IGNORE: Information Gap-based False Negative Loss Rejection for Single Positive Multi-Label Learning

GyeongRyeol Song, Noo-ri Kim, Jin-Seop Lee, and Jee-Hyong Lee*

Sungkyunkwan University, Suwon, South Korea
{thd7524, pd99j, wlstjq0602, john}@skku.edu

Abstract. Single Positive Multi-Label Learning (SPML) is a method for a scarcely annotated setting, in which each image is assigned only one positive label while the other labels remain unannotated. Most approaches for SPML assume unannotated labels as negatives ("Assumed Negative", AN). However, with this assumption, some positive labels are inevitably regarded as negative (false negative), resulting in model performance degradation. Therefore, identifying false negatives is the most important with AN assumption. Previous approaches identified false negative labels using the model outputs of assumed negative labels. However, models were trained with noisy negative labels, their outputs were not reliable. Therefore, it is necessary to consider effectively utilizing the most reliable information in SPML for identifying false negative labels. In this paper, we propose the **I**nformation **G**ap-based False **N**egative L**O**ss **RE**jection (**IGNORE**) method for SPML. We generate the masked image that all parts are removed except for the discriminative area of the single positive label. It is reasonable that when there is no information of an object in the masked image, the model's logit for that object is low. Based on this intuition, we identify the false negative labels if they have a significant model's logit gap between the masked image and the original image. Also, by rejecting false negatives in the model training, we can prevent the model from being biased to false negative labels, and build more reliable models. We evaluate our method on four datasets: Pascal VOC 2012, MS COCO, NUSWIDE, and CUB. Compared to previous state-of-the-art methods in SPML, our method outperforms them on most of the datasets.

Keywords: Multi-Label Learning · Single Positive Multi-Label Learning · Information Gap · False Negative Rejection

1 Introduction

Multi-label learning is an approach that trains a model using images with multiple labels, aiming to build a model that is able to predict all relevant labels for the images with multiple labels. Unlike multi-class learning, where the model aims to

* Corresponding author

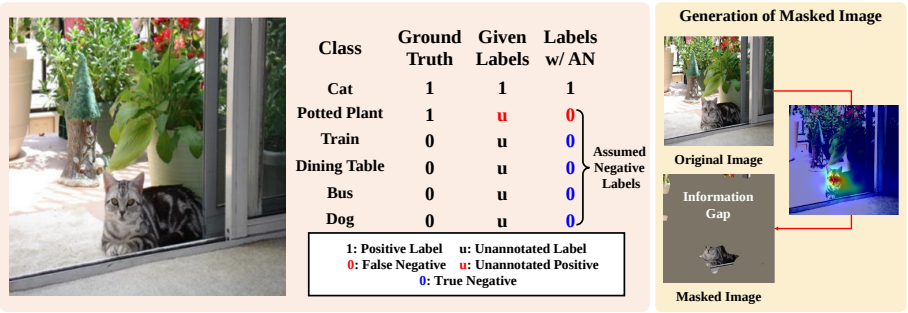


Fig. 1: Example of an image for Single Positive Multi-Label Learning (SPML) and our main idea for SPML. Ground truth is 1 if there is a corresponding object, otherwise 0. In SPML, only one positive label is given, and all the other labels are not given, i.e., unannotated. In Assumed Negative (AN) assumption, all unannotated labels are assumed to be negative. Here, false negative labels mean the unannotated positive labels, which are incorrectly assumed to be negative. To identify the false negative labels, we first extract the discriminative areas of the single positive label and generate the masked image. Then, we utilize the information gap between the original image and the masked image.

predict only a single label for the input image, multi-label learning is much more challenging because the model aims to predict multiple labels associated with the input image. Moreover, multi-label learning is more practical and realistic than multi-class learning, as the majority of real-world images contain multiple objects [2,3,17,20,22,31]. To achieve effective multi-label learning, similar to other supervised learning tasks, a fully labeled dataset is required. However, real-world images typically contain tens to thousands of different classes, and fully annotating all of them is extremely labor-intensive and practically unfeasible.

To reduce labor-intensive costs, researches on Single Positive Multi-Label Learning (SPML) have been increasing recently [5,12,13,26,29,33]. As shown in Fig. 1, in SPML, only one positive label is given and the others are unannotated for each images. To deal with unannotated labels in SPML, most approaches [12,13,26,29] assumed unannotated labels as negative (Assumed Negative, AN). This assumption has the advantage of providing supervision for all unannotated labels. Also, this assumption is reasonable because most multi-label datasets are dominated by negatives. However, by assuming all unannotated labels as negatives, it inevitably provides incorrect supervision for actually positive labels among the unannotated ones. In other words, these labels are false negative labels. These false negative labels can consistently cause confusion during model training, and impair the model's generalization ability.

Therefore, most studies [5,12,13,26,29] are focusing on identifying false negatives in SPML. They usually identify false negatives by utilizing model output's confidence or losses. In more detail, they identify the labels with abnormally high values of the model's outputs [26,29] or losses [12,13] as false negative labels.

