IPLab 智能感知与信息处理实验室

论文分享会

余海军　2025 年 9 月 3日

#弱监督多标签分类

Semantic-Aware Representation Blending for Multi-Label Image Recognition with Partial Labels

Tao Pu [1], **Tianshui Chen** [2], **Hefeng Wu** [1], **Liang Lin** [1*]

[1] Sun Yat-Sen University, [2] Guangdong University of Technology
putao3@mail2.sysu.edu.cn, tianshuichen@gmail.com, wuhefeng@gmail.com, linliang@ieee.org

Abstract

Training the multi-label image recognition models with partial labels, in which merely some labels are known while others are unknown for each image, is a considerably challenging and practical task. To address this task, current algorithms mainly depend on pre-training classification or similarity models to generate pseudo labels for the unknown labels. However, these algorithms depend on sufficient multi-label annotations to train the models, leading to poor performance especially with low known label proportion. In this work, we propose to blend category-specific representation across different images to transfer information of known labels to complement unknown labels, which can get rid of pre-training models and thus does not depend on sufficient annotations. To this end, we design a unified semantic-aware representation blending (SARB) framework that exploits instance-level and prototype-level semantic representation to complement unknown labels by two complementary modules: 1) an instance-level representation blending (ILRB) module blends the representations of the known labels in an image to the representations of the unknown labels in another image to complement these unknown labels. 2) a prototype-level representation blending (PLRB) module learns more stable representation prototypes for each category and blends the representation of unknown labels with the prototypes of corresponding labels to complement these labels. Extensive experiments on the MS-COCO, Visual Genome, Pascal VOC 2007 datasets show that the proposed SARB framework obtains superior performance over current leading competitors on all known label proportion settings, i.e., with the mAP improvement of 4.6%, 4.6%, 2.2% on these three datasets when the known label proportion is 10%. Codes are available at https://github.com/HCPLab-SYSU/HCP-MLR-PL.

Introduction

Multi-label image recognition (MLR) (Chen et al. 2019d,b; Wu et al. 2020), which aims to find out all semantic labels from the input image, is a more challenging and practical task compared with the single-label counterpart. Due to the complexity of the input images and output label spaces, collecting a large-scale dataset with complete multi-label annotation is extremely time-consuming. To deal with this issue, recent works tend to study the task of multi-label image recognition with partial labels (MLR-PL), in which merely a few positive and negative labels are provided whereas other labels are unknown (see Figure 1). MLR-PL is more practical to real-world scenarios because it does not require complete multi-label annotations for each image.

	[a]	[b]
person	1	0
giraffe	-1	0
car	1	1
zebra	1	0
boat	-1	-1
⋮	⋮	⋮
elephant	-1	0

Figure 1: An MLR image with complete labels [a], partial labels [b], in which 1 represents the corresponding category exists, -1 represents it does not exist, and 0 represents it is unknown.

Previous works (Sun et al. 2017; Joulin et al. 2016) simply ignore the unknown labels or treat them as negative, and they adopt traditional MLR algorithms to address this task. However, it may lead to poor performance because it either loses some annotations or even incurs some incorrect labels. More recent works (Durand, Mehrasa, and Mori 2019; Huynh and Elhamifar 2020) propose to train classification or similarity models with given labels, and use these models to generate pseudo labels for the unknown labels. Despite achieving impressive progress, these algorithms depend on sufficient multi-label annotation for model training, and they suffer from obvious performance drop if decreasing the known label proportion to a small level.

Fortunately, a specific label c that is unknown in one image I^n may be known in another image I^m. We can extract the information of label c from image I^m, blend this information to image I^n, and in this way complement the unknown label c for image I^n. Previous works (Zhang et al. 2017) utilize mixup algorithm to blend two images and generate a new image with semantic information from both images to help regularize training single-label recognition models. However, a multi-label image generally has multiple semantic objects scattering over the whole image, and

*Tao Pu and Tianshui Chen contribute equally to this work and share first authorship. Corresponding author is Liang Lin.

论文基本信息

- 论文题目：
  - **Semantic-Aware Representation Blending for Multi-Label Image Recognition with Partial Labels**
  - 使用部分标签进行多标签图像识别的语义感知表示融合技术

- 作者信息：中山大学
- 会议分区：2022-AAAI（CCFA）

摘要

Training the multi-label image recognition models with partial labels, in which merely some labels are known while others are unknown for each image, is a considerably challenging and practical task. To address this task, current algorithms mainly depend on pre-training classification or similarity models to generate pseudo labels for the unknown labels. However, these algorithms depend on sufficient multi-label annotations to train the models, leading to poor performance especially with low known label proportion. In this work, we propose to blend category-specific representation across different images to transfer information of known labels to complement unknown labels, which can get rid of pre-training models and thus does not depend on sufficient annotations. To this end, we design a unified semantic-aware representation blending (SARB) framework that exploits instance-level and prototype-level semantic representation to complement unknown labels by two complementary modules: 1) an instance-level representation blending (ILRB) module blends the representations of the known labels in an image to the representations of the unknown labels in another image to complement these unknown labels. 2) a prototype-level representation blending (PLRB) module learns more stable representation prototypes for each category and blends the representation of unknown labels with the prototypes of corresponding labels to complement these labels. Extensive experiments on the MS-COCO, Visual Genome, Pascal VOC 2007 datasets show that the proposed SARB framework obtains superior performance over current leading competitors on all known label proportion settings, i.e., with the mAP improvement of 4.6%, 4.6%, 2.2% on these three datasets when the known label proportion is 10%. Codes are available at https://github.com/HCPLab-SYSU/HCP-MLR-PL.

介绍本文的研究任务：用部分标注的标签去训练多标签分类模型时一个极具挑战性和实际性的任务。

简述现有算法：为了解决这个问题，现有算法主要利用预训练模型或半监督模型为”unknown label（未发现标签）”生成伪标签。

介绍现有方法的局限性：然而，这些方法需求足够多的”known label（已知标签）”来训练模型，因此在标签稀缺的情况下效果不佳。

介绍本文的思路：为了解决这个问题，本文提出将“在不同图像中将类别特定（category-specific）的表征进行融合，将已知标签的信息传播到未知（未发现）标签中作为补充，进而避免对预训练模型的依赖”

介绍本文的具体设计：具体来说，本文提出了一种语义感知的表征融合（SARB）方法，通过俩个补充模型分别将实例级别和原型级别的表征进行融合，从而补充未知标签。1）实例级别表征融合（ILRB）：将一个图像中已知标签的表征融合到另外一个图像相应的未知标签中。2）原型级别表征融合（PLRB）：为每个类别得到一个稳定的原型表征，在未知标签里融合相应的原型表征进行补充。

实验结果与开源：实验表明，本文的方法要优于现有主要竞争对手……

引言读解

第一段：引入**部分标注的多标签分类任务**的重要性

Multi-label image recognition (MLR) (Chen et al. 2019d,b; Wu et al. 2020), which aims to find out all semantic labels from the input image, is a more challenging and practical task compared with the single-label counterpart. Due to the complexity of the input images and output label spaces, collecting a large-scale dataset with complete multi-label annotation is extremely time-consuming. To deal with this issue, recent works tend to study the task of multi-label image recognition with partial labels (MLR-PL), in which merely a few positive and negative labels are provided whereas other labels are unknown (see Figure 1). MLR-PL is more practical to real-world scenarios because it does not require complete multi-label annotations for each image.

