Decoupled Adaptation for Cross-Domain Object Detection

Junguang Jiang, Baixu Chen, Jianmin Wang, Mingsheng Long

Tsinghua University

Problem

- Object detection in the real world suffers from performance drop due to variance in viewpoints, object appearance, backgrounds, illumination, image quality, *etc.*
- *Domain adaptation* aims to transfer a detector from a source domain $\mathcal{D}_s$, where sufficient training data is available, to a target domain $\mathcal{D}_t$ where only unlabeled data is available.

Challenges

- Data Challenge: What to adapt in the object detection task is unknown.

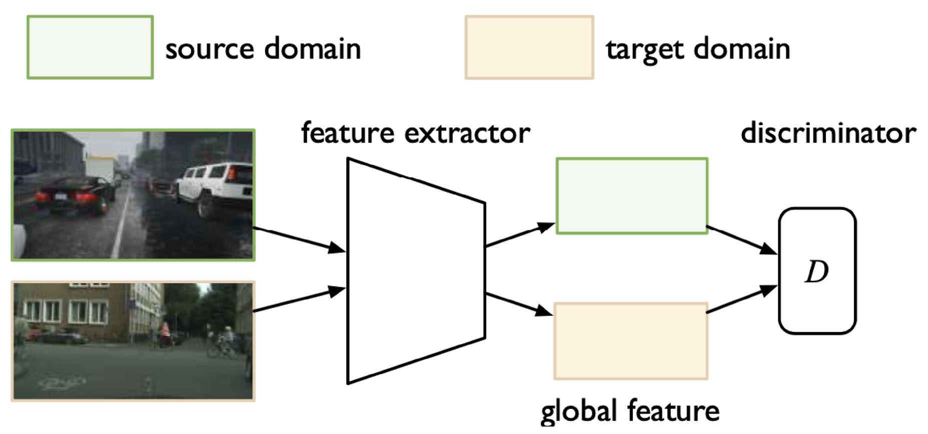

Figure 1:Global adapt

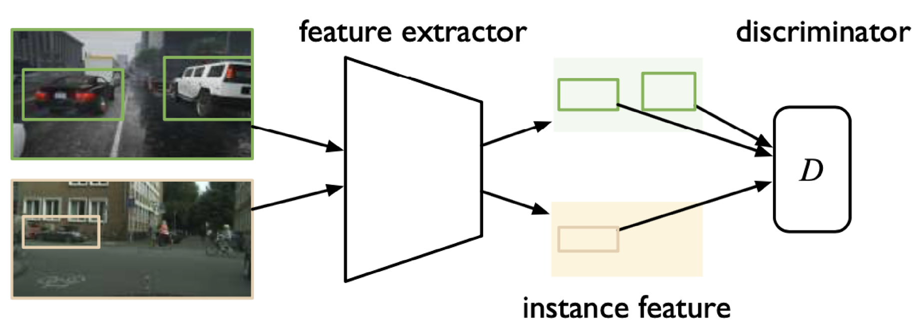

Figure 2:Instance adapt

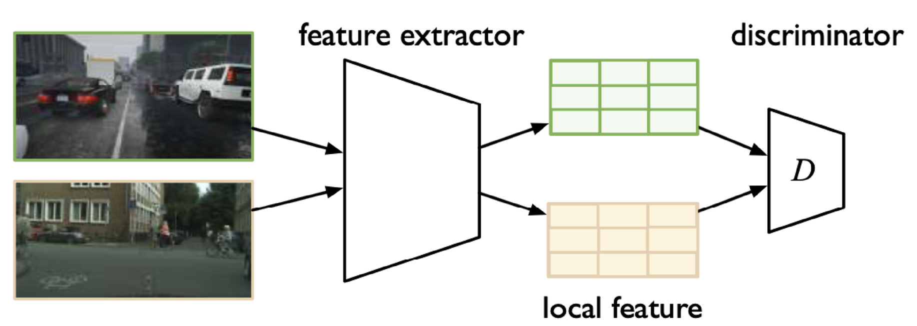

Figure 3:Local adapt

- Architecture Challenge: Introducing domain discriminators into the detector architecture might deteriorate the discriminability of detectors.
- Task Challenge: As a multi-task learning problem, it's difficult to adapt detectors for different tasks at the same time.

D-adapt Framework

Procedures:

1. Decouple the original cross-domain detection problem into several sub-problems.
2. Design domain adaptors to solve each sub-problem.
3. Coordinate the relationships between different adaptors and the detector.

