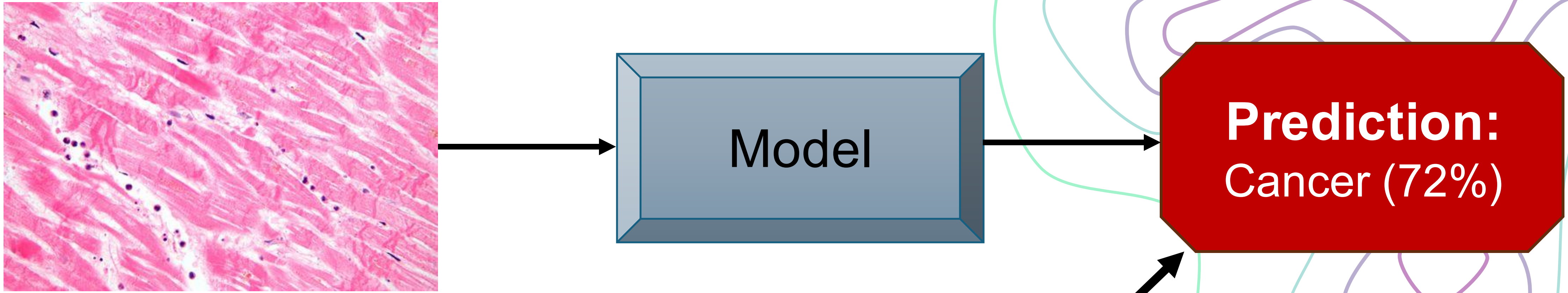
Conformal Object Detection by Sequential Risk Control

Léo Andéol – October 17th 2025

Institute of Mathematics of Toulouse - SNCF

Problem

- „All models are wrong”: Neural Networks make mistakes, predictions are uncertain guesses

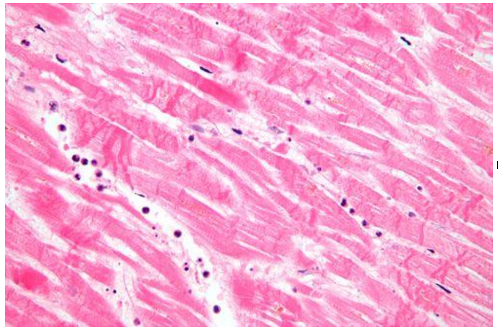


Is this certain ?

Confidence is required before taking action

Conformal Prediction I

- Conformal Prediction is a <u>Post-hoc</u> framework for <u>guaranteed</u> uncertainty quantification



Model → **Prediction:** Cancer (72%) → Conformal Prediction → **Trustworthy Predictions:** { Cancer, **Heart Attack** }

It works for <u>any model</u>

Including black-box neural networks!

-> even pre-trained!

It produces prediction sets <u>guaranteed</u> to contain the <u>ground truth</u> with <u>**arbitrarily high probability**</u>!

Conformal Prediction II

At what cost?

Computational:

Negligible

(for post-hoc „Split" CP, others require re-training)

Data:

