Refine and Redistribute: Multi-Domain Fusion and Dynamic Label Assignment for Unbiased Scene Graph Generation (Supplementary Document)

Yujie Zang[1], Yaochen Li[2], Yuan Gao[3], Yimou Guo[4], Wenneng Tang[5], Yanxue Li[6], Meklit Atlaw[7]
School of Software Engineering, Xi'an Jiaotong University
[1,3,4,5,6]{zyj1176422374,gaoyuan419,guoym,twn29004,yanxue_li}@stu.xjtu.edu.cn
[2]yaochenli@mail.xjtu.edu.cn, [7]meklitm29@gmail.com

In this supplementary document, we provide additional experimental analysis and visualization results to better demonstrate our framework, which are not included in the main paper.

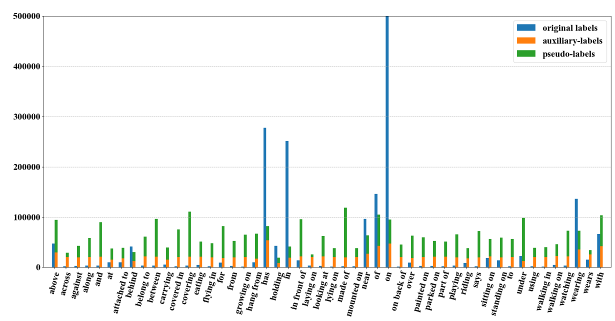

Figure 1. The instance distribution of VG, where the blue bars on the left show the original quantities, while the right bars represent DLA adjustments. Orange auxiliary-labels ensure steady instances for each category. Green pseudo-labels offer extra samples for fine-grained categories and filter redundant ones for coarse-grained categories.

Hyper-parameters	SGDet		
β	mR@20	mR@50	mR@100
$\beta = 0.3$	10.5	14.1	16.6
$\beta = 0.4$	10.6	14.3	16.0
$\beta = 0.5$	12.0	15.7	17.3
$\beta = 0.6$	11.9	15.6	17.8
$\beta = 0.7$	**12.4**	**16.1**	**18.6**
$\beta = 0.8$	11.4	15.1	17.3

Table 1. Analysis on VG for the impact of the label recombination coefficient β, which is tested within the scope of the common mixup method.

1. Additional Implementation Details

1.1. Analysis for Hyper-parameters

As discussed in the main text, during the process of recombining pseudo-labels, we consider that it is essential for the original labels to play a dominant role. The potential labels adopted from the unannotated data act more like reliable perturbations, enhancing the training diversity. The experimental results in Tab. 1 indicate that the effect decreases significantly when β is below 0.5, while it gradually approaches the optimal value above 0.5. Ultimately, we chose 0.7 as the experimental parameter.

2. Additional Visualization Demonstration

2.1. Redistribution of DLA

We illustrate the data distribution before and after label assignment in Fig. 1. Initially, the distribution exhibits significant imbalance, with many tail classes containing only a few instances that are barely noticeable in the histogram. Following the generation of siamese samples and the expansion of the sample space, we then derive auxiliary-labels, which are provided to each class through conditional sampling, offering a batch of sufficient instances to train the classifier. Subsequently, the classifier selectively filters high-confidence potential samples, enabling the recombination of pseudo-labels from unannotated data. This process further introduces more out-of-distribution instances to enrich each class. Ultimately, our adjusted labels significantly alleviate the long-tailed problem, achieving a more balanced training environment.

2.2. Qualitative Results

We present additional visual results in Fig. 2 to validate the effectiveness of our approach. In the first column, we display the predictions generated by MDF, which capture the majority of accurate inferences. By incorporating DLA, as shown in the second column, we notably enhance our ability to classify tail predicates. Moreover, our system

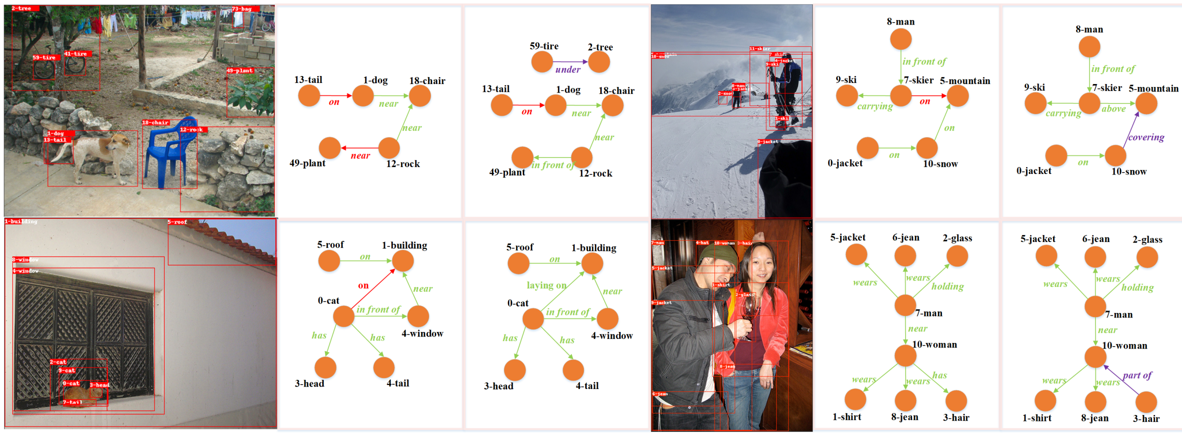

Figure 2. The visualization results for MDF and MDF+DLA. Green predicates represent correct matches with the ground-truth, while red ones are incorrect. Purple predicates represent acceptable predictions generated by our model (not in ground-truth).

adeptly uncovers previously unannotated labels with meaningful implications.