Weak-shot Object Detection through Mutual Knowledge Transfer

Xuanyi Du, Weitao Wan, Chong Sun, Chen Li
WeChat, Tencent

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

(A) Contributions

1. We develop the Mutual Knowledge Transfer scheme which bridges the gap between the source and target datasets for the Weakly Supervised Object Detection (WSHOD) task.
2. We design the Knowledge Transfer (KT) loss to improve the training of the multiple instance learning (MIL) module by exploiting the knowledge from a fully-annotated dataset, which facilitates the detection performance on novel categories.
3. We propose the Consistency Filtering (CF) method which effectively removes inaccurate pseudo labels for the refinement of the proposal generator.

(B) Overview of the proposed Mutual Knowledge Transfer scheme

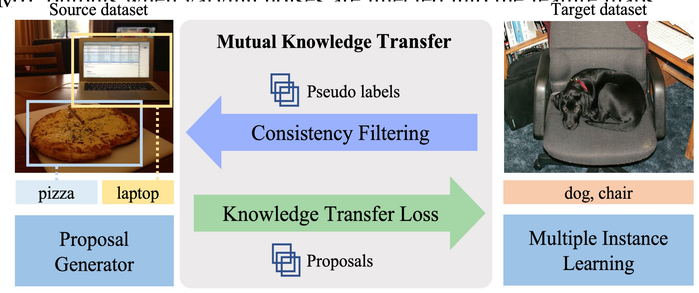
Weak-shot Object Detection methods exploit a fully-annotated source dataset to facilitate the detection performance on the target dataset which only contains image-level labels for novel categories.

- Knowledge Transfer (KT) loss

It performs knowledge transfer from the source (S) dataset to the target (T) dataset by constraining the training of the MIL module, enforcing the predicted objectness score and class entropy of the MIL module to be consistent with the predictions of the proposal generator (PG).

- Consistency Filtering (CF) method

It discovers the inaccurate pseudo labels by evaluating the stability of the MIL outputs when varying noises are injected into the feature maps.



(C) Knowledge Transfer (KT) loss

Basically, the KT loss transfers the knowledge of both class distribution entropy and objectness from the PG module to the MIL module.

The formulation is defined below, in which $\tilde{S}$ and $\dot{S}$ are predicted by the MIL module which are explained in the paper. $\mathcal{H}()$ computes the information entropy of the class distribution. O_i is the objectness score predicted by the PG module. R is the number of proposals.

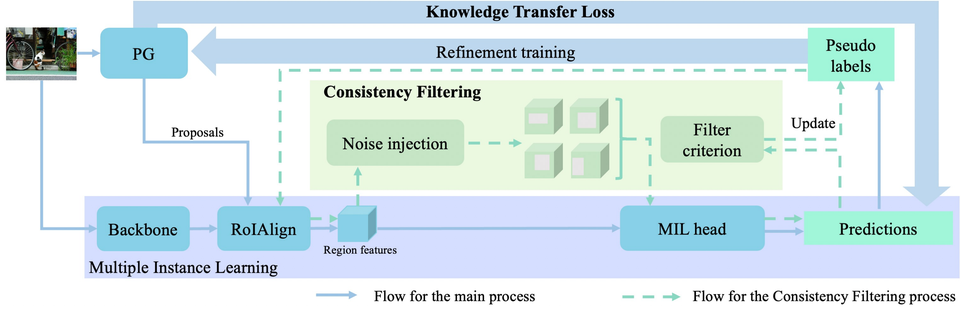
$$\mathcal{L}_{kt} = \mathcal{L}_{ent} + \mathcal{L}_{obj}$$
$$= \frac{1}{R}\sum_{i=1}^{R}\{[\mathcal{H}(\tilde{S}_i^d) - (1-O_i)\log C]^2 + (\max_j \dot{S}_{ij}^d - O_i)^2\}.$$

(D) Consistency Filtering (CF) method

Motivation: (1) the quality of pseudo labels can be further improved by CF; (2) object knowledge regarding the novel categories in the T dataset can boost the training of the PG module when transferred through CF.

The RoIAlign features (with spatial size 7 × 7) of the pseudo boxes are obtained,

in which random regions of feature maps are replaced with noises. After repeatedly injecting random noises to N random regions, the evaluated pseudo box will be removed if all the N predictions meet the proposed filter criterion.



(E) Experimental Results

Pascal VOC 2007 is adopted as the target dataset. Either MS COCO 2017 or ILSVRC 2013 detection dataset is adopted as the source dataset.