	[a]	[b]
person	1	0
giraffe	-1	0
car	1	1
zebra	1	0
boat	-1	-1
⋮	⋮	⋮
elephant	-1	0

多标签图像识别任务：MLR任务旨在从输入图像中提取所有语义标签，相较于单标签任务，MLR是一个更有挑战性且实用性的任务。

挖坑：由于输入图像和输出标签空间的复杂性，收集具有完整多标签注释的大规模数据集非常耗时。

引出本文任务：为解决此问题，近年来研究趋于研究部分标注的多标签图像识别任务（MLR-PL）。

介绍任务设定：如图[b]，MLR-PL设定中仅仅提供少数正标签和负标签，而其他标签未知。

MLR-PL更贴近实际场景，因为它不要求每张图像都拥有完整的多标签注释。

引言读解

第二段：介绍**传统算法**和**现有算法**的缺陷

Previous works (Sun et al. 2017; Joulin et al. 2016) simply ignore the unknown labels or treat them as negative, and they adopt traditional MLR algorithms to address this task. However, it may lead to poor performance because it either loses some annotations or even incurs some incorrect labels. More recent works (Durand, Mehrasa, and Mori 2019; Huynh and Elhamifar 2020) propose to train classification or similarity models with given labels, and use these models to generate pseudo labels for the unknown labels. Despite achieving impressive progress, these algorithms depend on sufficient multi-label annotation for model training, and they suffer from obvious performance drop if decreasing the known label proportion to a small level.

传统算法：传统算法只是简单地忽略未知标签或者假定其为负标签，这样都会带来**较差的性能，因为它们要么缺失了标注信息、要么引入了错误的标签**。

现有算法：最近工作提出先用已知标签数据训练出分类模型，并用此模型为未知标签生成伪标签。

现有算法的缺陷（挖坑）：尽管这些算法取得了令人印象深刻的进展，但它们**依赖于足够的多标签注释来进行模型训练，如果将已知标签比例降低到很小的水平，它们的性能就会明显下降**。

引言读解

第三段：介绍**本文方法**思路

Fortunately, a specific label c that is unknown in one image I^n may be known in another image I^m. We can extract the information of label c from image I^m, blend this information to image I^n, and in this way complement the unknown label c for image I^n. Previous works (Zhang et al. 2017) utilize mixup algorithm to blend two images and generate a new image with semantic information from both images to help regularize training single-label recognition models. However, a multi-label image generally has multiple semantic objects scattering over the whole image, and simply blending two images lead to confusing semantic information. In this work, we design a unified semantic-aware representation blending (SARB) framework that learns and blends category-specific feature representation to complement the unknown labels. This framework does not depend on pre-trained models, and thus it can perform consistently well on all known label proportion settings.

介绍思路启发：在n图像中未知的标签c，可能在另外一张m图像是已知存在的，我们可以从m图像中提取标签c的信息，并将其融合到n图像中，这样就为n图像补充了未知的标签c。

单标签场景做法：先前有工作使用mixup算法将两张图像融合，生成一张包含两者语义信息的新图像，从而帮助正则化单标签识别模型的训练。

单标签场景算法无法直接运用到多标签场景：然而，多标签图像通常包含多个语义对象分布在整张图像中，简单地将两张图像混合往往导致语义信息混乱。

本文SARB框架：我们设计了一个 **语义感知的表征融合（SARB）框架**，能够学习并融合类别特定的特征表示，用来补全未知标签。该框架不依赖于预训练模型，并且在已知标签比例不同的情况下都能保持稳定的性能。

引言读解

第四段：具体介绍**本文设计**

Specifically, we first introduce a category-specific representation learning (CSRL) module (Chen et al. 2019b; Ye et al. 2020) that incorporates category semantics to guide generating category-specific representations. An instance-level representation blending (ILRB) module is designed to blend the representations of the known label c in one image I^m to the representations of the corresponding unknown label c in another image I^n. In this way, image I^n can also contain the information of label c and thus this label is complemented. This module can generate diverse blended representations to facilitate the performance but these diverse representations may also lead to unstable training. To solve this problem, a prototype-level representation blending (PLRB) module is further proposed to learn more robust representation prototypes for each category and blend the representation of unknown labels with the prototypes of the corresponding categories. In this way, we can simultaneously generate diverse and stable blended representations to complement the unknown labels and thus facilitate the MLR-PL task.

具体来说，我们首先引入了一个**类别特定表征学习（CSRL）模块**，该模块利用类别语义来指导生成类别特定的表征。（类别解耦）

设计了一个**实例级表征融合（ILRB）模块**，用于将m图像中已知标签c的表征融合到另一张n图像中对应未知标签的表征上。这样一来，图像n也能包含标签的信息，从而实现标签补全。

ILRB能够生成多样化的融合表征以提升性能，但过于多样化的表征可能导致训练不稳定。（**PLRB的动机**）

为了解决这一问题，进一步提出了**原型级表征融合（PLRB）模块**，用于为每个类别学习更鲁棒的表征原型，并将未知标签的表征与对应类别的原型进行融合。这样可以同时生成多样化且稳定的融合表征，用于补全未知标签

