

文章编号: 2095-4980(2020)02-0306-07

基于集成分类的暴恐图像自动标注方法

严 靛, 周 欣, 何小海*, 熊淑华, 卿 粼波

(四川大学 电子信息学院, 四川 成都 610065)

摘 要: 为减少暴恐图像对社会发展和青少年成长造成的不利影响, 本文提出一种基于集成分类的暴恐图像自动标注方法, 辅助筛除网页中的暴恐信息。该方法将暴恐图像的标注视作多标签分类问题, 利用迁移学习训练多个子网络, 然后通过集成学习对子网络的输出进行融合, 同时在融合过程中针对各个标签在不同网络上的准确率进行权重分配, 最后经过一系列矩阵运算得到图像的标注结果。实验结果表明, 与传统机器学习算法相比, 本文方法在准确率和召回率上都有较大提升, 并改善了样本不均衡所造成的不同标签类别上模型标注精确度差异较大的问题。

关键词: 图像标注; 多标签分类; 集成学习; 权重分配; 样本不均衡

中图分类号: TN957.52; TP391.41 **文献标志码:** A **doi:** 10.11805/TKYDA2019157

Violent image annotation using ensemble learning

YAN Liang, ZHOU Xin, HE Xiaohai*, XIONG Shuhua, QING Linbo

(School of Electronic Information, Sichuan University, Chengdu Sichuan 610065, China)

Abstract: In order to reduce the negative impact of the horror image on social development and adolescent growth, a violent image annotation algorithm based on ensemble learning is proposed, assisting in screening out the horror information in the webpage. The annotation of violent image is considered as a multi-label classification problem in this method. Multiple sub-networks are trained through transfer learning, and then the ensemble learning is introduced to fuse the outputs of sub-networks. In the process of fusion, weights are allocated according to the precision of each label on different networks, thus the annotation result is obtained through a series of matrix operations. The experimental results show that the proposed method achieves a great improvement in precision and recall than traditional machine learning algorithm, and also improves the problem that the precision of model annotation on different labels varies greatly due to the label category imbalance.

Keywords: image annotation; multi-label classification; ensemble learning; weight allocation; label category imbalance

随着互联网社交平台的快速发展以及数码设备在日常生活中的普遍使用, 人们每天接触的图像和视频信息呈指数式增长。丰富的多媒体信息在给人们的日常通信和检索带来便利的同时, 其中部分暴恐图像也给社会和谐和青少年的健康成长带来了负面影响。如何有效管理这些数据成为一项亟需解决的问题。图像自动标注技术因其自动给图像添加反映其内容的文本特征信息的特点, 逐渐成为图像分析领域的关键技术之一^[1]。目前, 图像自动标注方法主要分为两类: 基于相关模型的图像标注方法和基于分类思想的图像标注方法。基于相关模型的图像标注方法首先对图像进行划分, 然后计算各分割区域内底层图像特征和高层语义关键词之间的联合概率分布, 最后通过建立概率相关模型对待标注图像进行标注。刘凯等人提出一种利用深度玻尔兹曼机与典型相关分析的图像标注算法(Deep Boltzmann Machine-Canonical Correlation Analysis, DBM-CCA)^[2], 其中 DBM 实现从底层图像特征向高层语义概念的转变, CCA 实现对子空间映射关系的建立, 两者相互配合实现对图像的标

收稿日期: 2019-04-30; 修回日期: 2019-06-17

基金项目: 国家自然科学基金资助项目(61871278); 四川省成都市产业集群协同创新资助项目(2016-XT00-00015-GX); 四川省科技计划资助项目(2018HH0143)

作者简介: 严 靛(1996-), 女, 在读硕士研究生, 主要研究方向为图像处理、模式识别和视频通信。email:419965079@qq.com

*通信作者: 何小海 email:hxh@scu.edu.cn

注。基于相关模型的方法受图像分割技术的影响较大，而效果较差的图像分割会造成图像信息的丢失，因此该类标注方法的准确率都不是很高^[1]。基于分类思想的图像标注方法把每个标注关键词视作一个类别，通过构建底层视觉特征到高层语义标注的映射函数来实现图像到文本的转化。主要方法有：基于支持向量机(Support Vector Machine, SVM)的方法^[3-4]，基于 K 近邻(K-Nearest Neighbor, KNN)的方法^[5-7]，基于决策树的方法^[8-9]和基于神经网络的方法^[10-12]。分类器的质量将直接影响这类标注方法的结果。