Table 1. mAP comparisons with state-of-the-art methods on VOC 2007 test set

Method	aero	bike	bird	boat	bottl	bus	car	cat	chair	cow	table	dog	horse	mbik	pers.	plant	sheep	sofa	train	tv	mAP
Pure WSOD:																					
WSDDN-Ens [1]	46.4	58.3	35.5	25.9	14.0	66.7	53.0	39.2	8.9	41.8	26.6	38.6	44.7	59.0	10.8	17.3	40.7	49.6	56.9	50.8	39.3
OICR-Ens+FR [29]	65.5	67.2	47.2	21.6	22.1	68.0	68.5	35.9	5.7	63.1	49.5	30.3	64.7	66.1	13.0	25.6	50.0	57.1	60.2	59.0	47.0
PCL-Ens+FR [28]	63.2	69.9	47.9	22.6	27.3	71.0	69.1	49.6	12.0	60.1	51.5	37.3	63.3	63.9	15.8	23.6	48.8	55.3	61.2	62.1	48.8
ICMWSD+FR [24]	66.4	69.1	58.9	32.5	27.6	71.5	73.1	66.2	32.8	75.4	47.4	53.7	63.3	71.7	34.8	28.5	57.4	54.7	62.5	67.1	55.7
CASD+FR [11]	66.6	81.3	58.4	33.5	31.6	75.7	55.2	68.3	36.8	59.5	61.0	52.9	65.4	72.0	29.1	29.4	65.7	54.2	74.5	70.7	57.1
WSHOD:																					
MSD [18]	70.5	69.2	53.3	43.7	25.4	68.9	68.7	56.9	18.4	64.2	15.3	72.0	74.4	65.2	15.4	25.1	53.6	54.4	45.6	61.4	51.1
OICR+UBBR [15]	59.7	44.8	54.0	36.1	29.3	72.1	67.4	70.7	23.5	63.8	31.5	61.5	63.7	61.9	37.9	15.4	55.1	57.4	69.9	63.6	52.0
Zhong et al. (single scale) [36]	64.4	45.0	62.1	42.8	42.4	73.1	73.2	76.0	28.2	78.6	28.5	75.1	74.6	67.7	57.5	11.6	65.6	55.4	72.2	61.3	57.8
Zhong et al.+ [36]	64.8	50.7	65.5	45.3	46.4	75.7	74.0	80.1	31.3	77.0	26.2	79.3	74.8	66.5	57.9	11.5	68.2	59.0	74.7	65.5	59.7
Zhong et al. (distill,vgg16)+ [36]	62.6	56.1	64.5	40.9	44.5	74.4	76.8	80.5	30.6	75.4	25.5	80.9	73.4	71.0	59.1	16.7	64.1	59.5	72.4	68.0	59.8
Zhong et al. (distill)+ [36]	65.5	57.7	65.1	41.3	43.0	73.6	75.7	80.4	33.4	72.2	33.8	81.3	79.6	63.0	59.4	10.9	65.1	64.2	72.7	67.2	60.2
TraMaS (single scale) [21]	65.6	53.7	67.4	47.2	46.9	76.3	76.6	81.7	33.0	76.9	29.3	80.9	76.8	66.2	61.1	12.6	65.8	58.9	74.4	66.7	60.9
TraMaS+ [21]	66.5	58.7	68.3	47.7	47.0	76.3	78.0	81.1	33.9	77.8	30.9	80.1	78.0	66.2	63.0	15.1	69.2	60.2	76.1	68.1	62.1
TraMaS (distill,vgg16)+ [21]	67.8	59.9	67.9	48.9	47.5	75.4	78.2	79.3	33.1	76.4	32.1	78.8	77.4	68.3	63.1	18.4	70.0	59.9	76.2	69.3	62.4
TraMaS (distill)+ [21]	68.6	61.1	69.6	48.1	49.9	76.3	77.8	80.9	34.9	77.0	31.1	80.9	78.5	66.3	64.0	19.1	69.1	62.3	74.4	69.1	62.9
Ours (single scale)	64.8	56.2	67.8	48.8	52.0	76.5	78.1	82.0	33.4	77.9	24.7	82.6	73.3	74.0	69.0	15.1	70.7	65.3	78.6	66.6	**62.9**
Ours+	68.5	57.6	68.5	47.3	50.9	79.2	78.4	81.8	34.7	77.5	23.1	81.8	74.3	73.0	69.6	15.9	70.8	62.3	78.2	69.1	**63.1**
Ours (distill,vgg16)+	64.9	64.6	69.4	44.9	48.3	72.0	81.4	80.9	38.7	74.5	26.4	79.3	75.3	74.2	72.1	20.2	65.5	62.3	76.4	69.6	**63.0**
Ours (distill)+	68.3	64.6	71.7	48.5	50.6	77.1	80.9	80.6	39.7	81.0	28.0	81.0	76.2	72.4	72.0	21.9	70.9	66.0	79.3	68.8	**65.0**

Table 2. Changing the source dataset to ILSVRC-179.

Methods	mAP	CorLoc
MSD (vgg16) [18]	47.5	65.3
Zhong et al. [36]	56.5	/
TraMaS (vgg16) [21]	57.8	74.1
TraMaS [21]	58.3	74.8
Ours (vgg16)	**58.9**	**75.7**
Ours	**60.4**	**77.5**

Table 3. Ablation studies for the proposed modules.

KT loss	CF	mAP (%)
-	-	60.9
+	-	61.6
-	+	62.1
+	+	62.9

Visualizations of pseudo boxes removed by the CF method. The first row is the ground truth. the cyan boxes are the corresponding ground-truth for the categories of the removed boxes, and the yellow ones are ground-truth for other categories.. The second row shows the removed boxes in cyan dashed boxes and the kept pseudo labels in red solid boxes.