Large Loss Matters in Weakly Supervised Multi-Label Classification - Supplementary Material -

Youngwook Kim[1*] Jae Myung Kim[2*] Zeynep Akata[2,3,4] Jungwoo Lee[1,5]

[1]Seoul National University [2]University of Tübingen
[3]Max Planck Institute for Intelligent Systems [4]Max Planck Institute for Informatics [5]HodooAI Lab

A . Hyperparameter settings

Throughout our experiments, we report the performance of the model with the highest mAP in the validation set.

For §4.1, we search hyperparameter Δ_{rel} in $\{0.1, 0.2, 0.3, 0.4, 0.5\}$ and learning rate by dividing the range between values in $\{0.01, 0.001, 0.0001, 0.00001\}$ into quarters. We train the model for 10 epochs in "End-to-end" scheme. In "LinearInit." we train the model for 25 epochs with freezing feature extractor and then fine-tune the best model for 10 epochs. Also, we use the same validation set with previous work [7] which use 20% of training set for validation.

For §4.2, we train the model for 20 epochs and use the learning rate 2e-5. We search the hyperparameter Δ_{rel} in {0.005, 0.01}. Both for §4.1 and §4.2, we set the learning rate of the last fully connected layer as 10 times larger than feature extractor's learning rate.

B . Experiments for absolute variant

In this section, we conduct experiments using the *absolute* value of loss as a large loss criterion, rather than using a *relatively* largest value in a mini-batch. That is, $R(t) = R_0 - t \cdot \Delta_{abs}$ where R_0 is an initial threshold and Δ_{abs} determines the speed of decrease of the threshold. Using this criterion, we convert our methods LL-R, LL-Ct, and LL-Cp into LL-R$_{abs}$, LL-Ct$_{abs}$, and LL-Cp$_{abs}$. We perform hyperparameter search for R_0 in $\{1.0, 1.5, 2.0\}$ and Δ_{abs} in $\{0.1, 0.15, 0.2\}$. We compare the performance of absolute variants with Naive AN and LL-relative, which is the highest performance value among LL-R, LL-Ct, and LL-Cp with the relative threshold.

Results. Table 1 showes the quantitative results in artificially created partial label datasets with fine-tuning the model's entire weight from the beginning. We observe that these absolute variants perform similarly to LL-relative. Especially for NUSWIDE dataset, LL-Ct$_{abs}$ has a +6.6 mAP gain compared to Naive AN, +0.3 mAP gain compared to LL-relative.

Table 2 represents the quantitative results in OpenImages V3 dataset with real partial label. Compared to Naive AN, our absolute variants show a +0.3 $\sim$ +0.4 mAP performance gain.

Method	End-to-end VOC	COCO	NUSWIDE	CUB
Naive AN	85.1	64.1	42.0	19.1
LL-relative (Ours)	**89.2**	**71.0**	48.3	**20.4**
LL-R$_{abs}$ (Ours)	89.0	70.3	47.4	20.1
LL-Ct$_{abs}$ (Ours)	89.0	70.3	**48.6**	19.8
LL-Cp$_{abs}$ (Ours)	88.7	70.5	48.0	19.8

Table 1. **Quantitative results in artificially created partial label datasets.**

Method	G1	G2	G3	G4	G5	All Gs
Naive AN	77.1	78.7	81.5	84.1	88.8	82.0
LL-relative (Ours)	77.7	79.3	82.1	84.7	89.4	**82.6**
LL-R$_{abs}$ (Ours)	77.9	78.8	81.7	84.1	88.9	82.3
LL-Ct$_{abs}$ (Ours)	77.5	79.1	81.8	84.4	89.3	82.4
LL-Cp$_{abs}$ (Ours)	77.7	79.0	81.7	84.3	88.9	82.3

Table 2. **Quantitative results in OpenImages V3 dataset with real partial label.**

C . Model Explanation

In this section, we visualize the Class Activation Map (CAM) result of ground truth categories for some of the COCO dataset test images. We compare the mapping results with Naive AN and ROLE. As shown in Figure 1, our LL-Ct has the capability of capturing the location of an object more than previous methods. This indicates that our method's explanation is better aligned to human's explanation.

