

# 基于深度学习框架的流媒体图像多标签分类

王大林

(重庆幼儿师范高等专科学校 重庆 404047)

**摘要** 考虑到存在未知类别的大规模流媒体图像多标签分类问题,提出一种基于深度学习框架的多标签分类方法。为了检测图像中是否含有新类标签,提出一种递归类检测器,它通过对图像特征和多个标签之间的关系进行有效编码来学习。为提升方法对大规模数据集处理能力,通过假设新的类图像在特征空间中远离已知类,从而有效地实现分类器和检测器交替学习的批处理模式。实验结果验证了该方法对大规模未知类流媒体图像多标签分类有效性。

**关键词** 卷积神经网络 多标签 流媒体图像 检测器

中图分类号 TP183 TP391.41

文献标志码 A

DOI:10.3969/j.issn.1000-386x.2024.08.032

## MULTI-LABEL CLASSIFICATION OF STREAMING MEDIA IMAGES BASED ON DEEP LEARNING FRAMEWORK

Wang Dalin

(Chongqing Preschool Normal College, Chongqing 404047, China)

**Abstract** Considering the problem of multi-label classification of large-scale streaming images with unknown classes, a multi-label classification method based on a deep learning framework is proposed. To detect whether an image contains a new class of labels or not, a recursive class detector was proposed, which learned by efficiently encoding the relationship between image features and multiple labels. To enhance the method's ability to handle large-scale datasets, a batch mode of learning the classifier and detector alternately was effectively implemented by assuming that the new class media images were far away from the known classes in the feature space. The experimental results verify the effectiveness of the method for multi-label classification of large-scale unknown class streaming media images.

**Keywords** Convolution neural network Multi-Label Streaming media image Detector

## 0 引言

随着时代的不断发展,计算机视觉已经在各个领域发挥了极其重要的作用<sup>[1-2]</sup>。作为计算机视觉领域的基础任务,多标签图像分类在实际使用过程中明显比单一图像分类更具研究价值<sup>[3]</sup>。

传统的多标签图像分类问题已经研究了多年,其中训练图像和测试图像具有相同的标签集,图像通常用多个标签进行标记<sup>[4]</sup>。Chen等<sup>[5]</sup>通过目标检测与定位,利用SVM学习图像中物体之间的上下文信息,有效提高了分类精度。Wei等<sup>[6]</sup>利用CNN从图像中

提取局部特征,将多标签分类问题转化为多实例学习问题。Wang等<sup>[7]</sup>使用CNN提取图像特征,RNN描述标签之间的依赖关系。虽然上述研究有效解决了多标签图像分类问题,但是传统的多标签图像分类的标签是已知的。在一个实际的图像识别系统中,该系统可能包含先前学习阶段不存在的新类。而CNN-RNN方法不能直接应用于解决训练中未知类的问题。

解决未知类别图像分类问题通常需要学习分类器和检测器,分别用于多标签图像分类和新类检测。Zhu等<sup>[8]</sup>采用支持向量机和随机森林分别学习多类分类器和异常检测器,对于检测器,它考虑了不同标签之间的依赖关系,即输入随机森林与连接的特征及其相关的

标签。Mu 等<sup>[9]</sup>采用矩阵剖分的方法来解决这个问题,将两个低维矩阵草图分别用于检测新出现的类和分类已知的类。然而,这些研究大多是针对小规模机器学习数据集进行的,这些数据集并没有显示出对大规模图像分类的能力。

针对上述方法存在的问题,提出一种基于未知流媒体图像的大规模多标签分类方法。

## 1 方法介绍

### 1.1 深度学习框架

提出在统一的深度学习框架中交替学习多标签图像分类器和新类检测器,流程如图 1 所示,  $O$  和  $Z$  是系统中的特征映射。BN 表示批处理规范化,1 536 是 Inception V4 分类器中完全连接层的长度<sup>[10]</sup>。不同颜色的流动表示不同的阶段。在开始时,使用具有已知标签的训练图像初始化分类器  $x_0 = [x_{-n+1}, x_{-n+2}, \dots, x_0] \in x$  和检测器。然后,通过未知流媒体图像来更新系统。当一个新的图像进入系统时,通过分类器获得预测的标签,然后由新类检测器检查,以确定是否存在新类。新的图像和标签被输入到图像缓冲区中装满后,利用图像更新分类器和检测器。

