Bayesian Detector Combination for Object Detection with Crowdsourced Annotations

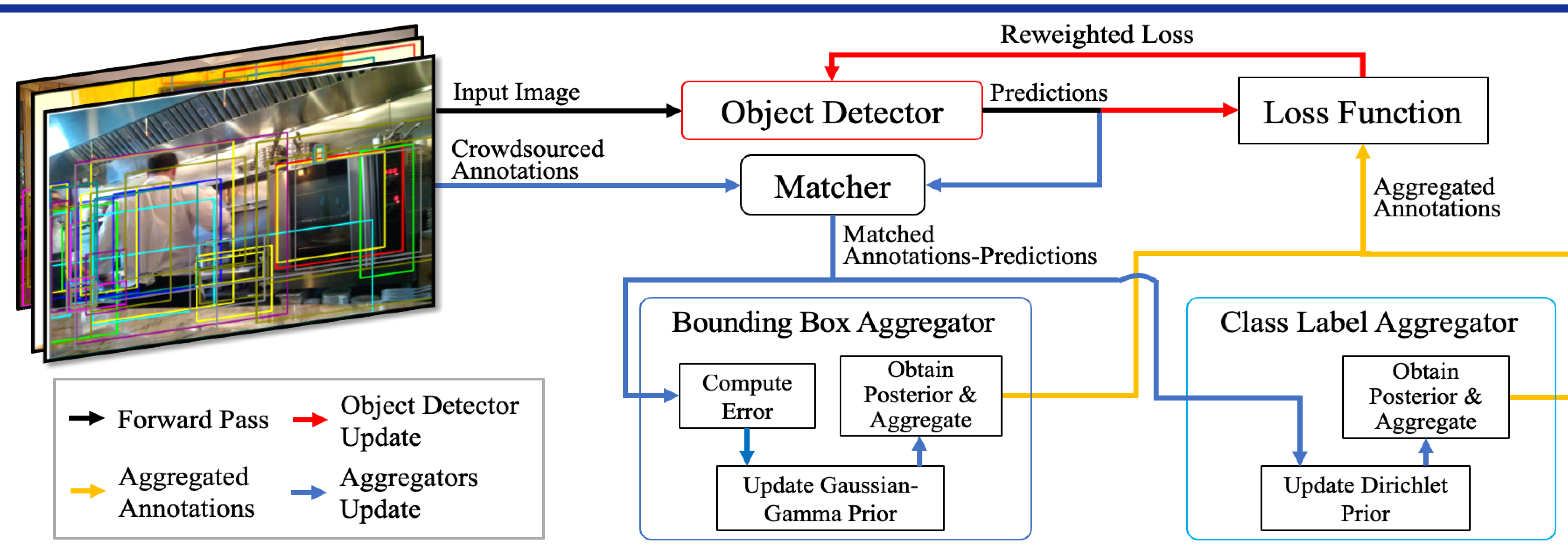
UNIVERSITY OF SURREY

UNIVERSITY OF OXFORD

Zhi Qin Tan, Olga Isupova, Gustavo Carneiro, Xiatian Zhu, and Yunpeng Li

KING'S College LONDON

Code & Dataset Available at:



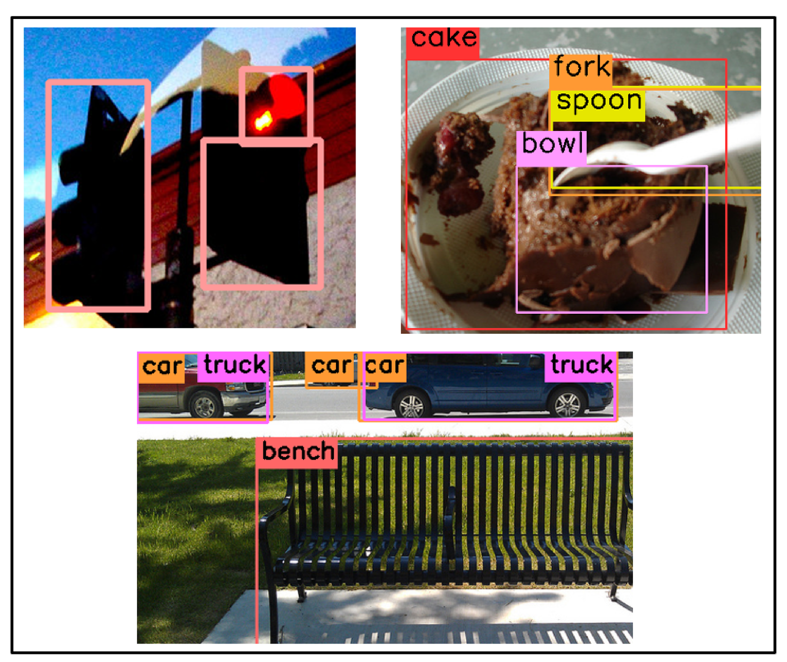
Our Contributions

- Proposed Bayesian Detector Combination (BDC), a ***model-agnostic*** framework to simultaneously infer:
  1. the annotation quality of each annotator,
  2. the consensus bounding boxes,
  3. and soft labels