鉴于卷积神经网络(Convolutional Neural Network, CNN)在图像分类问题中的强大优势，本文依据分类思想解决图像的标注问题，并将图像标注技术应用到防暴领域，通过集成多个 CNN 网络构建暴恐图像标注模型。同时考虑到样本不均衡的问题，将重点研究如何利用集成学习提高模型对每类标签的标注准确率。

1 暴恐图像自动标注算法

集成学习的一般流程是先训练多个“个体学习器”，然后采用取平均值或多数投票的方式对多个学习器的输出结果进行融合，从而实现目标任务的学习。Bagging 算法^[13]是并行式集成学习的著名代表，它通过对训练数据采用有放回的随机抽样构造多个独立的采样集，并训练出多个独立并行的个体学习器，最后实行多数投票原则得到最终分类结果。本文借鉴 Bagging 算法的思想对多个分类子网络(即集成学习中的个体学习器)进行集成，从而实现暴恐图像的标注。暴恐图像自动标注算法框图如图 1 所示。

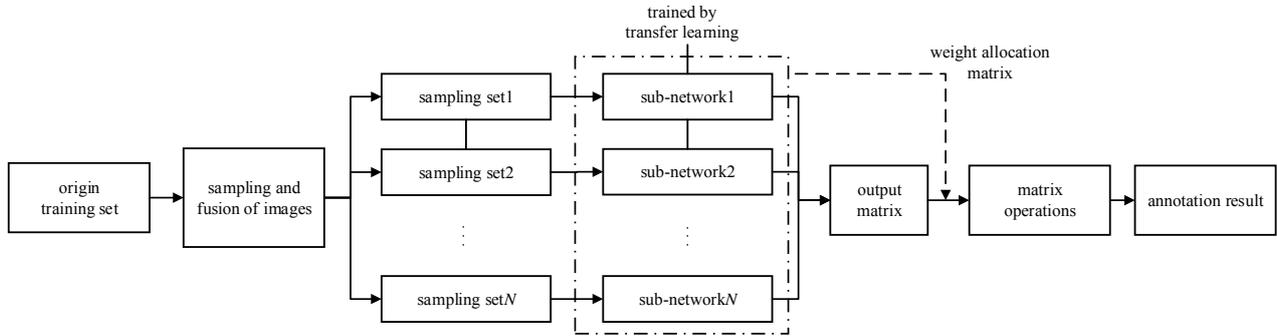


Fig.1 Block diagram of automatic annotation algorithm for violent images

图 1 暴恐图像自动标注算法框图

首先对数据集进行采样与融合，构造 N 个独立的采样集。由于原始数据集本身存在样本不均衡的问题，比如火、烟等标签出现频率明显高于其他标签。为避免 Bagging 算法中有放回的随机抽样可能会加剧样本不均衡的问题，本文在样本处理上先分别对每类标签所包含的图像采取有放回的随机抽样，再将各个标签类别下的采样集合并成一个数据平衡的训练集，重复 N 次后得到 N 个独立的采样集。接着，本文引入迁移学习来训练 N 个分类子网络，并提出了一种既能加快训练速度又能提高网络准确率的训练方式。最后，在结合模块，本文提出权重分配的结合策略，对各个子网络的输出结果进行融合，训练得到的 N 个分类子网络在各标签上都有不同的准确率，由此可以计算出一个权重分配矩阵，输出矩阵与权重分配矩阵通过一系列矩阵运算后得到最终的标注结果。

2 暴恐图像自动标注算法的实现

2.1 分类子网络的训练

图像标注问题可以通过多标签分类思想来处理，假定训练集图像 $X = \{x_i\}, i=1,2,\dots,n$ ，每幅图像 x_i 与 L 个语义标签相关联，并且第 i 幅图像的标签向量为 $y_i = (y_{i1}, y_{i2}, \dots, y_{iL}) \in \{0,1\}^L$ 。如果第 i 幅图像属于第 j 个语义类别，则 y_i 的第 j 个元素 $y_{ij} = 1$ ，否则 $y_{ij} = 0$ 。多标签分类的任务是从训练集 $I = \{(x_i, y_i) | 1 \leq i \leq n\}$ 中学习一个从特征空间到标签空间的映射： $f(x) \rightarrow y$ ，能够将新的未标注图像 x 映射到正确的语义关键词 y 上。