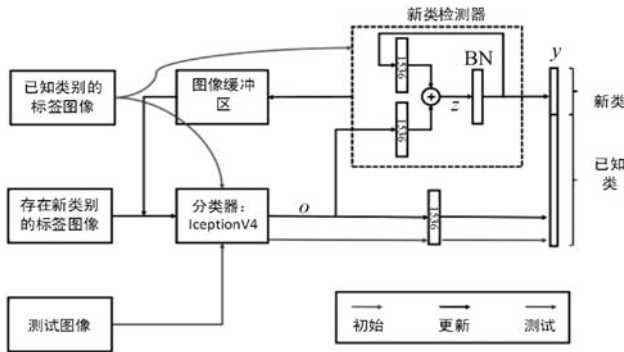


图 1 系统结构

将特征空间表示为  $x$ 。在开始时,即  $t=0$ ,初始训练实例为具有  $n$  个已标记实例,且  $y_0 = \{c_1, c_2, \dots, c_l\}$  是标签初始训练数据集。 $Y_0 = [y_{-n+1}, y_{-n+2}, \dots, y_0]$  是初始训练数据的对应标签。

每次假设只有一个实例进入系统。当一个新的实例在时间  $t$  出现时,一直观察实例直到时间  $t$  时为  $x_t = [x_{t-1}, x_t] \in x$ ,且  $x_t$  用表示时间  $t$  的标记实例,如果一个新标签是从  $x_t$  开始,则  $y_t = y_{t-1} \cup [c_a]$ ,其中  $c_a$  在时间  $t$  和  $a = |y_{t-1}| + 1$  时新发现的标签,否则  $y_t = y_{t-1}$ 。

### 1.2 学习多标签图像分类器

使用卷积神经网络(CNN)作为多标签图像分类器。本文选择了具有 softmax 损失的 CNN 模型 Inception-V4。因为标签设置  $Y$  可能在不同的时间  $t$  变化,当检

测到一个新的类时,CNN 的输出大小需要相应地改变。一种方法是,一旦检测到一个新的类并将其添加到训练集中,就会追加 CNN 输出,这对在这种动态环境中学习分类器不是很方便。为了避免 CNN 结构的变化,在开始时,为 CNN 输出分配了足够大的空间,以覆盖可能的新类。

这种 CNN 结构可以用新的标签来预测图像,这与零样本学习的主题有关。在零样本学习中,通常假设已知标签和新标签的图像共享一些公共属性,这些属性是由模型从训练图像中学习的。通过使用 CNN,这些属性可以在 CNN 内在结构中的不同类之间传递。另一点是,在标签向量中为可能的新类保留足够的空间。给定一个测试图像,如果在为新类保留的标签向量中有非零值,则说明可能包含新类,但是,需要由下面的新类检测器进一步检查。

### 1.3 检测器

在实际的多标签图像分类系统中,它可能会遇到具有未知类的图像。提出使用一种新的类检测器来区分新的图像是否与已知的类相一致,然后再更新分类系统。给了新图像  $x_t$  在  $t$  时,新类检测器被定义为  $f$ ,这是 CNN 中的一个标准模块。存在如下关系:

$$f(x_t) = \begin{cases} 1 & x_t \text{ 有新类} \\ -1 & \text{其他} \end{cases} \quad (1)$$

新类检测器起到异常检测的作用,其中具有新类的图像则说明是异常。在异常检测中,计算正常实例的分布,其中通常使用一个球来紧密地绑定实例。其中球的中心是已知实例的平均值,最大距离边界实例可以选择为半径。如果一个新的实例从球中掉出来,它被检测为异常。在大多数情况下,数据分布是以无监督的方式计算的。

进一步提出用监督的方法对已知类的图像分布进行建模,以学习新类检测器。在学习过程中,目前只有带有已知标签的图像。换句话说,不知道哪些图像有未知的新类。监督信息来自已知类中的类信息。对于多标签分类,自动学习检测器计算包含特征空间中已知类的图像特征的多个球。

具体来说,从 CNN 中最后一个池化层的输出中学习检测器的参数。给定图像特性  $Z \in \mathbf{R}^{D \times K}$ ,其中  $D$  和  $K$  分别代表通道数与图像数,第  $d$  个通道球的平均值  $\mu_d$  半径  $\sigma_d$  计算表达式为:

$$\mu_d = \frac{1}{k} \sum_{i=1}^K z_{di} \quad (2)$$

$$\sigma_d = \left( \frac{1}{K} \sum_{i=1}^K (z_{di} - \mu_d)^2 \right)^{\frac{1}{2}} \quad (3)$$