引言读解

第五段：介绍本文贡献

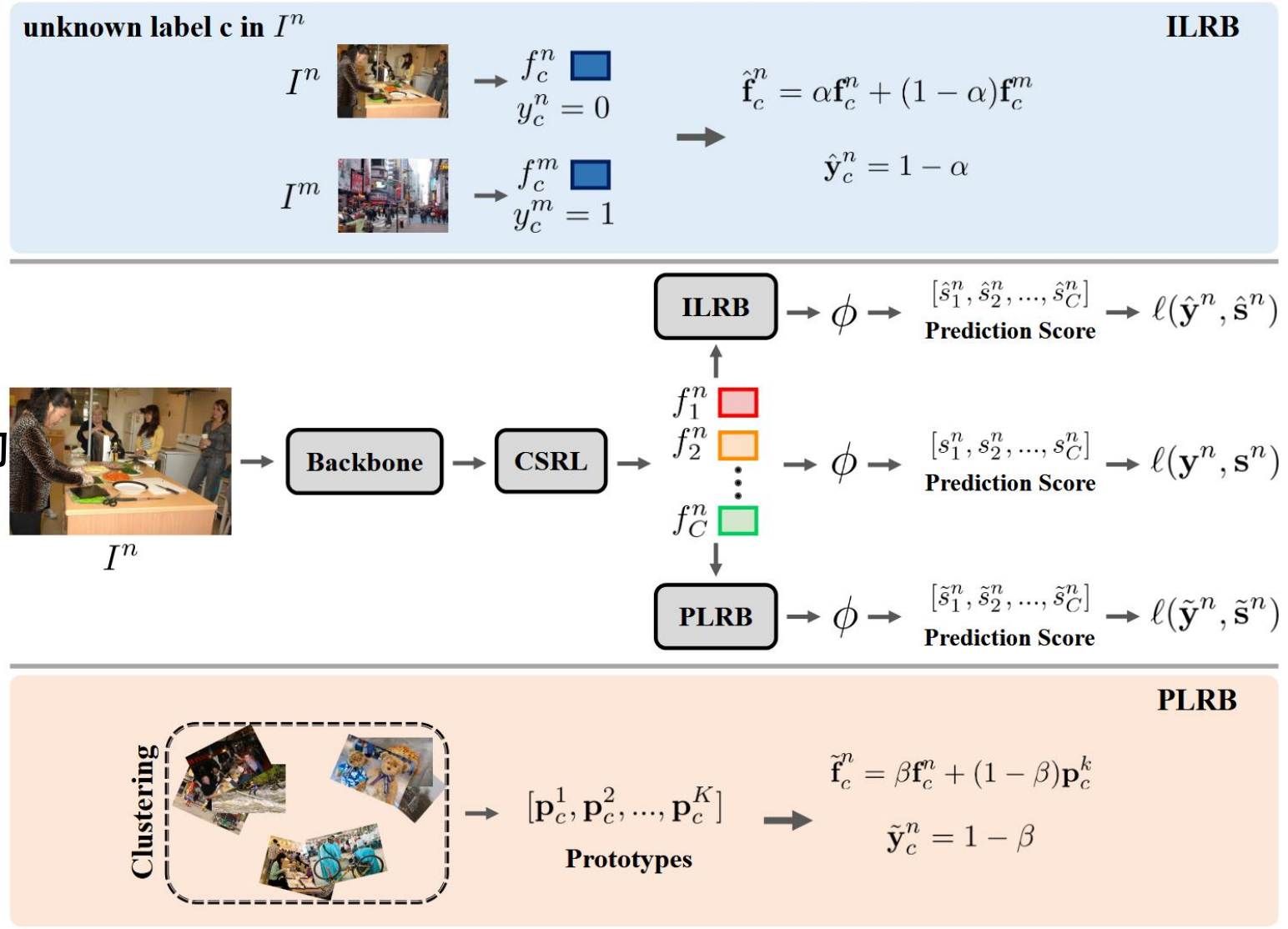
The contributions of this work are summarized into three folds: 1) We propose a semantic-aware representation blending (SARB) framework to complement unknown labels. It does not depend on pre-trained models and performs consistently well on all known label proportion settings. 2) We design the instance-level and prototype-level representation blending modules that generate diverse and stable blended feature representation to complement unknown labels. 3) We conduct extensive experiments on several large-scale MLR datasets, including Microsoft COCO (Lin et al. 2014), Visual Genome (Krishna et al. 2016) and Pascal VOC 2007 (Everingham et al. 2010), to demonstrate the effectiveness of the proposed framework. We also conduct ablative studies to analyze the actual contribution of each module for profound understanding.

贡献一：提出了一种**语义感知的表征融合框架**来补全未知标签，此框架不依赖预训练模型，在所有已知标签比例设置下表现稳定。

贡献二：设计了**实例级和原型级表征融合模块**，可以生成多样且稳定的融合特征表征，以补全未知标签。

贡献三：我们在多个大规模多标签识别（MLR）数据集上进行了广泛的实验，包括 Microsoft COCO、Visual Genome以及 Pascal VOC 2007，以验证所提出框架的有效性。我们还进行了消融实验，以深入分析每个模块的实际贡献。

方法学-算法流程图

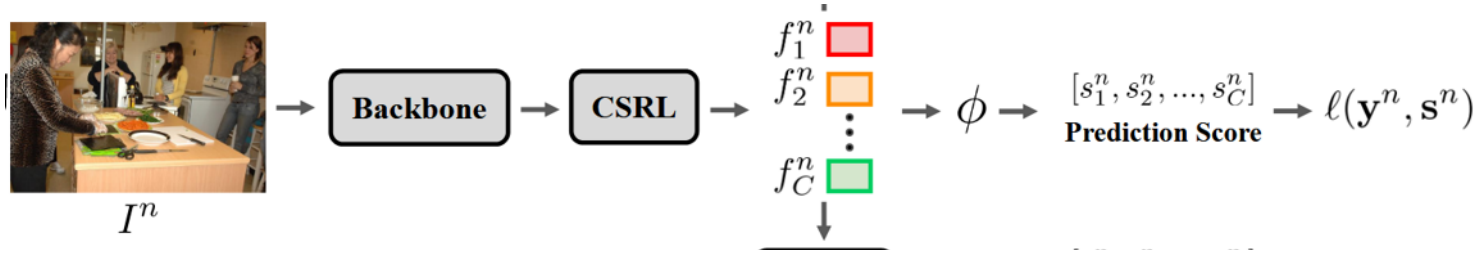


CSRL类别特定表征学习
（类别解耦表征）

ILRB实例级别融合

PLRB原型级别融合

方法学　Backbone 和 CSRL



整体流程

- 给定一张训练图像 I^n，先用 **主干网络（backbone network）** 提取全局特征图 f^n。
- 然后引入 **类别特定表征学习（CSRL）模块**，利用类别语义生成类别特定的表征。

- 类别特定表征的生成：

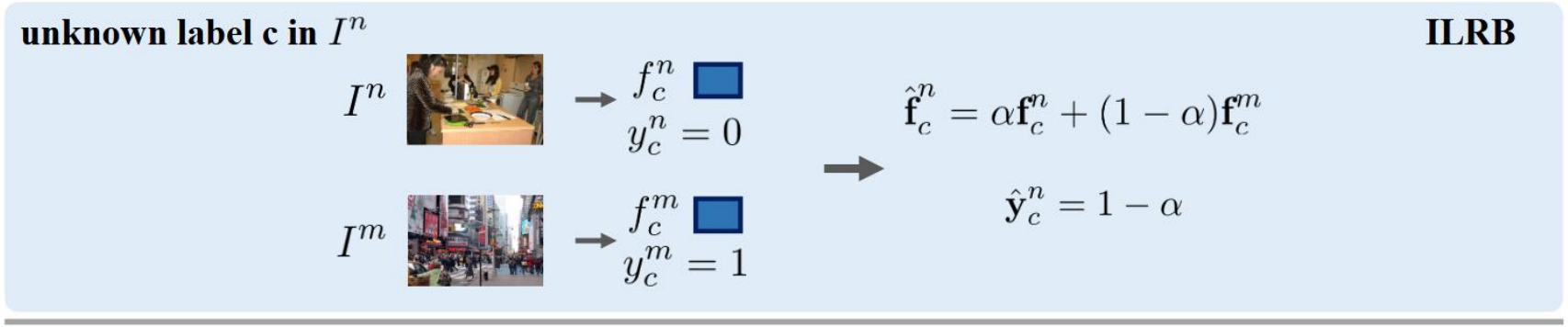
$$[f_1^n, f_2^n, \cdots, f_C^n] = \phi_{csrl}(f^n)$$

  其中，C 表示类别数。
- 基于这些类别特定表征，利用门控神经网络（GNN）和线性分类器，再经过 **sigmoid 函数** 计算出各类别的概率分数：

$$[s_1^n, s_2^n, \cdots, s_C^n] = \phi([f_1^n, f_2^n, \cdots, f_C^n])$$

- CSRL 模块可以通过不同算法实现，例如：
  - **语义解耦（semantic decoupling）** (Chen et al. 2019b)
  - **语义注意力机制（semantic attention）** (Ye et al. 2020)
- 本文借鉴了 Chen 等的工作，采用 **门控神经网络 + 线性分类器 + sigmoid** 的组合。

