Supplementary Material for "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, Ecole normale supérieure, CNRS, PSL Research University [5]MPI for Intelligent Systems

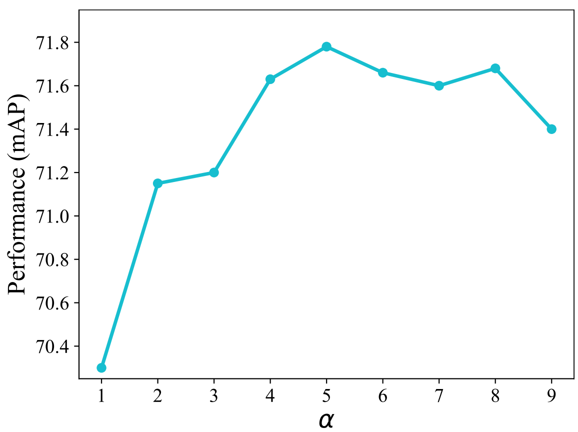

Figure 1. **Hyperparameter sensitivity with respect to α.**

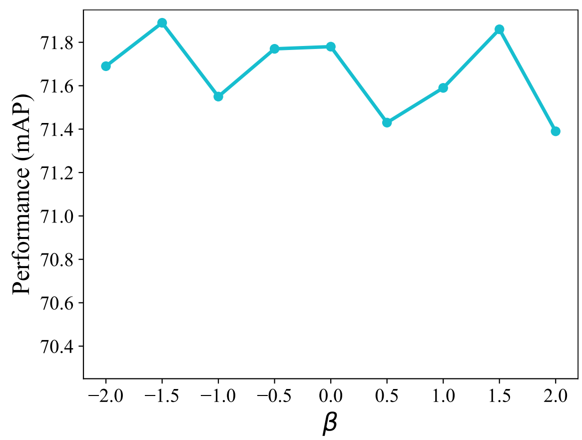

Figure 2. **Hyperparameter sensitivity with respect to β.**

A . Hyperparameter sensitivity

In this section, we check the hyperparameter sensitivity of our proposed BoostLU. All experiments are conducted for LL-Ct + BoostLU in a COCO dataset. Figure 1 shows the experimental results for various α with fixed $\beta = 0$. Note that $\alpha = 1$ refers to the case of the original LL-Ct since positive attribution scores are not scaled. When α exceeds 1, performance rises as the attribution score damaged by the false negative begins to be compensated. The performance gradually increases and peaks at $\alpha = 5$. Figure 2 shows the performance trend for various β with fixed $\alpha = 5$. According to the results, the value of β does not significantly affect the model's performance. These two figures represent that our BoostLU is generally robust to its hyperparameters α and β.

B . Additional discussion about Table 1

When LL-R and BoostLU are used in training, but BoostLU is not used in inference (fifth row), the network is *optimized* via BoostLU-activated attribution scores. So after training, pre-activated attribution scores for positive labels would become smaller, even though false negatives are further alleviated. It leads to worse model performance, even lower than when BoostLU is not used in training, but only LL-R is used in training (fourth row). From this result, we can confirm the importance of applying BoostLU in inference to obtain performance gain.

*Corresponding author.