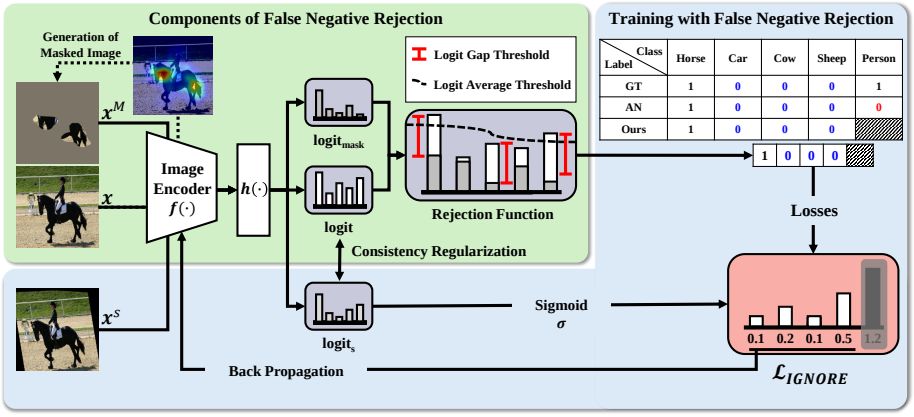


Fig. 2: Overview of the proposed method. In our method, we generate a masked image through the single positive label. Then, we utilize the logit gap between the masked image and the original image to identify false negatives and reject them in training ($\mathcal{L}_{IGNORE}$).

Because it implies that unannotated labels could be positive labels. However, the confidences or losses for the false negative labels are not reliable because models could learn them with noisy labels. To address this issue, it is necessary to consider robust criterion that are insensitive to noisy labels.

Since the single positive labels are the most reliable information in SPML, the model's output for the single positive label is reliable . Therefore, we can capture the discriminative areas [11, 15, 28, 32, 34] of objects and generate a masked image that includes only the discriminative areas of the single positive label (see Fig. 1). The masked image contains only the information corresponding to the single positive label. On the other hand, the original image contains information corresponding to ground truth labels. It is reasonable that when there is no information of an object in the image, the model's output for that object is low. Therefore, for false negatives, the model's output for the original image is expected to be higher than the model's output for the masked image. By utilizing the model's output gap resulting from the information gap between the images, false positive labels would be effectively identified.

In this paper, we propose the **I**nformation **G**ap-based false **N**egative L**O**ss **RE**jection (**IGNORE**) method for SPML. By rejecting the false negatives in the model training, we can prevent the model from being biased to false negative labels, and build more reliable models. We evaluate our method on various benchmark datasets such as PASCAL VOC 2012 [9], MS COCO [16], NUSWISE [4], and CUB [27]. Our method also outperforms most of the state-of-the-art methods in SPML.

2 Related Works

Problems with AN loss. In SPML, Assume Negative (AN) is the main approach that provides supervision to unannotated labels. Since multi-label datasets are dominated by negatives [21], assuming unannotated labels as negatives can be a reasonable and straightforward approach in situations where most of the label information is unknown. Also, Binary Cross-Entropy (BCE) loss in AN is referred to as AN loss. With AN loss, losses of all unannotated labels are calculated through the negative term of BCE. Thus, the model erroneously learns the unannotated positive labels (false negative labels) as negative labels. This confusion in multi-label learning results in bias to false negative labels and model performance drop. There are researches that point out issues caused by false negative label learning with AN loss [13,33]. In training with AN loss, BoostLU [13] observed that the attribution score of Class-Activation Map (CAM) is damaged due to false negative labels. To address this issue, they proposed an activation function that can boost the damaged CAM. EM [33] pointed out the issue that AN loss incurs the high loss gradient propagation problem for false negative labels. They proposed an entropy maximization loss that suppresses the gradient for unannotated labels with high prediction, without assuming unannotated labels as negatives.

In summary, learning false negative labels causes confusion in model training by continuously propagating the incorrect information to the model and affects the model's output. Therefore, it is necessary to consider the false negative labels when utilizing AN loss in SPML.

Approaches and Challenges for False Negatives. There are approaches that reject the identified false negative labels or regard them as positive labels. PLC [29] adopted a model prediction-based pseudo-labeling method. While training with AN loss, they generated pseudo-labels for unannotated labels using a fixed threshold. Using these pseudo-labels, They trained the model with unannotated labels once again. SCL [26] identified labels as false negatives with significantly high values of model predictions, and other methods [12, 13] used losses to identify false negatives. These methods proposed utilizing identified false negatives as positives in training or rejecting them from training.

In multi-label learning, usually the number of negatives overwhelms the number of positives in each sample. Therefore, it is challenging to identify as many false negatives as desired with high precision. Additionally, if the model incorrectly learns actual negative labels as positive, it can lead to significant bias in the model for those labels. As a result, generally rejecting the identified false negatives from training showed better performance.

Despite consideration of false negatives, however, because these methods are based on the AN loss, the model is easy to learn incorrect label information. The model learns assumed negative labels with incorrect label information, the distinct model outputs associated with false negative labels (e.g., high loss or high

probability) gradually become more similar to those of true negatives. That is, it becomes more challenging to identify false negatives from true negative labels using the information from assumed negative labels. Therefore, it is necessary to identify false negatives through new criterion rather than relying on the information from assumed negative labels.