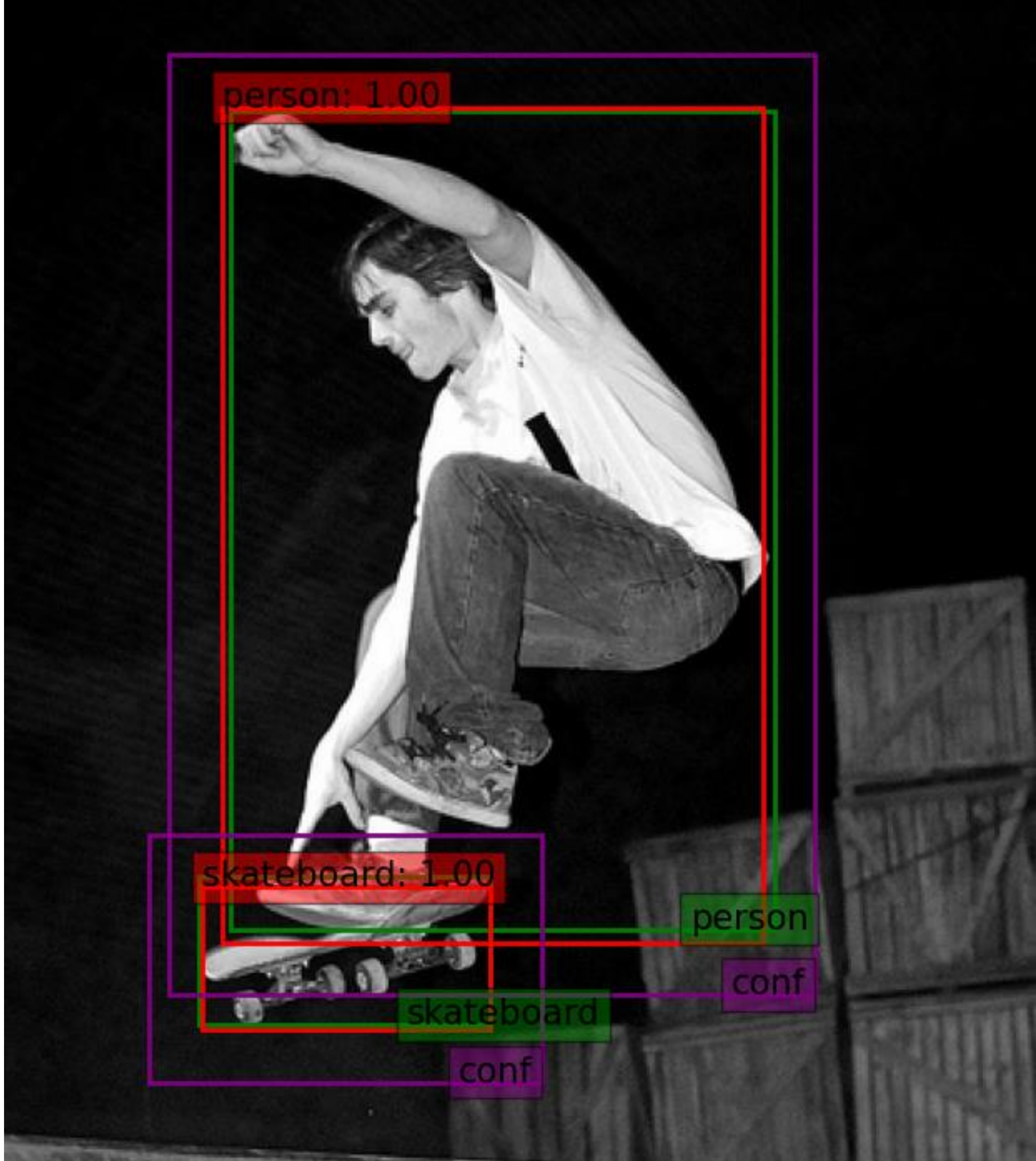
Requires an additional data-split for **calibration**.

Assumptions:

- **None** on the distribution
- **None** on the model
- Only requires **exchangeable / i.i.d. data**