鉴于深度学习在图像处理领域取得的显著成就，本文选用卷积神经网络拟合多标签映射函数。AlexNet^[14]参数数量小，训练速度快，在小数量的图像数据库上有较好的分类准确性。由于自制的暴恐图像数据集图片数量较少，直接从几千张图像中学习 AlexNet 的网络参数很难达到较高的准确率。因此，本文在分类子网络的训练过程中引入迁移学习，将 AlexNet 在 ImageNet^[15]分类任务中训练得到的参数应用到暴恐图像标注任务中，其模型结构图如图 2 所示。

AlexNet 网络结构共有 8 层, 其中 C1~C5 代表前 5 层卷积层, 用作视觉特征提取器; fc6,fc7 和 fc8 代表后 3 层全连接层, 用作分类器^[11]。在 ImageNet 分类任务中, fc8 层的输出通过一个 1 000 维的 softmax 层转化为 1 000 类不同标签的概率分布, 常见的迁移学习训练方法是将 fc8 层的输出由 1 000 维改为目标任务的类别数进行训练。在暴恐图像标注任务中, 为了提高子网络的标注质量, 本文将 AlexNet 的预训练信息全部保留下来, 提出一种在 fc8 层后额外添加一个输入为 1 000 维, 输出为标签维数的全连接层 fc9 进行训练的训练方式。实验结果表明, 这种训练方式在加快训练速度的同时也提高了子网络的标注准确率。

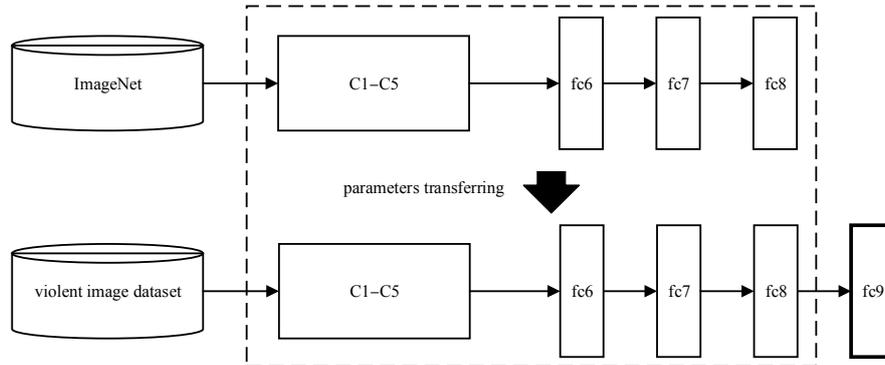


Fig.2 Model structure of sub-networks

图 2 分类子网络模型结构

最后一个全连接层 fc9 的输出通过一个 Sigmoid 函数转换为每一类标签的预测概率, Sigmoid 函数表示为:

$$h(x) = \frac{1}{1 + \exp(-x)} \quad (1)$$

本文使用 Sigmoid 交叉熵作为子网络的损失函数, 计算每一类标签预测概率值和真实值之间的误差再求均值, 公式可表示为:

$$\text{loss}(x, y) = -\frac{1}{N} \sum_{i=1}^N [y_i \log h(x_i) + (1 - y_i) \log(1 - h(x_i))] \quad (2)$$

式中: N 表示标签类别数; x 表示 fc9 层的输出结果, 则 x_i 表示子网络对第 i 类标签的预测概率, $0 \leq x_i \leq 1$, y 表示真实标签, $y_i \in \{0, 1\}$ 。

2.2 对结合模块的改进

集成学习的结合模块大多采用取平均值或多数投票原则对个体学习器的输出结果进行融合。在暴恐图像标注问题中, 考虑到每幅图像存在多个相关标签, 对同一标签类别, 子网络的学习能力各不相同。取均值和多数投票原则的结合策略很难充分发挥各子网络的长处, 甚至会加剧标注模型在各标签类别上的标注精确度差异。因此, 本文在结合模块提出一些改进, 根据每个分类子网络对同一标签不同的分类准确率进行权重分配, 进一步提高模型在每类标签上的标注精确度并减少方差。假设每幅图像与 L 个标签类别相关, 需要对 N 个子网络集成, 每类标签在不同网络上的分类准确率由矩阵 $\mathbf{P}_{L \times N}$ 表示, 其中 p_{ij} 表示第 i 类标签在第 j 个网络上的准确率。通过式(4)计算出第 i 类标签对第 j 个网络的权重 w_{ij} , 从而得到权重分配矩阵 $\mathbf{W}_{L \times N}$ 。

$$\mathbf{P}_{L \times N} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ p_{L1} & p_{L2} & \cdots & p_{LN} \end{bmatrix} \quad (3)$$

$$w_{ij} = \frac{\exp(p_{ij})}{\sum_{m=1}^N \exp(p_{im})} \quad (4)$$

$$\mathbf{W}_{L \times N} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1N} \\ w_{21} & w_{22} & \cdots & w_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ w_{L1} & w_{L2} & \cdots & w_{LN} \end{bmatrix} \quad (5)$$