这两个方程实际上具有与批处理归一化相同的形式, 该方法是 CNN 中的标准模块。为了测试一个新的图像  $x$ , 在 CNN 中计算了它的特征映射  $z$ , 仅当:

$$f: \sum_{d=1}^D [ |Z_d - \mu_d| > \lambda \sigma_d ] > \gamma D \quad (4)$$

式中:  $[\cdot]$  表示如果命题满足:  $\lambda > 0$  和  $0 \leq \gamma \leq 1$ , 则返回 1。新型检测器还需要考虑多标签关系, 进一步编码标签向量与特征一起作为输入, 从而建立一个树型检测器。然而, 两种不同信息的简单关联可能会引入学习检测器的偏差。为了以一种更自然的方式联合两种信息, 提出一种递归的新类检测器, 其中图像特征和多标签信息都被迭代编码为:

$$Z^s = W_0 o + W_y y^{s-1} + b \quad (5)$$

$$y^s = \text{softmax}(W_z BN(z^s)) \quad (6)$$

式中:  $BN(\cdot)$  表示归一化函数;  $z^s$  表示状态信息。

在递归新类检测器中, 它具有两个输入: 一个是来自上一层的  $o$  特征向量, 固定用于所有迭代。另一个是来自同一图像上次时间  $y^{s-1}$  的标签向量。  $y^s$  是当前时间的标签向量。  $W_0$ 、 $W_y$  和  $W_z$  是关联的权重矩阵,  $b$  是偏差。请注意, 此检测器可以为已知类中的图像建模特征分布。均值和半径都是使用监督的方式学习的。通过选择一个大的  $D$  值, 检测球可以提供正常实例特征空间的完整集合。

## 1.4 更新模型

交替地更新多标签图像分类器和新类检测器。在一开始, 使用一个预先训练过的 CNN, 并在训练图像上使用已知标签对其进行微调。然后, 得到初始的多标签分类器和新类检测器。注意, 标签向量中为可能的新类预留了足够的存储单元。因此, 通过学习已知标签的图像, 这些存储单元是零。接下来, 使用带有未知标签的图像来更新系统, 其标签可能包含新的类。每次图像进入系统时, 它分别通过多标签分类器和新类检测器。通过分类器将具有预测标签的新图像添加到图像缓冲区中。如果检测器检测到包含新标签的新图像, 则使用标签更新系统; 否则新标签将被丢弃。一旦缓冲区满了, 其中的图像将用于更新分类器。通过使用图像缓冲区, 可以有效地学习 CNN。学习分类器和检测器如算法 1 所示。

### 算法 1 学习分类器和检测器

输入: 已知类  $X$  的训练图像  $x_0$  以及它们的标签  $y_0$ ; 未知的图像伴随着新的类出现在流中。

输出: 多标签图像分类器和新类检测器。

1. 用  $x_0$  和  $y_0$  初始化多标签图像分类器和新类检测器
2. for  $t = 1, 2, 3, \dots$  do

3. 一幅未知的图像  $x_t$  进入系统
4. if 图像缓冲区不满, 并且图像流不停止 then
5. 预测当前分类器类标签为  $x_t$
6. 使用当前检测器进行检查。  $x_t$  包含新的类或不包含式(3)
7. if 检测到一个新的类 then
8. 使用新类的标签
9. else
10. 新类标签被丢弃
11. end if
12. 将  $x_t$  以及其预测的标签放入图像缓冲区
13. else
14. 在批处理模式下使用缓冲区中的图像更新分类器
15. for  $i_{ter} = 1, 2, \dots$ , do
16. 更新递归检测器以及缓冲器的图像
17. end for
18. 清空图像缓冲区
19. end if
20. end for
21. end

在 CNN 训练中, 缓冲区大小与批处理大小相同。需要以迭代方式更新循环的新类检测器。当图像缓冲区已满时, 它们将被用于使用分类器预测的标签, 从而更新检测器。在每次迭代过程中, 对于每幅图像, 其最后一步的输出作为检测器的输入, 以学习图像特征和多标签之间的关系。检测器在其中反复学习  $M$  次。

## 2 实验与分析

### 2.1 数据集

在实验中, 在几个基准数据集上评估了所提出的方法。将每个数据集中 80% 的图像作为训练图像分为两部分, 一半用作已知类, 一半用作新类。新类图像中的标签不会出现在已知类的图像中。随机选择不同分数的新类, 并评估学习的分类在测试图像上的性能。本文使用的多标签图像数据集如下:

