Conformal Prediction for Spatial Uncertainty in Object Detection

Yantong.LU
2024/09/13

Content

01 Introduction
Spatial Uncertainty VS Semantic Uncertainty

Spatial Uncertainty

the uncertainty regarding whether the ***bounding boxes*** generated by object detection models accurately locate and cover the objects

Semantic Uncertainty

the uncertainty of whether object detection models ***classify objects*** into specific categories

Issues:

(1) Confidence scores for class categorization are often not calibrated. (2) There is a lack of uncertainty estimation for the deviation of bounding boxes[2]

[1] D. Hall *et al.*, 'Probabilistic Object Detection: Definition and Evaluation', presented at the Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2020, pp. 1031–1040. [2] Y. Gal, 'Uncertainty in deep learning', 2016.

Why Conformal Prediction not …?

Compared to other Uncertainty Quantification Methods[2]

Provide statistical rigorous uncertainty sets/intervals for the predictions of such models.

Distribution-free: No specific requirement
Other methods may require data assumptions like on Bernoulli or Gaussian distributions….

easy-to-understand and easy-to-use post-processing method applied after the model's output, it does not require any modifications to the internal structure of the model

[2] A. N. Angelopoulos and S. Bates, 'A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification', Cornell University Library, arXiv.org, Ithaca, 2022. doi: 10.48550/arxiv.2107.07511.

Example 1:Conformal Prediction for animal recognition model

Semantic Uncertainty

Suppose we set an acceptable error rate at α

All of additional prediction sets would guarantee to cover the true categories in the ground truth with a coverage rate of $1 - \alpha$

[2] A. N. Angelopoulos and S. Bates, 'A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification', Cornell University Library, arXiv.org, Ithaca, 2022. doi: 10.48550/arxiv.2107.07511

Example 2:Conformal Prediction for ‘Pedestrian Detection’

Baseline work[3]

Spatial Uncertainty: the focus of my paper

Suppose we set an acceptable error rate at α

All of conformalized bounding boxes based on predicted boxes would guarantee to cover the true boxes in the ground truth with a coverage rate of $1-\alpha$

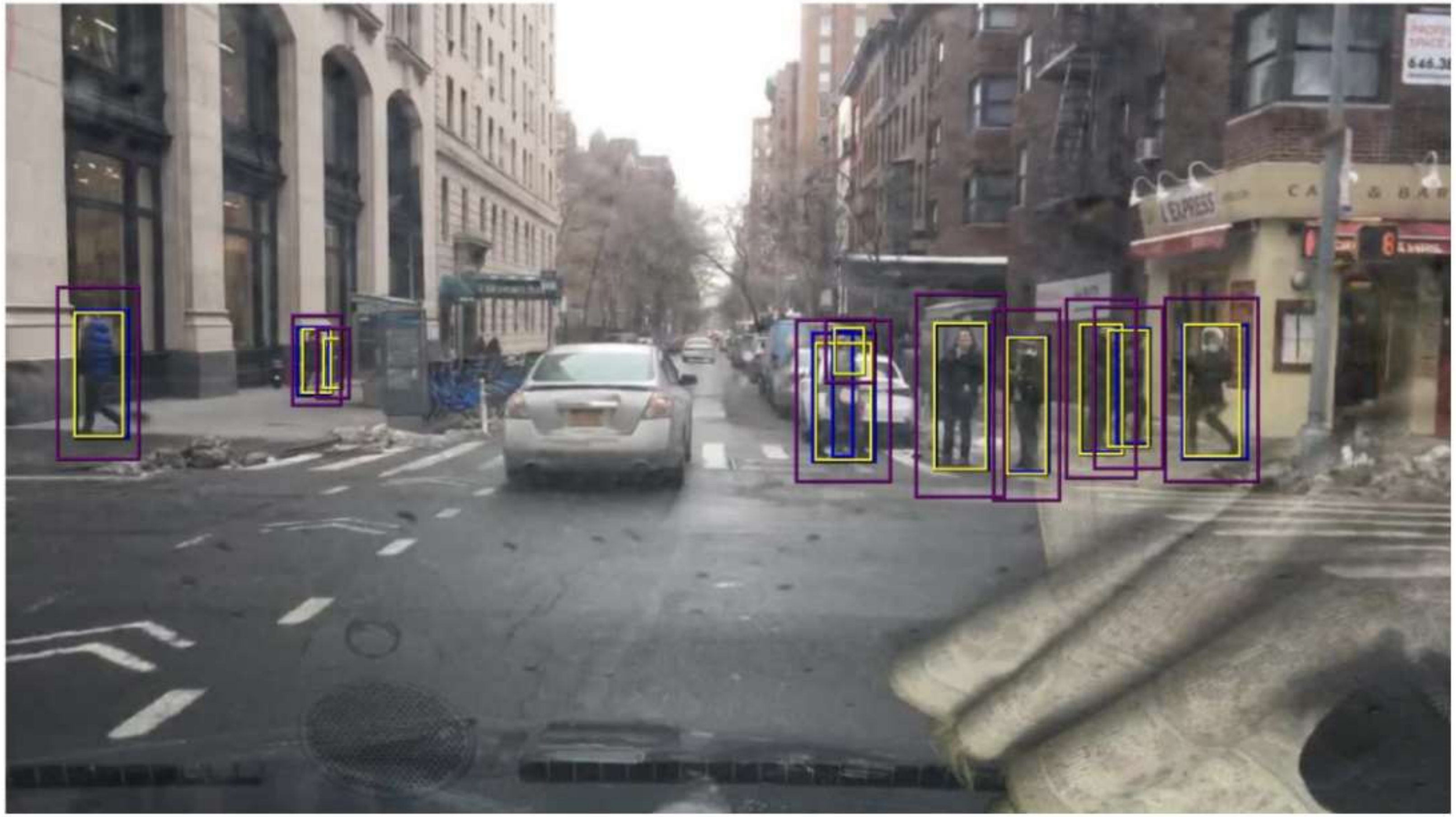


Figure 2 Conformalization example (*Coordinate-Adaptation*, risk level $\alpha = 0.1$) on a BDD100k image with Ground Truth, Inference and Conformalized boxes.

[3] 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 *Computer Safety, Reliability, and Security. SAFECOMP 2022 Workshops*, M. Trapp, E. Schoitsch, J. Guiochet, and F. Bitsch, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2022, pp. 316–329. doi: 10.1007/978-3-031-14862-0_23.

Main contributions of my paper:

(1) Fined-tuned model vs pretrained model

(2) New conformalization Methods:

Coordinate-Adaptation
Area-Difference
Binary-Search

(3)New evaluation metric:

Expansion

Overview of Methods

Method (*this paper*)	Specific coverage $(1-\alpha)$
Coordinate-Wise	x_{min}
	x_{max}
	y_{min}
	y_{max}
Box-Wise	*Coordinate-difference* w/o Bonferroni
	Coordinate-difference with Bonferroni
	Coordinate-adaptation w/o Bonferroni
	Coordinate-adaptation with Bonferroni
	Area-difference
Image-Wise	*Binary-search*

The Core Process of Conformal Prediction

Algorithm 1 Split conformal prediction

1: **Input:** data $\mathcal{D} \subset \mathcal{X} \times \mathcal{Y}$, prediction algorithm $\mathcal{A}$, tolerated miscoverage $\alpha \in (0,1)$

2: **Output:** Prediction set $\hat{C}(X_{n+1})$ for test sample $(X_{n+1}, Y_{n+1}) \in \mathcal{D}_{test}$

3: **Procedure:**

4: Split data $\mathcal{D}$ into two non-intersecting subsets: a training set $\mathcal{D}_{train}$ and calibration set $\mathcal{D}_{cal}$