方法学 ILRB



动机：直观地说，图像 In 中的未知标签 c 可能在另一幅图像 Im 中是已知的。**ILRB 模块的目的是将图像 Im 中标签 c 的信息融合到图像 In 中，从而使图像 In 也能拥有已知标签 c。**为了实现这一目的，我们融合了属于同一类别的不同图像的表征，将一幅图像的已知标签转移到另一幅图像的未知标签上。

背景

给定两张训练图像 I^n 和 I^m，它们分别有：

- 语义表征向量：$[f_1^n, f_2^n, \ldots, f_C^n]$、$[f_1^m, f_2^m, \ldots, f_C^m]$
- 标签向量：$y^n = \{y_1^n, y_2^n, \ldots, y_C^n\}$、$y^m = \{y_1^m, y_2^m, \ldots, y_C^m\}$

目标：通过融合两个图像的语义表征与标签，补全未知标签。

融合过程（以类别 c 为例）

1. **语义表征融合**

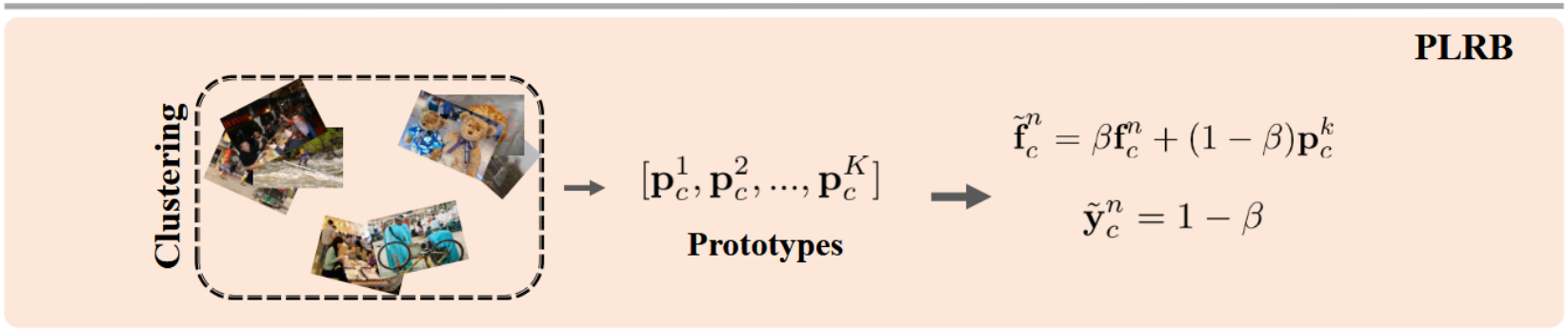
$$\hat{f}_c^n = \begin{cases} \alpha f_c^n + (1-\alpha) f_c^m & \text{若 } y_c^n = 0,\ y_c^m = 1 \\ f_c^n & \text{其他情况} \end{cases}$$

2. **标签融合**

$$\hat{y}_c^n = \begin{cases} 1-\alpha & \text{若 } y_c^n = 0,\ y_c^m = 1 \\ y_c^n & \text{其他情况} \end{cases}$$

其中，α 是一个可学习参数，初始值设为 0.5。

方法学 PLRB



动机：虽然 ILRB 模块可以明显提高性能，但它可能会干扰训练过程，因为它会生成许多不同的融合表示进行训练，尤其是当已知标签比例较低时。针对这一问题，我们进一步设计了一个 PLRB 模块，该模块通过学习为每个类别生成更稳定的表示原型，并将图像 In 中未知标签的表示与相应类别的原型进行融合。

- 原型用于描述每个类别的整体表征。
- 对于类别 c：
  1. 先收集所有含有已知标签 c 的图像。
  2. 提取这些图像的类别表征，得到向量集合：$[f_c^1, f_c^2, \ldots, f_c^{N_c}]$。
  3. 使用 **K-means 聚类** 将这些特征向量聚类为 K 个原型：

$$P_c = \{p_c^1, p_c^2, \ldots, p_c^K\}$$

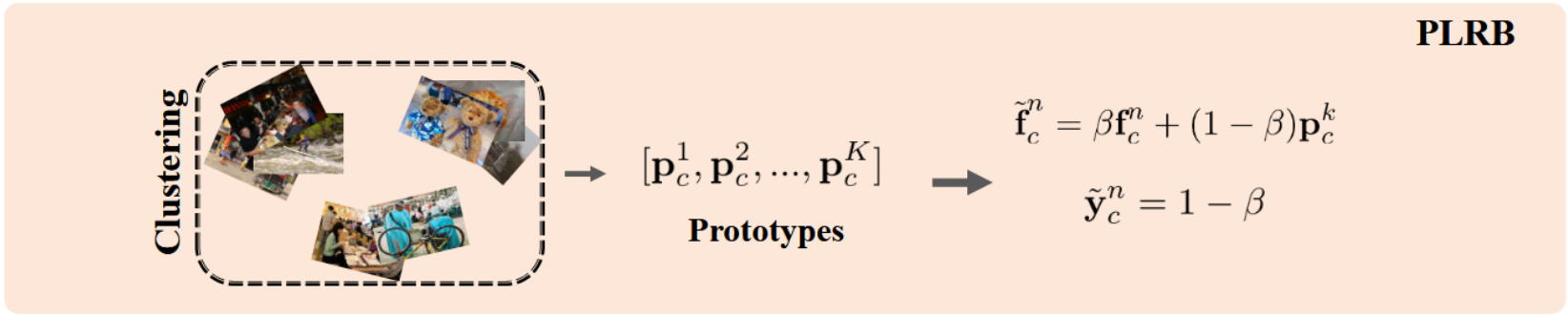
- 假设同一类别的表征应该彼此接近，不同类别的表征应远离。
- 因此，定义对比损失：
  - 若两张图像 I^n, I^m 都含有类别 c，则**增大相似性**（余弦相似度）。
  - 否则，**减小相似性**。

$$\ell_c^{n,m} = \begin{cases} 1 - \cos(f_c^n, f_c^m), & y_c^n = 1, y_c^m = 1 \\ 1 + \cos(f_c^n, f_c^m), & \text{其他情况} \end{cases}$$

- 总损失：

$$L_{cst} = \sum_{n=1}^{N} \sum_{m=1}^{N} \sum_{c=1}^{C} \ell_c^{n,m}$$

方法学　PLRB



原型级表征融合（PLRB）

- 给定一张输入图像 I^n，其表征向量为 $[f_1^n, f_2^n, \ldots, f_C^n]$，标签向量为 y^n。
- 随机选取一个 **未知标签** c（即 $y_c^n = 0$）。
- 从类别 c 的原型集合 P_c 中随机选择一个原型 p_c^k。
- 融合规则：
  - 表征：

$$\tilde{f}_c^n = \begin{cases} \beta f_c^n + (1-\beta) p_c^k, & y_c^n = 0 \\ f_c^n, & \text{其他情况} \end{cases}$$

  - 标签：

$$\tilde{y}_c^n = \begin{cases} 1-\beta, & y_c^n = 0 \\ y_c^n, & \text{其他情况} \end{cases}$$

其中 β 是可学习参数。

方法学　优化方案

部分交叉熵损失：按照已有工作，我们采用 **部分二元交叉熵损失（partial binary cross entropy loss）** 作为监督网络的目标函数。