(1) VOC2007<sup>[6]</sup> 有 20 个对象类, 每幅图像都有多个标签。具有 2 501 幅训练图像、2 510 幅验证图像和 4 952 幅测试图像。

(2) VOC2012<sup>[8]</sup> 有 20 个对象类, 每幅图像都有多个标签。具有 5 717 幅训练图像、5 823 幅验证图像和 10 991 幅测试图像。

(3) MS COCO<sup>[10]</sup> 有一组 82 081 幅图片的训练集和一组 40 137 幅图片的验证集。数据集有 80 个公共对象类别。由于其测试集不提供地面标签, 因此只在其验证集上进行评估。

设置一个基线方法,在训练阶段便已知所有类,并且利用本文提出的 CNN 模型,为了对模型进行微调,从方差为 1.0 和均值为 0.0 的截断高斯分布中采样最后一个完全连接层的初始权值。学习率为所有层设置一个固定值 0.000 01。网络的训练过程在 50 个时代内收敛。式(4)中的两个参数分别为: $\lambda = 3, \gamma = 0.95$ 。所有的实验都是在一个带有 NvidiaTitan Xp GPU 和 64 GB 内存的 PC 上进行。在测试中,运行时速度约为 50 帧/s。将平均精度作为评价指标对提出的方法进行评估。

## 2.2 在 VOC2012 和 VOC2007 数据集上的评估

首先评估提出的方法在 VOC2012 数据集上的性能。预先训练的 CNN 模型进行微调,进一步使用数据增强来提高性能。从输入图像中提取 10 种不同的尺度。

表 1 用于数据增强。对不同补丁大小的多标签图像分类性能进行评价。所有列车/车辆图像用于培训。可以观察到,补丁越大,可以获得更高的 mAP。当补丁

大小超过 500, mAP 的增长速度变慢。为了提升计算效率,本文选择 500 作为输入补丁大小。

表 1 VOC2012 进行多标签图像分类结果

输入补丁大小	250	300	350	400	450	500
平均精度/%	86.6	87.5	88.0	89.1	89.4	89.7

将提出方法与 VOC2012 数据集上的其他方法进行了比较。结果如表 2 所示,在 VOC2012 的 Trainval 集上对 CNN 进行了微调,并在其测试集上进行了评估。可以观察到,基线方法的平均精度 92.4% 优于大多数现有的方法。RCP 融合方法将 RCP 与 AGS 方法相结合,其准确率高于提出的方法,但是相对于它的单一 RCP 方法,提出的方法具有相当的精度。在模型中,如果采用不同的方法进行融合,则可以获得更高的精度。接下来,将在训练阶段使用一小部分类作为新类时,对大规模多标签图像分类问题上提出的方法进行评估。

表 2 不同方法在 VOC2012 数据集上的结果(%)

方法	飞机	自行车	鸟	船	瓶子	公共汽车	汽车	猫	椅子	牛	桌子
NUS-PS <sup>[6]</sup>	97.3	84.2	80.8	85.3	60.8	89.9	86.8	89.3	75.4	77.8	75.1
VGG16-SVM <sup>[10]</sup>	99.0	88.8	95.9	93.8	73.1	92.1	85.1	97.8	79.5	91.1	83.3
VGG19-SVM <sup>[8]</sup>	99.1	88.7	95.7	93.9	73.1	92.1	84.8	97.7	79.1	90.7	83.2
FN-VGG16 <sup>[9]</sup>	99.2	92.5	96.8	94.4	81.0	93.2	92.3	98.2	82.9	94.3	82.2
RCP <sup>[10]</sup>	99.3	97.6	98.0	96.4	79.3	93.8	96.6	97.1	78.0	88.7	87.1
RCP-融合 <sup>[11]</sup>	99.8	94.5	98.1	96.1	85.5	96.1	95.5	99.0	90.2	95.0	87.8
基线方法	99.4	95.6	96.6	95.0	80.2	93.8	91.6	98.4	84.7	94.2	86.0

