Boosting Weakly Supervised Object Detection via Learning Bounding Box Adjusters

Bowen Dong1, Zitong Huang1, Yuelin Guo1, Qilong Wang2, Zhenxing Niu3, Wangmeng Zuo1,4
1Harbin Institute of Technology 2Tianjin University 3Alibaba Damo Academay 4Pazhou Lab, Guangzhou

2021 ICCV OCTOBER 11-17 VIRTUAL

Motivation:

- In terms of **localization performance**, there remains a huge gap between WSOD methods and their fully-supervised counterparts.
- The class-agnostic object localization ability of an object detector can be **transferred into objects from novel categories**.
- We can pretrain a detector with **public well-annotated datasets**, and then guide training of WSOD using the **pretrained detector**.

Problem Setting:

- During LBBA training, a **publicly available** well-annotated auxiliary dataset $\mathbb{X}^{\text{aux}}$ is used to supervise LBBA g.
- During LBBA-boosted WSOD, $\mathbb{X}^{\text{aux}}$ is abandoned, we only use **pretrained LBBA** g and image-level labels $\mathbf{y}$ from **any** weakly annotated dataset $\mathbb{X}$ to supervise WSOD network f.

Baseline WSOD network:

- Baseline network: OICR with bounding box regression branch.
- Training objective function of our baseline network:

$$\mathcal{L}_{\text{wsod}} = \mathcal{L}_{\text{wsddn}} + \mathcal{L}_{\text{r}} + \mathcal{L}_{\text{rpn-cls}} + \mathcal{L}_{\text{rpn-det}} + \mathcal{L}_{\text{det}}$$

Learnable Bounding Box Adjuster (LBBA):

- We adopt Faster R-CNN with ResNet-50 backbone as well as a class-agnostic bounding box regression branch as our LBBA.
- LBBA is trained on fully-annotated **auxiliary dataset**.

Training of LBBA:

- **Initialization: optimizing parameter of LBBA by fully supervised object detection**
- **E-Step:**

$$\theta_g = \arg\min_{\theta_g} \mathcal{L}_{\text{bba}}(\{\mathbf{b}^{\text{aux}}\}, g(\mathbf{I}^{\text{aux}}, \mathbb{P}^{\text{aux}}; \theta_g))$$

Where $\mathcal{L}_{\text{bba}} = \sum_{\mathbf{p}^{\text{aux}} \in \mathbb{P}^{\text{aux}}} \text{Smooth}_{L1}(\mathbf{b}^{\text{aux}}, \hat{\mathbf{b}}^{\text{aux}}; \theta_g)$

- **M-Step:**

$$\theta_{f^{\text{aux}}} = \arg\min_{\theta_{f^{\text{aux}}}} (\mathcal{L}_{\text{wsod}} + \mathcal{L}_{\text{bbr}})(\hat{\mathbf{b}}^{\text{aux}}, f^{\text{aux}}(\mathbf{I}^{\text{aux}}, \mathbb{P}^{\text{aux}}; \theta_{f^{\text{aux}}}))$$

Where $\mathcal{L}_{\text{bbr}} = \sum_{\mathbf{p}^{\text{aux}} \in \mathbb{P}^{\text{aux}}} \text{Smooth}_{L1}(\hat{\mathbf{b}}^{\text{aux}}, \tilde{\mathbf{b}}^{\text{aux}}; \theta_f)$

- After T stages training on auxiliary dataset, we obtain a series of adjusters for final LBBA-boosted WSOD training.

