# Bayesian Detector Combination for Object Detection with Crowdsourced Annotations

Zhi Qin Tan, Olga Isupova, Gustavo Carneiro, Xiatian Zhu, Yunpeng Li
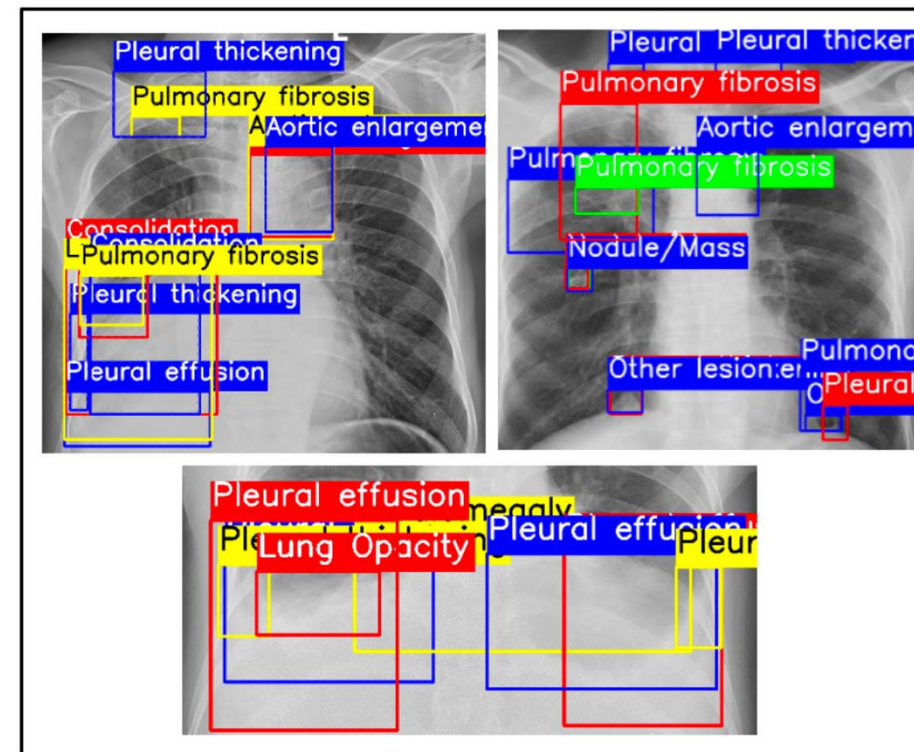
Poster Session: 1st October 2024, 10:30 – 12:30

Noisy Crowdsourced Object Annotations

- Often difficult and expensive to obtain accurate annotations.

- High disagreements observed in complex domains due to high interobserver variability; challenging to achieve consensus.
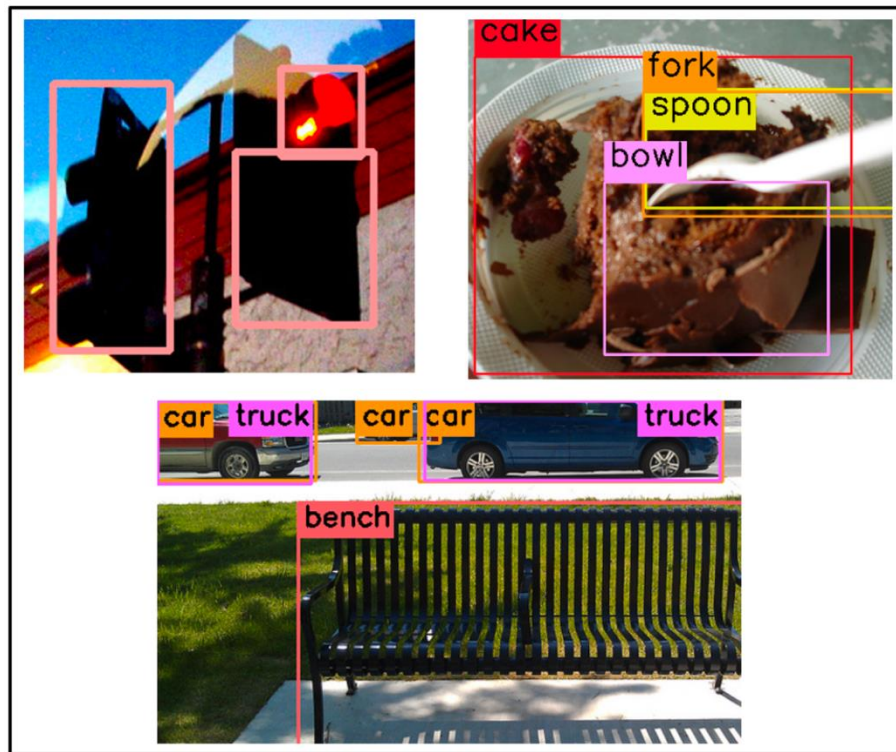
Noisy annotations in MSCOCO

Disagreements in VinDr-CXR

Limitations of Existing Solutions

Algorithmic limitations:

- Majority voting: Assumes equal annotator annotation accuracy;

- Crowd R-CNN [1]: Not generalisable to other object detectors;

- WBF-EARL [2]: Requires annotators' proficiency levels.