5: Train a machine learning model on the training set:

$$\hat{f}(\cdot) \leftarrow \mathcal{A}(\mathcal{D}_{train})$$

6: Define a score function $s : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, and the learned model $\hat{f}(\cdot)$ is evaluated by s on $\mathcal{D}_{cal}$, resulting in a set of nonconformity scores

$$S = \{s(\hat{f}(X_i), Y_i)\}_{i=1}^{n} = \{s_i\}_{i=1}^{n}$$

Where score s_i represents a notion of dissimilarity (nonconformity) between the prediction $\hat{f}(X_i)$ and the ground truth Y_i, we calculate the nonconformity S on all data points of $\mathcal{D}_{cal}$, n is the size of $\mathcal{D}_{cal}$

7: Define a conformal quantile $\hat{q}$ as:

$\lceil (n+1)(1-\alpha)/n \rceil$-th quantile of S, *which means* $\lceil (n+1)(1-\alpha)/n \rceil$-th largest value among the nonconformity S

Under exchangeability of $\mathcal{D}_{cal} \cup \{(X_{n+1}, Y_{n+1})\}$, the conformal quantile $\hat{q}$ is a finite sample-corrected quantile ensuring target coverage $(1-\alpha)$ by its construction

8: Given a new input $(X_{n+1}, Y_{n+1}) \in \mathcal{D}_{test}$, instead of simply outputting a $\hat{f}(X_{n+1})$, a conformal prediction set $\hat{C}(X_{n+1})$ for X_{n+1} can be defined as

$$\hat{C}(X_{n+1}) = \{y \in \mathcal{Y} : s(\hat{f}(X_{n+1}), y) \le \hat{q}\}$$

Step1:
Designing Nonconformity Scores Across Different Dimensions
Step2:
Calculating the Conformal Quantile
Step3:
Generating Prediction Sets

Each Step in 3 dimensions：
Coordinate-Wise
Box-Wise
Image-Wise

Conformal prediction in bounding box

A more intuitive understanding:
The core of conformal prediction (CP) lies in capturing the "errors" between the predicted box and the true box.

The key to capturing this critical information is designing an appropriate nonconformity score function.

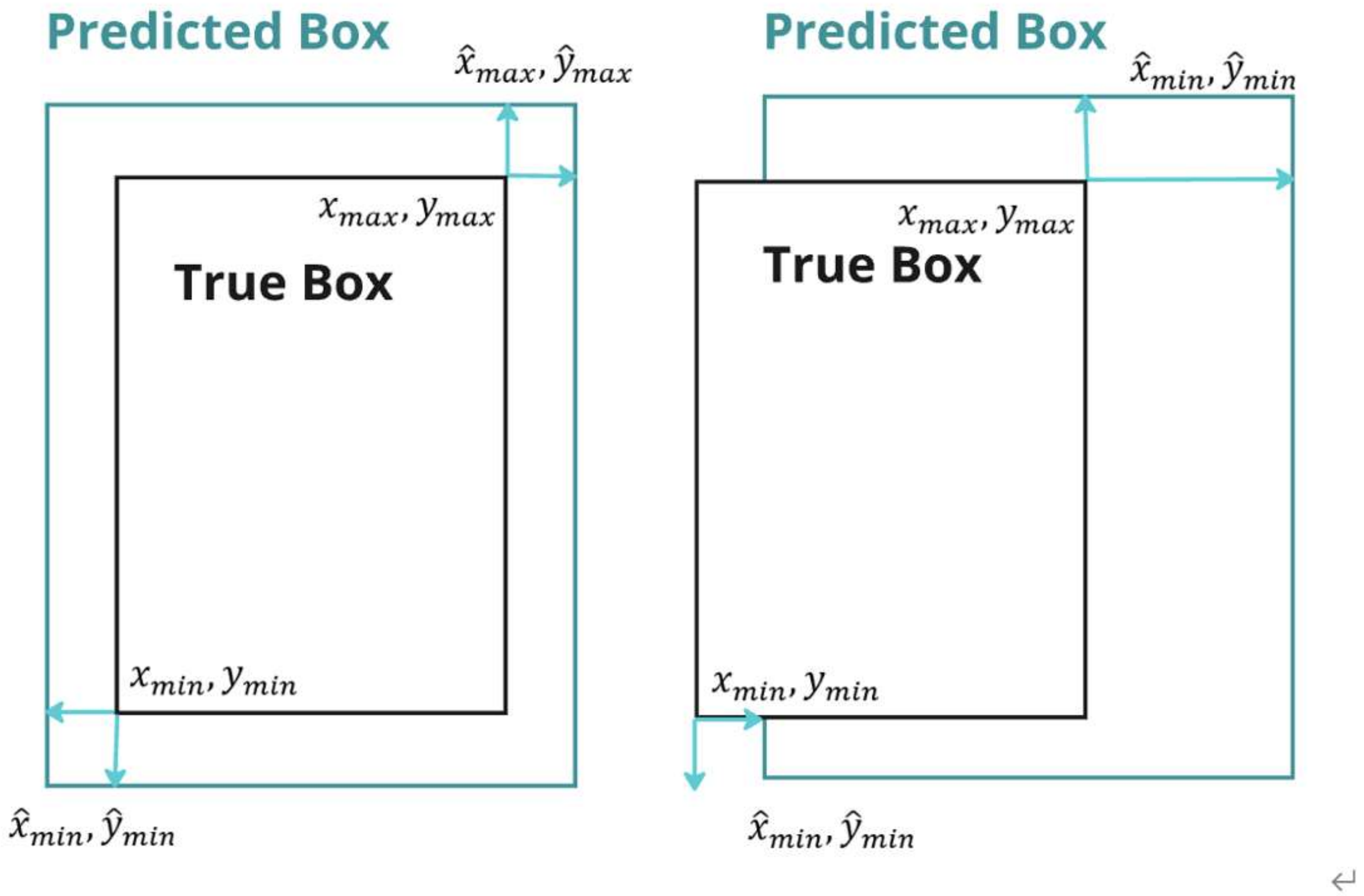


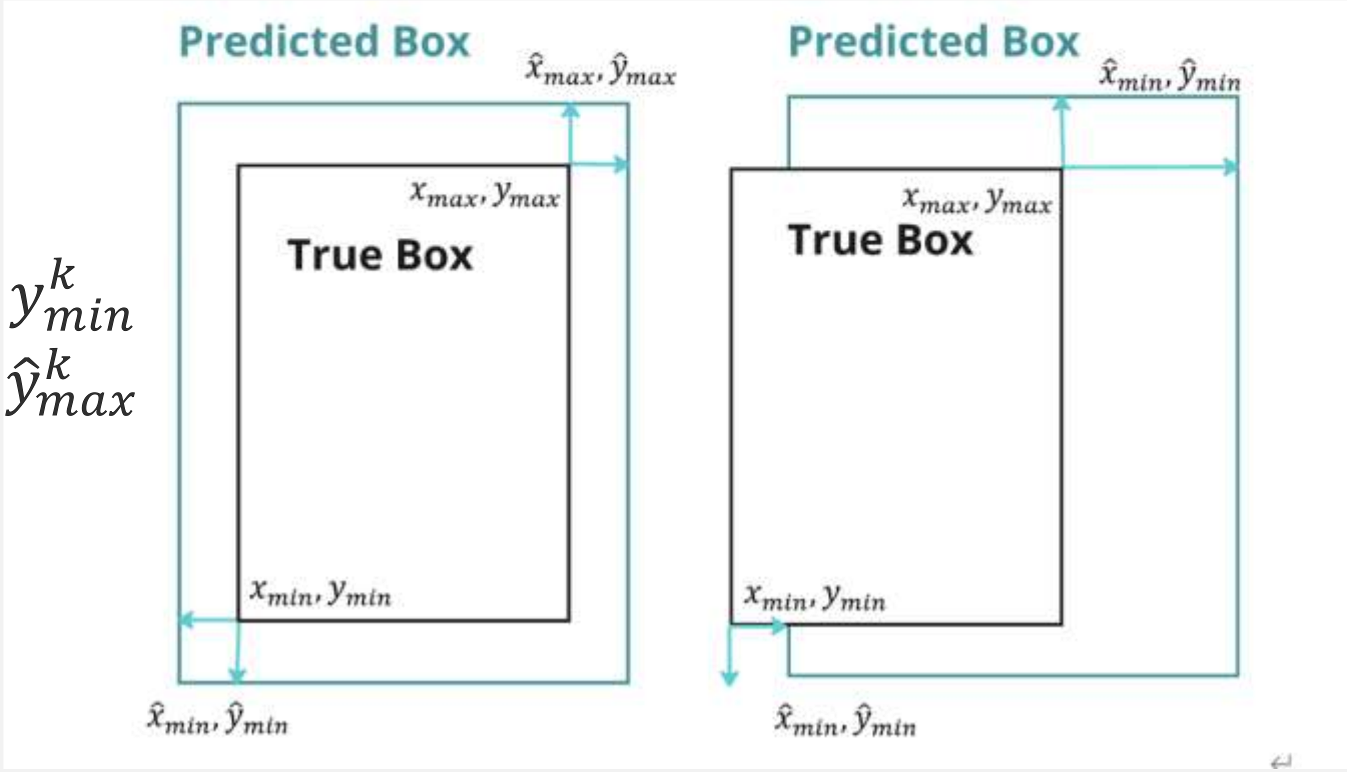
Figure 6 Possible scenarios of deviation between true box and predicted boxes