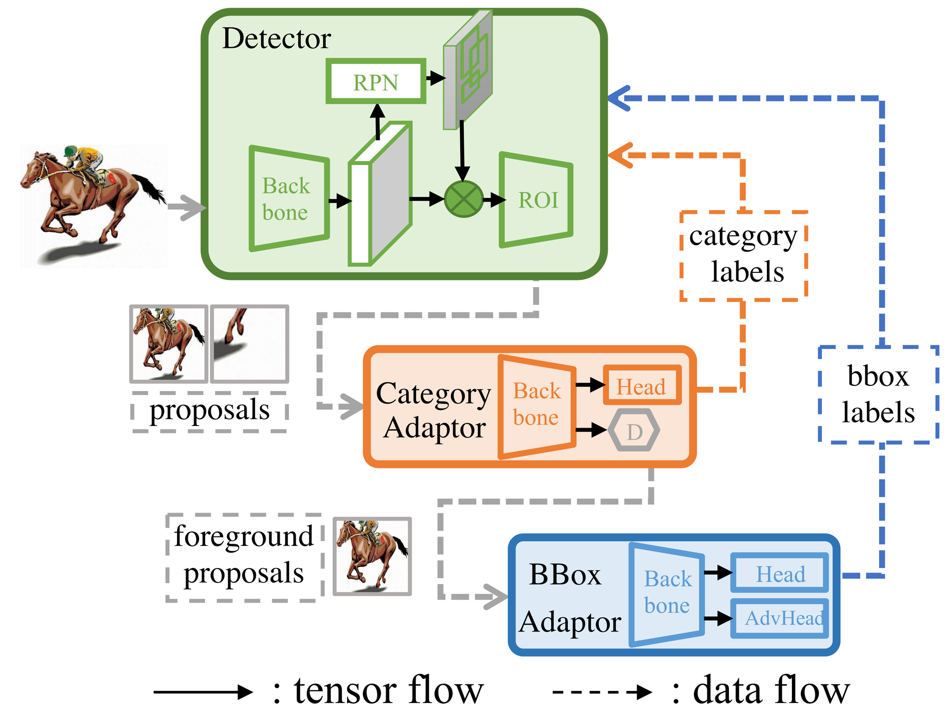

Figure 4:D-adapt Framework.

Meanings: Different parts have independent parameters, data distributions, and training losses.

Summary:

1. Architecture challenge ⇒ Decouple the adversarial adaptation from the training of the detector by introducing a parameter-independent category adaptor.
2. Task challenge ⇒ Introduce another bounding box adaptor that's decoupled from both the detector and the category adaptor.
3. Data challenge ⇒ Adjust the object-level data distribution for specific adaptation tasks.

Contact Information

- Paper: Decoupled Adaptation for Cross-Domain Object Detection
- Code: Transfer-Learning-Library
- Email: JiangJunguang1123@outlook.com

Category Adaptation

- **Challenge**: The input data distribution doesn't satisfy the low-density separation assumption well ⇒ impede the adversarial alignment.
- **Solution**: Use the confidence of each proposal to discretize the input space. The objective of the discriminator D is,

$$\begin{aligned}\max_D \mathcal{L}^{\text{cls}}_{\text{adv}} =& \mathbb{E}_{\mathbf{x}_s\sim\mathcal{D}_s^{\text{prop}}} w(\mathbf{c}_s)\log[D(\mathbf{f}_s,\mathbf{g}_s)] \\ &+\mathbb{E}_{\mathbf{x}_t\sim\mathcal{D}_t^{\text{prop}}} w(\mathbf{c}_t)\log[1-D(\mathbf{f}_t,\mathbf{g}_t)],\end{aligned} \tag{1}$$

where $\mathbf{f}=F^{\text{cls}}(\mathbf{x})$ is the feature and $\mathbf{g}=G^{\text{cls}}(\mathbf{f})$ is the category prediction.