Evaluation limitation:

- Prior works used private synthetic crowdsourced datasets constructed under different setups;

- Cannot compare their results directly.

[1] Hu and Meina. Crowd R-CNN: An object detection model utilizing crowdsourced labels. In *ICVISP*, 2020.
[2] Le et al. Learning from multiple expert annotators for enhancing anomaly detection in medical image analysis. *IEEE Access*, 11, 2023.

Bayesian Detector Combination (BDC)

Legend:
- → Forward Pass
- → Object Detector Update
- → Aggregated Annotations
- → Aggregators Update
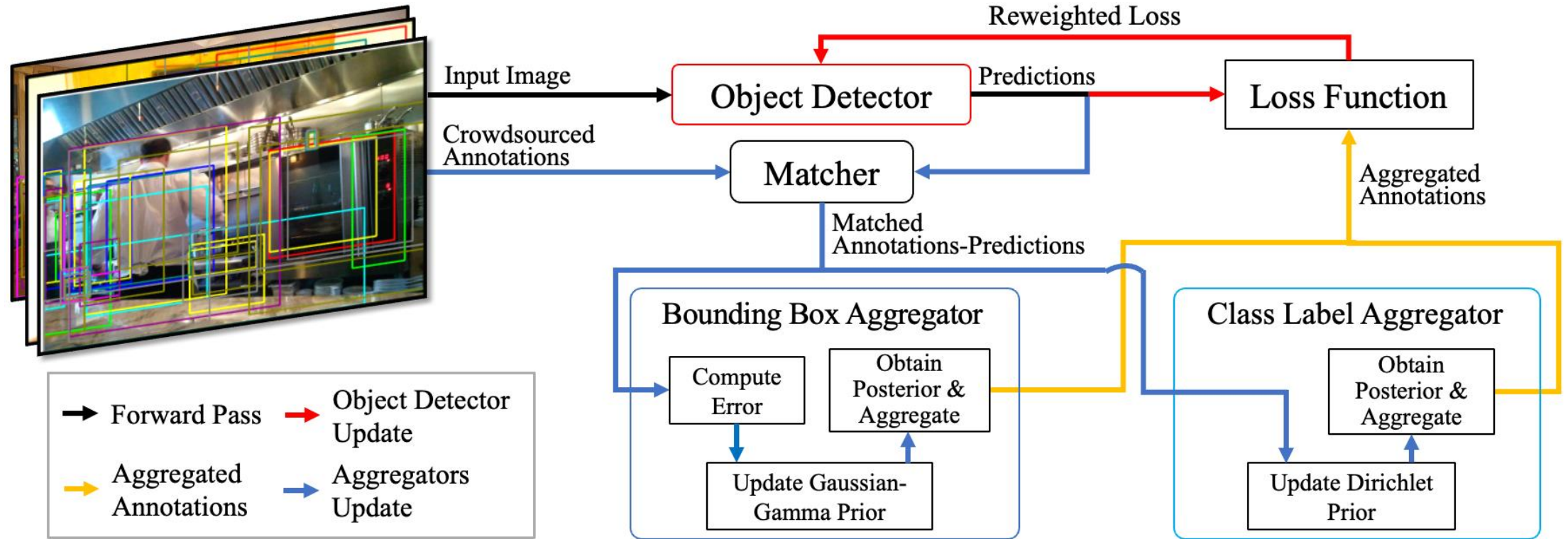
Model-agnostic framework to simultaneously infer:
1. the annotation quality of each annotator,
2. the consensus bounding boxes,
3. and soft labels

from noisy crowdsourced object annotations *without any additional inputs*.