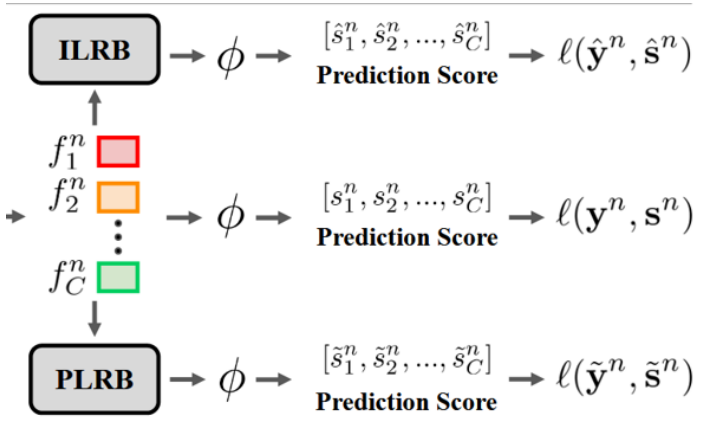
具体来说，给定预测的概率得分向量 $s^n = \{s_1^n, s_2^n, \ldots, s_C^n\}$ 以及已知标签的真实值 y^n，目标函数定义为：

$$\ell(y^n, s^n) = \frac{1}{\sum_{c=1}^{C} |y_c^n|} \sum_{c=1}^{C} \left[1(y_c^n = 1) \log(s_c^n) + 1(y_c^n = -1) \log(1 - s_c^n) \right]$$

其中，$1[\cdot]$ 是指示函数，条件成立时取值 1，否则为 0。

同样地，我们采用该损失来监督 **ILRB** 和 **PLRB** 模块，即 $\ell(\hat{y}^n, \hat{s}^n)$ 和 $\ell(\tilde{y}^n, \tilde{s}^n)$。因此，最终的分类损失是对所有样本的三部分损失求和：

$$L_{cls} = \sum_{n=1}^{N} \left[\ell(y^n, s^n) + \ell(\hat{y}^n, \hat{s}^n) + \ell(\tilde{y}^n, \tilde{s}^n) \right]$$



最后，将分类损失和对比损失结合，得到最终损失函数：

$$L = L_{cls} + \lambda L_{cst}$$

其中，λ 是一个平衡参数，用于保证对比损失 L_{cst} 的规模与分类损失 L_{cls} 相当。由于 L_{cst} 通常远大于 L_{cls}，实验中设置 $\lambda = 0.05$。

实验-实施细节

为了公平比较，我们采用 ResNet-101 作为主干网络提取全局特征图，并用在 ImageNet 数据集上预训练的参数初始化，同时随机初始化新增层的参数。训练时，前 91 层 ResNet-101 的参数固定，其他层采用端到端的方式训练。

训练过程:
- 优化器：Adam
- 批大小：16
- 动量：0.999 和 0.9
- 权重衰减：5×10^{-4}
- 初始学习率：10^{-5}，每 10 个 epoch 降低 10 倍
- 总共训练 20 个 epoch

训练策略:
- 从第 5 个 epoch 开始使用 ILRB 和 PLRB 模块
- 每 5 个 epoch 重新计算各类别的原型

推理阶段:
- 移除 ILRB 和 PLRB 模块
- 输入图像统一调整为 448×448

数据增强方面:
- 输入图像先缩放到 512×512
- 随机选择 {512, 448, 384, 320, 256} 中的一个作为裁剪大小
- 裁剪后再缩放到 448×448
- 同时使用随机水平翻转

实验-数据集

我们在 **MS-COCO** (Lin et al. 2014)、**Visual Genome** (Krishna et al. 2016) 和 **Pascal VOC 2007** (Everingham et al. 2010) 上进行实验，用于公平对比。

MS-COCO

- 包含 80 个日常生活**类别**。
- 训练集：82,801 张图像
- 验证集：40,504 张图像

Pascal VOC 2007

- 包含 **9,963 张图像**，来自 **20 个类别**。

Visual Genome

- 共 **108,249 张图像**，涉及 **80,138 个类别**。
- 大多数类别样本很少 → 选取 **200 个最常见类别**，构建 **VG-200 子集**。
- 因为没有官方训练/验证划分：
  随机选取 10,000 张图像作为测试集
  剩余 98,249 张图像作为训练集
- 该划分将公开供进一步研究。

实验-实验设定

所有的数据集均具有完整标注。为了模拟部分标注，参考前人工作 (Durand et al. 2019; Huynh & Elhamifar 2020)：**随机丢弃部分正负标签，制造“部分标注”数据集。**

在这项工作中，丢弃标签的比例从 90% 到 10% ，因此已知标签的比例为 10% 到 90%。

为了进行公平的比较，我们采用了所有类别的**平均精度（mAP）**来评估不同比例的已知标签。为了进行更全面的评估，我们还计算了所有比例的平均 mAP。

此外，我们还遵循之前大多数 MLR 作品（Chen 等人，2019b）的做法，采用整体（overall）和每类（per）精度、召回率、F1-measure（**即 OP、OR、OF1、CP、CR 和 CF1**）进行更全面的评估。

实验-SOTA对比实验

为了评估所提出的 SARB 框架的有效性，我们将其与传统的 MLR 算法和当前的 MLR-PL 算法进行了比较。

MLR算法
（改动：将BCE换为partial-BCE）

MLR-PL算法

Datasets	Methods	Avg. mAP	Avg. OP	Avg. OR	Avg. OF1	Avg. CP	Avg. CR	Avg. CF1
MS-COCO	SSGRL	74.1	86.3	64.8	73.9	82.1	58.4	68.1
	GCN-ML	74.4	85.2	64.2	73.1	81.8	58.9	68.4
	KGGR	75.6	84.0	65.6	73.7	81.4	60.9	69.7
	Curriculum labeling	60.7	87.8	51.0	61.9	60.9	40.4	48.3
	partial-BCE	74.7	**86.7**	64.7	74.0	**83.1**	58.9	68.8
	Ours	**77.9**	86.6	**68.6**	**76.5**	82.9	**64.1**	**72.2**
VG-200	SSGRL	39.7	69.9	25.9	37.8	45.3	18.3	26.1
	GCN-ML	39.3	64.1	28.2	38.7	44.6	18.2	25.6
	KGGR	41.5	64.5	30.5	41.2	54.8	25.8	33.6
	Curriculum labeling	28.4	66.4	15.4	23.6	20.4	7.6	10.9
	partial-BCE	39.8	69.7	24.6	36.1	44.3	18.1	25.7
	Ours	**45.6**	**70.1**	**33.2**	**45.0**	**56.8**	**27.8**	**37.4**
Pascal VOC 2007	SSGRL	89.5	91.2	**84.4**	87.7	87.8	**81.4**	84.5
	GCN-ML	88.9	92.2	83.0	87.3	89.7	80.1	84.6
	KGGR	89.7	90.5	82.9	86.5	88.5	81.4	84.7
	Curriculum labeling	84.1	92.7	78.2	83.8	79.5	71.7	75.4
	partial-BCE	90.0	91.8	84.3	87.9	88.8	81.3	84.8
	Ours	**90.7**	**93.0**	83.6	**88.4**	**90.4**	81.1	**85.9**