3 Proposed Method

In this section, we propose the method of identifying and rejecting false negatives based on the information gap between the masked image and the original image. Firstly, we introduce the preliminary in Sec. 3.1. Then, we explain the proposed method in the following sections.

3.1 Preliminary

For single positive multi-label learning tasks, a dataset $\mathcal{D} = \{\mathbf{x}_n, \mathbf{y}_n\}_{n=1}^N$ is given, where $\mathbf{x}_n$ is an image and $\mathbf{y}_n \in \mathcal{Y} = \{u, 1\}^C$ is the vector of labels. N is the number of images, and C is the number of classes. The $\mathbf{y}_n$ consists of a single positive label and unknown labels. Except for the single positive, the others are unannotated labels. The single positive label is indicated by 1 and unannotated labels are indicated by u.

In SPML, most approaches assumed unannotated labels as negative, and trained models using AN (Assume Negative) loss as follows:

$$\begin{aligned}\mathcal{L}_{AN} = -\frac{1}{BC}\sum_{i=1}^{B}\sum_{j=1}^{C}[\mathbb{I}[y_{ij}=1]\cdot\log(p_{ij}) \\ +\mathbb{I}[y_{ij}=u]\cdot\log(1-p_{ij})]\end{aligned} \tag{1}$$

where y_{ij} is the label of the j-th class for the i-th image, and p_{ij} is the the model's prediction for y_{ij}. B is the batch size, C is the number of classes, and $\mathbb{I}$ is the indicator function that returns 1 if the condition is true and 0 otherwise. Since the AN loss is useful to give supervision to unannotated labels, our approach also based on the AN loss. However, the AN loss inevitably provides incorrect supervision for actually positive labels among the unannotated ones. Therefore, it is necessary to consider that effectively utilizing the most reliable information in SPML to minimize false negative problems.

3.2 Overview of Proposed Method

In this paper, we propose our **I**nformation **G**ap-based false **N**egative L**O**ss **RE**jection (**IGNORE**) method for SPML. The method for identifying false negative labels based on the information gap proceeds through the following steps. Firstly, we generate the masked image that all parts are removed except the discriminative area of the single positive label (see Fig. 2). Then, we compare

the model's logit gap between the masked image and the original image. If there is an object in the masked-out areas of the masked image, there is likely to be a substantial gap in the model's logit between the masked image and the original image. Because the object is absent in the masked image, in contrast to the original image. Based on this intuition, we identified the false negative labels if they have a significant model's output gap between the two images. The details of IGNORE's components are discussed in Sec. 3.3. As shown in Fig. 2, after identifying the false negative labels, we reject the losses of the corresponding negative labels to prevent the model from learning incorrect label information.

3.3 Components of False Negative Loss Rejection

In this section, we first introduce the process of generating the masked images that provide information to identify false negatives in our proposed method. And then, we introduce the logit gap threshold and the logit average threshold that compose our information gap-based false negative loss rejection, IGNORE.

Generation of Masked Image. We identify and reject the false negative labels based on the information gap between the masked image and the original image. In this section, we discuss the method of generating the masked images through the single positive labels. It is well-known that the class-activation map (CAM) [23, 32] focuses on the discriminative areas of objects [11, 15, 28, 34]. Furthermore, BoostLU [13] reported that the model trained with AN loss can extract reliable CAM compared to the CAM extracted by a model trained with fully-labeled dataset. Therefore, we extract the CAM for the single positive labels from the original image $\mathbf{x}$. The detailed explanation for extracting the CAM is in the supplementary material. Based on the CAM, we generate the mask as follows:

$$M_{ij} = \begin{cases} 1 & \text{if } A_{ij} > \gamma \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where A and M indicate the CAM and the mask, respectively. Also, γ is a CAM threshold to capture discriminative area for the positive label. By multiplying this mask with the original image $\mathbf{x}$ element-wise, we can generate a masked image $\mathbf{x}^M$ where the areas only related to single positive labels are preserved, and the rest of the image is masked out as follows:

$$\mathbf{x}^M = \mathbf{x} \odot M \tag{3}$$

The masked image is used to establish the Logit Gap Threshold for identifying the false negative labels.

Logit Gap Threshold. We design the criterion for identifying the false negative labels based on the logit gap between the masked image $\mathbf{x}^M$ and the original image $\mathbf{x}$. The logit gap G is defined as follows:

$$G_{ij} = \left| h(f(\mathbf{x}_i))_j - h(f(\mathbf{x}_i^M))_j \right| \tag{4}$$

where f and h denotes image encoder and classification head, respectively. We define the logit for the image as $h(f(\mathbf{x}))$, and the logit for the j-th class of the i-th image as $h(f(\mathbf{x}_i))_j$. Among the assumed negative labels, the labels with a high G are more likely to be false negatives.

