+06**	1.41e+08±8.60e+06
0.95	9.11e+06±4.91e+05	6.44e+06±3.23e+05	8.29e+06±4.79e+05	7.49e+06±3.99e+05	**6.22e+06±3.24e+05**
0.90	1.62e+06±2.76e+04	9.84e+05±1.55e+03	1.36e+06±6.67e+03	1.21e+06±8.24e+01	**8.88e+05±5.10e+03**
0.80	2.46e+05±2.75e+02	1.40e+05±1.23e+03	1.93e+05±6.70e+01	1.73e+05±4.26e+01	**1.27e+05±6.38e+02**
0.70	8.48e+04±3.86e+02	4.45e+04±2.48e+02	6.06e+04±6.23e+02	5.45e+04±5.08e+02	**4.09e+04±3.34e+02**
0.60	4.07e+04±4.19e+02	1.82e+04±3.31e+01	2.54e+04±8.34e+01	2.29e+04±3.05e+01	**1.74e+04±9.66e+00**
0.50	2.29e+04±1.07e+02	8.41e+03±2.42e+01	1.21e+04±6.42e+01	1.08e+04±8.34e+01	**8.10e+03±4.18e+01**
0.40	1.42e+04±9.66e+00	4.07e+03±2.33e+01	5.97e+03±1.05e+01	5.33e+03±3.30e+01	**3.96e+03±3.40e+01**
0.30	9.32e+03±5.08e+01	1.92e+03±1.55e+01	2.91e+03±1.57e+01	2.58e+03±1.73e+01	**1.89e+03±1.62e+01**
0.20	6.39e+03±1.62e+01	8.22e+02±3.30e+00	1.28e+03±2.23e+00	1.12e+03±4.36e+00	**8.13e+02±3.06e+00**
0.10	4.53e+03±2.36e+01	2.57e+02±1.08e+00	4.09e+02±2.99e+00	3.51e+02±4.16e+00	**2.55e+02±2.57e-01**

Table 8: Empirical efficiency (mean and std.) for YOLO v8's residuals on BDD100K.

For the real datasets, as detailed in Tables 7 & 8, the "Empirical DE-CCP" approach yields the best efficiency results, the "Max additive" approach comes in second but it performs worse when $1-\epsilon^g = 0.99$. As shown on Figure 3, on the KITTI dataset (on the first subplot, from left to right), when $1-\epsilon^g = 0.99$ all the curve abrutly shoot up and the curve of the "Max additive" approach shoots up the most. As detailed in Table 7, the

efficiency curves of all the other approaches stay below 7×10^9 but the efficiency curve of the "Max additive" approach reaches 10^{10}.

On the BDD100K dataset (the fourth subplot), we also observe abrupt increases when $1 - \epsilon^g = 0.99$ but, in contrast with the results on the KITTI, the efficiency curves only shoot up by two orders of magnitude (at most), as detailed in Table 8. The "Max additive" approach still performs worse when $1 - \epsilon^g = 0.99$ but its results are closer to the results of the other methods.

4.3. Results discussion

In this section, we interpret the results of our experiments, we discuss the importance of this study, and we draw recommendations to alleviate its limitations.

As reported in the previous section, the calibration results of the "Max additive" approach are not highly affected by the presence or the absence of correlations in the residuals. However, its efficiency results are highly affected by heteroscedasticity (on the synthetic datasets). As described in section 4.1, the synthetic residuals are generated such that the range of one dimension is significantly higher than the ranges of all the other dimensions. In these cases, the "Max additive" approach simply discards the dissimilarity data of the dimensions with small ranges as shown in equation 3, and this is very inefficient because a very big correction $\alpha_s^1 = \cdots = \alpha_s^4$ is always applied to each dimension of each predicted bounding box when inferring the corresponding interval (as described in Equations 1 & 2).

In contrast, the efficiency results of the "Bonferroni" approach and the copula-based approaches are not affected by dissimilarities with a dimension that has a larger range than all the other dimensions. This can be explained by the fact that these methods compute a different conformal quantile α_s^t for each dimension t: a dimension with a very large range doesn't make them discard all the other dimensions. Moreover, the copula-based approaches are more efficient than the "Bonferroni" approach. They explicitly model the dependencies and they directly minimize the sizes of the inferred bounding box intervals.

The "Empirical DE-CCP" is shown to be flexible w.r.t. the type of dependency and it has the best efficiency results on the real datasets (on KITTI and on BDD100K, using YOLO v8's residuals). However, it is also shown to be slightly over-conservative (above the calibration line) in the case of correlated dissimilarities and slightly under-conservative in the case of independent residuals. In the case of YOLO v8's residuals, it is nearly perfectly calibrated on the BDD100K dataset but it is also slightly under-conservative on the KITTI dataset. These over-conservative results can be explained by the fact that Equation 11 allows the DE optimization algorithm to find solutions that are above the calibration line: the copula constraint is expressed as an inequality instead of an equality. This explanation is supported by the fact that the DE optimization algorithm did not find any feasible solution when the copula constraint was expressed as an equality in our experiments. The under-conservative results can be explained by the fact that the "Empirical DE-CCP" approach depends on the amount of data that is used for calibration. This explanation is supported by the results of the "Gumbel DE-CCP" approach because it is closer to the calibration line than the "Empirical DE-CCP" in the case of independent dissimilarities, and it is also supported by the (near perfect calibration) results on the larger dataset (BDD100K).

