WEAKLY- AND SEMI-SUPERVISED OBJECT LOCALIZATION

Zhen-Tang Huang, Yan-He Chen, Mei-Chen Yeh

Department of Computer Science and Information Engineering

NATIONAL TAIWAN NORMAL UNIVERSITY

Weakly Supervised Object Localization (WSOL)

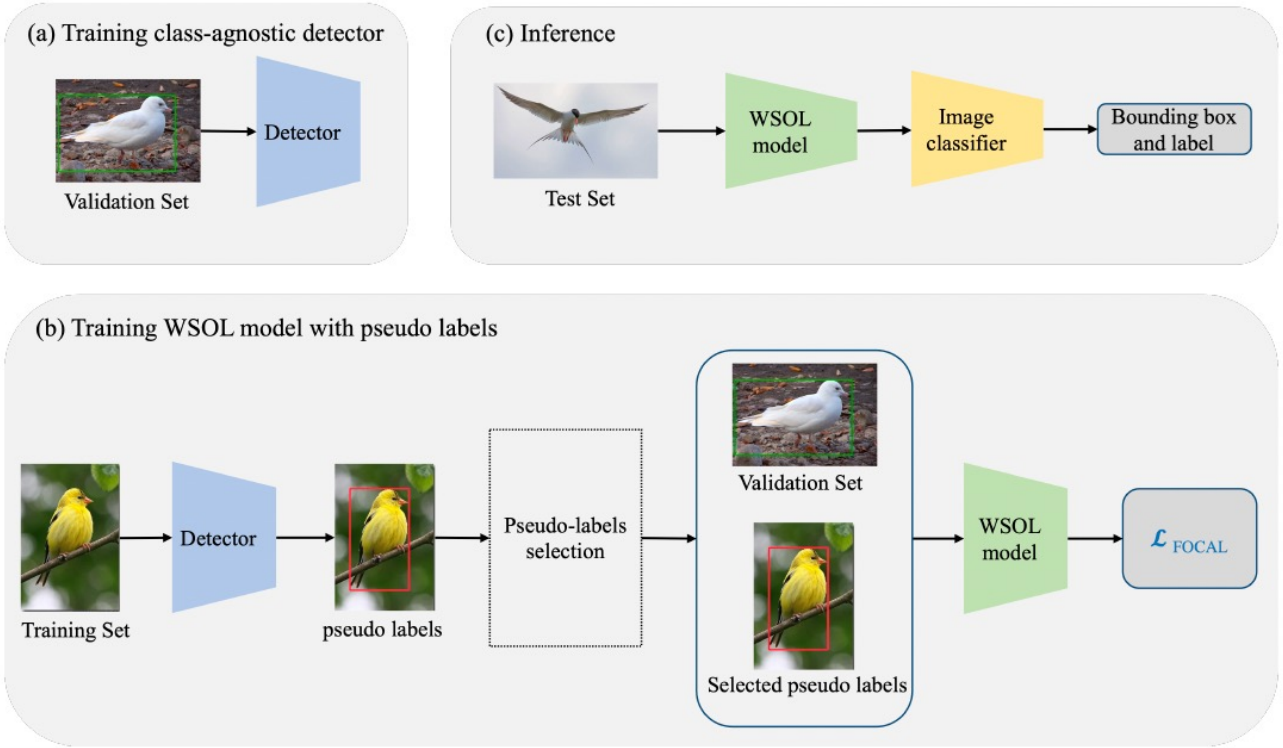
- The training set contains single-label images with image-level labels
  - A WSOL model is trained to localize objects of interest under the setting in which only class labels are given.
- Cost of data annotation
  - WSOL only needs image-level labels and can significantly reduces the cost of data annotation, being essential to numerous real-world applications.
- Learning under uncertainty
  - The training set in WSOL doesn’t have ground truth bounding boxes, but the task is to find the foreground locations.
- Learning with imbalanced data
  - Classes have different number of samples with different difficulty.

Motivation

- There is no unified protocol for WSOL before 2020.
- Choe et al. proposed a WSOL benchmark containing a small validation set with full supervision, which is available for parameter search and bounding box regression.
- Under this protocol, we wonder how to incorporate both weak and full supervision to improve WSOL performance.