As shown in Fig. 2, in the masked image, the object of the single positive label retains some of its information, but any information of the other objects is completely eliminated in the masked image. The value of G for the other objects in the image will be larger than the value of G for the single positive label. Based on this observation, we set the threshold τ using the average G value of the single positive labels. The τ is defined as follows:

$$\tau^t = \begin{cases} 0 & \text{if } t = 0 \\ \lambda \tau^{t-1} + (1-\lambda) \frac{\sum_{j=1}^{C} \sum_{i=1}^{B} I[y_{ij}=1] G_{ij}}{B} & \text{otherwise} \end{cases} \tag{5}$$

where B denotes the batch size. The gap threshold τ is adaptively updated at each iteration t using Exponential Moving Average (EMA). λ is a hyperparameter that determines the momentum decay of the EMA.

Logit Average Threshold. In this section, we introduce logit average threshold. We identify the false negatives that are actually present in the original image with the logit gap resulting from the information gap between the images. Therefore, even if a large logit gap occurs for labels that are not actually present in the original image, we do not need to identify these labels as false negatives labels with low logit values in the original image are likely to be true negatives that did not originally exist in the original image. To prevent these rejection of true negatives in our rejection method, we propose the logit average threshold.

It is evident that the logits of true negative labels are relatively low. Therefore, we only consider the rejection of the labels with a logit higher than the average of the logits for each class. We set the threshold μ as the average logit of assumed negative labels for each class. We select the labels with logits higher than the threshold μ. Through this process, we only reject the selected labels. Additionally, μ is self-adaptively updated using Exponential Moving Average (EMA) at each iteration. The threshold μ_j^t for the j-th class at time step t is as follows:

$$\mu_j^t = \begin{cases} 0 & \text{if } t = 0 \\ \lambda \mu_j^{t-1} + (1-\lambda) \frac{1}{N_j} \sum_{i=1}^{N_j} \mathbb{I}[y_{ij} = u] \cdot h(f(\mathbf{x}_i))_j & \text{otherwise} \end{cases} \tag{6}$$

where N_j denotes the number of assumed negative labels for the j-th class.

3.4 Training with False Negative Loss Rejection

Information Gap-based False Negative Loss Rejection. In this section, we describe our rejection function that incorporates both the logit average threshold and the logit gap threshold. The rejection function R incorporating the logit

average threshold and the logit gap threshold is defined as follows:

$$R_{ij} = \begin{cases} 0 & \text{if } h(f(\mathbf{x}_i))_j > \mu_j^t \text{ and } G_{ij} > \tau^t \\ 1 & \text{otherwise} \end{cases} \tag{7}$$

The rejection function R rejects the labels with G higher than the τ and a logit higher than μ. In other words, it identifies the labels with high logit values for the original image and a significant logit gap between the two images, then rejects them from the training.

By incorporating this function into the negative term of the AN loss, identified false negatives are rejected from the model training. Our IGNORE method that combined with AN loss and the rejection function R is as follows:

$$\begin{aligned} \mathcal{L}_{IGNORE} = -\frac{1}{S}\sum_{i=1}^{B}\sum_{j=1}^{C}(&\mathbb{I}[y_{ij}=1]\cdot\log(p_{ij}^s) \\ &+\mathbb{I}[y_{ij}=u]\cdot R_{ij}\cdot\log(1-p_{ij}^s)) \end{aligned} \tag{8}$$

where S denotes the sum of the single positive labels and the labels that are not rejected negative labels. p_{ij}^s denotes $\sigma(h(f(\mathbf{x}_i^s)_j)$ and σ denotes sigmoid function. $\mathbf{x}^s$ denotes strong augmented image. It is well known that strong augmentation improves the generalization ability of deep learning models across most datasets [1, 6, 7, 24, 30]. Therefore, we utilize the strong augmented images to enhance model generalization.

Single Positive Binary Classification (SPBC). We utilize the CAM of the single positive labels. Therefore, it is necessary to train the model with correct supervision for the single positive labels. The model learns given single positive labels through binary classification for a single positive class. Single Positive Binary Classification (SPBC) is defined as follows:

$$\mathcal{L}_{SPBC} = -\frac{1}{B}\sum_{i=1}^{B}\sum_{j=1}^{C}\mathbb{I}[y_{ij}=1]\cdot\log(p_{ij}) \tag{9}$$

Consistency Regularization. Our method identifies the false negatives based on the model's logits. Therefore, our model needs to output consistent logits on similar images. To ensure such consistency, we take into account the consistency regularization between original images and strongly augmented images used in our framework. We propose the consistency regularization to the logits between original images and strongly augmented images, enabling the model to have consistency across differently augmented images. This ensures that the model produces consistent logits for different views of the same image. Our consistency regularization is defined as follows:

$$\mathcal{L}_{reg} = -\frac{1}{BC}\sum_{i=1}^{B}\sum_{j=1}^{C}|h(f(\mathbf{x}_i)_j - h(f(\mathbf{x}_i^s))_j|^2 \tag{10}$$

Overall loss. The overall loss is as follows:

$$\mathcal{L} = \lambda_{SPBC} \cdot \mathcal{L}_{SPBC} + \mathcal{L}_{IGNORE} + \lambda_{reg} \cdot \mathcal{L}_{reg} \tag{11}$$

where λ_{SPBC} and λ_{reg} are hyperparameters that denote the weights assigned to the respective loss terms.

