Proceedings of Machine Learning Research 266:1–4, 2025 14th Symposium on Conformal and Probabilistic Prediction with Applications

Sequential Conformal Risk Control for Safe Railway Signaling Detection

Léo Andéol LEO.ANDEOL@MATH.UNIV-TOULOUSE.FR
Institute of Mathematics of Toulouse, University of Toulouse, Toulouse, France
DTIPG, SNCF, Saint-Denis, France

Thomas Masséna THOMAS.MASSENA@IRIT.FR
Institut de Recherche en Informatique de Toulouse, University of Toulouse, Toulouse, France
DTIPG, SNCF, Saint-Denis, France

Editor: Khuong An Nguyen, Zhiyuan Luo, Harris Papadopoulos, Tuwe Löfström, Lars Carlsson and Henrik Boström

Abstract

As machine learning becomes a more common tool in industry, its needs for certification increase. Conformal Prediction (Vovk et al., 2005), a framework for construction of prediction sets with tight coverage guarantees at any desired error rate, is an ideal tool for this purpose. However, adapting conformal methods to complex computer vision pipelines and providing appropriate guarantees is still a challenging task. Indeed, conformal approaches to object detection are often restricted to subtasks: often localization, and sometimes classification. In this study, we apply the comprehensive framework from Andéol et al. (2025) to the safety-critical task of railway signaling detection.

Keywords: Conformal Risk Control, Sequential Conformal Risk Control, Object Detection, Railway Signaling

1. Introduction

Conformal Prediction has been extensively studied in many classical tasks such as regression in Papadopoulos et al. (2002) and classification in Romano et al. (2020); Sadinle et al. (2019). Furthermore, industrial interest has recently motivated conformal applications in complex computer vision settings, such as object detection in Andeol et al. (2023); Timans et al. (2025) and semantic segmentation Mossina et al. (2024); Brunekreef et al. (2024). Recently, Andéol et al. (2025) proposed a comprehensive approach to apply conformal prediction to the complete object detection procedure applicable to any state-of-the-art model, including YOLO (Redmon et al., 2016) and DETR (Carion et al., 2020). In this work, we apply this method to the global task of railway signal detection, expanding on the previous work of Andeol et al. (2023) that only covered the localization of objects.

2. Methodology

The object detection task commonly consists of three subtasks: confidence thresholding, localization and classification. We use the SeqCRC (Sequential Conformal Risk Control) of Andéol et al. (2025), an extension of Conformal Risk Control (Angelopoulos et al., 2024) that allows to control risks of multiple parameters chosen sequentially, as required in conformal object detection.

More formally, we consider a deterministic object detector $f : \mathcal{X} \to (\mathcal{B} \times \Sigma^{K-1} \times [0,1])^{N_{\text{pred}}}$ that outputs N_{pred} predictions, constituted of a bounding box in $\mathcal{B} = \mathbb{R}_+^4$, a

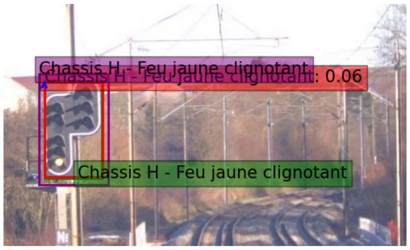

Figure 1: An image showing the conformalized bounding box in purple, the prediction in red, and the ground truth in green.

		Confidence		Localization		Classification		Global
Model	mAP	SS	Risk	SS	Risk	SS	Risk	Risk
YOLO11x	0.879	1.29	0.000	1.25	0.004	1.01	0.001	0.0043
YOLO12x	0.895	1.27	0.001	1.21	0.005	**1.00**	0.002	0.0051
YOLO12x+	**0.911**	**1.21**	0.001	**1.18**	0.005	**1.00**	0.001	0.0052
YOLO12x++	0.887	1.24	0.000	1.24	0.004	1.01	0.001	0.0049

Table 1: SeqCRC performance on object detection models. Set sizes (SS) indicate average set size for confidence filtering and classification, and average stretch in localization. All risks are controlled at target levels. Bold is best.

classification distribution over K classes, and a confidence score. After Non-Maximum Suppression, we construct three nested prediction sets: confidence $\Gamma^{\text{cnf}}_{\lambda^{\text{cnf}}}(x)$, localization $\Gamma^{\text{loc}}_{\lambda^{\text{cnf}},\lambda^{\text{loc}}}(x)$, and classification $\Gamma^{\text{cls}}_{\lambda^{\text{cnf}},\lambda^{\text{cls}}}(x)$, parameterized by $\lambda^{\text{cnf}} \in \Lambda^{\text{cnf}}$, $\lambda^{\text{loc}} \in \Lambda^{\text{loc}}$, and $\lambda^{\text{cls}} \in \Lambda^{\text{cls}}$ respectively. We assume that $(X_1, Y_1), \ldots, (X_n, Y_n), (X_{\text{test}}, Y_{\text{test}})$ are i.i.d., where each Y_i represents the set of ground-truth objects on image X_i. The corresponding loss functions $L^{\text{cnf}}(\lambda^{\text{cnf}})$, $L^{\text{loc}}(\lambda^{\text{cnf}}, \lambda^{\text{loc}})$ and $L^{\text{cls}}(\lambda^{\text{cnf}}, \lambda^{\text{cls}})$ are assumed to be $[0, 1]$-valued, non-increasing in each parameter and right-continuous. We then obtain the following guarantee:

Theorem 1 (Theorem 2 from Andéol et al. (2025)) *Let $\alpha^{\text{cnf}} \geq 0$ and $\alpha^{\text{loc}}, \alpha^{\text{cls}} \geq \alpha^{\text{cnf}} + \frac{1}{n+1}$. Then, if the aforementioned assumptions hold true, and if $L_i^{\text{cnf}}(\bar{\lambda}^{\text{cnf}}) \leq \alpha^{\text{cnf}}$ almost surely, we have that the parameters λ^{cnf}, λ^{loc} and λ^{cls} are well defined and*

$$\mathbb{E}\left[L^{\text{cnf}}_{\text{test}}(\lambda^{\text{cnf}}_{+})\right] \leq \alpha^{\text{cnf}}, \quad \mathbb{E}\left[\max\left\{L^{\text{loc}}_{\text{test}}(\lambda^{\text{cnf}}_{+}, \lambda^{\text{loc}}_{+}), L^{\text{cls}}_{\text{test}}(\lambda^{\text{cnf}}_{+}, \lambda^{\text{cls}}_{+})\right\}\right] \leq \alpha^{\text{tot}} = \alpha^{\text{loc}} + \alpha^{\text{cls}}.$$

3. Experiments

In these experiments we use YOLOv11 and YOLOv12 from Jocher et al. (2023); Tian et al. (2025). Models used a 640x640 resolution by default, replaced by 1080x1080 or 1280x1280 for YOLOv12+ and YOLOv12++ respectively. Then, we conduct our work using the TASV railway signaling dataset. It includes 37 classes and 12925 images, and is split as follows: 10% for validation, 10% for calibration and 10% for testing. The model training is conducted on the remaining 70% of the dataset. We set $\alpha^{\text{cnf}} = 3{\cdot}10^{-3}$, $\alpha^{\text{loc}} = 5{\cdot}10^{-3}$ and $\alpha^{\text{cls}} = 5{\cdot}10^{-3}$ for a global $\alpha^{\text{tot}} = 10^{-2}$. Following Andéol et al. (2025), we use the box_count_recall loss for confidence and the pixelwise loss for localization. The prediction sets are multiplicative for localization and employ Least Ambiguous Classifiers (LAC, Sadinle et al. (2019)) for the classification task.

4. Discussion

Our results demonstrate the successful application of Sequential Conformal Risk Control to railway signaling detection. All models achieve risk control below the strict target levels ($\alpha^{\text{tot}} = 10^{-2}$), with YOLO12x+ achieving the best balance of accuracy (mAP = 0.911) and efficiency (minimal set sizes). Future work will aim for even lower error rates by enhancing the base models, incorporating temporal information from video, and designing domain-specific conformal losses and sets tailored to railway safety requirements.

References

Leo Andeol, Thomas Fel, Florence de Grancey, and Luca Mossina. Confident object detection via conformal prediction and conformal risk control: an application to railway signaling. In *Proceedings of the Twelfth Symposium on Conformal and Probabilistic Prediction with Applications*, volume 204, pages 36–55. PMLR, 2023. URL https://proceedings.mlr.press/v204/andeol23a.html.

Léo Andéol, Luca Mossina, Adrien Mazoyer, and Sébastien Gerchinovitz. Conformal object detection by sequential risk control. *arXiv preprint arXiv:2505.24038*, 2025.

Anastasios Nikolas Angelopoulos, Stephen Bates, Adam Fisch, Lihua Lei, and Tal Schuster. Conformal risk control. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=33XGfHLtZg.

Joren Brunekreef, Eric Marcus, Ray Sheombarsing, Jan-Jakob Sonke, and Jonas Teuwen. Kandinsky conformal prediction: efficient calibration of image segmentation algorithms. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4135–4143, 2024.

Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *ECCV 2020*, pages 213–229. Springer, 2020.

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

Luca Mossina, Joseba Dalmau, and Léo Andéol. Conformal semantic image segmentation: Post-hoc quantification of predictive uncertainty. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3574–3584, 2024.

Harris Papadopoulos, Kostas Proedrou, Volodya Vovk, and Alex Gammerman. Inductive confidence machines for regression. In *Machine Learning: ECML 2002*, 2002.

Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE CVPR*, June 2016.

Yaniv Romano, Matteo Sesia, and Emmanuel Candes. Classification with valid and adaptive coverage. In *Advances in Neural Information Processing Systems*, volume 33, pages 3581–3591, 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/244edd7e85dc81602b7615cd705545f5-Paper.pdf.

Mauricio Sadinle, Jing Lei, and Larry Wasserman. Least ambiguous set-valued classifiers with bounded error levels. *Journal of the American Statistical Association*, 114(525): 223–234, 2019. doi: 10.1080/01621459.2017.1395341. URL https://doi.org/10.1080/01621459.2017.1395341.

Yunjie Tian, Qixiang Ye, and David Doermann. Yolov12: Attention-centric real-time object detectors. *arXiv preprint arXiv:2502.12524*, 2025.

Alexander Timans, Christoph-Nikolas Straehle, Kaspar Sakmann, and Eric Nalisnick. Adaptive bounding box uncertainties via two-step conformal prediction. In Aleš Leonardis, Elisa Ricci, Stefan Roth, Olga Russakovsky, Torsten Sattler, and Gül Varol, editors, *Computer Vision – ECCV 2024*, pages 363–398, Cham, 2025. Springer Nature Switzerland.

Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Algorithmic learning in a random world*. Springer, 2005.