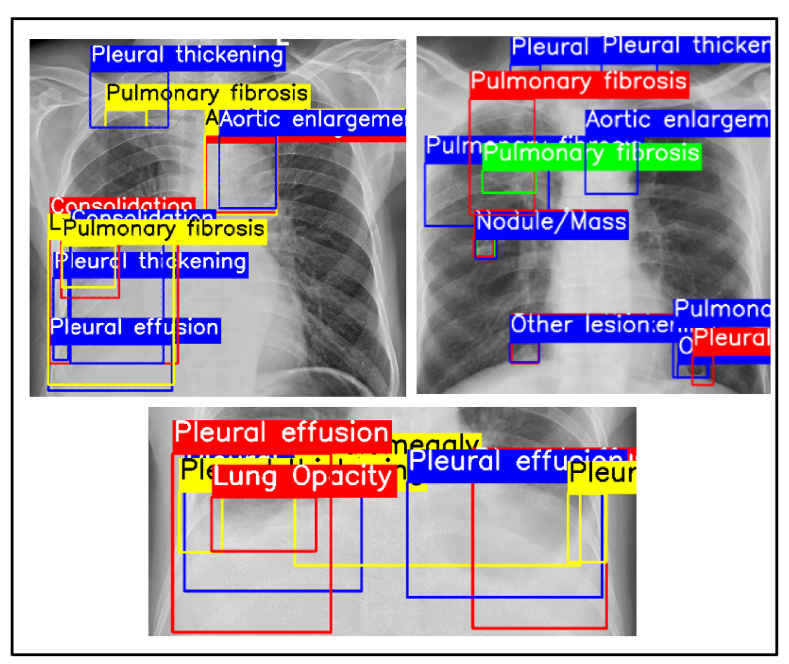
  from noisy crowdsourced object annotations ***without any additional inputs***.
- Introduced a benchmark to ***systematically evaluate*** BDC and previous methods using synthetic datasets with crowdsourced annotations simulating varying crowdsourcing scenarios.
- Demonstrated ***superior performance, scalability and robustness*** of BDC with extensive experiments.

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

This can result in ***multiple noisy, inconsistent object annotations*** originating from multiple annotators per image.

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.

Matching annotations to model predictions

Optimal prediction for each annotation is found by minimising:

$$\hat{y}^*_m = \underset{\hat{y}_n \in \hat{y}}{\arg\min}\, \mathcal{L}_{match}(\hat{y}_n, y_m) \ ,$$

$$\mathcal{L}_{match}(\hat{y}_n, y_m) = -\hat{p}_{n(c_m)} + \lambda_1 \mathcal{L}_{IoU}(\hat{b}_n, b_m) + \lambda_2 ||\hat{b}_n - b_m||_1 \ .$$

- One-to-many matching
- Local minimum matching cost

Modelling annotators' annotations as distributions

Bounding Box Aggregator

- Scaling and translation errors of each annotator modelled using ***Gaussian*** distributions with ***Gaussian-Gamma*** conjugate prior:

$$p(\epsilon_m | k_m = k, \mu, \sigma) = \mathcal{N}(\mu^k, \sigma^k) \ .$$

$$\epsilon_m = \left[\hat{b}^*_{m(1)} - b_{m(1)},\ \hat{b}^*_{m(2)} - b_{m(2)},\ \hat{b}^*_{m(3)} \div b_{m(3)},\ \hat{b}^*_{m(4)} \div b_{m(4)}\right] \ .$$

- Annotations are corrected with the posterior mean:

$$b_m := \left(b_m + \left[\mu^k_{(1)},\, \mu^k_{(2)},\, 0,\, 0\right]\right) \odot \left[1,\, 1,\, \mu^k_{(3)},\, \mu^k_{(4)}\right] \ .$$

- All annotations matched to the same prediction are aggregated using the posterior precision as weight.