令 $\mathbf{W}_i = [w_{i1}, w_{i2}, \dots, w_{iN}]$ 表示第 i 类标签在每一个子网络上的权重分布, 则权重分配矩阵 $\mathbf{W}_{L \times N}$ 可以表示为:

$$\mathbf{W}_{L \times N} = [\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_N]^T \quad (6)$$

式中上标 T 表示矩阵转置。

同理，令 $O_i = [o_{i1}, o_{i2}, \dots, o_{iN}]$ 表示输入图像的第 i 类标签在每个子网络上的预测概率值，将各子网络的输出结果组合为输出矩阵 $O_{L \times N}$ ，可表示为：

$$O_{L \times N} = [O_1, O_2, \dots, O_N]^T \tag{7}$$

输入图像最终的标签预测概率向量 $Prob_{1 \times L}$ 通过权重分布 W_i 和原始预测概率值 O_i 的乘积计算得到：

$$Prob_{1 \times L} = [W_1 \cdot O_1^T, W_2 \cdot O_2^T, \dots, W_L \cdot O_L^T] \tag{8}$$

标签预测概率向量 $Prob_{1 \times L}$ 表示的是标注模型预测 L 个标签类别出现在图像中的概率，预测概率值在 0 到 1 之间，本文设定阈值 0.5，大于阈值的所有标签类别即模型最终的标注结果。

不同于简单的取均值或多数投票原则，基于权重分配的结合方式使每个子网络在其分类准确率高的标签类别上发挥更大的作用，不仅增强标注模型的泛化能力，还改善了样本不均衡会降低模型标注质量的问题。

3 实验结果和分析

3.1 数据集

为验证本文标注方法的有效性，本文构建了暴恐图像数据集进行实验，同时还在公共数据集 Corel 5K 上进行实验，下文将分别介绍两个数据集的详细情况。

1) 暴恐图像数据集的构建

由于目前还没有公开的暴恐图像数据库，本文实验是在自制的暴恐图像数据集上训练和测试的。本文利用互联网爬虫技术从特定的网页和社交网络中爬取与暴恐信息相关的图片，再通过人工搜索对数据进行补充，同时也搜集了部分生活场景图片作为非暴恐图像用于测试。共搜集到 1 800 张暴恐图像和 300 张非暴恐图像，其中 1 500 张暴恐图像用于训练，300 张暴恐图像和 300 张非暴恐图像用于测试。暴恐图像数据集的部分暴恐样本示例和非暴恐样本示例分别如图 3(a)和图 3(b)所示。



Fig.3 Examples of violent image dataset

图 3 暴恐图像数据集样本示例

该暴恐图像数据集包含人群、火、烟、舰船、装甲车、飞机、刀、枪支、血腥和火炮这 10 种暴恐元素。它们在数据集的出现频率如表 1 统计。每类暴恐元素在数据集的出现频率均不少于 200，平均每幅暴恐图像包含 1.4 个暴恐元素。

表 1 暴恐元素出现频率统计表

Table1 Frequency statistics of violent elements

| violent elements | crowd | fire | smoke | warship | armored vehicle | airplane | knife | gun | blood | cannon |
|------------------|-------|------|-------|---------|-----------------|----------|-------|-----|-------|--------|
| frequency | 225 | 368 | 339 | 214 | 234 | 211 | 227 | 229 | 214 | 210 |

2) Corel 5K 数据集

Corel 5K 图像集共包含科雷尔(Corel)公司收集整理的 5 000 幅图片，它已成为图像自动标注的重要标准数据集，被广泛应用于标注算法性能的比较。Corel 5K 数据集涵盖公共汽车、恐龙、海滩等 50 个语义主题，每个主题有 100 幅图片，每幅图片手动标注 1 到 5 个标签，数据集中共包含 260 类标签。本文使用 4 500 张图片进行训练，余下的 500 张图片用于测试。