方法	狗	马	自行车	个人	植物	羊	沙发	火车	电视	平均精度
NUS-PSL	83.0	87.5	90.1	95.0	57.8	79.2	73.4	94.5	80.7	82.2
VGG16-SVM	97.2	96.3	94.5	96.9	63.1	93.4	75.0	97.1	87.1	89.0
VGG19-SVM	97.3	96.2	94.3	96.9	63.4	93.2	74.6	97.3	87.9	89.0
FN-VGG16	97.4	97.3	95.9	98.7	72.9	95.1	77.7	97.5	90.8	91.5
RCP	97.1	96.3	95.4	99.1	82.1	93.6	82.2	98.4	92.8	92.2
RCP-融合	98.7	98.4	97.5	99.0	80.1	95.9	86.5	98.8	94.6	<b>94.3</b>
基线方法	98.2	98.3	96.7	98.3	74.5	96.2	79.1	98.4	92.1	92.4

首先评估多标签图像分类时,不同分数的类在训练阶段被选择为新类。从 VOC2012 数据集的 Trainval 集中选择了不同的类分数。表 3 结果表明,新类数量越少,分类结果越好。例如,当分数为 0.1 时,意味着在训练阶段有 18 个已知类和 2 个未知新类,其中 mAP 为 92.0%。与提出方法的结果 92.4% 相比,下降了 0.4 个百分点。当分数从 0.1 变为 0.3 时,性能不会降低太多。当分数为 0.1 时,如果分别将随机选取植物

和绵羊两类作为新类,它们的性能有很大的不同。植物的准确性远低于提出方法的结果,这是因为 CNN 从已知的类中捕获很少的植物共享信息。当分数较高时,某些分类的成绩就变得很糟糕。例如,当分数为 0.4 时,牛、自行车和电视的分类精度仅为 10% ~ 20% 左右,远低于其他类别。在这个数据集上,可以看到,当发现一小部分的新类(< 30%)并用于更新学习系统时,此方法是有效的。

表 3 不同的类分数在 VOC2012 数据集分类结果(%)

分数	飞机	自行车	鸟	船	瓶子	公共汽车	汽车	猫	椅子	牛	桌子
0.1	99.2	95.4	96.1	94.8	79.7	93.9	90.9	98.4	84.2	94.0	85.4
0.2	98.7	97.4	97.2	94.2	76.9	96.5	86.6	99.2	77.4	94.7	76.9
0.3	98.9	97.9	96.5	94.3	78.9	96.7	85.4	98.8	79.9	88.1	76.8
0.4	99.4	94.5	95.5	94.1	56.3	93.0	85.4	97.8	80.9	3.9	78.3
0.5	5.9	89.1	93.7	90.0	73.3	49.5	90.2	96.4	80.2	86.6	78.9
分数	狗	马	自行车	人	盆栽	羊	沙发	火车	电视	平均精度	
0.1	98.0	98.0	96.7	98.0	72.5	95.6	78.9	98.4	91.4	<b>92.0</b>	
0.2	99.4	95.2	96.0	96.6	68.4	96.1	77.6	99.2	85.7	90.5	
0.3	99.4	90.9	67.9	95.6	65.9	97.3	73.3	99.5	83.1	88.3	
0.4	97.1	93.7	35.3	94.7	58.6	25.7	74.5	95.8	18.5	73.4	
0.5	96.0	89.5	82.5	95.1	58.2	24.4	76.0	98.0	5.7	72.9	

新类的分类失败结果如图 2 所示。大多数分类失败发生在图像中的对象很小的时候,就像盆栽植物一样,分类器很难捕获和区分它。这可能是由于新类检测器无法很容易地检测到来自小型新对象的激活,从而导致分类器无法识别到它。