4 Experiments

In this section, we present our experimental results and compare them with previous SPML approaches. Furthermore, we conduct an ablation study to analyze why our method is effective in SPML. We use mean average precision (mAP) as the comparison metric with previous approaches.

4.1 Experimental Setup

Datasets. We use four benchmark datasets to demonstrate our experimental results: PASCAL VOC 2012 [9], MS COCO 2014 [16], NUSWIDE [4], and CUB [27]. Following the settings of previous research [5], we create a single positive multi-label setting for the four benchmark datasets by randomly retaining one positive label and removing all other annotations. In particular, for the CUB dataset [27], we use bird attributes as target classes instead of bird categories, following the setup of ROLE [5]. For each dataset, we reserve 20% of the train set as a validation set, and select the best model based on its performance on the validation set.

Models. For a fair comparison with previous approaches [5,12,13,19,29,33], we conduct experiments with two models [10, 17]. One of the models is a ResNet-50 [10] that is pre-trained on ImageNet [8]. For ResNet-50, We employ the architecture of a convolutional layer followed by a 1x1 convolutional layer and global average pooling, the same as BoostLU [13].

The other one is a model with a transformer [25] architecture called Q2L [17]. For the image encoder of Q2L [17], we utilize the ResNet-50 [10] that is pre-trained on ImageNet [8]. We explain the details on how to obtain CAMs for each model in Supplementary details.

Hyperparameter Settings. For Q2L, we set batch sizes in {8, 16}, learning rates in {1e-4, 1e-5}. We use AdamW [18] as optimizer and OneCycleLR as scheduler. For ResNet-50, we set batch sizes in {8, 12, 16}, learning rates in {1e-4, 1e-5}, and use Adam [14] as optimizer. Each image is resized to 448×448. Only random HorizontalFlip is used for original images, while random horizontal flip and RandAugmentation [7] are used for strong augmentated images CAM threshold γ is set to 0.4 for ResNet-50 and 0.8 for Q2L, respectively. In Q2L, we set the CAM threshold to 0.8. And, we set loss weights λ_{reg} to 0.5, and λ_{SPBC} to 0.5 except COCO dataset. On COCO dataset, we set loss weights λ_{reg} to 0.1,

Table 1: Experimental results of our method with Q2L model on four SPML benchmark datasets. Metric presented in the table is mAP (mean Average Precision). Additionally, bold font indicates the highest performance, while underlining indicates the second highest performance. Our method demonstrates the highest performance on most datasets.

Method	Dataset			
	VOC	COCO	NUS	CUB
AN	87.6	72.3	48.5	18.6
WAN [5]	89.2	73.5	48.5	22.5
EPR [5]	88.8	72.7	49.3	23.1
AN-LS [5]	88.0	70.9	47.1	16.3
ROLE [5]	88.1	69.6	44.5	14.2
EM [33]	89.2	73.2	48.7	22.2
EM + APL [33]	89.2	73.1	48.6	**23.6**
PLC + LAC [29]	89.6	75.6	51.1	23.3
Ours	**90.1**	**76.3**	**52.8**	<u>23.4</u>

and λ_{SPBC} to 0.5. In ResNet-50, we set the CAM threshold in $\{0.4, 0.5\}$. And, we set the loss weights λ_{reg} to 0.5, and λ_{SPBC} to 0.1 except CUB dataset. On CUB dataset, we set loss weights λ_{reg} to 0.5, and λ_{SPBC} to 0.5. Considering that the updates for the threshold τ might not be sufficient in the early stages of training, we set the warm-up phase. In warm-up phase, our model is trained with AN loss.

4.2 Experimental Results

Performance with Q2L model. We compare our experimental results with recent approaches for the Q2L model on four datasets: Weak Assume Negative (WAN) [5], Expected Positive Regularization (EPR) [5], Assume Negative-Label Smoothing (AN-LS) [5], ROLE [5], Entropy-maximization (EM) with Asymmetric Pseudo-Labeling (APL) [33], and Pseudo-Labeling Consistency Regularization (PLC) with Label-Aware Global Consistency Regularization (LAC) [29].

As shown in Tab. 1, our method outperforms recent approaches on the VOC, COCO, and NUSWIDE datasets. In particular, on the NUSWIDE dataset, our method achieved an mAP of 1.7, higher than the previous best-performed approach. For CUB, our method do not achieve the best performance, but it shows the second-best results. The gap in performance with the best-performed approach is very small, so it can be considered comparable.

Performance with ResNet-50 model. For ResNet-50, we compare the experimental results with the following recent approaches: Label-Smoothing (LS) [19],

Table 2: Experimental results of our method with ResNet-50 model on four SPML benchmark datasets. Detailed ingredient of the table (metrics, bold, and underline) is the same as in Table 1.