LBBA-boosted WSOD:

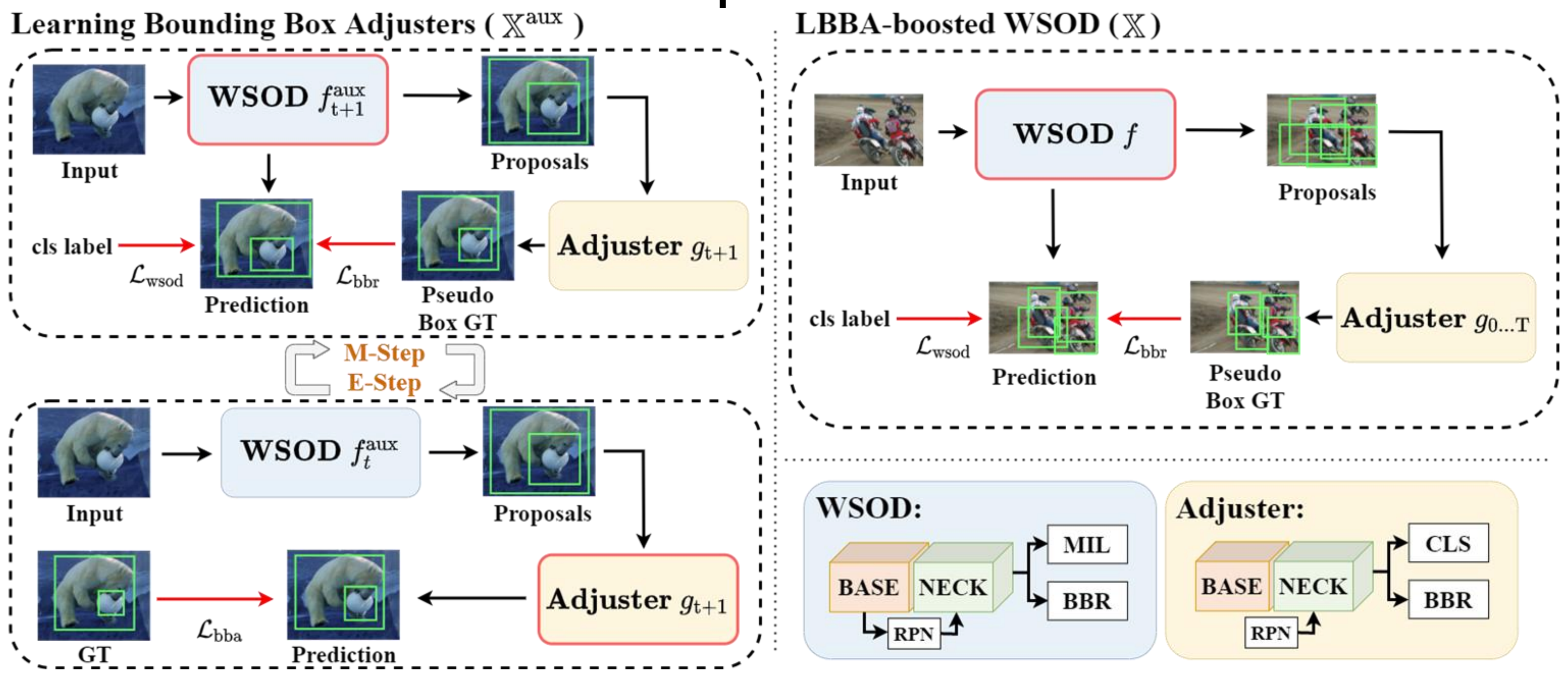
- Now, our WSOD network is supervised by both image level labels as well as a series of learnable bounding box adjusters.
- **Optimizing WSOD on target dataset:**

$$\theta_f = \arg\min_{\theta_f} (\mathcal{L}_{\text{wsod}} + \mathcal{L}_{\text{bbr}})(\hat{\mathbf{b}}, f(\mathbf{I}, \mathbb{P}; \theta_f))$$

Where $\mathcal{L}_{\text{bbr}} = \sum_{\mathbf{p} \in \mathbb{P}} \text{Smooth}_{L1}(\hat{\mathbf{b}}, \tilde{\mathbf{b}}; \theta_f)$

- Both training with the last LBBA and training with all LBBAs progressively are feasible.