Step 1:
Designing Nonconformity Scores Across Different Dimensions

A. Coordinate-Wise

Referencing baseline, the nonconformity score of for Coordinate-Wise can simply be obtained by calculating the differences between the coordinates of the true box and the predicted box, as given in Equation

$$R^k_{x_{min}} = \hat{x}^k_{min} - x^k_{min} \qquad R^k_{y_{min}} = \hat{y}^k_{min} - y^k_{min}$$
$$R^k_{x_{max}} = x^k_{max} - \hat{x}^k_{max} \qquad R^k_{y_{max}} = y^k_{max} - \hat{y}^k_{max}$$



Step 1:
Designing Nonconformity Scores Across Different Dimensions

B. Box-Wise

"*coordinate-difference*"

$$R^k_{box} = \left(\hat{x}^k_{min} - x^k_{min}, \hat{y}^k_{min} - y^k_{min}, x^k_{max} - \hat{x}^k_{max}, y^k_{max} - \hat{y}^k_{max}\right)$$

The calculated errors are therefore larger for bigger objects and smaller for smaller objects, which may cause conformal prediction calculations to be disproportionately influenced by larger objects, potentially neglecting smaller ones.

"*coordinate- adaptation*"

$$R^k_{box} = \left(\frac{\hat{x}^k_{min} - x^k_{min}}{\hat{w}^k}, \frac{\hat{y}^k_{min} - y^k_{min}}{\hat{h}^k}, \frac{x^k_{max} - \hat{x}^k_{max}}{\hat{w}^k}, \frac{y^k_{max} - \hat{y}^k_{max}}{\hat{h}^k}\right)$$

where $\hat{w}^k$ and $\hat{h}^k$ represent the width and height of the predicted box, respectively.

Step 1:
Designing Nonconformity Scores Across Different Dimensions

B. Box-Wise

In "*coordinate-difference*“ methods, baseline found that under a 95% coverage requirement $(1-\alpha)$, only 86% of the real boxes were covered by the conformalized boxes due to not consider all four coordinates together

Baseline apply a stricter control: Bonferroni correction : $\alpha/4$ for each coordinate.
Inspired by it, If I do not use the Bonferroni correction but intuitively consider the areas of the predicted and true boxes, can the formula ensure actual coverage by the four coordinates? Therefore, the "*area difference*" method are proposed:

$$\hat{A}^k = (\hat{x}^k_{max} - \hat{x}^k_{min}) \times (\hat{y}^k_{max} - \hat{y}^k_{min})$$

$$A^k = (x^k_{max} - x^k_{min}) \times (y^k_{max} - y^k_{min})$$

$$R^k_{box} = \frac{\hat{A}^k - A^k}{\hat{A}^k}$$

Step 2:
Calculating the Conformal Quantile

We obtain an array $R_{all} = \{R^1, \dots, R^k\}$, that includes the "deviations" between all predicted and actual bounding boxes. Upon sorting this array, we can determine the conformal quantile:

(1) For the *Coordinate-Wise* nonconformity score, defined as

$$q_{1-\alpha} = \lceil (n_c + 1)(1 - \alpha) \rceil - th\ element\ of\ R_{all}$$

(2)For the box-wise nonconformity score without Bonferroni correction (Methods include *Coordinate-Difference w/o Bonferroni*, *Coordinate-Adaptation w/o Bonferroni*, *Area-Difference*):

$$q_{1-\alpha} = \lceil (n_{c/box} + 1)(1 - \alpha) \rceil - th\ element\ of\ R_{all}$$

Step 2:
Calculating the Conformal Quantile

(3) For box-wise methods with Bonferroni correction applied (*Coordinate-Difference with Bonferroni*, *Coordinate-Adaptation with Bonferroni*):

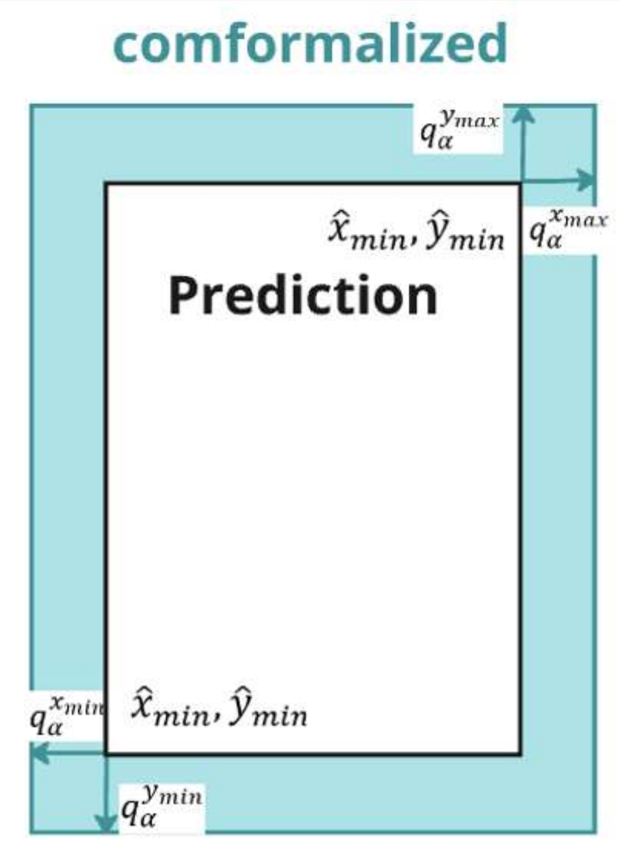
$$q_{1-\frac{\alpha}{4}} = \left\lceil (n_{box}+1)\left(1-\frac{\alpha}{4}\right)\right\rceil \ -th\ element\ of\ R^{c}_{all}, c \in \{x_{min}, y_{min}, x_{max}, y_{max}\}$$

α is the pre−set error rate, and $1-\alpha$ or $1-\frac{\alpha}{4}$ is the desired coverage rate. n_{box} represents the number of bounding boxes in the calibration set.

The quantile calculated in the this chapter can essentially be referred to as the margin[4] applied to the predicted boxes.

[4] A. Timans, C.-N. Straehle, K. Sakmann, and E. Nalisnick, ‘Adaptive Bounding Box Uncertainties via Two-Step Conformal Prediction’, Mar. 11, 2024, *arXiv*: arXiv:2403.07263. doi: 10.48550/arXiv.2403.07263.

