Appendices

A. Training Details

Unless stated otherwise, all experiments were conducted with the following training configuration. As a default, we used the TResNet-M model [25], pre-trained on ImageNet-21k dataset [24]. The model was fine-tuned using Adam optimizer [14] and 1-cycle cosine annealing policy [27] with a maximal learning rate of 2e-4 for training OpenImages and MS-COCO, and 6e-4 for training LVIS. We used true-weight-decay [22] of 3e-4 and standard ImageNet augmentations. For fair comparison to previously published results on OpenImages V3, we also trained a ResNet-101 model, pre-trained on ImageNet.

In the OpenImages experiments we used the following hyper-parameters: $\eta = 0.05$, $K = 200$, $\gamma_u = 7$, $\gamma_- = 2$ and $\gamma_+ = 1$. In LVIS we used: $\gamma_u = 1$, $\gamma_- = 0$ and $\gamma_+ = 0$.

B. Soft Label Prior

Herein, we will explore a soft alternative for integrating the label prior in the loss. We follow equation (3) and define the un-annotaetd weights by

$$\omega_c = \exp(-\alpha \hat{P}_r(c)) \tag{12}$$

where α is the decay factor. In Table 5 we compare the soft label prior to the configuration used in section 4.1.

Method	mAP(C)	mAP(O)
P-ASL, *Selective*	**86.72**	93.57
P-ASL, *Selective* (soft)	86.62	**93.59**

Table 5. **OpenImages (V6) results using soft label prior.** We used $\alpha = 10$.

As the soft label prior provided with lower mAP(C) results, we did not use it in our experiments.

C. Results on MS-COCO

In this section, we will present the results obtained on a partially annotated version of MS-COCO, based on the fixed per class (FPC) simulation scheme. Note that in this experiment, the class distribution measured by the number of annotations is no longer meaningful, as all classes have the same number of annotations. The mAP results, as well as the average precision (AP) scores for the class "Person", are presented in Figure 10. The *Negative* mode produces higher mAP (computed over all the classes) compared to the *Ignore* mode. However, as the frequent class "Person" is present in most of the images, the *Negative* mode is inferior, especially in the cases of a small number of annotations.
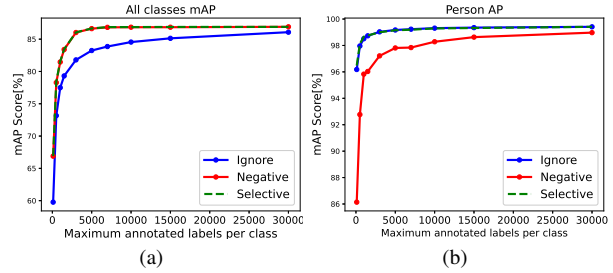

(a)　　　　　　　　(b)

Figure 10. **Results on MS-COCO (FPC).** .

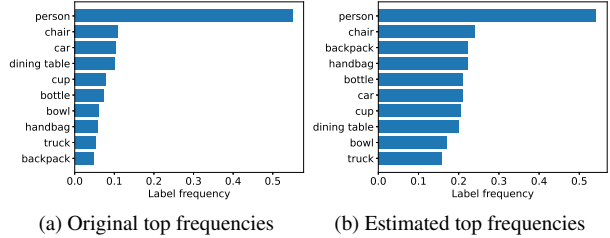(a) Original top frequencies　　(b) Estimated top frequencies

Figure 11. **Class frequency estimation in MS-COCO.** Top frequent classes measured by (a) original class distribution and (b) estimated class distribution. The estimated top 10 frequent classes are included in the original top classes.

Using the *Selective* approach, top results can be achieved for both mAP and the person AP. In Figure 11, we show the top 10 frequent classes obtained using our procedure for estimating the class distribution as described in section 4.2, and compared them to those obtained using the original class frequencies in MS-COCO. As can be seen, most of the frequent classes measured by the original distribution are also highly ranked by our estimator.

D. Frequent Classes in OpenImages

We add more results of the class distribution estimated by our approach (detailed in 4.2) for OpenImages dataset. See Figure 12.