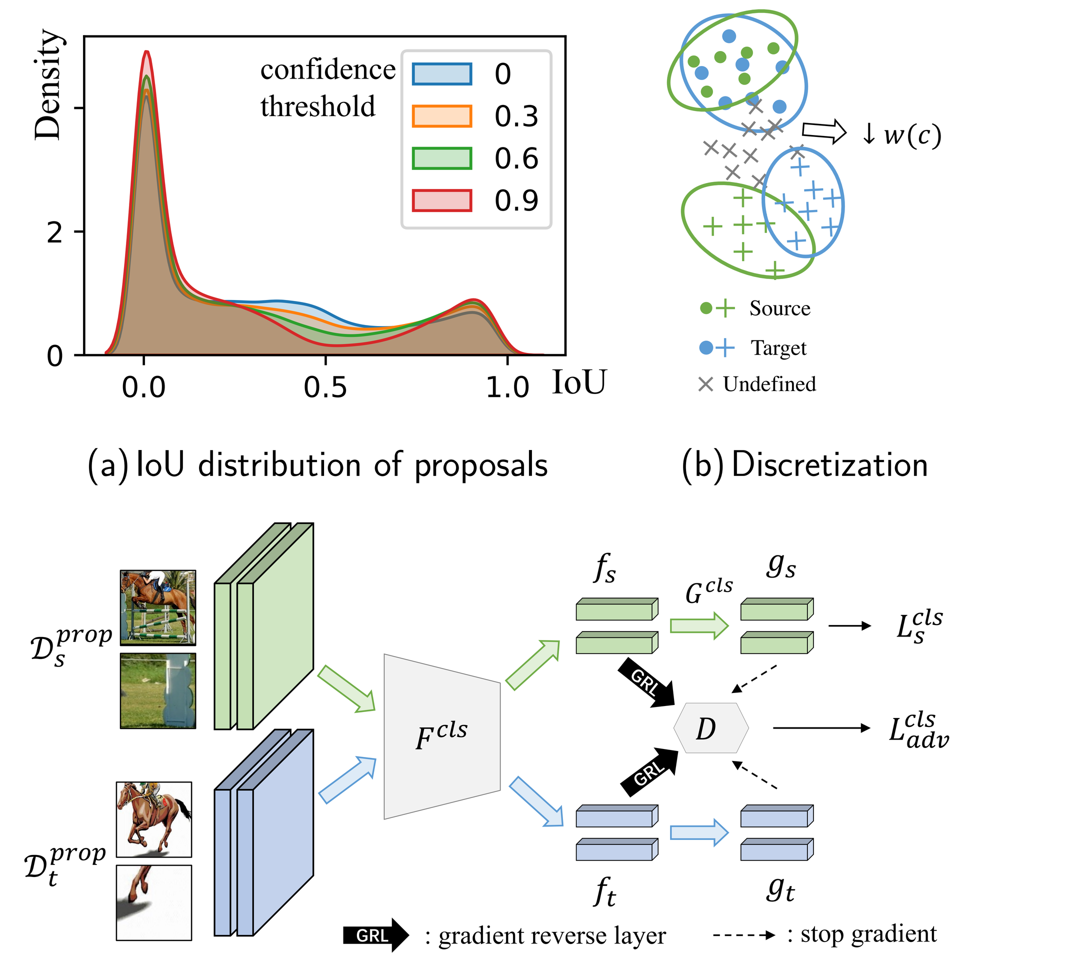

Figure 5:Architecture of the category adaptor.

Bounding Box Adaptation

Introduce an adversarial regressor network $G^{\text{reg}}_{\text{adv}}$ to maximize its disparity with the main regressor on the target domain while minimizing the disparity on the source domain to measure the discrepancy across domains.

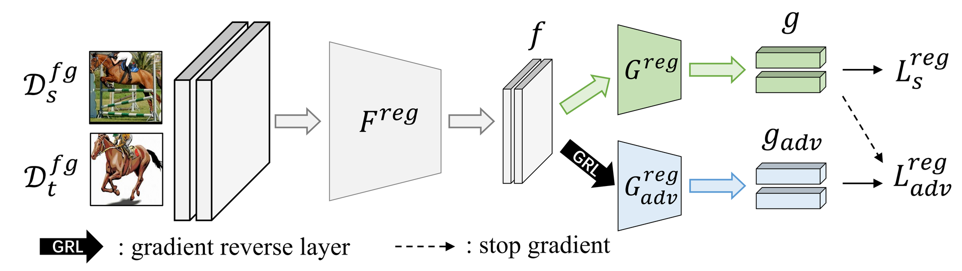

Figure 6:Architecture of the bounding box adaptor.

Quantitative Results

D-adapt achieves state-of-the-art results on four cross-domain object detection tasks and yields 17% and 21% relative improvement on benchmark datasets *Clipart1k* and *Comic2k* in particular.