Step 3:
Generating Prediction Sets



Coordinate-wise: For each new coordinate in x_{mi} … ven the model's predicted coordinates $\hat{x}_{min}, \hat{y}_{min}, \hat{x}_{max}, \hat{y}_{max}$ … fined as:

$$\hat{C}_\alpha(x_{min}) = [\hat{x}_{min} - q_\alpha^{x_{min}}, +\infty) \qquad \hat{C}_\alpha(y_{\dots} \dots \infty)$$

$$\hat{C}_\alpha(x_{max}) = (-\infty, \hat{x}_{max} + q_\alpha^{x_{max}}] \qquad \hat{C}_\alpha(y_{\dots} \dots {}^{max}]$$

Box-wise: For new data points X_{n+1}, the model predic…

$(\hat{x}_{min}, \hat{y}_{min}, \hat{x}_{max}, \hat{y}_{max})$

For *Coordinate-Difference w/o Bonferroni Method*:

$$\hat{C}_\alpha(X_{n+1}) = \left\{\hat{x}_{min} - q_\alpha^{x_{min}}, \hat{y}_{min} - q_\alpha^{y_{min}}, \hat{x}_{max} + q_\alpha^{x_{max}}, \hat{y}_{max} + q_\alpha^{y_{max}}\right\}$$

For *Coordinate-Difference with Bonferroni Method:*

$$\hat{C}_\alpha(X_{n+1}) = \left\{\hat{x}_{min} - q_{1-\frac{\alpha}{4}}^{x_{min}}, \hat{y}_{min} - q_{1-\frac{\alpha}{4}}^{y_{min}}, \hat{x}_{max} + q_{1-\frac{\alpha}{4}}^{x_{max}}, \hat{y}_{max} + q_{1-\frac{\alpha}{4}}^{y_{max}}\right\}$$

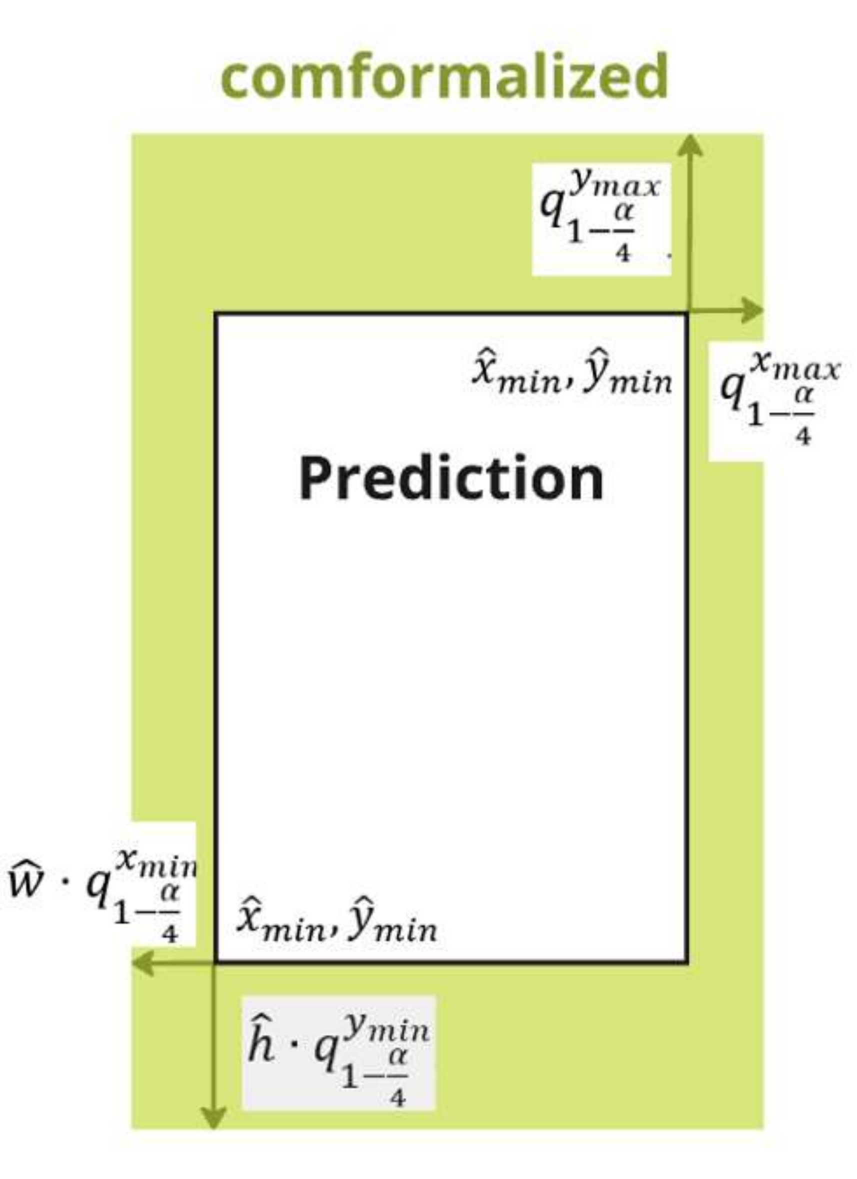
Step 3:
Generating Prediction Sets

For *Coordinate-Adaptation with Bonferroni Method* :

$$\hat{C}_\alpha(X_{n+1}) = \begin{Bmatrix} \hat{x}_{min} - \hat{w} \cdot q^{x_{min}}_{1-\frac{\alpha}{4}}, \hat{y}_{min} - \hat{h} \cdot q^{y_{min}}_{1-\frac{\alpha}{4}}, \\ \hat{x}_{max} + \hat{w} \cdot q^{x_{max}}_{1-\frac{\alpha}{4}}, \hat{y}_{max} + \hat{h} \cdot q^{y_{max}}_{1-\frac{\alpha}{4}} \end{Bmatrix}$$

For *Coordinate-Adaptation w/o Bonferroni Method*

$$\hat{C}_\alpha(X_{n+1}) = \begin{Bmatrix} \hat{x}_{min} - \hat{w} \cdot q^{x_{min}}_{1-\alpha}, \hat{y}_{min} - \hat{h} \cdot q^{y_{min}}_{1-\alpha} \\ \hat{x}_{max} + \hat{w} \cdot q^{x_{max}}_{1-\alpha}, \hat{y}_{max} + \hat{h} \cdot q^{y_{max}}_{1-\alpha} \end{Bmatrix}$$



Step 3:
Generating Prediction Sets

For *Area-Difference Method:*

when calculating the conformal quantile, $q_{1-\alpha}$ is actu[...] [...]e. My design method is finding a margin θ to distribute this $q_{1-\alpha}$ eve[...] [...] box.

The area of the conformalized boxes can be repr[...] [...]$2\theta)(\hat{y}_{max} - \hat{y}_{min} + 2\theta)$. Similarly, in the error calculation, we als[...] [...]values equal. Therefore, the calculation formula can be expressed as:

$$\frac{[(\hat{x}_{max} - \hat{x}_{min} + 2\theta)(\hat{y}_{max} - \hat{y}_{mi} \ldots}{\hat{A}}$$

From this, we solve for the margin θ, which can b[...] [...]ncy of $O(1)$.