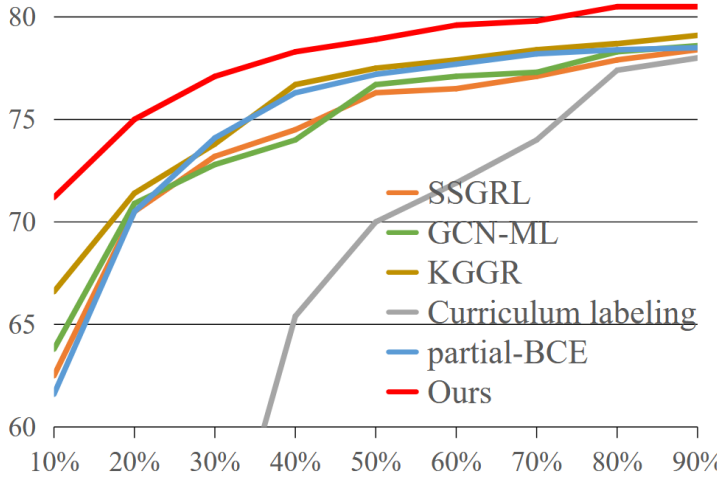
实验-在COCO上表现

- 与目前最先进的算法相比，我们的 SARB 框架获得了最佳性能。（如表+图）
- （列数据）如表 1 所示，该算法的 **mAP、OF1 和 CF1** 平均值分别为 77.9%、76.5% 和 72.2%，比之前表现最好的 KGGR 算法分别高出 2.3%、2.8% 和 2.5%。如图所示，在**所有已知标签比例**设置中，SARB 框架也能获得更好的 mAP。
- 值得注意的是，当**已知标签比例降低时，SARB 框架的性能提升更为明显**。（如图）例如，当使用 90% 和 10% 的已知标签时，mAP 与之前的最佳 KGGR 算法相比分别提高了 1.4% 和 4.6%。
- （摆结论）这些比较表明，SARB 框架可以**适应不同的比例设置，因为它不依赖于预先训练的模型**。（回应挖坑）

MLR算法
（改动：将BCE换为partial-BCE）

MLR-PL算法

Datasets	Methods	Avg. mAP	Avg. OP	Avg. OR	Avg. OF1	Avg. CP	Avg. CR	Avg. CF1
MS-COCO	SSGRL	74.1	86.3	64.8	73.9	82.1	58.4	68.1
	GCN-ML	74.4	85.2	64.2	73.1	81.8	58.9	68.4
	KGGR	75.6	84.0	65.6	73.7	81.4	60.9	69.7
	Curriculum labeling	60.7	87.8	51.0	61.9	60.9	40.4	48.3
	partial-BCE	74.7	**86.7**	64.7	74.0	**83.1**	58.9	68.8
	Ours	**77.9**	86.6	**68.6**	**76.5**	82.9	**64.1**	**72.2**

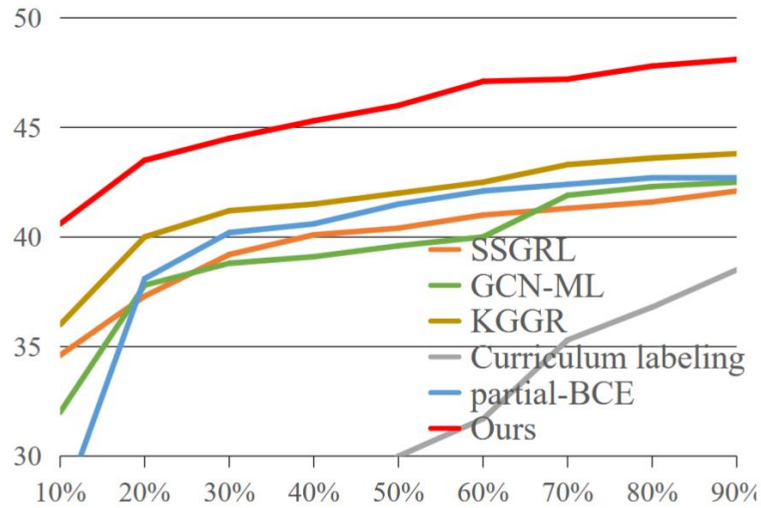


(a)

实验-在VG-200上表现

- （数据背景+找补）VG200 是一个更具挑战性的基准，**涵盖的类别更多**。因此，目前的工作性能相当差。
- （列数据）如表，之前性能最好的 KGGR 算法的 mAP、OF1 和 CF1 平均值分别为 41.5%、41.2% 和 33.6%。在这种情况下，我们的 SARB 框架表现出**更明显的性能提升**。其平均 mAP、OF1 和 CF1 分别为 45.6%、45.0% 和 37.4%，比 KGGR 算法分别高出 4.1%、3.8% 和 3.8%。
- （列数据）如图，与现有算法相比，我们发现我们的框架在所有已知标签比例设置下的 mAP 提高了 3.3% 以上。

Datasets	Methods	Avg. mAP	Avg. OP	Avg. OR	Avg. OF1	Avg. CP	Avg. CR	Avg. CF1
MS-COCO	SSGRL	74.1	86.3	64.8	73.9	82.1	58.4	68.1
	GCN-ML	74.4	85.2	64.2	73.1	81.8	58.9	68.4
	KGGR	75.6	84.0	65.6	73.7	81.4	60.9	69.7
	Curriculum labeling	60.7	87.8	51.0	61.9	60.9	40.4	48.3
	partial-BCE	74.7	**86.7**	64.7	74.0	**83.1**	58.9	68.8
	Ours	**77.9**	86.6	**68.6**	**76.5**	82.9	**64.1**	**72.2**
VG-200	SSGRL	39.7	69.9	25.9	37.8	45.3	18.3	26.1
	GCN-ML	39.3	64.1	28.2	38.7	44.6	18.2	25.6
	KGGR	41.5	64.5	30.5	41.2	54.8	25.8	33.6
	Curriculum labeling	28.4	66.4	15.4	23.6	20.4	7.6	10.9
	partial-BCE	39.8	69.7	24.6	36.1	44.3	18.1	25.7
	Ours	**45.6**	**70.1**	**33.2**	**45.0**	**56.8**	**27.8**	**37.4**
Pascal VOC 2007	SSGRL	89.5	91.2	**84.4**	87.7	87.8	**81.4**	84.5
	GCN-ML	88.9	92.2	83.0	87.3	89.7	80.1	84.6
	KGGR	89.7	90.5	82.9	86.5	88.5	81.4	84.7
	Curriculum labeling	84.1	92.7	78.2	83.8	79.5	71.7	75.4
	partial-BCE	90.0	91.8	84.3	87.9	88.8	81.3	84.8
	Ours	**90.7**	**93.0**	83.6	**88.4**	**90.4**	81.1	**85.9**