This study is significant because it is the first application of copula-based conformal prediction to bounding box regression and because it fills the knowledge gap regarding the comparison of this approach to state-of-the-art conformal bounding box prediction methods (on two real autonomous driving benchmarks). Also, although De Grancey et al. (2022) had mentioned the possibility of using absolute values in the dissimilarity score to infer bounding box intervals, instead of simply enlarging the predicted bounding box as in De Grancey et al. (2022) & Andéol et al. (2023), this is the first study that puts this idea in practice in the context of object detection.

Inferring bounding box intervals has multiple practical advantages but it is also theoretically more rigorous. A properly sized and aligned inner box is more useful when a robot needs to reach a target object with guarantees (when a drone lands on a non-rectangular spot, for example) whereas the outer box is more useful when a robot needs to avoid an obstacle with guarantees (when an autonomous car parks in an surrounded slot, for example). From a theoretical perspective (Couso and Dubois, 2014), a ground truth bounding box is an ontic set and therefore inferring a single larger box that contains it does not properly quantify the uncertainty in its prediction. In fact, an epistemic set over a set (an ontic set that represents an upper bound and another ontic set that represents a lower bound) is needed to quantify the uncertainty in an ontic set: a bounding box interval that contains the ground truth bounding box is the proper representation of the uncertainty in its prediction.

Nonetheless, since object detection is not always single-class nor solely comprised of the bounding box regression task, this study is limited by the fact that it does not jointly address multi-class object classification & localization (bounding box regression). Solving this problem would require considering a generic multi-target problem (Waegeman et al., 2019), where a classification task is mixed with a (multivariate) regression task. The challenge here would be to mix heterogeneous tasks, yet since the copula-based approaches aggregate quantiles of conformal scores rather than the conformal scores themselves, extending the idea seems feasible [8]. An even more challenging task would be to integrate the recognition of a varying number of objects across images, possibly needing to mix conformal approaches with tools such as random finite sets (Vo and Vo, 2013).

5. Conclusion

This paper has applied copula-based split conformal prediction approaches to bounding box regression and it has compared their results to state-of-the-art conformal bounding box prediction methods. It has shown that the empirical copula-based method is more efficient and more robust to heteroscedastic residuals even when one dimension has a very large spread. This non-parametric method is also shown to be nearly perfectly calibrated on the BDD100K benchmark and slightly under-conservative on the KITTI benchmark. Further development should include conformal predictions for the classification task (as well) and rely on an efficiency optimization algorithm that can find feasible solutions for equality copula constraints.

8. This contrasts with the max-additive approach, which implicitly assumes commensurability of the computed non-conformity scores or residuals.

6. Acknowledgements

This research was supported by the UTC Foundation & the Safe AI Chair.

References

Moloud Abdar, Farhad Pourpanah, Sadiq Hussain, Dana Rezazadegan, Li Liu, Mohammad Ghavamzadeh, Paul Fieguth, Xiaochun Cao, Abbas Khosravi, U Rajendra Acharya, et al. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information fusion*, 76:243–297, 2021.

Léo Andéol, Thomas Fel, Florence De Grancey, and Luca Mossina. Confident object detection via conformal prediction and conformal risk control: an application to railway signaling. In *Conformal and Probabilistic Prediction with Applications*, pages 36–55. PMLR, 2023.

J Martin Bland and Douglas G Altman. Multiple significance tests: the bonferroni method. *Bmj*, 310(6973):170, 1995.

Daniel Bok, Copulae The, and team. Copulae, March 2024. URL https://github.com/DanielBok/copulae.

Wei Chen, Jinjin Luo, Fan Zhang, and Zijian Tian. A review of object detection: Datasets, performance evaluation, architecture, applications and current trends. *Multimedia Tools and Applications*, pages 1–59, 2024.

Inés Couso and Didier Dubois. Statistical reasoning with set-valued information: Ontic vs. epistemic views. *International Journal of Approximate Reasoning*, 55(7):1502–1518, 2014.

Florence De Grancey, Jean-Luc Adam, Lucian Alecu, Sébastien Gerchinovitz, Franck Mamalet, and David Vigouroux. Object detection with probabilistic guarantees: A conformal prediction approach. In *International Conference on Computer Safety, Reliability, and Security*, pages 316–329. Springer, 2022.