Ultimately, the prediction set can be defined as:

$$\hat{C}_\alpha(X_{n+1}) = \{\hat{x}_{min} - \theta, \hat{y}_{min} - \theta, \hat{x}_{max} + \theta, \hat{y}_{max} + \theta\}$$

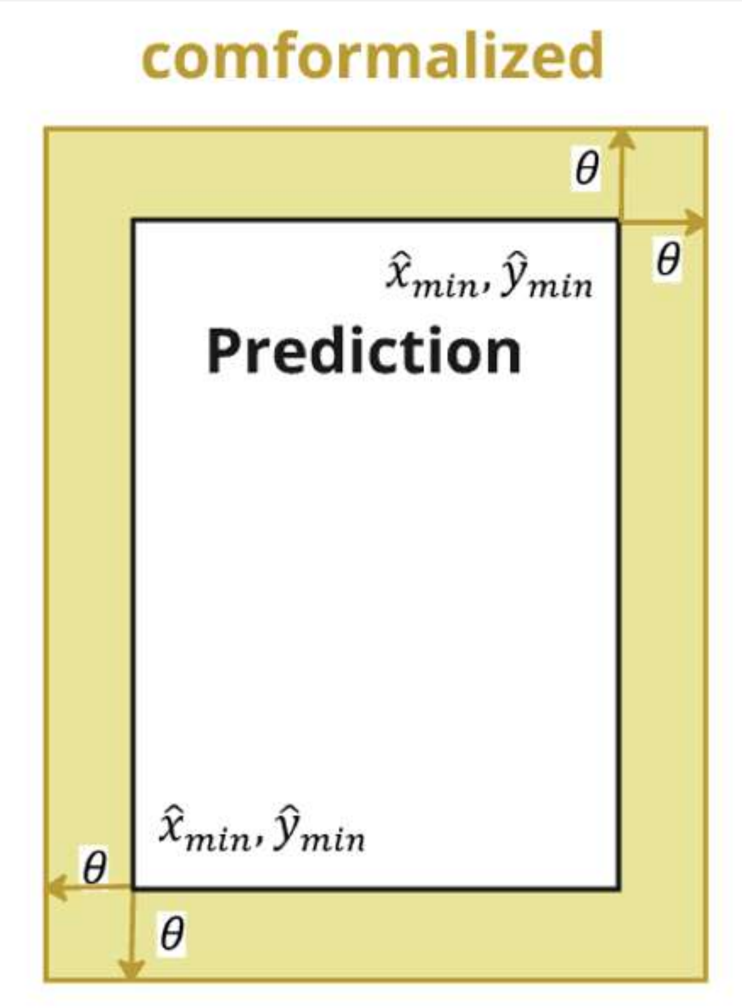
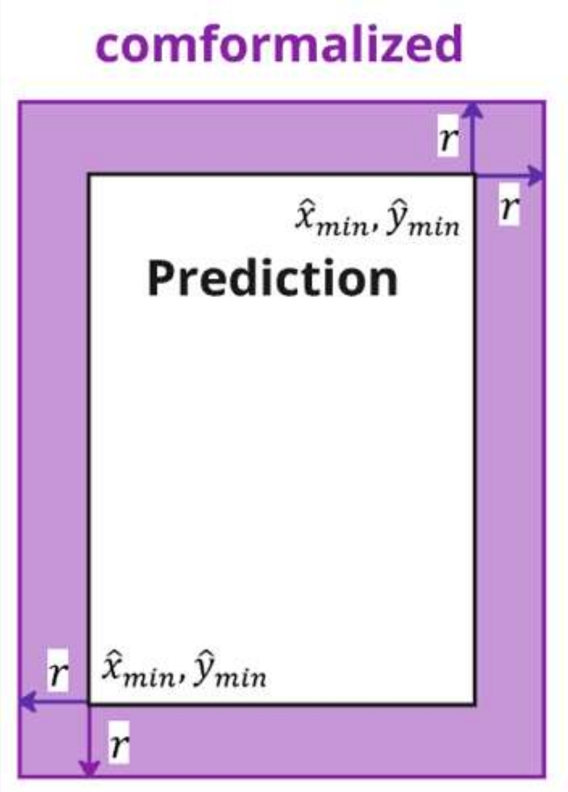


Image-wise Step 1-3

Since baseline only briefly mentioned the nonconformity score function without providing any formulas or experimental procedures, the following section aims to supplement these theoretical details. Specifically, it addresses how to find the smallest boundary $r \geq 0$ such that in the dataset, for $1 - \alpha$ of the images, at least $1 - \beta$ of the true boxes are correctly covered by the conformalized box expanded by the margin r

If $\alpha = 0.1$, $\beta = 0.25$ are set, it means that at least 75% of the true boxes in at least 90% of the images in the dataset need to be covered by the conformalized boxes.



Algorithm 2 *Binary-search* Image-wise Conformal Prediction

1.**Input:** Image dataset $D \in X \times Y$, prediction algorithm A, tolerated miscoverage $\alpha \in (0,1)$, and non-coverage rate per image $\beta \in (0,1)$.

2.**Output:** Prediction set $\hat{C}(X_{n+1})$ for each test image $(X_{n+1}, Y_{n+1}) \in D_{test}$, ensuring that at least $1-\beta$ of the true boxes are covered with a margin $r \geq 0$ for at least $1-\alpha$ proportion of the images.

3.**Procedure:**

4.Divide the dataset D into two non-intersecting subsets: a calibration set D_{cal} and a test set D_{test}.

5. **Initialize** binary search parameters:

$left = 0$

$right = 1000$

Maximum iterations N for binary search

6.**While** $N > 0$:

$\lambda = (left + right)/2$ (midpoint of current left and right)

Generate conformalized boxes for each X_{n+1} in D_{cal}:

$$\hat{C}(X_{n+1}) = \{\hat{x}_{min} - \lambda, \hat{y}_{min} - \lambda, \hat{x}_{max} + \lambda, \hat{y}_{max} + \lambda\}$$

Compute the coverage: Evaluate the proportion of images where at least $1-\beta$ of the true boxes are covered by $\hat{C}(X_{n+1})$

If this proportion $\geq 1-\alpha$, set $right = \lambda$ (reduce λ to decrease coverage).

Else, **set** $left = \lambda$ (increase λ to enhance coverage

$N = N - 1$

7.**Finalize** λ as the optimal margin r after N iterations.

8.For a new input image $(X_{n+1}, Y_{n+1}) \in D_{test}$, output the prediction set:

$$\hat{C}(X_{n+1}) = \{\hat{x}_{min} - r, \hat{y}_{min} - r, \hat{x}_{max} + r, \hat{y}_{max} + r\}$$

we abstract the nonconformity score function and the process of finding the quantile into a binary search, updating boundaries, and seeking the minimal boundary r.

Given that the algorithm strictly ensures coverage rates, this method can be considered a variant of conformal prediction (while also guaranteeing coverage rates).

Actually, this algorithm is a procedure involves initially setting a conformalized boxes $\hat{C}(X_{n+1}) = \{\hat{x}_{min} - \lambda, \hat{y}_{min} - \lambda, \hat{x}_{max} + \lambda, \hat{y}_{max} + \lambda\}$ and then **iteratively adjusting this bounding box to continually test the coverage rate.**

Evaluation Metrics

Observed Coverage, follows the baseline, and measures the proportion of new conformalized bounding boxes that completely cover the actual ground truth bounding boxes across the entire test set.

$$observed\ coverage = \frac{\sum_{i=1}^{n} Y_i \in \hat{C}(X_i)}{n_{truebox}}$$

Limits of only this metric:

there is a possibility that the new conformalized boxes are significantly larger than the original predicted boxes, which could increase the coverage rate but contradict our goal of precisely locating the actual objects.

Therefore, we hope for the conformalized boxes to fit the original model's predicted boxes as closely as possible while ensuring a certain *Observed Coverage* rate.

I propose a new evaluation metric called "*Expansion*".