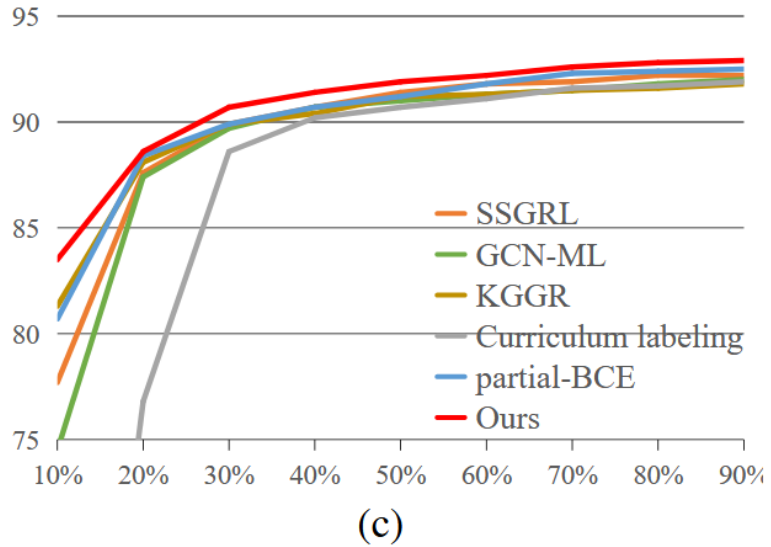


(b)

实验-在VOC上表现

- （数据背景）Pascal VOC 2007 是评估多标签图像识别最广泛使用的数据集。由于该数据集仅涵盖 20 个类别，是一个简单得多的数据集，目前的算法也能取得相当不错的性能。
- （列数据）如表 1 所示，之前性能最好的 KGGR 算法的 mAP、OF1 和 CF1 平均值分别为 41.5%、41.2% 和 33.6%。在这种情况下，我们的 SARB 框架表现出更明显的性能提升。它的平均 mAP、OF1 和 CF1 分别为 45.6%、45.0% 和 37.4%，比 KGGR 算法分别高出 4.1%、3.8% 和 3.8%。
- （列数据）现有算法相比，我们发现我们的框架在所有已知标签比例设置下的 mAP 提高了 3.3%。

Datasets	Methods	Avg. mAP	Avg. OP	Avg. OR	Avg. OF1	Avg. CP	Avg. CR	Avg. CF1
MS-COCO	SSGRL	74.1	86.3	64.8	73.9	82.1	58.4	68.1
	GCN-ML	74.4	85.2	64.2	73.1	81.8	58.9	68.4
	KGGR	75.6	84.0	65.6	73.7	81.4	60.9	69.7
	Curriculum labeling	60.7	87.8	51.0	61.9	60.9	40.4	48.3
	partial-BCE	74.7	**86.7**	64.7	74.0	**83.1**	58.9	68.8
	Ours	**77.9**	86.6	**68.6**	**76.5**	82.9	**64.1**	**72.2**
VG-200	SSGRL	39.7	69.9	25.9	37.8	45.3	18.3	26.1
	GCN-ML	39.3	64.1	28.2	38.7	44.6	18.2	25.6
	KGGR	41.5	64.5	30.5	41.2	54.8	25.8	33.6
	Curriculum labeling	28.4	66.4	15.4	23.6	20.4	7.6	10.9
	partial-BCE	39.8	69.7	24.6	36.1	44.3	18.1	25.7
	Ours	**45.6**	**70.1**	**33.2**	**45.0**	**56.8**	**27.8**	**37.4**
Pascal VOC 2007	SSGRL	89.5	91.2	**84.4**	87.7	87.8	**81.4**	84.5
	GCN-ML	88.9	92.2	83.0	87.3	89.7	80.1	84.6
	KGGR	89.7	90.5	82.9	86.5	88.5	81.4	84.7
	Curriculum labeling	84.1	92.7	78.2	83.8	79.5	71.7	75.4
	partial-BCE	90.0	91.8	84.3	87.9	88.8	81.3	84.8
	Ours	**90.7**	**93.0**	83.6	**88.4**	**90.4**	81.1	**85.9**



(c)

实验-消融实验CSRL

- （CSRL方法）CSRL 模块用于提取类别特定的特征表征，可采用不同算法实现：语义解耦 (SD) (Chen et al.)语义注意机制 (**SAM**) (Ye et al. 2020)。
- （基线方法）传统 Mixup 算法通过 位置级融合生成新样本，增强训练。本文实现了两个基线算法：**IP-Mixup**：在图像空间做融合，**FM-Mixup**：在特征空间做融合。（不解耦，直接融合）
- （结论）SD 和 SAM 性能接近， SD 略优于 SAM，因此后续实验全部采用 **SD 实现 CSRL 模块**。
- （结论）这两种 Mixup 与 **SSGRL 基线性能相近**，因为简单融合无法带来额外信息。

- （结论）与 **基于 CSRL 的 SARB** 相比：

IP-Mixup 在三个数据集上 mAP 分别下降 3.6%、5.9%、1.0%

FM-Mixup 在三个数据集上 mAP 分别下降 3.8%、6.0%、1.1%

Methods \ Datasets	MS-COCO	VG-200	VOC2007
Ours w/ SAM	77.6	45.4	90.6
Ours w/ SD	77.9	45.6	90.7
IP-Mixup	74.3	39.7	89.7
FM-Mixup	74.1	39.6	89.6
SSGRL	74.1	39.7	89.5
Ours ILRB	77.3	44.9	90.2
Ours ILRB fixed α	76.9	44.5	89.8
Ours PLRB	77.3	44.9	90.4
Ours PLRB fixed β	76.9	44.6	90.2
Ours	77.9	45.6	90.7

实验-消融实验ILRB

- 为了分析 **ILRB 模块** 的实际贡献，作者进行了只使用该模块的实验（称为 **Ours ILRB**），并与 **SSGRL 基线**在 **MS-COCO、VG-200 和 Pascal VOC 2007** 三个数据集上进行了对比。
- ILRB 模块包含一个关键参数 **α**，用于控制 **实例级融合的比例**。为了验证 α 的可学习性带来的贡献，作者对比了使用固定值 α=0.5 的情况。
- （结论）实验表明，Ours ILRB在各项数据表现均优于SSGRL，mAP分别提升了3.2、5.2、0.7。
- （结论）实验表明， Ours ILRB在各项数据表现均优于Ours ILRB fixed **α**。说明 **固定 α 会降低性能**，而 **自适应 α 更优**。

Methods \ Datasets	MS-COCO	VG-200	VOC2007
Ours w/ SAM	77.6	45.4	90.6
Ours w/ SD	77.9	45.6	90.7
IP-Mixup	74.3	39.7	89.7
FM-Mixup	74.1	39.6	89.6
SSGRL	74.1	39.7	89.5
Ours ILRB	77.3	44.9	90.2
Ours ILRB fixed α	76.9	44.5	89.8
Ours PLRB	77.3	44.9	90.4
Ours PLRB fixed β	76.9	44.6	90.2
Ours	77.9	45.6	90.7

实验-消融实验PLRB

- 同样地，**PLRB 模块** 也是框架中的关键部分。为了分析其有效性，作者使用了只使用PLRB的实验，并与基线SSGRL对比。同时也针对可学习参数β进行的实验。
- （结论）加入 PLRB 后，相较于SSGRL，mAP分别提升了3.2、5.2、0.9。
- （结论）在 **训练损失曲线可视化（图 4）** 中可以看到：**没有 PLRB** → loss 波动明显（不稳定），**加入 PLRB** → loss 曲线更加平滑（训练稳定），**根据之前的分析，PLRB 模块有助于生成稳定的融合表示，从而补充未知标签并使训练更加稳定。**
- （结论）自适应 β 优于固定 β。