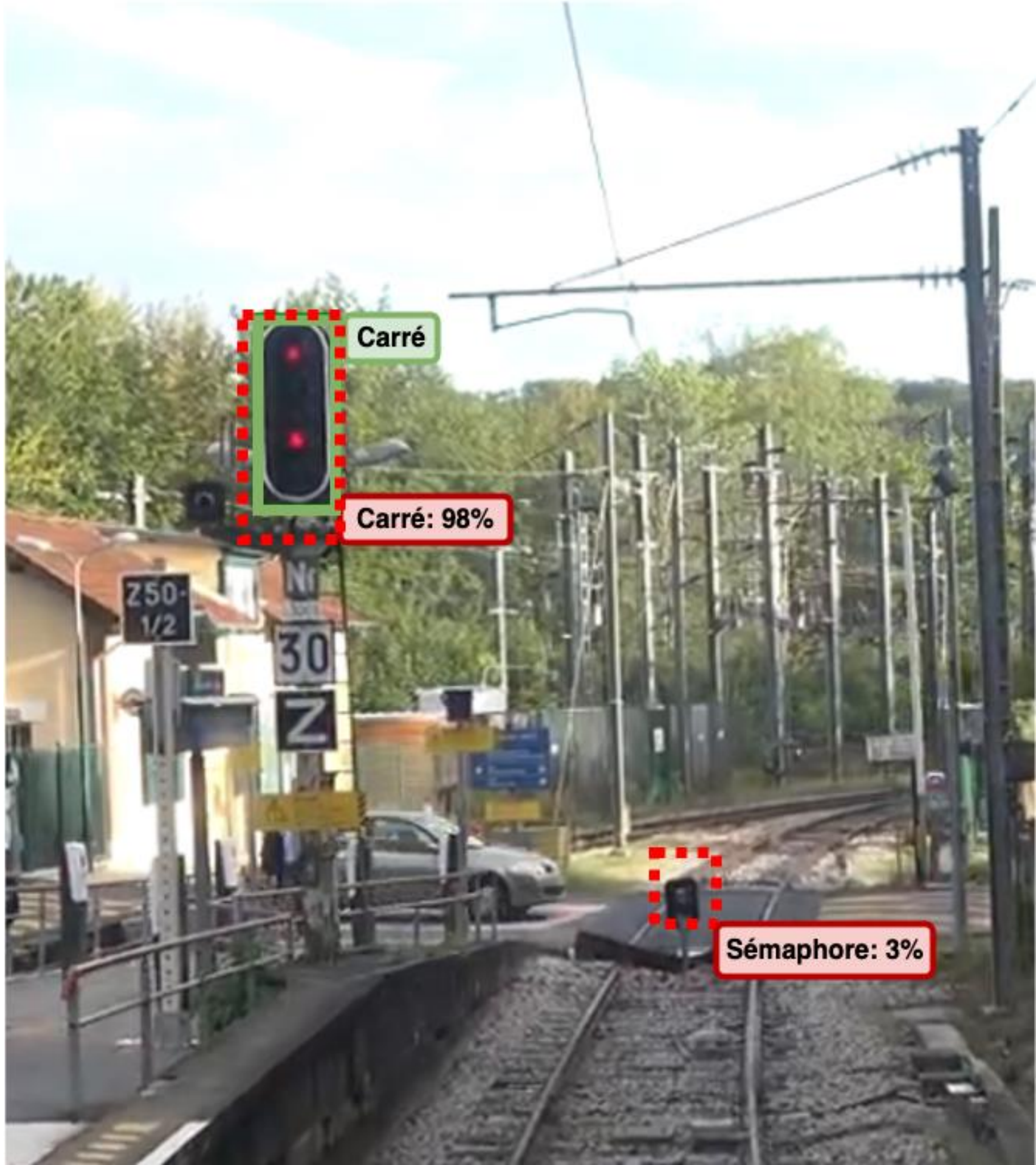
Conformal Object Detection

With Sequential Risk Control



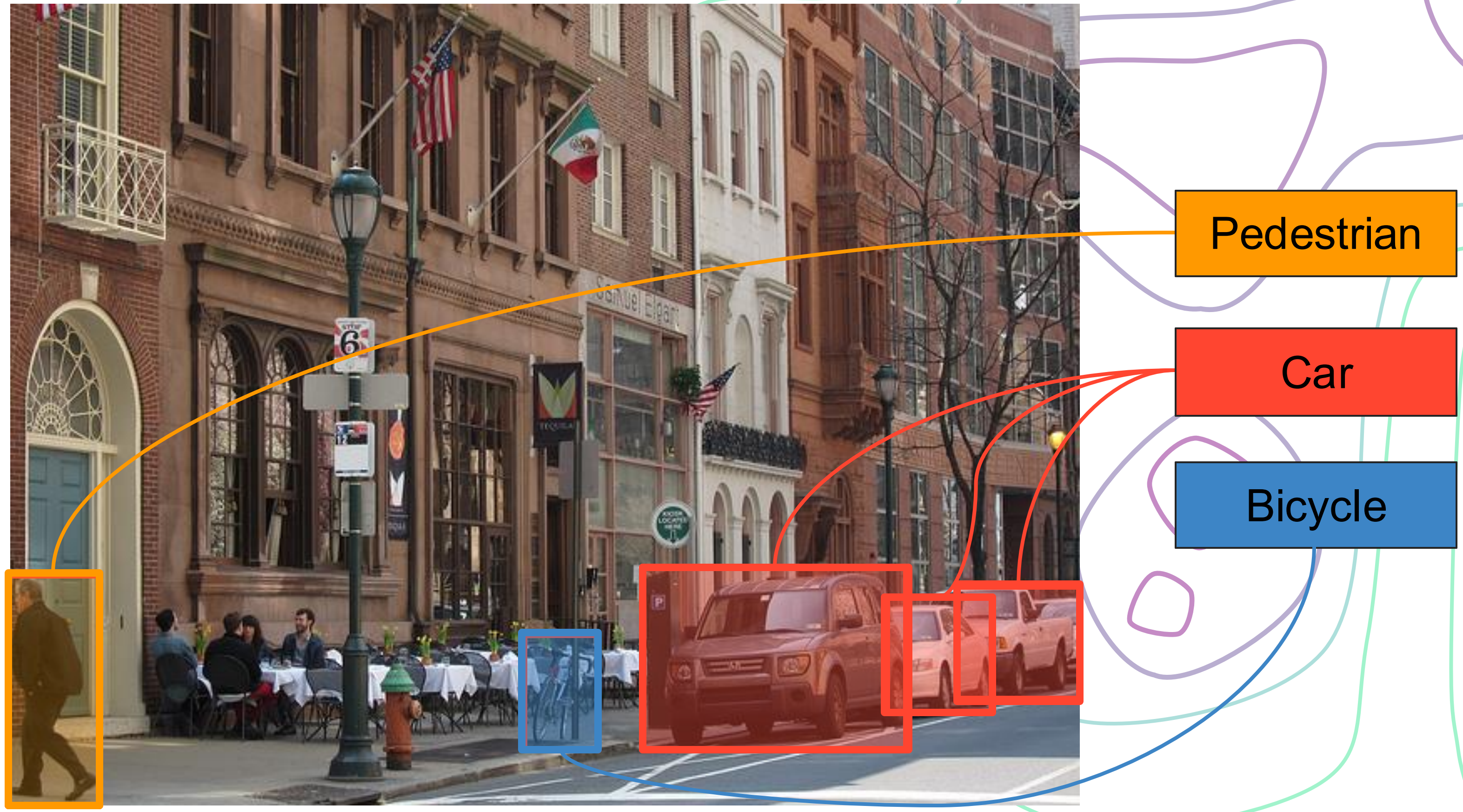
Industrial Objective: **Safe** Railway Signal Detection

- Why ?
  - Safety
  - Availability
  - Towards Autonomous Trains
- How ?
  - Using Neural Networks !



Object Detection: Object Localization & Classification

- Based on Deep Learning Models
- Combines Classification & Localization tasks for an unknown amount of objects