Evaluation Metrics

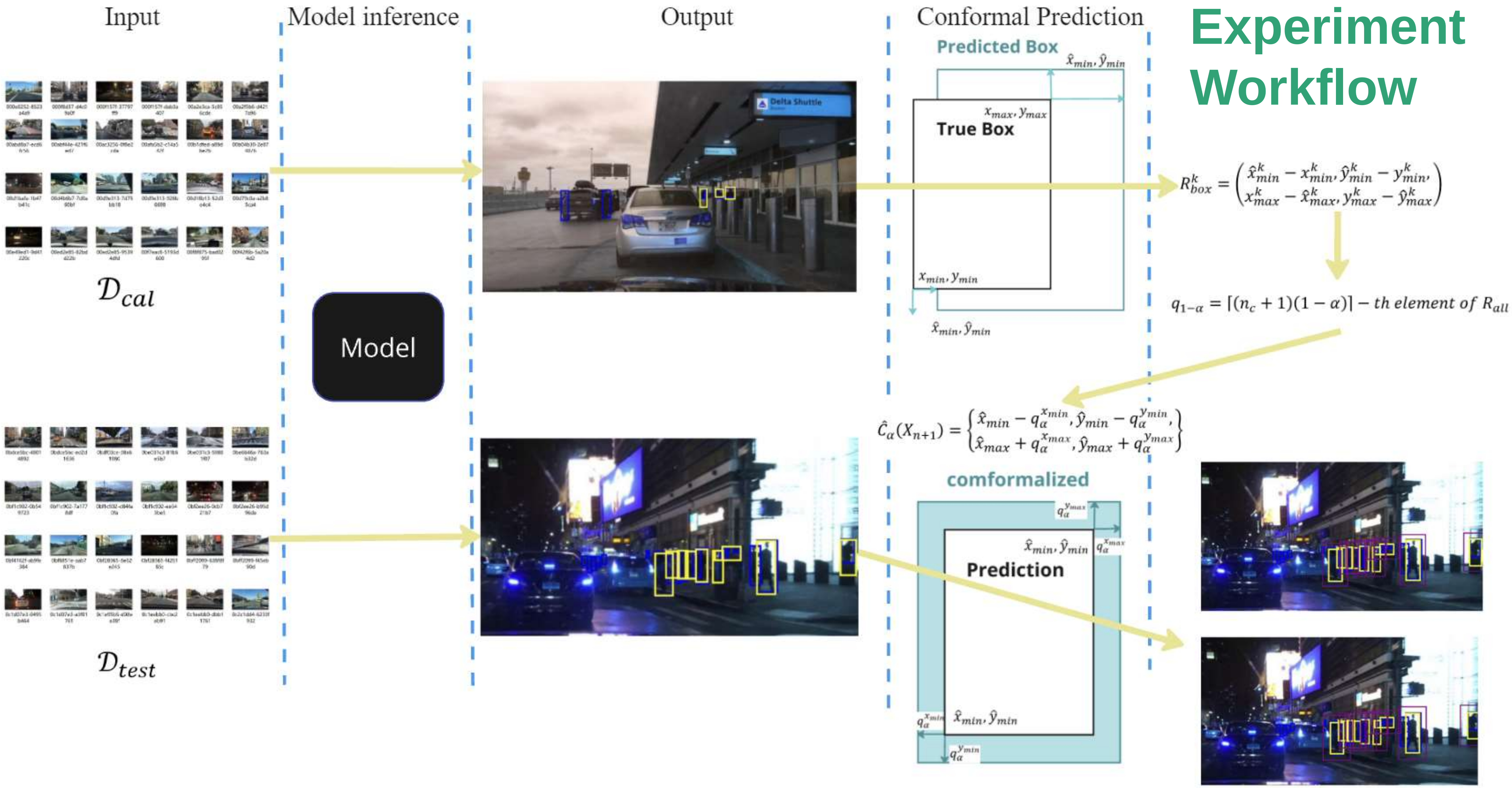
The *Expansion* metric can be defined as the difference in area between the conformalized boxes and the predicted boxes, divided by the area of the predicted boxes.

$$E = \frac{\sum_{i=1}^{n} \sum_{j}^{n_{box}} A\left(C\left(X_i\right)\right)_j - \sum_{i=1}^{n} \sum_{j}^{n_{box}} A\left(f\left(X_i\right)\right)_j}{\sum_{i=1}^{n} \sum_{j}^{n_{box}} A\left(f\left(X_i\right)\right)_j}$$

Limits of this two metrics:

It's worth noting that both *Expansion* and *Coverage* are calculated as "average expansion" and "average coverage," for instance, while some images may have many true boxes not covered by conformalized boxes, the *Coverage* can be compensated by the number of successful covers in other images.

Experiment Workflow



Results----my research

Table 1 Evaluation of observed coverage on $\mathcal{D}_{test}$

Method ($this\ paper$)	Specific coverage $(1-\alpha)$	0.7	0.9	0.95
		Observed Coverage (pretrained/fine-tuned)		
Coordinate-Wise	x_{min}	0.68/0.74	0.89/0.91	0.94/0.96
	x_{max}	0.71/0.74	0.93/0.93	0.95/0.96
	y_{min}	0.70/0.70	0.88/0.91	0.95/0.96
	y_{max}	0.70/0.73	0.92/0.93	0.90/0.95
Box-Wise	*Coordinate-difference* w/o Bonferroni	0.41/0.39	0.70/0.71	0.78/0.80
	Coordinate-difference with Bonferroni	0.75/0.76	0.92/0.91	0.96/0.96
	Coordinate-adaptation w/o Bonferroni	0.45/0.55	0.81/0.80	0.89/0.90
	Coordinate-adaptation with Bonferroni	0.73/0.78	0.91/0.93	0.95/0.97
	Area-difference	0.55/0.63	0.80/0.83	0.88/0.93
Image-Wise	*Binary-search*	0.78/0.79	0.93/0.94	0.95/0.96

Results----baseline

Table 2 Baseline work of de Grancey et al. Observed Coverage on $\mathcal{D}_{test}$

Method (*de Grancey*)	Specific coverage $(1-\alpha)$	0.7	0.9	0.95
		Observed Coverage (Only Pretrained)		
Coordinate-Wise	x_{min}	0.76	0.91	0.96
	x_{max}	0.78	0.91	0.96
	y_{min}	0.70	0.92	0.95
	y_{max}	0.71	0.91	0.95
Box-Wise	*Coordinate-difference* w/o Bonferroni	0.35	0.73	0.86
	Coordinate-difference with Bonferroni	0.79	0.92	0.96