Table 1:Results from VOC to Comic (ResNet-101).

Method	bike	bird	car	cat	dog	prsn	mAP
Source Only	32.5	12.0	21.1	10.4	12.4	29.9	19.7
DA-Faster	31.1	10.3	15.5	12.4	19.3	39.0	21.2
SWDA	36.4	21.8	29.8	15.1	23.5	49.6	29.4
MCAR	47.9	20.5	37.4	20.6	24.5	**53.6**	33.5
Instance Adapt	39.5	17.7	26.5	27.3	22.4	48.4	30.3
Global Adapt	31.9	15.7	30.3	21.3	17.1	37.9	25.7
D-adapt	**52.4**	**25.4**	**42.3**	**43.7**	**25.7**	53.5	**40.5**
Oracle	42.2	35.3	31.9	46.2	40.9	70.9	44.6

Table 2:Ablations on category adaptation. VOC to Clipart.

metric	ours	w/o condition	w/o bg proposals				w/o weight	w/o adaptor
			source target					
mIoU^{cls}	38.2	36.9	-	36.6	33.6	25.1	17.2	12.6
mAP	43.5	41.7	-	41.7	38.8	36.5	33.3	28.0

Table 3:Ablations on box adaptation. VOC to Clipart.

metric	Ours	w/o DD	w/o adaptor
mIoU^{reg}	0.631	0.598	0.531
mAP	45.0	44.4	43.5

Qualitative results

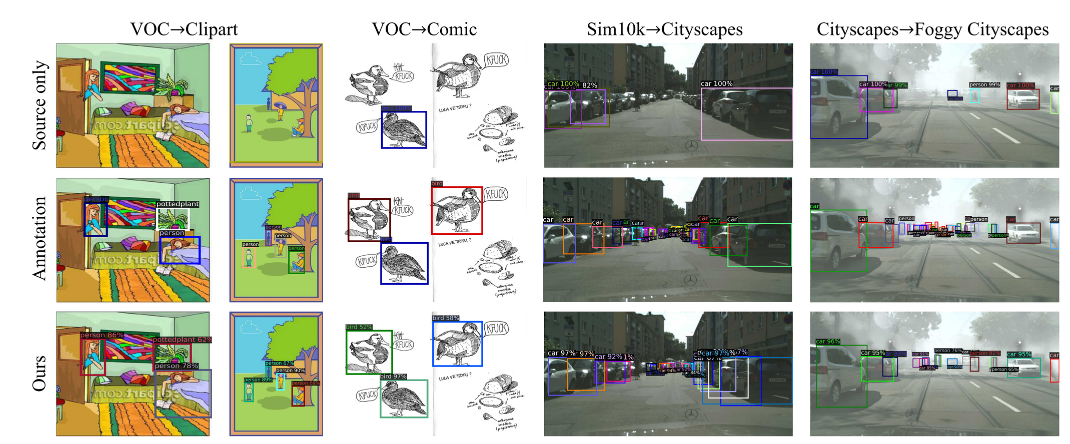

Figure 7:Qualitative results on the target domain.

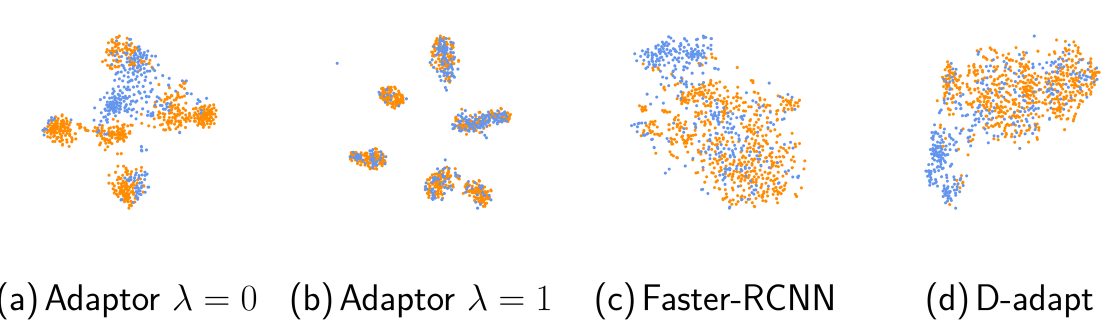

Figure 8:T-SNE visualization of features on task VOC → Comic2k.