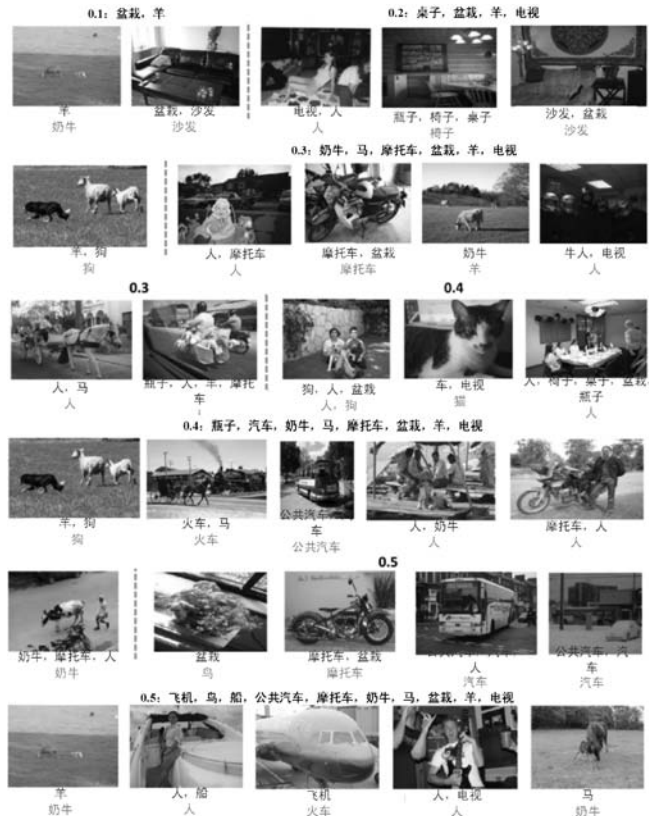


图 2 PASCAL VOC 2012 数据集失败分类结果

进一步评估在 VOC2007 数据集上提出方法的性能。图 3 显示了评估结果。从 VOC2007 中的训练集中随机选择不同数量(0~5)的新类。当新类数为 0 时,结果为基准。可以看出,新类数量在 0~5 范围内,其与基准的差值在 0.05 以内,说明新类数量的变化对

准确率的影响较小。

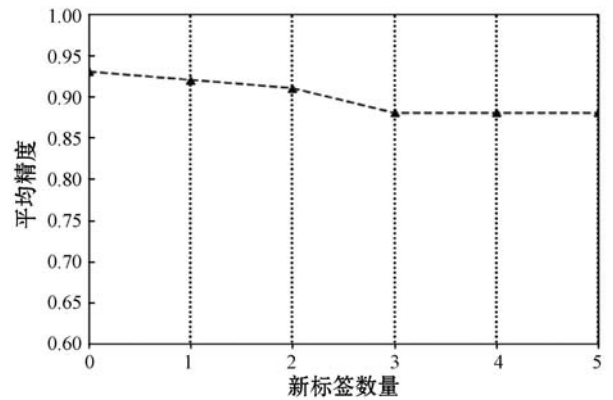


图 3 VOC2007 数据集上分类结果

### 2.3 迭代性能研究

本文的一个贡献是提出的递归新类检测器,其中最重要的因素是迭代数  $M$ 。在 VOC2012 上评估了不同的  $M(0 \sim 4)$  数,结果如表 4 所示。

表 4 不同  $M$  的 VOC2012 数据集分类结果(%)

新类数	迭代数 $M$				
	0	1	2	3	4
1	88.1	90.5	91.6	<b>92.0</b>	92.0
2	85.3	89.2	90.1	<b>92.1</b>	92.1
3	87.8	89.9	90.5	<b>91.8</b>	91.8
0(基准方法)	—	—	92.4	—	—

一般来说,多迭代可以提高检测精度,进一步提高分类精度。当  $M = 0$  时,这意味着上一次的输出不会被输入到此时的投入中,这是性能最差的。评估结果验证了这一结论,即特征和标签可以在多标签图像分类中相互受益。经过 3 次迭代,分类结果稳定在 92.0%。本实验以相同分数的新类重复三次。每次

选择的小说类都是不同的。可以看到,在大多数情况下,结果在 3 次迭代后变稳定。稳定的结果接近基线方法结果,这验证了所提出的新类检测器的效率。因此,本文对检测器使用了 3 次迭代。

表 5 比较了所提出的递归新类检测器效果。Mu 等使用了一种基于完全随机树的方法。为了公平比较,只使用它们的检测器来检测新的类别,而使用 CNN 进行分类。除此之外,还与文献[11]提出的分类方法 SENCForest 以及文献[12] MuENL 方法进行对比,以 VOC2012 数据集作为实验对象。CNN 中最后一个 FC 层的特征被用作图像表示准确性。实验结果显示提出的方法平均精度比较高,这可以解释为:(1) 使用了基于 CNN 的模型,比图像分类更有代表性;(2) 新类检测器使用递归结构对图像特征和标签信息进行编码,比所使用的检测器更有效。

表 5 比较递归新类检测器和检测方法

方法	分数				
	0.1	0.2	0.3	0.4	0.5
本文方法	92.0	90.5	88.3	73.4	72.9
SENCFores + 提出的检测器	90.1	89.2	86.4	70.0	67.2
SENCForest	88.2	87.1	84.8	68.4	65.1
MuENL	89.0	88.3	86.1	69.2	66.1