Choe, J., et al. Evaluation for weakly supervised object localization: Protocol, metrics, and datasets. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2022.

Model Architecture



Pseudo-Labels Selection

- The proposed pseudo-labels selection algorithm prioritizes high confidence pseudo label.
- Calculating the minimal confidence score for all bounding boxes in one image is important.

Algorithm 1 Selecting Reliable Pseudo Labels

Input:

Samples $\mathcal{S}$: each sample x_i has a class label y_i, n_i pseudo boxes $\{\mathbf{b}_j\}_{j=1}^{n_i}$ and n_i confidence scores $\{s_j\}_{j=1}^{n_i}$;
A threshold γ;
The number of selected samples per class N;

Output:

Selected samples $\mathcal{F}$;

Initialize $\mathcal{F} \leftarrow \emptyset$
for $x_i \in \mathcal{S}$ **do**
  Calculate $m_i \leftarrow \min_j\{s_j\}$;
end for
Sort samples by descending order of m_i
for $x_i \in$ sorted $\mathcal{S}$ **do**
  let n be the no. of selected samples whose label is y_i;
  if $m_i \geq \gamma$ and $n < N$ **then**
    Put x_i into $\mathcal{F}$;
  end if
end for
return $\mathcal{F}$;

Experimental Results

Table 1. Comparison with state-of-the-arts on ImageNet.

Method	GT Loc	Top-1 Loc	Top-5 Loc
FSL [3] (*CVPR*'20)	66.30	-	-
PSOL [10] (*CVPR*'20)	66.28	55.31	64.18
SLT-Net [21] (*CVPR*'21)	67.60	55.70	65.40
SCG [22] (*CVPR*'21)	65.05	49.56	61.32
SPOL [11] (*CVPR*'21)	69.02	**59.14**	**67.15**
Zhang *et al.* [9] (*ICASSP*'22)	65.40	50.10	-
Kim *et al.* [23] (*CVPR*'22)	69.89	53.76	65.75
Zhu *et al.* [24] (*CVPR*'22)	70.27	55.84	-
Wu *et al.* [25] (*CVPR*'22)	72.00	52.97	66.59
Supervised baseline	61.45	46.73	56.44
SSL w. cross entropy loss	67.57	50.35	61.22
SSL w. focal loss	**74.72**	54.09	66.51

Table 2. Comparison with state-of-the-arts on CUB.

Method	GT Loc	Top-1 Loc	Top-5 Loc
FSL [3] (*CVPR*'20)	92.00	-	-
PSOL [10] (*CVPR*'20)	93.01	77.44	89.51
SLT-Net [21] (*CVPR*'21)	87.60	67.80	-
SCG [22] (*CVPR*'21)	72.14	53.59	66.50
SPOL [11] (*CVPR*'21)	96.46	80.12	**93.44**
Zhang *et al.* [9] (*ICASSP*'22)	82.32	61.85	-
Kim *et al.* [23] (*CVPR*'22)	93.17	70.83	88.07
Zhu *et al.* [24] (*CVPR*'22)	81.83	66.65	-
Wu *et al.* [25] (*CVPR*'22)	95.13	77.25	90.08
Supervised baseline	96.81	79.70	91.77
SSL w. cross entropy loss	97.96	**80.89**	92.80
SSL w. focal loss	**98.39**	79.96	92.77
Supervised baseline*	95.75	79.55	91.04
SSL w. cross entropy loss*	95.01	78.72	90.23
SSL w. focal loss*	96.05	79.57	91.15

Analysis of reliable bounding box selection

- Calculating the minimum confidence score can avoid the situation where the foreground is treated as the background.

Table 3. Analysis of reliable bounding box selection

Method	min. conf.	GT Loc	Top-1	Top-5
SSL w.	✗	64.95	48.81	59.12
cross entropy	✓	67.57	50.35	61.22
SSL w.	✗	72.76	53.22	65.11
focal loss	✓	74.72	54.09	66.51

Conclusion

- We present a new WSOL method based on SSL.
- Our model outperforms past methods on ImageNet and CUB in GT-known Loc. It also has a good generalization ability.
- One possible future direction is to elaborate pseudo labels more effectively.