Bridging the Gap between Model Explanations in Partially Annotated Multi-label Classification

Youngwook Kim[1], Jae Myung Kim[2], Jieun Jeong[1,3], Cordelia Schmid[4], Zeynep Akata[2,5], Jungwoo Lee[1,3]

[1]Seoul National University [2]University of Tübingen [3]HodooAI Lab [4]Inria [5]MPI for Intelligent Systems

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Problem Definition

	full annotation	partial annotation
person	1	1
horse	1	
cat	0	0
dog	0	
truck	0	

Motivation

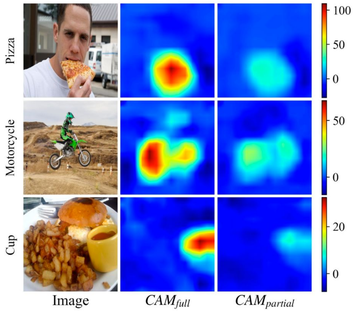
person	1	1
horse		0
cat	0	0
dog		0
truck		0

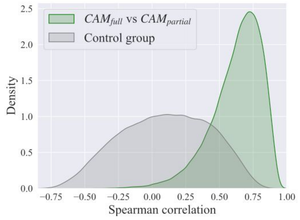
Baseline : **A**ssume **N**egative

=> Inducing **label noise (i.e., false negative)**

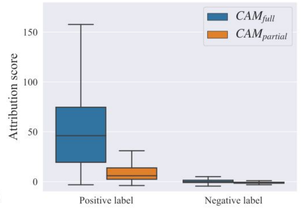
Q. How do **false negative** labels affect **model explanation?**

Analysis on Model Explanation

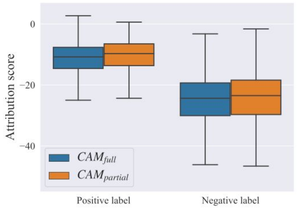




(a) Similarity of CAM by Spearman correlation



(b) Top-ranking attribution score



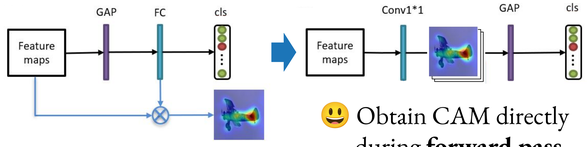
(c) Bottom-ranking attribution score

Model's **localization ability** is not much damaged by **false negative labels**.

=> **Bridge the Gap** !

Proposed Method : BoostLU

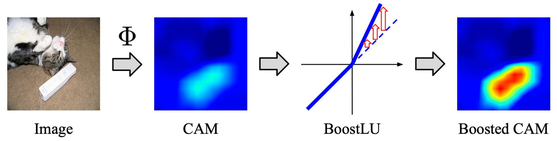
1) Modify the last FC layer into 1x1 conv



😃 Obtain CAM directly during **forward pass**

2) Apply BoostLU element-wisely on CAM

$$\text{BoostLU}(x) = max(x, \alpha x)$$



3) When to apply BoostLU

Rejecting losses from false negative labels

BoostLU in inference	BoostLU in training	LL-R in training	Performance VOC	Performance COCO
			86.10	64.58
✓			87.31	66.27
✓	✓		86.73	65.33
✓	✓	✓	**89.27**	**72.82**

① 😃 Performance increase **without additional training**

② 😃 **Boosted gradient** $\frac{\partial g}{\partial M_{ij}} = \begin{cases} \alpha/HW, & M_{ij} \geq 0 \\ 1/HW, & M_{ij} < 0 \end{cases}$

③ 😓 **Side effect** by false negative labels

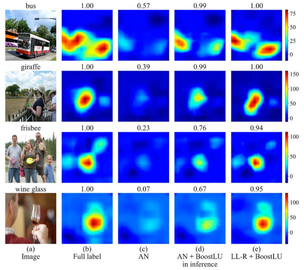
😄 **Alleviated side effect**

😄 **Synergy effect** with large loss modification scheme

Results

Single positive label setting

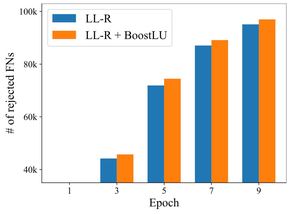
Methods	VOC	COCO	NUS	CUB
Full Label	89.42	76.78	52.08	30.90
AN	85.89	64.92	42.27	18.31
LS [30]	87.90	67.15	43.77	16.26
ASL [33]	87.76	68.78	46.93	18.81
ROLE [11]	87.77	67.04	41.63	13.66
ROLE + LI [11]	88.26	69.12	45.98	14.86
EM [50]	89.09	70.70	47.15	20.85
EM + APL [50]	89.19	70.87	47.59	**21.84**
LL-R [21]	88.27	70.70	48.76	19.56
+ BoostLU (Ours)	**89.29**	**72.89**	**49.59**	19.80
LL-Ct [21]	87.79	70.29	48.08	19.06
+ BoostLU (Ours)	88.61	71.78	48.37	19.25
LL-Cp [21]	87.44	70.27	47.92	19.21
+ BoostLU (Ours)	87.81	71.41	48.61	19.34



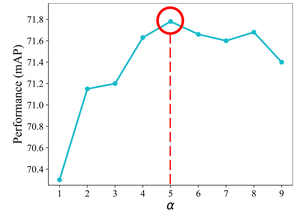
Large-scale partial label setting (Openimages)

Methods	Group 1	Group 2	Group 3	Group 4	Group 5	All Classes
CNN-RNN [39]	68.76	69.70	74.18	78.52	84.61	75.16
Curriculum Labeling [13]	70.37	71.32	76.23	80.54	86.81	77.05
IMCL [17]	70.95	72.59	77.64	81.83	87.34	78.07
P-ASL [2]	73.19	78.61	85.11	87.70	90.61	83.03
LL-R [21]	77.76	79.07	81.94	84.51	89.36	82.53
+ BoostLU (Ours)	79.28	80.81	83.32	85.63	90.27	83.86
LL-Ct [21]	77.76	79.18	81.97	84.46	89.51	82.58
+ BoostLU (Ours)	79.43	80.75	83.41	85.70	90.41	83.94
LL-Cp [21]	77.49	79.22	81.89	84.51	89.18	82.46
+ BoostLU (Ours)	79.53	81.04	83.40	85.85	90.39	**84.04**

Synergy effect



HP Sensitivity



References

[1] Zhou et al. Learning Deep Features for Discriminative Localization. CVPR 2016
[2] Zhang et al. Adversarial Complementary Learning for Weakly Supervised Object Localization. CVPR 2018
[3] Kim et al. Large Loss Matters in Weakly Supervised Multi-Label Classification. CVPR 2022