Method	Dataset			
	VOC	COCO	NUS	CUB
AN	85.89	64.92	42.27	18.31
LS [19]	87.90	67.15	43.77	16.26
ASL [21]	87.76	68.78	46.93	18.81
ROLE [5]	87.77	67.04	41.63	13.66
ROLE + LI [5]	88.26	69.12	45.98	14.86
EM [33]	89.09	70.70	47.15	20.85
EM + APL [33]	<u>89.19</u>	70.87	47.59	<u>21.84</u>
LL-R [12]	88.27	70.70	48.76	19.56
+ BoostLU [13]	**89.29**	**72.89**	<u>49.59</u>	19.80
Ours	<u>89.19</u>	<u>72.00</u>	**50.08**	**21.90**

Assymetric loss (ASL) [21], ROLE [5], ROLE with LinearInit (LI) [5], Entropy-maximization (EM) with Asymmetric Pseudo-Labeling (APL) [33], Large Loss Rejection (LL-R) [12], and Large Loss Rejection (LL-R) [12] with BoostLU [13].

As shown in Tab. 2, our method exhibited the highest performances on the NUSWIDE and CUB dataset compared to recent approaches. On VOC and COCO, our method achieves the second-best performances. While not achieving the best performance on all datasets, We observe that our method with ResNet-50 performs well enough to be comparable to state-of-the-art approaches.

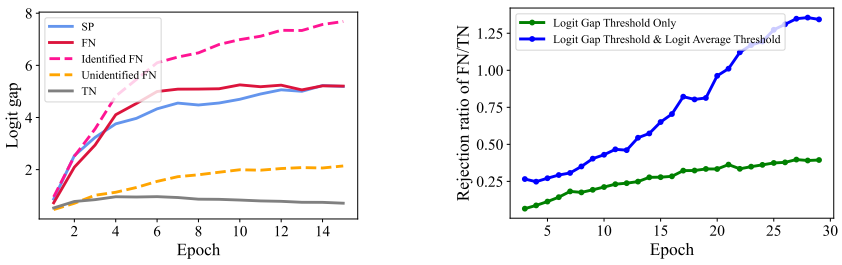
4.3 Ablation Studies

We analyze the effectiveness of our framework components, analysis of logit gap, effectiveness of logit average threshold, comparison for false negative identification, and sensitivity analysis of CAM threshold.

Table 3: Ablation study of our method with Q2L model.

Consistency Reg.	IGNORE + SPBC	w/ Logit Avg. Threshold	VOC	COCO	NUSWIDE	CUB
			87.9	73.5	49.2	19.7
✓			88.8	73.6	47.5	11.1
✓	✓		90.0	76.2	52.0	22.8
✓	✓	✓	**90.1**	**76.3**	**52.8**	**23.4**

Analysis of Proposed Components. To demonstrate the effectiveness of the components in our framework, we conduct an ablation study using the PASCAL VOC, COCO, NUSWIDE and CUB datasets. In Tab. 3, the first row is the case where the model is trained with only AN loss. As shown in Tab. 3, when all components are included, the highest performance is achieved. The logit average threshold shows meaningful improvements in NUSWIDE and CUB. Considering that NUSWIDE and CUB have a relatively large number of classes, it is evident that the logit average threshold works more effectively in datasets with a larger number of classes. Also, when only consistency regularization is applied without IGNORE and SPBC, there was a significant performance drop in NUSWIDE and CUB. This observation indicates that simply applying consistency regularization alone does not significantly improve performance. In contrast, applying it with IGNORE and SPBC results in notable performance improvements. It means that our regularization is mutually effective when combined with our IGNORE method.



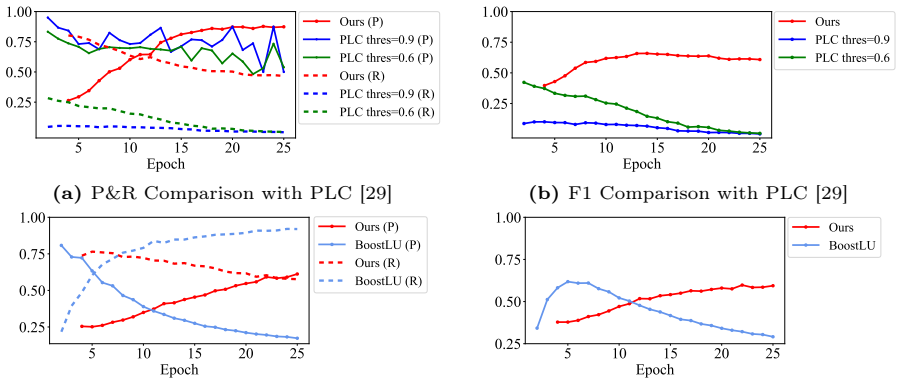
(a) Logit Gap Distribution **(b)** Ablation Study of Logit Average Threshold

Fig. 3: Analysis of Logit Gap Threshold and Logit Average Threshold on PASCAL VOC 2012. SP, TN, and FN indicate the single positive, true negative, and false negative, respectively.

Analysis of Logit Gap. In this section, we examine whether there is a significant difference in the logit gap between true negatives and false negatives. Additionally, by observing the logit gap of single positives, we validate the effectiveness of the logit gap threshold. Figure 3a depicts the average logit gap per epoch observed during model training with our framework. As shown in Fig. 3a, there is a significant difference in the gap between false negatives and true negatives. Furthermore, it is observed that the gap for the single positives is lower than that of the false negatives. In other words, during model training with our framework, the logit gap is a suitable criterion for identifying false negatives, and the logit gap threshold works effectively.