### 2.4 MS COCO 的评估结果

进一步评估了提出方法在 MS COCO 数据集上的性能,该数据集是一个多标签数据集,比 PASCALVOC 具有更大的规模。相关指标定义如式(7)所示,分类结果见表 6。

$$\begin{aligned}
 C_P &= \frac{1}{C} \sum_i \frac{N_i^c}{N_i^p} & O_P &= \frac{\sum_i N_i^c}{\sum_i N_i^p} \\
 C_R &= \frac{1}{C} \sum_i \frac{N_i^p}{N_i^g} & O_R &= \frac{\sum_i N_i^c}{\sum_i N_i^g} \\
 C_{F_1} &= \frac{2 \times C_P \times C_R}{C_P + C_R} & O_{F_1} &= \frac{2 \times O_P \times O_R}{O_P + O_R}
 \end{aligned} \tag{7}$$

表 8 更新系统 VOC 2012 数据集的分类结果(%)

分数	飞机	自行车	鸟	船	瓶子	公共汽车	汽车	猫	椅子	牛	桌子
0.1	88.8	99.3	95.4	96.3	95.8	85.7	93.9	91.2	98.4	87.3	94.0
0.2	77.1	98.8	97.4	97.3	94.2	78.3	96.5	87.2	99.4	82.4	94.7
0.3	77.0	98.9	97.9	96.5	94.3	78.9	96.7	96.1	98.8	80.2	88.1
0.4	78.3	99.4	94.5	95.5	94.1	60.1	93.0	85.4	97.8	80.9	21.8
0.5	78.9	31.2	89.1	93.7	91.3	73.3	49.5	96.4	96.4	80.2	86.6

表 6 MS COCO 上的分类结果

方法	CP	CR	CF <sub>1</sub>	OP	OR	OF <sub>1</sub>
WARP <sup>[13]</sup>	59.3	52.5	55.7	59.8	61.4	60.7
CNN-RNN <sup>[14]</sup>	66.0	55.6	60.4	69.2	66.4	67.8
Attention <sup>[15]</sup>	79.1	58.7	67.4	84.0	63.0	72.0
多证据	80.4	70.2	74.9	85.2	72.5	78.4
基线方法	84.8	71.5	77.6	86.5	74.6	80.1
0.1	84.6	68.7	75.8	85.7	70.8	77.6
本文方法/新类分数	0.2	78.1	59.1	69.7	83.8	64.9
0.3	68.9	51.5	61.0	81.7	58.7	68.3
0.4	65.3	49.1	56.1	80.6	50.2	61.9
0.5	62.3	37.4	46.9	78.8	41.2	54.1

式(7)中, C 表示标签数, N<sub>i</sub><sup>c</sup> 是图像数, 正确地预测了第 i 个标签, N<sub>i</sub><sup>p</sup> 是第 i 个标签的预测图像数量, N<sub>i</sub><sup>g</sup> 是第 i 个标签的地面真实图像数量。从结果可知本文方法的结果优于比较方法。

从其训练集中选择不同的分数作为新的类来学习分类器。即使在训练阶段也缺少一些类, 结果与现有的方法也能够进行比较。例如, 当分数为 0.1 时, 此方法的 CF<sub>1</sub> 优于多证据模型。由于 MS COCO 在训练集中有 80 个类别, 即使分数很小, 新类别的数量也比 PASCALVOC 大得多, 导致结果迅速下降。当分数从 0.2 变为 0.5 时, 结果远不如本文方法的结果。

### 2.5 使用测试图像更新系统

通过将测试图像视为具有未知类的图像更新识别系统来评估此方法, 这是一个自然的选择, 然而, 大多数现有的多标签图像分类方法不能以这种方式工作。只能从标签图像中学习。关于 VOC2012 的结果显示在表 8 可以看到, 使用测试图像可以帮助提高识别能力。当分数为 0.1 时, 即 2 个类作为未知的新类, 准确率(93.2%)甚至高于基线结果(92.4%)和不使用测试图像进行更新的结果(92.0%)。对于不同比例的训练图像选择为新类, 通过使用测试图像更新系统, 可以获得更高的识别能力。

续表 8

分数	狗	马	自行车	人	植物	羊	沙发	火车	电视	平均精度
0.1	98.0	98.1	96.7	98.0	79.2	95.6	82.1	98.4	91.5	93.2
0.2	99.5	95.2	96.1	96.7	73.4	96.1	77.6	99.2	86.3	91.2
0.3	99.4	92.4	75.9	95.6	70.9	97.3	73.3	99.5	83.1	89.0
0.4	97.1	93.7	46.3	94.7	72.6	45.2	74.5	95.8	38.5	78.0
0.5	96.2	89.5	82.5	95.1	58.2	35.2	76.0	98.0	15.2	75.3