*Equal contribution.

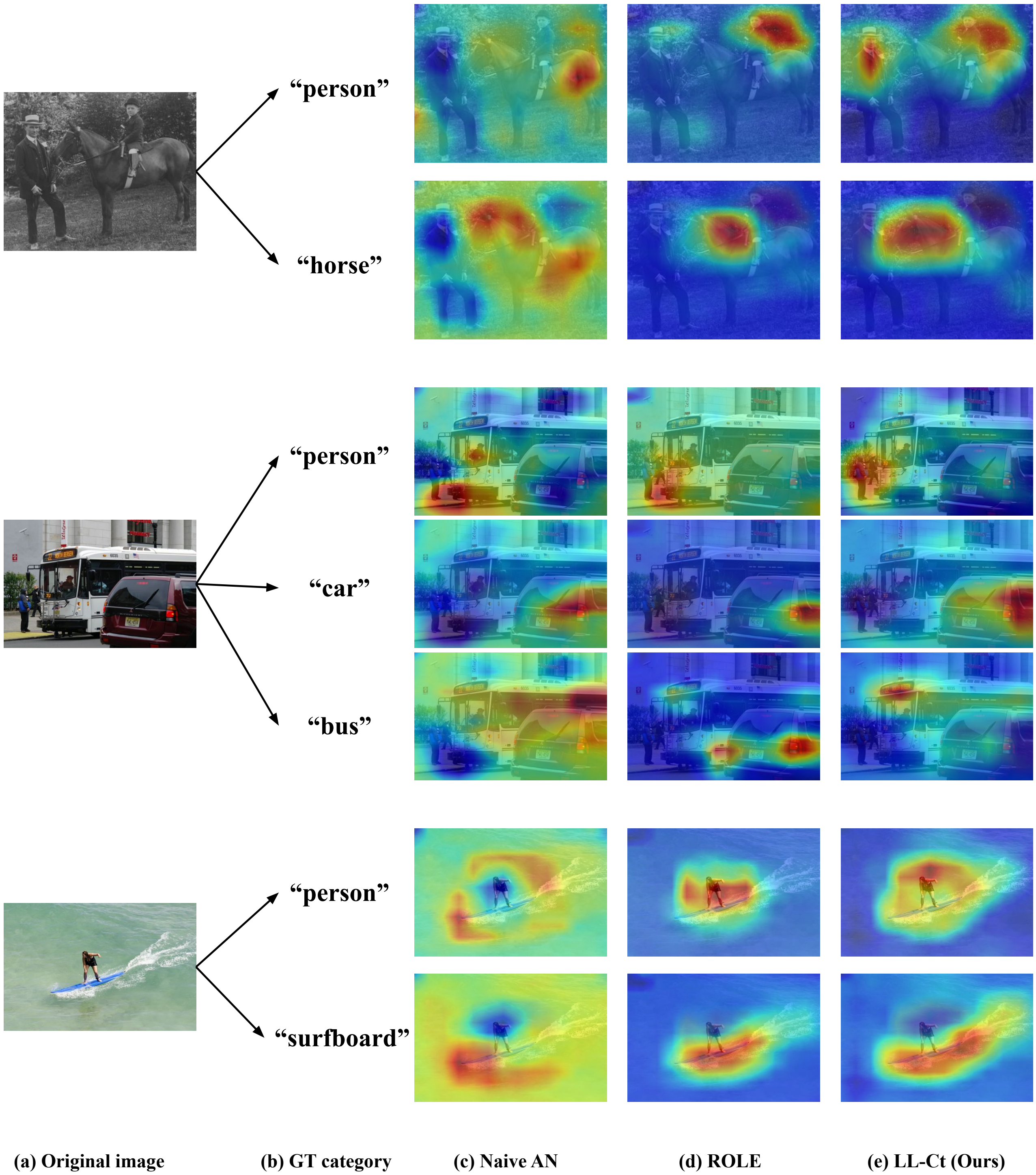

Figure 1. **Class Activation Mapping results in COCO test images.**