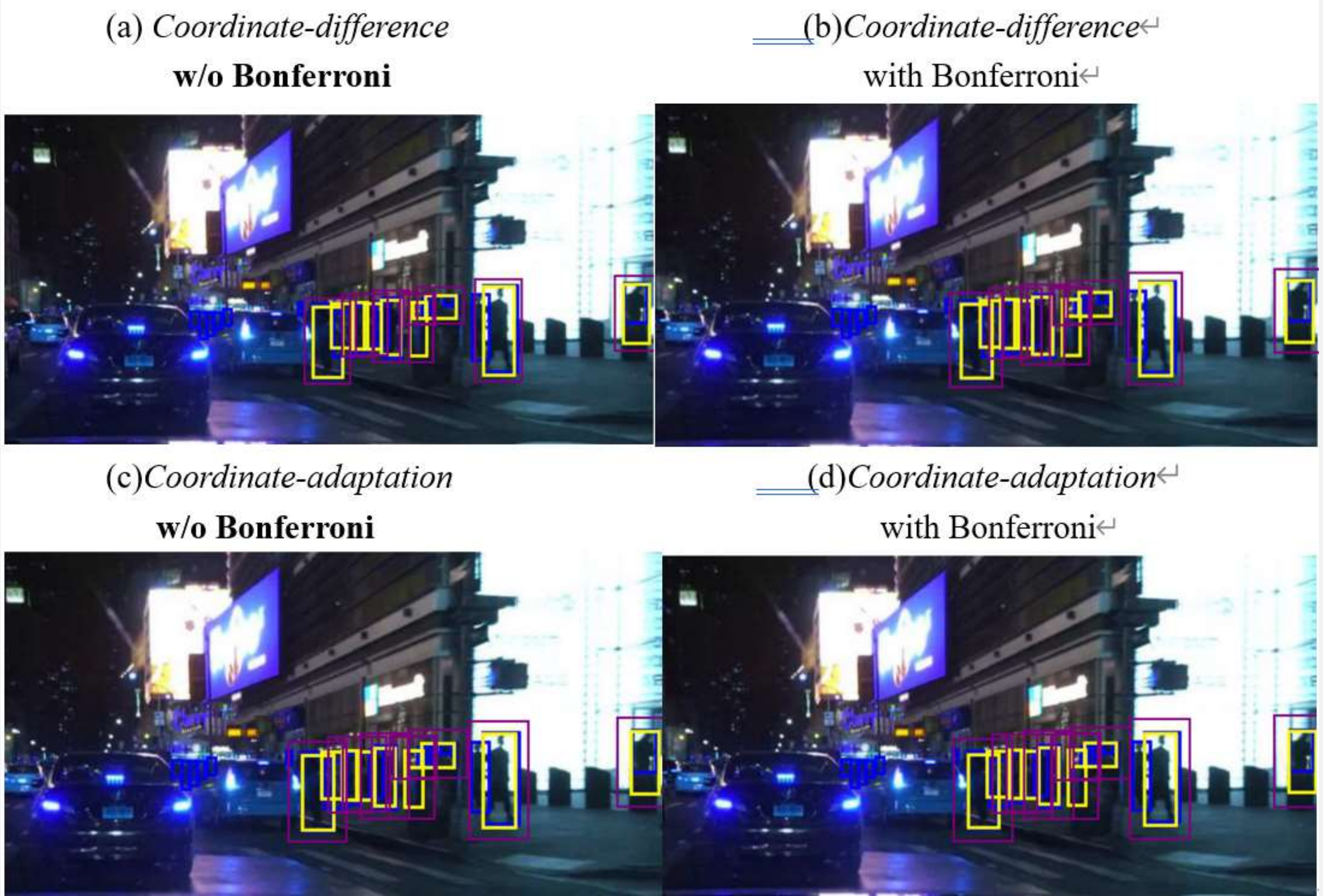
Results----my research

Table 3 Evaluation of Expansion on $\mathcal{D}_{test}$ for conformalization methods

Method	Specific coverage $(1-\alpha)$	0.7	0.9	0.95
		Expansion (pretrained/fine-tuned)		
Box-Wise	*Coordinate-difference* **w/o Bonferroni** *(de Grancey)*	0.15/0.17	0.32/0.27	0.38/0.32
	Coordinate-difference with Bonferroni *(de Grancey)*	0.37/0.28	0.45/0.35	0.49/0.41
	Coordinate-adaptation **w/o Bonferroni** *(this paper)*	0.39/0.21	0.49/0.31	0.61/0.42
	Coordinate-adaptation with Bonferroni *(this paper)*	0.48/0.29	0.53/0.36	0.67/0.48
	Area-difference *(this paper)*	0.52/0.44	0.58/0.50	0.78/0.70
Image-Wise	*Binary-search* *(this paper)*	12.25/8.54	15.64/10.94	22.64/16.84

Results

Images of conformalized box
(*including 6 methods*, risk level $\alpha = 0.1$, pretrained model) on a BDD100k image
with Ground Truth, Inference and Conformalized boxes.

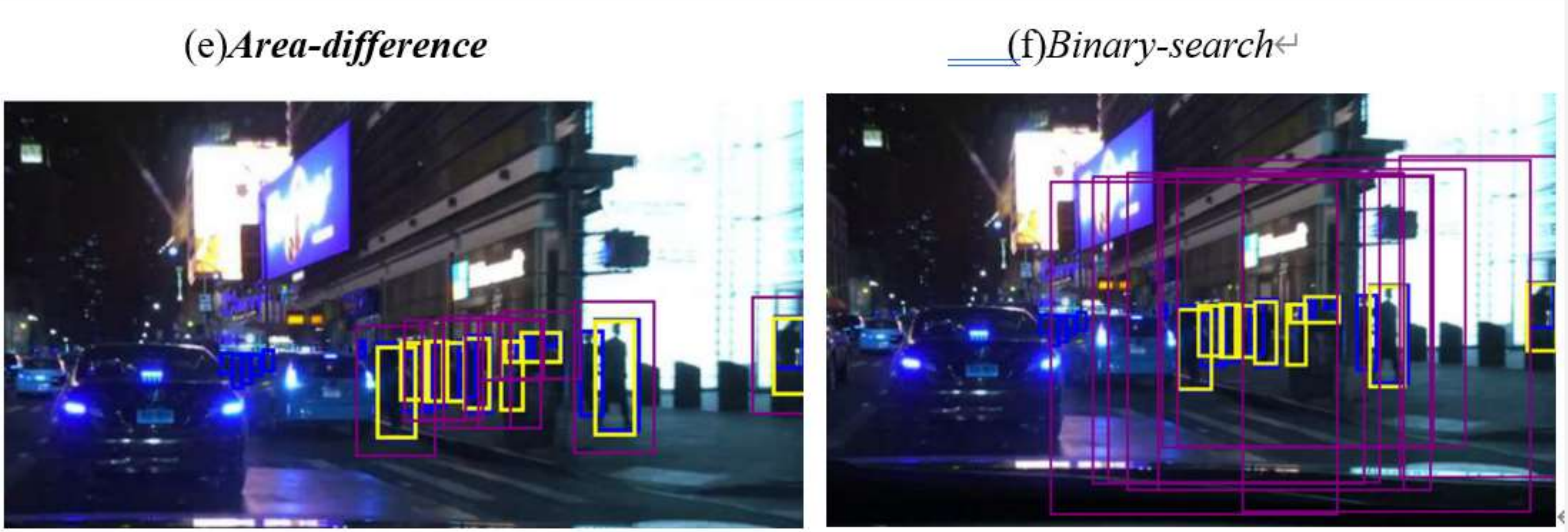


Results

Images of conformalized box

(*including 6 methods*, risk level $\alpha = 0.1$, pretrained model) on a BDD100k image

with Ground Truth, Inference and Conformalized boxes.



Analysis for Research Question in Coverage and Expansion

RQ1: What impact does a **fine-tuned** model have on the results of conformal prediction compared to a pretrained model?

For *Observed Coverage*:

Taking the methods *Coordinate-adaptation w/o Bonferroni* (0.45/0.55, 0.81/0.80, 0.89/0.90) and *Coordinate-adaptation with Bonferroni* (0.73/0.78, 0.91/0.93, 0.95/0.97) as examples: **The fine-tuned models exhibit the same or higher observed coverage rates compared to the pretrained models.**

Why? the fine-tuned models have higher precision, capable of recognizing objects that the pretrained models could not, thereby producing more conformalized boxes during the conformal prediction process, which can cover more real boxes.

For *Expansion*:

Considering *Coordinate-difference with Bonferroni* (de Grancey) (0.37/0.28, 0.45/0.35, 0.49/0.41), we observe that in most cases, **when the model's precision increases, the required Expansion decreases, and the generated conformalized boxes fit more closely to the predicted boxes.**

Why? fine-tuned model can more precisely locate each object, resulting in predicted boxes that closely match the actual boxes. Consequently, the calculated error is smaller, meaning that the conformalized box's margin is smaller, naturally leading to less expansion.

Analysis for Research Question in Coverage and Expansion

RQ2: Does ***Coordinate-Adaptation*** method provide better coverage and more accurately fit conformalized boxes to predicted boxes compared to the baseline *Coordinate-Difference* method?

For *Observed Coverage*:

with Bonferroni correction applied, the results for *Coordinate-Difference* (0.75/0.76, 0.92/0.91, 0.96/0.96) and *Coordinate-Adaptation* (0.73/0.78, 0.91/0.93, 0.95/0.97) **both ensuring strict probabilistic coverage** with almost no differences.

For *Expansion:*

the results for *Coordinate-Difference* with Bonferroni (de Grancey) are 0.37/0.28, 0.45/0.35, 0.49/0.41, whereas for *Coordinate-Adaptation* with Bonferroni (this paper) the results are 0.48/0.29, 0.53/0.36, 0.67/0.48. It is evident that at any preset coverage rate, ***Coordinate-Difference*** **with Bonferroni (de Grancey) performs better, requiring only 0.45/0.35 Expansion at a 0.9 coverage rate**, which closely aligns with the predicted boxes.

Of course, the Expansion for ***Coordinate-Adaptation*** **is not significantly high, ranking second** among all methods, with only a 0.1-0.2.As figures show, The conformalized boxes generated by *Coordinate-Adaptation* are only slightly larger.