通过使用 MS COCO 上的验证集继续更新系统,评估结果如表 9 所示。通过使用验证集更新学习系统,它可以获得比表 7 更好的结果。当分数很小,例如 0.1,其结果比基线结果要好一些,这证明了本文方法的有效性。

表 9 更新系统对 MSCOCO 的分类结果

分数	CP	CR	CF <sub>1</sub>	OP	OR	OF <sub>1</sub>
0.1	85.1	72.1	78.1	86.2	74.9	80.2
0.2	81.8	67.4	73.9	84.7	68.1	75.5
0.3	74.3	59.2	65.9	82.0	63.1	71.3
0.4	69.2	52.9	60.0	79.3	55.8	65.5
0.5	67.6	41.4	51.4	79.0	47.3	59.2

### 3 结 语

考虑到存在未知类别的大规模流媒体图像多标签分类问题,提出一种基于深度学习框架的多标签分类方法。通过数据集评估结果可得出如下结论:

(1) 新类检测器能够编码图像特征和标签之间的关系,有效提升分类器的分类精度。

(2) 新类检测器使用递归结构对图像特征和标签信息进行编码,比所有传统检测器对新类检测更有效。

(3) 分类精度随着新类别的数量的增加会下降,但是通过使用测试图像更新系统能够进一步提升分类精度。

(4) 相对于其他方法,本文方法在存在未知类别的大规模流媒体图像多标签分类问题上精度更高。

但是所提出的框架有一个问题:当系统遇到太多的新类时,例如在训练阶段,新类的数量大于已知类的数量,分类器的分类结果将更差。因此研究更有效的新类检测方法来识别图像是非常重要的。

### 参 考 文 献

[1] 蔡志锋,袁宝华,刘广海. 基于深度特征融合的图像分类算法的研究[J]. 计算机应用与软件,2020,37(10):175-179,243.  
[2] 郑欣悦,黄永辉. 基于 VAE 和注意力机制的小样本图像

分类方法[J]. 计算机应用与软件,2019,36(10):168-174.

- [3] Tang P, Wang X, Feng B, et al. Learning multi-instance deep discriminative patterns for image classification[J]. IEEE Transactions on Image Processing,2017,26(7):3385-3396.  
[4] 赵仲秋,季海峰,高隽,等. 基于稀疏编码多尺度空间潜在语义分析的图像分类[J]. 计算机学报,2014,37(6):1251-1260.  
[5] Chen Q, Song Z, Dong J, et al. Contextualizing object detection and classification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2015,37(1):13-27.  
[6] Wei Y, Xia W, Lin M, et al. HCP: A flexible CNN framework for multi-label image classification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2016,38(9):1901-1907.  
[7] Wang Z, Chen T, Li G, et al. Multi-label image recognition by recurrently discovering attentional regions[C]//2017 IEEE International Conference on Computer Vision (ICCV), 2017.  
[8] Zhu Y, Ting K, Zhou Z. Multi-label learning with emerging new labels[J]. IEEE Transactions on Knowledge and Data Engineering,2018,30(10):1901-1914.  
[9] Mu X, Ting K, Zhou Z. Classification under streaming emerging new classes: A solution using completely-random trees[J]. IEEE Transactions on Knowledge and Data Engineering,2017,29(8):1605-1618.  
[10] 高亮,潘积远,于佳平. 基于非负矩阵分解的类标签映射集成学习[J]. 科学技术与工程,2018,18(1):59-65.  
[11] 李校林,王成. 一种基于质心的多标签文本分类模型研究[J]. 计算机工程与科学,2020,42(6):1120-1126.  
[12] 郭利强,孟庆超. 基于多标签共享子空间学习和内核脊回归的空谱分类算法[J]. 光子学报,2020,49(5):121-133.  
[13] 张永,刘浩科,张洁. 基于类属特征和实例相关性的多标签分类算法[J]. 模式识别与人工智能,2020,33(5):439-448.  
[14] 徐香香. 基于神经网络的多标签文本分类[D]. 成都:电子科技大学,2020.  
[15] 肖琳,陈博理,黄鑫,等. 基于标签语义注意力的多标签文本分类[J]. 软件学报,2020,31(4):1079-1089.