Class Label Aggregator

- Integrated Bayesian classifier combination neural network [3].
- Modelled the annotated class labels of each annotator as ***multinomial distributions*** conditioning on the true object label:

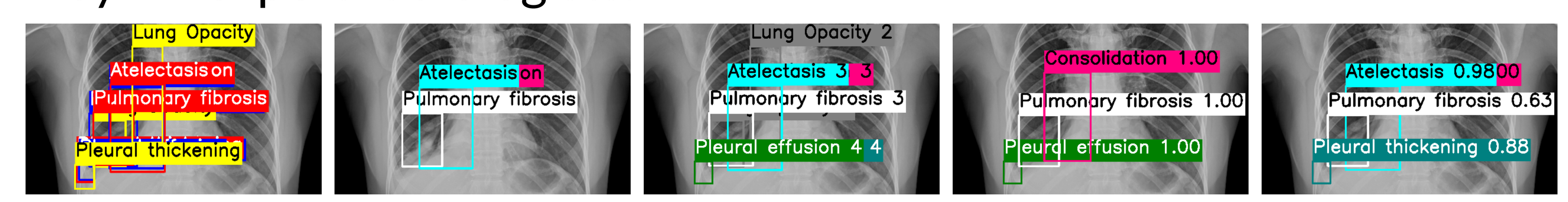
$$p(c_m | k_m = k, t_m = j, \pi) = \pi^k_{j, c_m} \ .$$

- Have a ***Dirichlet*** conjugate prior.
- The aggregated class label probability is computed as:

$$\rho_{n,j} = \exp\left(\ln \hat{p}_{n,j} + \sum_{(c,k) \in \tilde{\kappa}_n} \mathbb{E}_{\pi^k_j} \ln \pi^k_{j,c}\right) \ .$$

Experiments and Results

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



(a) NA (b) MV (c) WBF-EARL (d) Crowd R-CNN (e) BDC (ours)

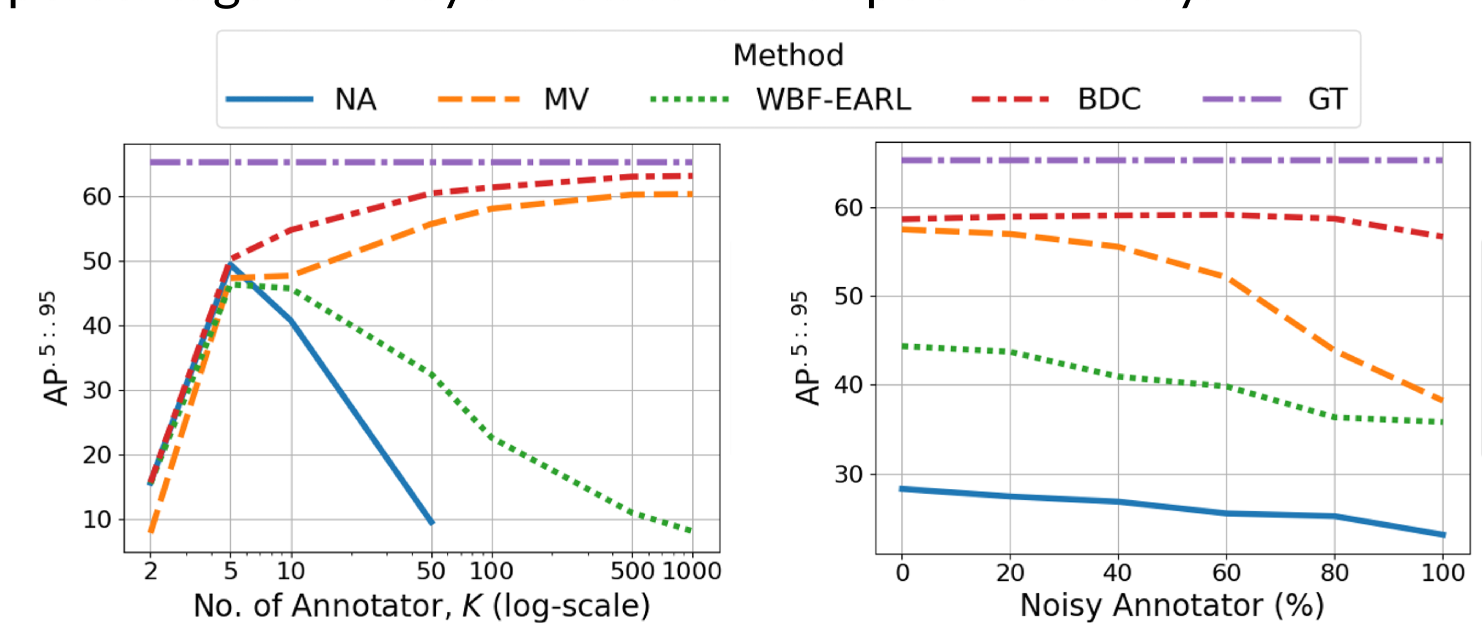
- Synthetic datasets: simulate various synthetic crowdsourcing settings with VOC and MSCOCO datasets.

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**

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**

Table: AP metrics for (left) VinDr-CXR and (right) COCO-FULL synthetic datasets with 10 synthetic annotators of varying annotating accuracies.

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



References

[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.

[3] Isupova et al. BCCNet: Bayesian classifier combination neural network. In *NeurIPS ML4D*, 2018.