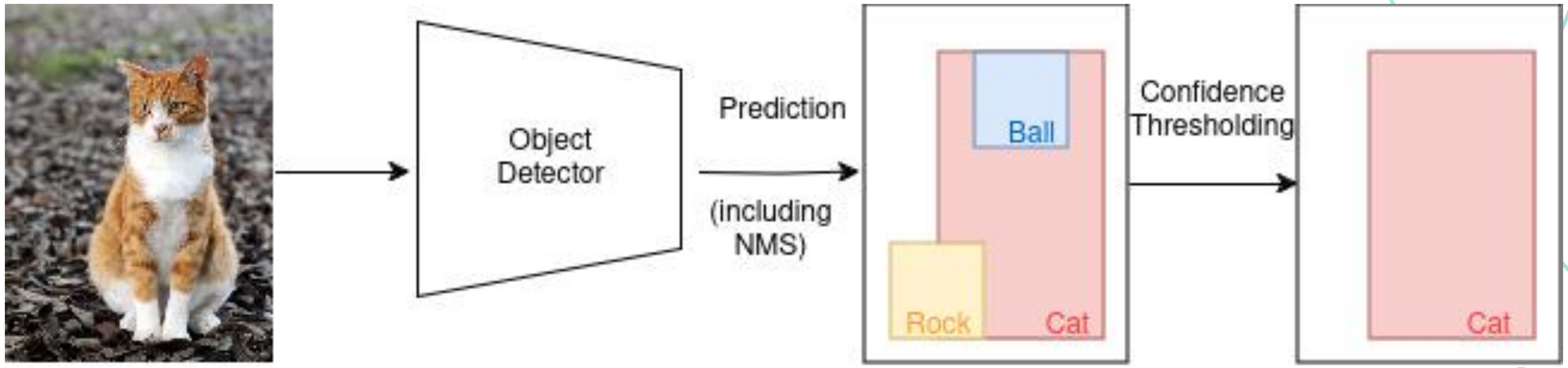
Andéol et al. (2023). Confident Object Detection, COPA 2023. PMLR 204:36-55

Object Detection

- Locate any target object via bounding boxes (regression)
- Classify its content (multi class classifications)
- Select the correct number of objects

(Previous Works)

(Our work)



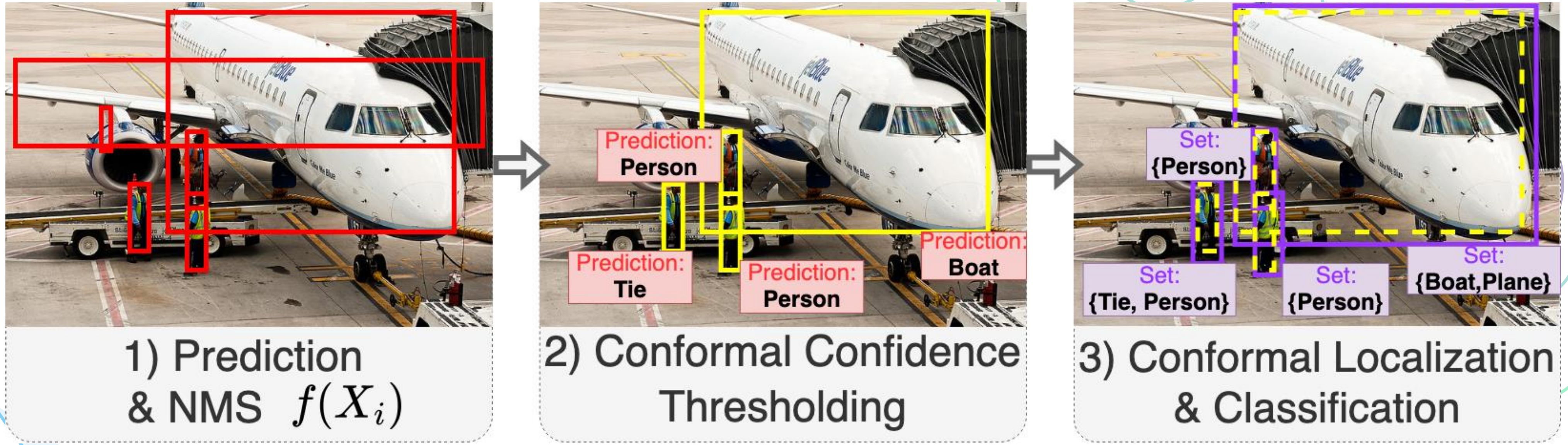
We propose a new theory for Sequential Conformal Risk Control, necessary to unlock the dependence here!