Analysis for Research Question in Coverage and Expansion

RQ3: Can the ***Area-difference*** method designed in this paper capture the interdependence of coordinates, ensure box-wise coverage, and potentially replace the Bonferroni correction method used in the baseline?

For *Observed Coverage*:

It is noticeable that, although this method did not meet the preset probabilistic requirements for both pretrained and fine-tuned models—with results being 0.55/0.63, 0.80/0.83, 0.88/0.93—it still performed closer to the coverage requirements compared to *Coordinate-difference w/o Bonferroni* (0.41/0.39, 0.70/0.71, 0.78/0.80) and *Coordinate-adaptation w/o Bonferroni* (0.45/0.55, 0.81/0.80, 0.89/0.90). This suggests that the Area-difference method **can to some extent capture the interdependence of coordinates, though this connection is still not as strict as methods with Bonferroni correction applied.**

For *Expansion:*

Results in the table for *Area-difference (this paper)* are 0.52/0.44, 0.58/0.50, 0.78/0.70. Given that *Area-difference* almost ensures coverage, switching to a higher precision model could potentially meet the coverage requirement.. Among all methods that nearly meet or satisfy coverage requirements, **it ranks third**.

Analysis for Research Question in Coverage and Expansion

RQ4: How do the *Observed Coverage* and *Expansion* performance of the **Image-wise Binary-Search** method designed in this paper measure up?

For *Observed Coverage*:

Data from Table 1 shows that the *Image-Wise Binary-Search* achieves *Coverage* rates of 0.78/0.79, 0.93/0.94, and 0.95/0.96. Compared to other methods. **Binary-Search shows best performance in coverage**. Observing Figure, although the conformalized boxes are significantly expanded, they **indeed cover objects that the model failed to recognize** (the blue true boxes in the middle of figure), indirectly addressing a common issue with other methods that lack measures for unrecognized objects, thereby enhancing the actual coverage rate.

For *Expansion:*

The Image-Wise Binary-Search (de Grance) displays *Expansion* of 12.25/8.54, 15.64/10.94, and 22.64/16.84. These results indicate **a substantial expansion, nearly ten times the size of the predicted boxes**. This suggests **that to ensure coverage, some boxes must expand significantly**.

Analysis for Research Question in Coverage and Expansion

For example, The image-wise method requires that 75%

must be covered. To meet this requirement, the conformal

continuously expanded until it covers these not re

demands for coverage.

Therefore, the constraints posed by de Grancey's met

Coverage, more true boxes must be included in an image

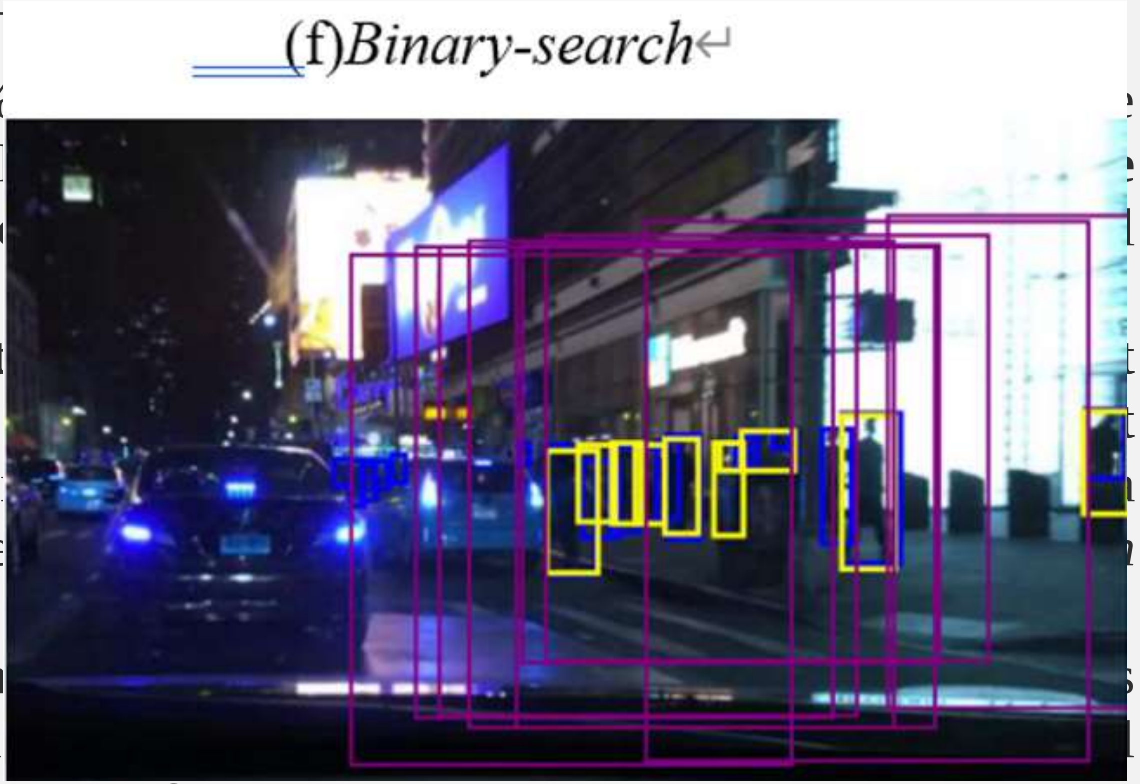
the image's edge, or if the model fails to detect an object e

lot to cover this true box to satisfy the probability of cove

metric to **explode**, particularly in image-wise cases.

Unlike box-wise, where coverage failures in difficult im

the average is maintained, image-wise strictly requires at l



boxes, leading to significant expansions. **Furthermore, because fine-tuned models can detect more objects, they require less Expansion.**

Analysis for Research Question in Coverage and Expansion

RQ5: Can *Expansion* metric effectively compare different nonconformity score functions that generate conformalized boxes? What is the relationship between this metric and another metric, *Observed Coverage*?

The *Expansion* metric serves as **an effective** indicator for measuring the size of conformalized boxes relative to predicted boxes. the *Coordinate-difference* with Bonferroni method (0.37/0.28, 0.45/0.35, 0.49/0.41), and the Coordinate-adaptation with Bonferroni, in this paper (0.48/0.29, 0.53/0.36, 0.67/0.48), showing that after calculating the expansion, slight differences are noticeable.

Connection between the metrics of Expansion and *Observed Coverage*, **the former complements the latter**. For a method that does not meet the specific coverage requirements, such as *Coordinate-difference without Bonferroni* (0.15/0.17, 0.32/0.27, 0.38/0.32), even if its conformalized boxes align very well with the predicted boxes, it is meaningless since ensuring specific *Coverage* is a critical aspect that distinguishes conformal prediction from other uncertainty quantification methods.

Moreover, as shown in table 3 with the example of Coordinate-difference with Bonferroni (0.37/0.28, 0.45/0.35, 0.49/0.41), as the specific coverage requirements increase**, *Expansion* shows a positive correlation with increasing specific *Coverage*.** This happens because as we impose stricter conditions, without changing the model accuracy, to cover more predicted boxes, the conformalized boxes must be enlarged.

Conclusion

The discrepancy between results from fine-tuned and pretrained models shows that **model precision significantly impacts *Observed Coverage***. Conformal prediction cannot replace the model training and fine-tuning process; objects which are not recognized by model and hence without predicted bounding boxes. It is difficult to cover unrecognized objects using traditional conformal prediction methods, leading to reduced coverage.

For objects that are not detected, how to cover them with conformalized boxes remains a challenge. The *image-wise Binary-Search* method introduced in this paper serves as an exploratory attempt. Although the method's Expansion is excessively high, it successfully identifies undetected objects. Future research could explore relaxing the overly strict probabilistic requirements of de Grancey et al. to reduce the method's Expansion.

Furthermore, **designing the nonconformity score function is absolutely a core task to conformal prediction**, as it pertains to the handling of all known information. Current methods are designed from an intuitive graphical perspective; hence, future research could explore deriving this function from internal model outputs or incorporating probabilistic computations.