Weakly Supervised Object Localization Using Things and Stuff Transfer — Supplemental Material—

Miaojing Shi[1,2] Holger Caesar[1] Vittorio Ferrari[1]

[1]University of Edinburgh [2]Tencent Youtu Lab

name.surname@ed.ac.uk

Abstract

The appendices provided in this supplemental material complement our paper in two aspects. We provide the proxy measures to detail the parameter learning process of our method. We conduct an ablation study to demonstrate the contribution of each component in the proposed system.

1. Proxy measures

We propose two proxy measures to jointly learn the score parameters by maximizing the performance over the entire validation set in $\mathcal{A}$ (Sec. 5.6):

1. Rank: the highest rank of any proposal whose intersection-over-union (IoU) with ground truth bounding box is > 0.5.

2. CorLoc@1: the percentage of images in which the highest scoring proposal localizes an object of the target class correctly (IoU > 0.5).

These two measures characterize well whether a proposal scoring function gives a higher score to the target objects than to other proposals. Hence they are good proxy measures for their usefulness within MIL. The behavior of the proposal scoring functions (Eqn. 5) depends on the class similarity measure used within them. Referring to Sec. 4.2, the guided similarities can be either appearance or semantic similarity (APP/SEM).

Results. We notice that roughly the same parameters are obtained from both criteria. Now we test how well they work on two of our target sets: ILSVRC-20 and COCO-07. We gradually add each proposal scoring scheme from Sec. 5.2 - 5.4 and denote them by +LW, +CW, and +AW in Fig. 1. Both Rank and CorLoc@1 are gradually improved: using APP we achieve the highest CorLoc@1: 22.9 on ILSVRC-20 and 6.2 on COCO-07; and the highest Rank: 0.94 on ILSVRC-20 and 0.89 on COCO-07. SEM is lower than APP: 19.2 and 4.3 in terms of CorLoc@1, and 0.94 and 0.83 in terms of Rank, on ILSVRC-20 and COCO-07, respectively. Comparing our proposal scoring schemes with a modern version of objectness [1],we see that both perform similarly well. In Sec. 7.2 and 7.3 we integrate our scheme with objectness and achieve a big improvement (+5%), which shows that both are complementary.

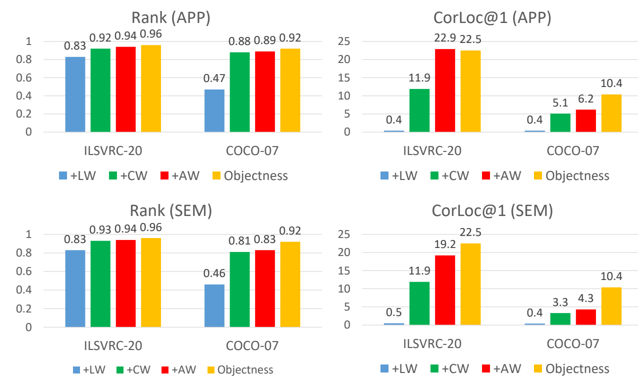

Figure 1: Proxy measures on ILSVRC-20 and COCO-07. Our transfer is guided by the appearance similarity (top: APP) as well as the semantic similarity (bottom: SEM). Performance is measured by Rank (left) and CorLoc@1 (right) (see Sec. 5.6).

2. Ablation study

We report here an ablation study to offer the justification of each component in our proposed system. We incorporate the LW, AW and CW scores (Sec. 5.2 -5.4) separately into the Basic MIL framework (Sec. 7.1). We report experiments on ILSVRC-20 in Table 1 in the same protocol as for the first table in the paper (guided by appearance similarity APP column). The three scores bring +0.9%, +3.3%, and +3.2% CorLoc on top of Basic MIL's 39.7%. This demonstrates that each individual score brings an improvement. Moreover, we also tried combining multiple

Method	LW	CW	AW	CorLoc
Basic MIL				39.7
	+			40.6
		+		43.0
			+	42.9
	+	+		44.0
		+	+	45.8
	+	+	+	47.6

Table 1: Ablation study on ILSVRC-20. LW: label weighting; CW: contrast weighting; AW: area weighting. We start from Basic MIL and incorporate LW, CW, AW, or any of their combination into it. We report the CorLoc.

scores: LW+CW reaches 44.0%, AW+CW reaches 45.8%, and using all three scores AW+LW+CW gives us the highest CorLoc 47.6%. Here we can see that LW brings an additional improvement of +1.8% when added to AW+CW. This shows that by carefully designing and integrating each component into our system, we are able to boost the overall performance over each individual component or any two of them.

References

[1] P. Dollar and C. Zitnick. Edge boxes: Locating object proposals from edges. In *ECCV*, 2014. 1