Pedro F Felzenszwalb, Ross B Girshick, David McAllester, and Deva Ramanan. Object detection with discriminatively trained part-based models. *IEEE transactions on pattern analysis and machine intelligence*, 32(9):1627–1645, 2009.

Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE conference on computer vision and pattern recognition*, pages 3354–3361. IEEE, 2012.

Christian Genest and Anne-Catherine Favre. Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of hydrologic engineering*, 12(4):347–368, 2007.

Abhishek Gupta, Alagan Anpalagan, Ling Guan, and Ahmed Shaharyar Khwaja. Deep learning for object detection and scene perception in self-driving cars: Survey, challenges, and open issues. *Array*, 10:100057, 2021.

Ali Harakeh, Michael Smart, and Steven 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. IEEE, 2020.

Eyke Hüllermeier and Willem Waegeman. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine learning*, 110(3):457–506, 2021.

Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics yolov8, 2023. URL https://github.com/ultralytics/ultralytics.

Fabian Kuppers, Jan Kronenberger, Amirhossein Shantia, and Anselm Haselhoff. Multivariate confidence calibration for object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 326–327, 2020.

Jing Lei, Max G'Sell, Alessandro Rinaldo, Ryan J Tibshirani, and Larry Wasserman. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113(523):1094–1111, 2018.

Zongyao Lyu, Nolan Gutierrez, Aditya Rajguru, and William J Beksi. Probabilistic object detection via deep ensembles. In *European Conference on Computer Vision*, pages 67–75. Springer, 2020.

Soundouss Messoudi, Sébastien Destercke, and Sylvain Rousseau. Copula-based conformal prediction for multi-target regression. *Pattern Recognition*, 120:108101, 2021.

Dimity Miller, Niko Sünderhauf, Haoyang Zhang, David Hall, and Feras Dayoub. Benchmarking sampling-based probabilistic object detectors. In *CVPR Workshops*, volume 3, page 6, 2019.

Roger B Nelsen. *An introduction to copulas*. Springer, 2006.

Kemal Oksuz, Tom Joy, and Puneet K Dokania. Towards building self-aware object detectors via reliable uncertainty quantification and calibration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9263–9274, 2023.

Harris Papadopoulos, Kostas Proedrou, Volodya Vovk, and Alex Gammerman. Inductive confidence machines for regression. In *Machine learning: ECML 2002: 13th European conference on machine learning Helsinki, Finland, August 19–23, 2002 proceedings 13*, pages 345–356. Springer, 2002.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

Tobias Riedlinger, Matthias Rottmann, Marius Schubert, and Hanno Gottschalk. Gradient-based quantification of epistemic uncertainty for deep object detectors. In *Proceedings*

of the IEEE/CVF Winter Conference on Applications of Computer Vision, pages 3921–3931, 2023.

Soja Salim and JS Jayasudha. A literature survey on estimating uncertainty in deep learning models: Ensuring safety in intelligent systems. In *2023 2nd International Conference on Computational Systems and Communication (ICCSC)*, pages 1–5. IEEE, 2023.

Yaser Sheikh and Mubarak Shah. Bayesian object detection in dynamic scenes. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 74–79. IEEE, 2005.

M Sklar. Fonctions de répartition à n dimensions et leurs marges. In *Annales de l'ISUP*, volume 8, pages 229–231, 1959.

Rainer Storn and Kenneth Price. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11:341–359, 1997.

Keke Su, Lihua Cao, Botong Zhao, Ning Li, Di Wu, and Xiyu Han. N-iou: better iou-based bounding box regression loss for object detection. *Neural Computing and Applications*, 36(6):3049–3063, 2024.

Sophia Huiwen Sun and Rose Yu. Copula conformal prediction for multi-step time series prediction. In *The Twelfth International Conference on Learning Representations*, 2023.

Pauli Virtanen, Ralf Gommers, Travis E Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature methods*, 17(3):261–272, 2020.

Ba-Tuong Vo and Ba-Ngu Vo. Labeled random finite sets and multi-object conjugate priors. *IEEE Transactions on Signal Processing*, 61(13):3460–3475, 2013.

Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Algorithmic Learning in a Random World*. Springer Nature, 2022.

Willem Waegeman, Krzysztof Dembczyński, and Eyke Hüllermeier. Multi-target prediction: a unifying view on problems and methods. *Data Mining and Knowledge Discovery*, 33: 293–324, 2019.

Simon Wenkel, Khaled Alhazmi, Tanel Liiv, Saud Alrshoud, and Martin Simon. Confidence score: The forgotten dimension of object detection performance evaluation. *Sensors*, 21 (13):4350, 2021.

Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2636–2645, 2020.

Ruiyao Zhang, Ping Zhou, and Tianyou Chai. Improved copula-based conformal prediction for uncertainty quantification of multi-output regression. *Journal of Process Control*, 129: 103036, 2023.

Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu, and Xindong Wu. Object detection with deep learning: A review. *IEEE transactions on neural networks and learning systems*, 30(11): 3212–3232, 2019.