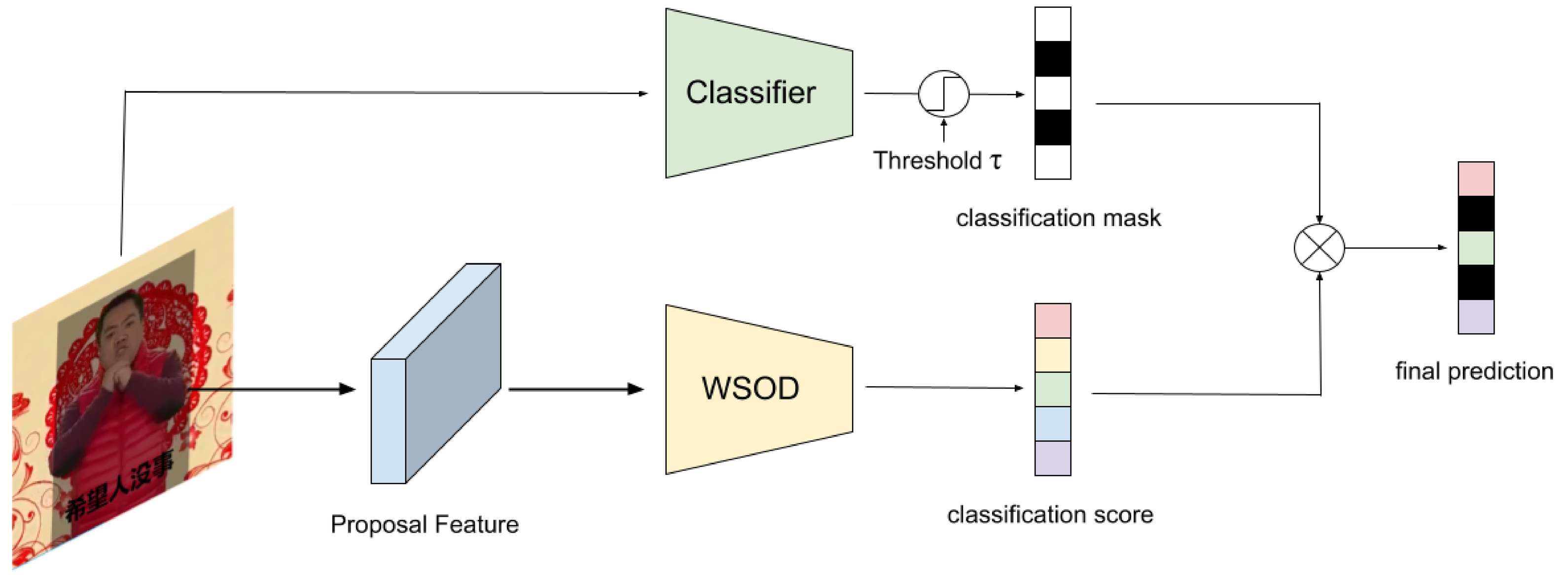
Pipeline



Visualization

Masking Strategy:

- We introduce an additional multi-label image classifier and present a classification score masking strategy.



Experiments:

- All LBBA methods **outperform the baseline** OICR and OICR+REG by improving the localization ability of WSOD methods.
- **LBBA with masking performs better**, which indicates that masking strategy can further improve detection ability of WSOD methods.
- Our LBBA method can be **generalized to multiple datasets**.

Method	AP50(VOC07)	AP50(COCO-20)	AP50(ILSVRC)
OICR	41.2	22.8	20.5
OICR+REG	51.4	23.9	22.4
LBBA w/o masking	55.8	27.5	28.0
LBBA w/ masking	**56.5**	**29.9**	**30.1**

Conclusion

- **Multi-stage learnable bounding box adjusters** are presented for improving localization performance of WSOD.
- An **EM-like multi-stage training algorithm**, are suggested to learn LBBAs specified for optimizing WSOD.
- An **effective masking strategy** is introduced to improve the accuracy of the proposal classification branch.
- LBBA performs **favorably against the state-of-the-art** WSOD methods and knowledge transfer models.