E. Frequent Classes in LVIS

In Figure 13 we plot the top frequent classes in LVIS, obtained by our estimator detailed in section 4.2. Also in LVIS, it can be seen that the most estimated frequent classes are related to common objects as "Person", "Shirt", "Trousers", "Shoe", etc.

F. Discussion on the Hyper-parameter Selection

Two main factors that should be considered when selecting the hyper-parameters are: (1) the number of classes, and (2) the class distribution in the data. Usually, as the number of classes increases, so does the number of un-annotated
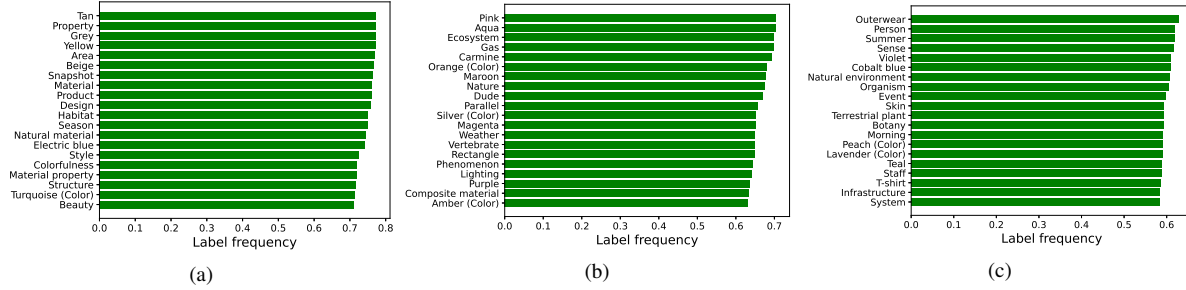
(a)   (b)   (c)

Figure 12. **Estimating the calss distribution in OpenImages.** Additional top 60 frequent classes as estimated by our approach.
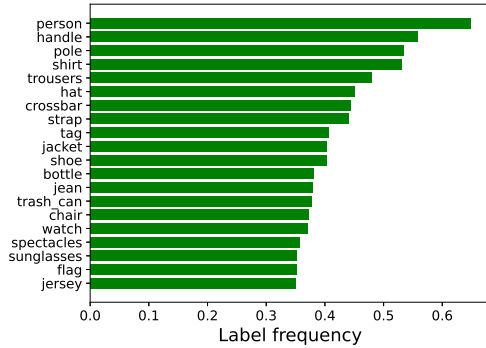

Figure 13. **Estimating the calss distribution in LVIS.** Top 20 frequent classes estimated by the *Ignore* model.

Figure 14. **Impact of the prior threshold η parameter.** mAP scores for different values of η on the OpenImages dataset.

	Ignore	Negative	Selective (our)
Scratch	79.1	81.3	83.0

Table 6. **mAP(C) for training OpenImages (V6) from scratch.**

H. Training from scratch

We conducted an experiment that demonstrates the impact of training OpenImages from scratch using the proposed approach. As can be seen in Table 6, using the *Selective* approach, a significant improvement is achieved also for the case of training without any pre-training. Interestingly, the *Ignore* mode converges slower and achieves lower results, probably because it effectively utilizes only a small portion of the data.
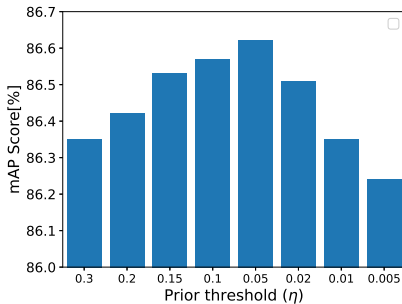
labels relative to the annotated ones. Thus, setting a higher decay rate for the un-annotated labels (γ^u) is desired to mitigate the negative-positive imbalance. Also, to preserve the impact of the *annotated* labels, γ^- and γ^+ should be lower. In addition, as the number of frequent classes is larger, a higher K and a lower η should be set to enable selectivity between labels.

G. Impact of the Prior Threshold Parameter

We show in Figure 14 the impact of the threshold parameter η on the mAP score for the OpenImages dataset.