Process: Visualization

Two stages:

1. Confidence Threshold Calibration

2. Simultaneously:
   1. Localization Margin Calibration
   2. Classification Threshold Calibration



Sequential Conformal Risk Control

1. Choose Confidence, Localization and Classification losses
2. Choose target error rates $\alpha^{\text{tot}} = \alpha^{\text{loc}} + \alpha^{\text{cls}}$
3. Compute Confidence Threshold

$$\lambda_+^{\text{cnf}} = \inf\left\{\lambda^{\text{cnf}} \in \Lambda^{\text{cnf}} : \frac{n\tilde{R}_n^{\text{cnf}}(\lambda^{\text{cnf}})}{n+1} + \frac{\tilde{B}^{\text{cnf}}}{n+1} \leq \alpha^{\text{cnf}}\right\}$$

4. Compute Localization Margin and Classification Threshold

$$\lambda_+^{\bullet} = \inf\left\{\lambda^{\bullet} \in \Lambda^{\bullet} : \frac{nR_n^{\bullet}(\lambda_-^{\text{cnf}}, \lambda^{\bullet})}{n+1} + \frac{B^{\bullet}}{n+1} \leq \alpha^{\bullet}\right\}$$

SeqCRC: Guarantee

Theorem:

Assume the calibration and test sample is drawn i.i.d. Then, we have

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

This guarantee holds **on average** over the calibration set and test example but is:

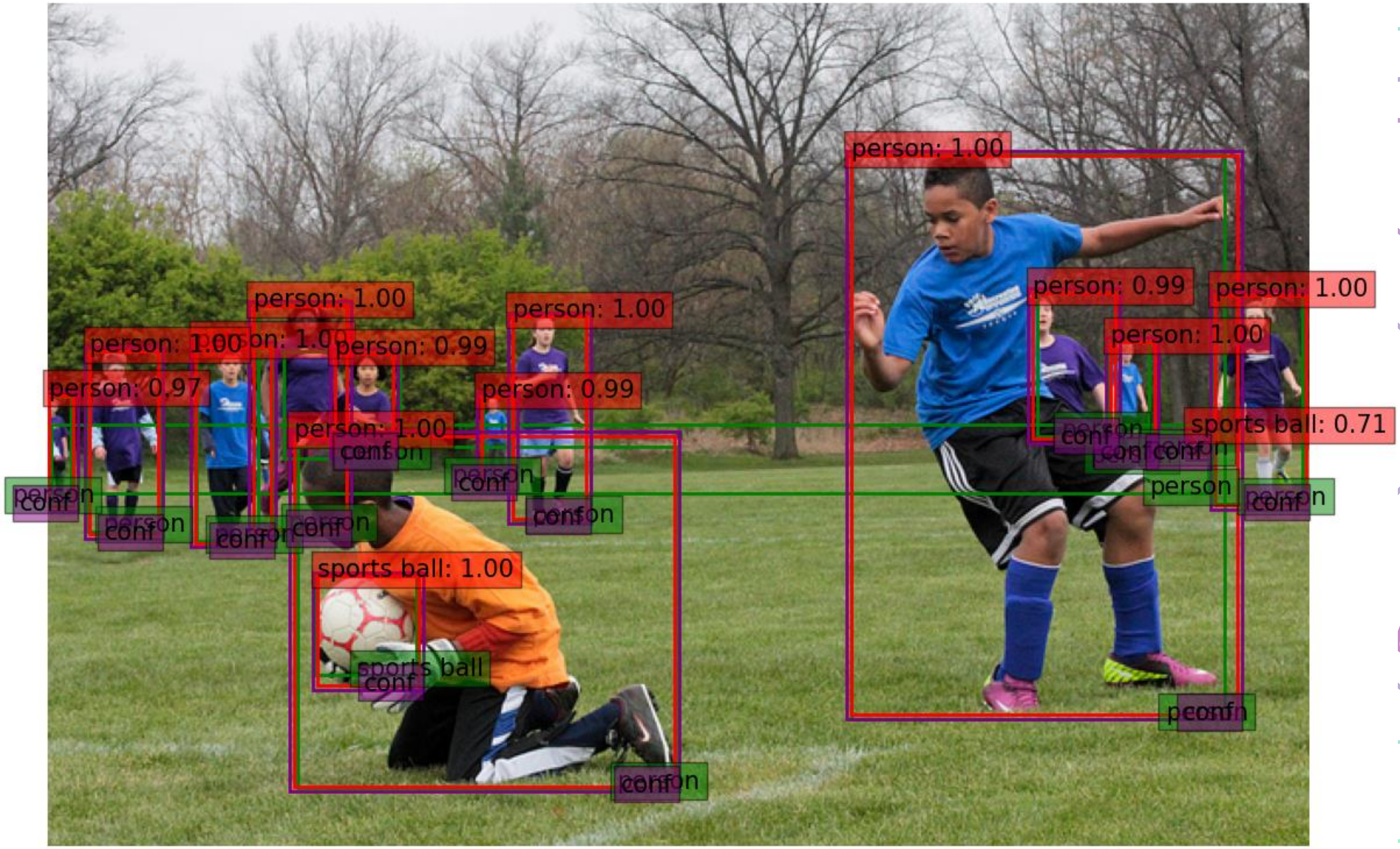
1. Distribution-free (valid for any distribution $P_{X,Y}$)
2. Finite-sample (valid for any $n_{\text{cal}} \geq 1/\alpha - 1$)
3. Model-agnostic (valid for any model f : pretrained, black-box, …)