Experiments and Results

1. Real-world datasets: VinDr-CXR: thoracic abnormalities dataset annotated by 17 expert radiologists.

2. Synthetic datasets: Simulate various crowdsourcing scenarios with VOC and MSCOCO dataset.

VinDr-CXR

Method	Test AP$^{.4}$		
	YOLOv7	FRCNN	EVA
NA	17.4	17.2	7.8
MV	13.9	16.3	8.2
Crowd R-CNN [1]	-	16.7	-
WBF-EARL [2]	16.4	17.0	8.4
BDC (ours)	**19.2**	**17.9**	**8.9**

COCO-FULL: 10 synthetic annotators with varying accuracy

Method	Test AP$^{.5}$		
	YOLOv7	FRCNN	EVA
NA	53.4	39.7	71.8
MV	61.9	55.6	74.8
Crowd R-CNN [1]	-	48.5	-
WBF-EARL [2]	55.6	51.9	74.7
BDC (ours)	**65.0**	**56.6**	**78.0**

[1] Hu and Meina. Crowd R-CNN: An object detection model utilizing crowdsourced labels. In *ICVISP*, 2020.

[2] Le et al. Learning from multiple expert annotators for enhancing anomaly detection in medical image analysis. *IEEE Access*, 11, 2023.

Experiments and Results

BDC **scales well** with the number of annotators and **is robust** to the percentage of noisy annotators with poor reliability.
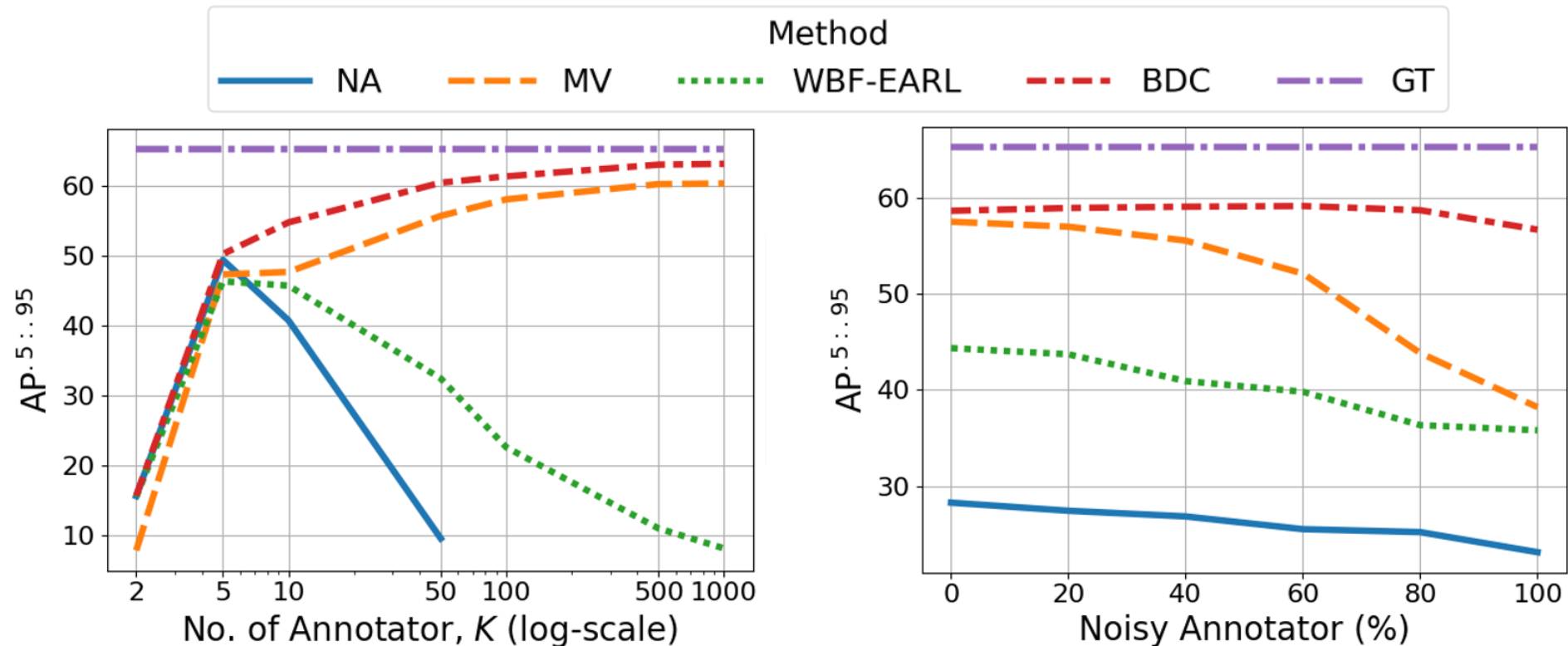
Thank you!

Code & Dataset
Available at:

https://t.ly/fDxrP