Supplementary Material for "Bridging the Gap between Model Explanations in Partially Annotated Multi-label Classification"

Youngwook Kim1
Cordelia Schmid4
Jae Myung Kim2
Zeynep Akata2,5
Jieun Jeong1,3
Jungwoo Lee1,3*

1Seoul National University  2University of Tübingen  3HodooAI Lab  4Inria, Ecole normale supérieure, CNRS, PSL Research University  5MPI for Intelligent Systems

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 β = 0. Note that α = 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 α = 5. Figure 2 shows the performance trend for various β with fixed α = 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.