Effectiveness of Logit Average Threshold. In the process of rejecting false negatives, to minimize incorrect rejections (rejections of true negatives), we pro-

posed the logit average threshold. To verify the effectiveness of the logit gap threshold, we examine in this section how the logit average threshold mitigates the rejection of true negatives. Figure 3b illustrates the rejection ratio starting from the end of the warm-up phase. The y-axis represents FN/TN ratio, where FN is the number of unannotated positives (False Negatives) rejected by our method, and TN is the number of true negatives rejected by our method. As shown in Fig. 3b, when logit average threshold is applied, FN/TN noticeably decreases. This indicates that our logit average threshold effectively mitigates the rejection of true negatives.



(a) P&R Comparison with PLC [29]

(b) F1 Comparison with PLC [29]

(c) P&R Comparison with BoostLU [13]

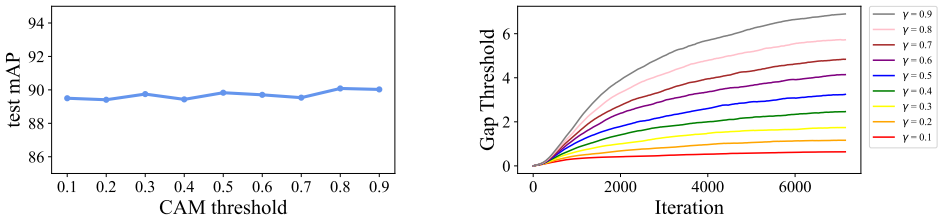
(d) F1 Comparison with BoostLU [13]

Fig. 4: Comparison with recent approaches. Experiments are conducted on PASCAL VOC 2012 dataset. Q2L model is used in comparison with PLC [29] and ResNet-50 is used in comparison with BoostLU [13]. BoostLU represents the case where BoostLU is applied to LL-R [12]. P and R represent precision and recall, respectively.

Comparison for False Negative Identification. We show our method's ability to identify false negatives through a comparison with recent approaches [13, 29]. We compare with PLC [29] and BoostLU [13], which include LL-R [12].

For PLC [29], precision is very high; however, as a trade-off, recall is very low (see Fig. 4a). Therefore, many false negative labels are not identified, and the model continuously learns these labels with incorrect label information. For BoostLU [13], as training progresses, recall becomes very high. Therefore, it could identify many false negative labels. However, due to the significant decrease of precision, a large number of true negative labels are also incorrectly identified as false negatives.

Compared to these methods, our approach does not exhibit a significant trade-off between precision and recall. Furthermore, as shown in Fig. 4b and 4d, F1 score of our approach exhibits a relatively high score and becomes increases, while the F1 scores of other methods becomes sharply decrease. This observation demonstrates that our method, based on logit gap threshold and logit average threshold, has an remarkable ability in identifying false negatives.



(a) Comparison of mAP according to CAM threshold

(b) Variation of the Gap threshold according to CAM threshold

Fig. 5: Analysis of CAM Threshold and Gap Threshold.

Sensitivity Analysis of CAM Threshold. We analyze the sensitivity of the model performance according to changes in the CAM threshold. Fig. 5a shows that the model's performance is not sensitive to the CAM threshold. Also, Fig. 5b indicates that as the CAM threshold γ increases, the logit gap threshold τ also automatically increases.

With a higher CAM threshold, the region exceeding the threshold in the CAM area decreases, causing more areas of the image to be obscured in the masked image. Consequently, the information gap between the masked image and the original image for a single positive label becomes large.

Our gap threshold is the average logit gap between the original image and the masked image for a single positive label. Therefore, as the CAM threshold increases, the gap threshold automatically increases as well. In other words, the gap threshold we propose can adapt dynamically to changes in the hyperparameter CAM threshold.

5 Conclusion

We proposed a new approach for identifying false negative labels in SPML. Our approach was based on the model's output gap resulting from the information gap between the images. With rejecting the false negative labels, we trained the models with minimizing bias to false negative labels. Through various experimental results, we demonstrated the superiority of our proposed for identifying false negative labels in assumed negative labels. Our approach showed performance improvements, and achieved state-of-the-art results in most datasets. Our proposed method generated the masked image with the CAM of the single positive label. Then, we identified false negatives with the information gap between the masked image and the original image. By rejecting identified false negatives, we prevented the model from being biased to false negative labels and built more reliable models. In future work, we would consider utilizing modules for more precise CAM extraction or leveraging the foundational model for generating the masked images.