Methods \ Datasets	MS-COCO	VG-200	VOC2007
Ours w/ SAM	77.6	45.4	90.6
Ours w/ SD	77.9	45.6	90.7
IP-Mixup	74.3	39.7	89.7
FM-Mixup	74.1	39.6	89.6
SSGRL	74.1	39.7	89.5
Ours ILRB	77.3	44.9	90.2
Ours ILRB fixed α	76.9	44.5	89.8
Ours PLRB	77.3	44.9	90.4
Ours PLRB fixed β	76.9	44.6	90.2
Ours	77.9	45.6	90.7

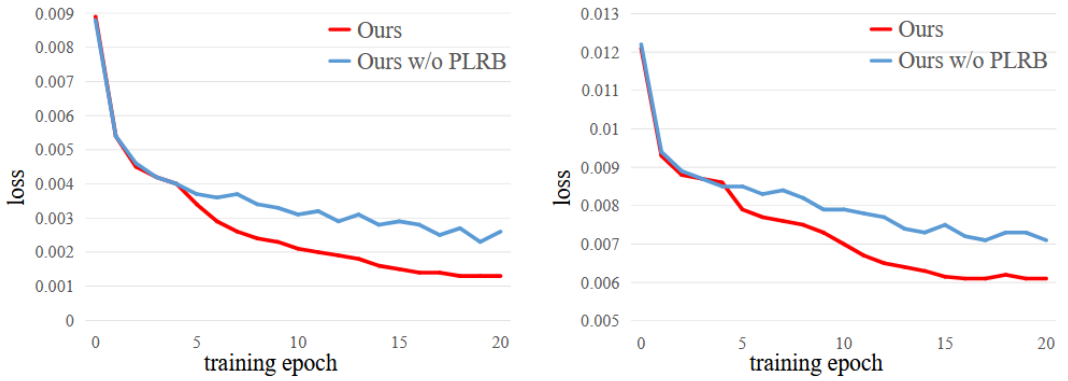


Figure 4: Analysis of the effect on PLRB. These experiments are conducted on MS-COCO (left) and VG-200 (right).

总结

- 在这项工作中，我们提出了一种新的视角：通过融合类别特定的表征来补充未知标签，从而解决 **MLR-PL 任务**。这种方法**不依赖于充分的标注**，因此在所有已知标签比例的设置下都能获得更优的性能。
- 具体来说，我们的方法包含两个核心模块：
- **ILRB 模块**：融合已知标签的 **实例级表示**，以补充对应未知标签的表示；
- **PLRB 模块**：学习并融合 **原型级表示**，以补充对应未知标签的表示。
- 这两个模块能够同时生成**多样化且稳定的融合表示**来弥补未知标签，从而促进 **MLR-PL 任务**的完成。
- 在 **MS-COCO、VG-200 和 Pascal VOC** 数据集上的大量实验验证了该方法相较于现有算法的优越性。

好词好句

- 1) Training the multi-label image recognition models with partial labels, in which merely some labels are known while others are unknown for each image, is a considerably challenging and practical task.
  - Considerably challenging 和 practical task：相当大的挑战和很实际的任务

- 2) Extensive experiments on the MS-COCO, Visual Genome, Pascal VOC 2007 datasets show that the proposed SARB framework obtains superior performance over current leading competitors on all known label proportion settings, i.e., with the mAP improvement of 4.6%, 4.6%, 2.2% on these three datasets when the known label proportion is 10%.
  - Extensive experiments on ... show that …obtains superior performance over current leading competitors ...
  - 在（数据集）上的大量实验表明，（我们 的方法）在性能上优于当前的主要竞争对手…

好词好句

- 3) Despite achieving impressive progress, these algorithms depend on sufficient multi-label annotation for model training, and they suffer from obvious performance drop if decreasing the known label proportion to a small level.
  - 尽管取得了显然进度，然而…（可以用在批判其他算法时）

- 4) However, a multi-label image generally has multiple semantic objects scattering over the whole image, and simply blending two images lead to confusing semantic information.
  - multiple semantic objects scattering over the whole image
  - 整个图像中散落了多个语义对象
  - 这个表达比较生动

好词好句

- 5) Our In this way, we can simultaneously generate diverse and stable blended representations to complement the unknown labels and thus facilitate the MLR-PL task.
  - 同时做到A and B(simultaneously)

- 6) This framework does not depend on pre-trained models, and thus it can perform consistently well on all known label proportion settings.
  - consistently well，这个副词可以用于说明自己方法鲁棒性好

好词好句

- 7) In this work, we present a new perspective to complement the unknown labels by blending category-specific feature representation to address the MLR-PL task.
  - 创新性描述
  - we present a new perspective to：为 xxx 提供了新的视角
- 8) It is noteworthy that the SARB framework obtains more obvious performance improvement when decreasing the known label proportions.
  - 值得注意的是…

好词好句

- 9) In this work, we propose to blend category-specific representation across different images to transfer information of known labels to complement unknown labels, which can get rid of pre-training models and thus does not depend on sufficient annotations.
  - 提出xx技术/思路，可以使用 we propose to
  - get rid of　摆脱了xxx（缺陷）
  - sufficient annotations：充分注释

- 10) As shown in Table 2, both two baseline algorithms achieve comparable performance with the SSGRL baselines as such simple blending can not provide additional information.
  - 解释俩种算法性能差不多时，用comparable performance with

观点论据

- 1) Due to the complexity of the input images and output label spaces, collecting a large-scale dataset with complete multi-label annotation is extremely time-consuming.
- **部分标注的多标签分类的动机/挖坑**
  - 由于输入图像和输出标签空间的复杂性，收集具有完整多标签注释的大规模数据集非常耗时。

观点论据

- 2) Previous works (Sun et al. 2017; Joulin et al. 2016) simply ignore the unknown labels or treat them as negative, and they adopt traditional MLR algorithms to address this task. However, it may lead to poor performance because it either loses some annotations or even incurs some incorrect labels.
- MLR-PL的传统方法缺点
- 假定负标签/忽略未知标签的缺陷
  - 之前的工作（Sun 等人，2017 年；Joulin 等人，2016 年）只是忽略未知标签或将其视为负标签，并采用传统的 MLR 算法来解决这一任务。然而，这可能会导致性能不佳，因为它会丢失一些注释，甚至产生一些错误的标签。

观点论据

- 3) Despite achieving impressive progress, these algorithms depend on sufficient multi-label annotation for model training, and they suffer from obvious performance drop if decreasing the known label proportion to a small level.
- MLR-PL的现有基于伪标签算法的缺陷
  - 尽管这些算法取得了令人印象深刻的进展，但它们依赖于足够的多标签注释来进行模型训练，如果将已知标签比例降低到很小的水平，它们的性能就会明显下降。

观点论据

- 4) However, a multi-label image generally has multiple semantic objects scattering over the whole image, and simply blending two images lead to confusing semantic information.
- **语义解耦（类别特征表征学习）的动机**
  - 然而，多标签图像通常会有多个语义对象散布在整个图像中，简单地混合两幅图像会导致语义信息混乱。

观点论据

- 5) This module can generate diverse blended representations to facilitate the performance but these diverse representations may also lead to unstable training. To solve this problem, a prototype-level representation blending (PLRB) module is further proposed to…
- **原型学习的动机**
  - 该模块（实例级别）可生成多种混合表征，以促进执行，但这些不同的表征也可能导致训练不稳定。