数据集：VOC2007

实验 1：增加信息的小尝试

编号	方法	结果
1	TagCLIP	mAP: 0.928024 F1: 0.803368, Precision: 0.881892, Recall: 0.737683
2	用红框把 TagCLIP 找到的目标区域在图片中圈起来，然后进行分类，prompt 没变	mAP: 0.881212 F1: 0.755885, Precision: 0.650241, Recall: 0.902516

实验 2：去掉 mask

编号	方法	结果
1	TagCLIP	mAP: 0.928024 F1: 0.803368, Precision: 0.881892, Recall: 0.737683
2	TagCLIP 去掉 CWR 模块中的 attention mask	mAP: 0.927974 F1: 0.777729, Precision: 0.883500, Recall: 0.694575

编号		mAP	F1	Precision	Recall
1	TagCLIP（ViT-B-16）	0.928024	0.803368	0.881892	0.737683
2	TagCLIP 去掉 mask（ViT-B-16）	0.927974	0.777729	0.883500	0.694575

实验 3：测试 crop 方法的上限

编号	方法	结果
1	TagCLIP	mAP: 0.928024 F1: 0.803368, Precision: 0.881892, Recall: 0.737683
2	用标签中的 bbox 把物体 crop 出来进行识别，不同 bbox 得到的结果最后用 max 函数合并	mAP: 0.959623 F1: 0.860597, Precision: 0.898844, Recall: 0.825472

编号		mAP	F1	Precision	Recall
1	TagCLIP（ViT-B-16）	0.928024	0.803368	0.881892	0.737683
2	使用 groundtruth 的 bbox（max、ViT-B-16）	0.959623	0.860597	0.898844	0.825472
	使用 groundtruth 的 bbox（max、RN50）	0.915671	0.786008	0.822964	0.752227
	使用 groundtruth 的 bbox（mean、RN50）	0.889354	0.777887	0.835994	0.727332

实验 4：统计预测多了和预测缺漏情况

预测多了，给出了不存在的物体	933 个测试用例
遗漏了	1365 个测试用例
图片总数	4952 个测试用例

实验 5：测试 CWR 模块中的不同阈值

编号	方法	结果
1	TagCLIP（原始阈值为 0.5）	mAP: 0.928024 F1: 0.803368, Precision: 0.881892, Recall: 0.737683
2	把 CWR 模块中的阈值改为 0.2	mAP: 0.920918 F1: 0.806649, Precision: 0.847697, Recall: 0.769392
3	把 CWR 模块中的阈值改为 0.3	mAP: 0.923840 F1: 0.810721, Precision: 0.864845, Recall: 0.762972
4	把 CWR 模块中的阈值改为 0.6	mAP: 0.928058 F1: 0.794955, Precision: 0.883389, Recall: 0.722615
5	把 CWR 模块中的阈值改为 0.8	mAP: 0.927465 F1: 0.772260, Precision: 0.870745, Recall: 0.693789

编号		mAP	F1	Precision	Recall
1	TagCLIP（原始阈值为 0.5）	0.928024	0.803368	0.881892	0.737683
2	CWR 阈值设为 0.2	0.920918	0.806649	0.847697	0.769392
3	CWR 阈值设为 0.3	0.923840	0.810721	0.864845	0.762972
4	CWR 阈值设为 0.6	0.928058	0.794955	0.883389	0.722615
5	CWR 阈值设为 0.8	0.927465	0.772260	0.870745	0.693789

TagCLIP 论文中 VOC2007 上分类的结果：

Method	Extra Training Data	VOC	COCO
Supervised specialist:			
SARB	10% Data	83.5	75.5
DualCoOp	10% Data	90.3	78.7
TAI-DPT	10% Data	93.3	81.5
Open-vocabulary generalist:			
TAI-DPT	COCO captions	88.3	65.1
CLIP†	None	79.5	54.2
CLIP	None	85.8	63.3
DPT†	None	83.4	59.6
DPT	None	86.2	64.3
CLIPSurgery	None	85.4	61.2
TagCLIP(Ours)	None	**92.8**	**68.8**

Table 2: Experimental results of multi-label classification. † represents not using softmax on classification scores.

TagCLIP 论文中 VOC2012 val set 上的消融实验：

Coarse Score	DMAR	CWR	mAP	mIoU
✓			85.4	30.9
✓		✓	88.0	55.2
✓	✓		93.9	63.7
✓	✓	✓	94.1	64.8