Acknowledgements. This work was partly supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2019-0-00421, AI Graduate School Support Program), the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2024-00352717), and institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.RS-2024-00437633, Development of Open-ended Alignment Fluxional AI for Everchanging Environment and Value)

References

1. Berthelot, D., Carlini, N., Cubuk, E.D., Kurakin, A., Sohn, K., Zhang, H., Raffel, C.: Remixmatch: Semi-supervised learning with distribution alignment and augmentation anchoring. arXiv preprint arXiv:1911.09785 (2019)
2. Beyer, L., Hénaff, O.J., Kolesnikov, A., Zhai, X., Oord, A.v.d.: Are we done with imagenet? arXiv preprint arXiv:2006.07159 (2020)
3. Chen, Z.M., Wei, X.S., Wang, P., Guo, Y.: Multi-label image recognition with graph convolutional networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5177–5186 (2019)
4. Chua, T.S., Tang, J., Hong, R., Li, H., Luo, Z., Zheng, Y.: Nus-wide: a real-world web image database from national university of singapore. In: Proceedings of the ACM international conference on image and video retrieval. pp. 1–9 (2009)
5. Cole, E., Mac Aodha, O., Lorieul, T., Perona, P., Morris, D., Jojic, N.: Multi-label learning from single positive labels. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 933–942 (2021)
6. Cubuk, E.D., Zoph, B., Mane, D., Vasudevan, V., Le, Q.V.: Autoaugment: Learning augmentation policies from data. arXiv preprint arXiv:1805.09501 (2018)
7. Cubuk, E.D., Zoph, B., Shlens, J., Le, Q.V.: Randaugment: Practical automated data augmentation with a reduced search space. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. pp. 702–703 (2020)
8. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
9. Everingham, M., Eslami, S.A., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes challenge: A retrospective. International journal of computer vision **111**, 98–136 (2015)
10. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
11. Jiang, P.T., Yang, Y., Hou, Q., Wei, Y.: L2g: A simple local-to-global knowledge transfer framework for weakly supervised semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16886–16896 (2022)
12. Kim, Y., Kim, J.M., Akata, Z., Lee, J.: Large loss matters in weakly supervised multi-label classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14156–14165 (2022)

13. Kim, Y., Kim, J.M., Jeong, J., Schmid, C., Akata, Z., Lee, J.: Bridging the gap between model explanations in partially annotated multi-label classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3408–3417 (2023)
14. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
15. Lee, S., Lee, M., Lee, J., Shim, H.: Railroad is not a train: Saliency as pseudo-pixel supervision for weakly supervised semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5495–5505 (June 2021)
16. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. pp. 740–755. Springer (2014)
17. Liu, S., Zhang, L., Yang, X., Su, H., Zhu, J.: Query2label: A simple transformer way to multi-label classification. arXiv preprint arXiv:2107.10834 (2021)
18. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
19. Lukasik, M., Bhojanapalli, S., Menon, A., Kumar, S.: Does label smoothing mitigate label noise? In: International Conference on Machine Learning. pp. 6448–6458. PMLR (2020)
20. Rajeswar, S., Rodriguez, P., Singhal, S., Vazquez, D., Courville, A.: Multi-label iterated learning for image classification with label ambiguity. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4783–4793 (2022)
21. Ridnik, T., Ben-Baruch, E., Zamir, N., Noy, A., Friedman, I., Protter, M., Zelnik-Manor, L.: Asymmetric loss for multi-label classification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 82–91 (2021)
22. Ridnik, T., Sharir, G., Ben-Cohen, A., Ben-Baruch, E., Noy, A.: Ml-decoder: Scalable and versatile classification head. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 32–41 (2023)
23. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision. pp. 618–626 (2017)
24. Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C.A., Cubuk, E.D., Kurakin, A., Li, C.L.: Fixmatch: Simplifying semi-supervised learning with consistency and confidence. Advances in neural information processing systems **33**, 596–608 (2020)
25. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
26. Verelst, T., Rubenstein, P.K., Eichner, M., Tuytelaars, T., Berman, M.: Spatial consistency loss for training multi-label classifiers from single-label annotations. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 3879–3889 (2023)
27. Wah, C., Branson, S., Welinder, P., Perona, P., Belongie, S.: The caltech-ucsd birds-200-2011 dataset (2011)
28. Wang, Y., Zhang, J., Kan, M., Shan, S., Chen, X.: Self-supervised equivariant attention mechanism for weakly supervised semantic segmentation. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2020)

29. Xie, M.K., Xiao, J., Huang, S.J.: Label-aware global consistency for multi-label learning with single positive labels. Advances in Neural Information Processing Systems **35**, 18430–18441 (2022)
30. Xie, Q., Dai, Z., Hovy, E., Luong, T., Le, Q.: Unsupervised data augmentation for consistency training. Advances in neural information processing systems **33**, 6256–6268 (2020)
31. Yun, S., Oh, S.J., Heo, B., Han, D., Choe, J., Chun, S.: Re-labeling imagenet: from single to multi-labels, from global to localized labels. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2340–2350 (2021)
32. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2921–2929 (2016)
33. Zhou, D., Chen, P., Wang, Q., Chen, G., Heng, P.A.: Acknowledging the unknown for multi-label learning with single positive labels. In: European Conference on Computer Vision. pp. 423–440. Springer (2022)
34. Zhou, T., Zhang, M., Zhao, F., Li, J.: Regional semantic contrast and aggregation for weakly supervised semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4299–4309 (2022)