Application : MSCOCO

- We work with the **validation split** (5k images) of the COCO 2017 dataset (Lin et al., 2014), with 80 classes, which we split into:
  - 2500 images for calibration
  - 2500 images for test (we cannot use the unlabelled COCO test data)

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

Results on MSCOCO: Correct Detections



Results on MSCOCO: Table

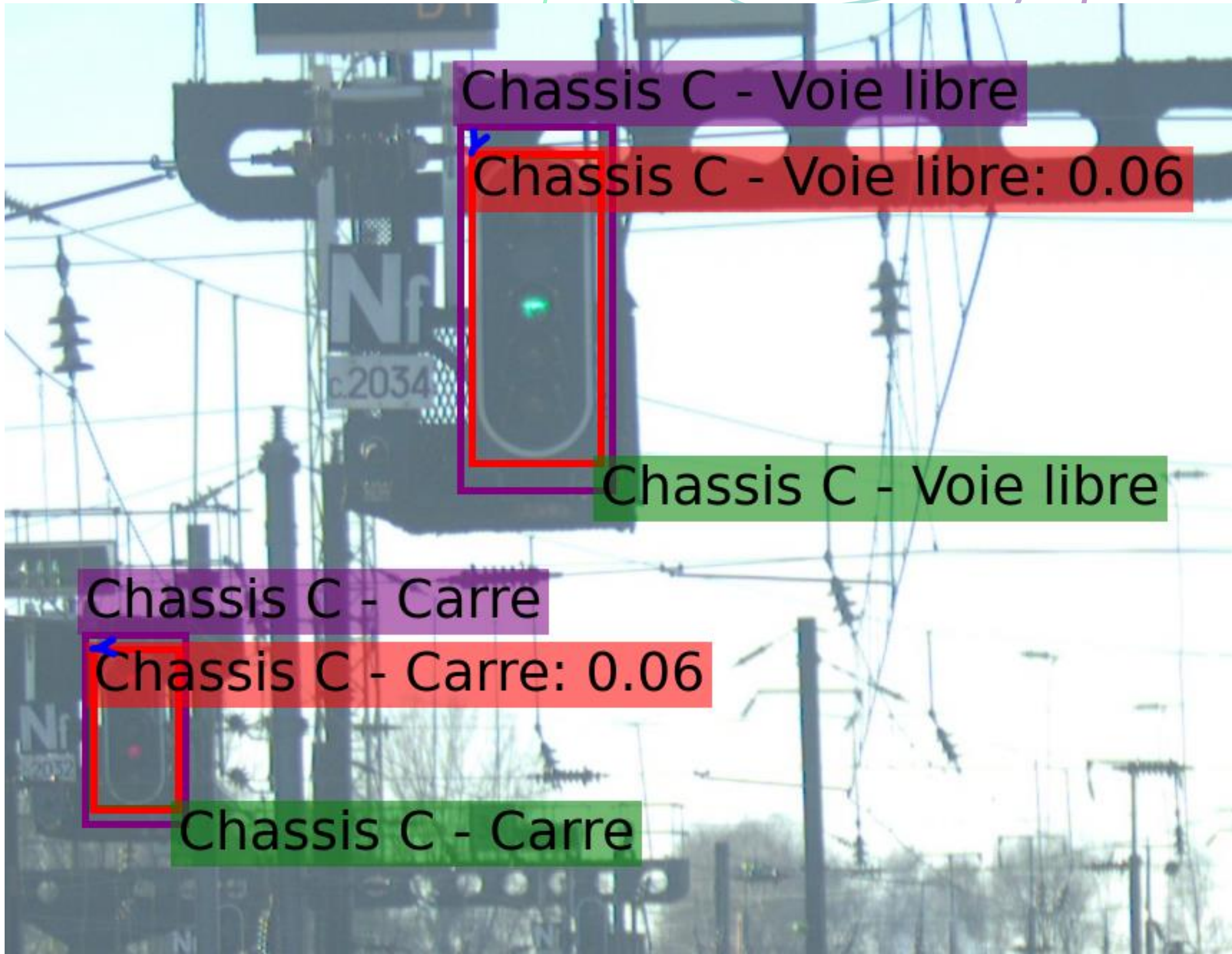
		Setting	Set Size	Task Risk	Global Risk
Loss	Cnf.	box_count_threshold	26.379	0.019	0.084
		box_count_recall	18.868	0.018	0.085
	Loc.	thresholded	1.626	0.045	0.097
		boxwise	1.578	0.049	0.098
		pixelwise	1.056	0.051	0.100
Pred. Sets	Loc.	additive	1.094	0.051	0.099
		multiplicative	1.056	0.051	0.100
	Cls.	aps	1.072	0.051	0.083
		lac	1.044	0.051	0.083

Results on MSCOCO: Issues

- Difficult to predict accurately a large number of objects
- No false positive penalty
- Many sources of error:
  - Detection
  - Annotation
  - Loss functions

Result on Railway Signaling Detection

- Guaranteed Global error rate: 1%
- Measured Error rate on Test: .5% (average)
- Average 18% increase of bounding box size
- Average 1.2 predictions per image
- Average 1.0 class(es) per set.



Thank you !

References

- Articles
  - Andéol et al., « Sequential Conformal Risk Control for Safe Railway Signaling Detection», COPA 2025 (PMLR)
  - Andéol et al., « Conformal Object Detection by Sequential Risk Control», Preprint
  - Andéol et al., « Confident Object Detection via Conformal Prediction and Conformal Risk Control: An Application to Railway Signaling », COPA 2023 (PMLR)
  - Angelopoulos et al., « Conformal Prediction: A Gentle Introduction », Found. And Trends in ML 2023
  - Vovk, « Conditional validity of inductive conformal predictors », ACML 2012
  - Angelopoulos et al., « Conformal Risk Control », ICLR 2024 Spotlight
  - Gerchinovitz, « Conformal prediction for object detection », Talk @ Stats du Sud 2024
- Libraries
  - PUNCC : DEEL Conformal Prediction Library : https://github.com/deel-ai/puncc
  - CODS : Conformal Object Detection & Segmentation : https://github.com/leoandeol/cods