3.2 评价指标

本文依据分类思想对暴恐图像进行自动标注，因此选用常用的分类指标平均准确率 P 、平均召回率 R 、调和指数 $F1$ 值和非暴恐图像误判率 Er 来评估本文模型的标注性能。计算公式如下，

$$P = \frac{1}{N} \sum_{i=1}^N \frac{A_i^n}{A_i^m} \tag{9}$$

$$R = \frac{1}{N} \sum_{i=1}^N \frac{A_i^n}{A_i^y} \tag{10}$$

$$F1 = \frac{2P \cdot R}{P + R} \tag{11}$$

$$Er = \frac{B_e}{B} \tag{12}$$

式中： A_i^y 表示第*i*类标签包含的图像数量； A_i^m 表示模型预测出包含第*i*类标签的图像数量； A_i^n 表示预测正确的图像数量； B 表示非暴恐图像的数量； B_e 表示被标注出暴恐元素的非暴恐图像数量(即被误判的非暴恐图像数量)。 $P,R,F1$ 值越大， Er 值越小，模型标注性能越好。

3.3 网络训练

对分类子网络的训练，实验采用一块 2 GB 显存的 Nvidia Quadro K620 GPU，代码基于 PyTorch 框架实现，采用 Adam 算法优化网络，学习率为 0.000 1，权重衰减参数设置为 0.005，batch size 设为 32。本文提出了一种保留 AlexNet 全部预训练信息，在网络最后额外添加一个全连接层进行训练的训练方式。本文将在不利用迁移学习进行训练、利用常见的迁移学习训练方式和利用本文提出的训练方式下生成的子网络进行对比，对比指标包括网络迭代一次所需时间、网络收敛时的训练损失值、网络训练完成后在测试集上的 P,R 和 $F1$ 值。对比结果如表 2 所示，在本文提出的训练方式下，网络迭代一次用时 29 s，比常见的迁移学习训练方式慢 4 s，但是相比不利用迁移学习进行训练的方式节省了 14 s，网络收敛时的损失值最低，且生成的子网络标注性能在各项指标上都表现较好。

表 2 不同训练方式对比

Table2 Comparison of different training methods

| training methods | time for each epoch/s | training loss | $P/\%$ | $R/\%$ | $F1/\%$ |
|---------------------------|-----------------------|---------------|--------|--------|---------|
| without transfer learning | 43 | 0.20 | 67 | 53 | 59 |
| common training method | 25 | 0.08 | 74 | 62 | 67 |
| proposed training method | 29 | 0.03 | 81 | 75 | 77 |

分类子网络的数量也是影响集成结果的一个重要参数，本文将子网络的数量设置在 2 到 10 之间进行对比测试，测试结果如图 4 所示。不同数量的分类子网络组成的标注模型准确率 P 均大于 78%，随着子网络数量的增加，召回率 R 呈现不断上升的趋势，当个体学习器数量为 7 时，准确率为 84%，召回率为 73%，均达到最高。因此本文选择训练 7 个子网络来构建暴恐图像标注模型。

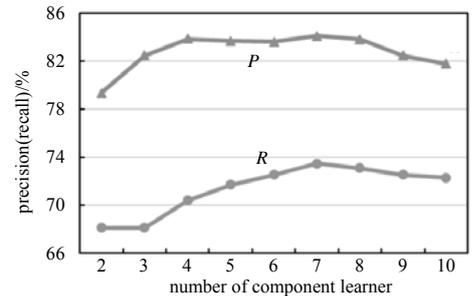


Fig.4 Contrast test results of sub-networks number
图 4 分类子网络数量对比测试结果图

3.4 实验结果

为了验证本文方法的有效性，本文选择经典机器学习算法 SVM、C4.5 决策树、KNN 进行对比试验。本文对暴恐图像数据集的每幅图像提取颜色特征(颜色矩)和纹理特征(方向梯度直方图、灰度共生矩)用于训练机器学习分类器，并采用十折交叉验证来评估分类器的性能。为了验证集成学习相较于单个网络在暴恐图像标注方法准确率上有提升，本文还利用单个 CNN 进行对比试验，CNN 选择 AlexNet 进行训练，Sigmoid 交叉熵作为损失函数。同时，为了验证本文标注方法不仅在暴恐领域有效，在一般场景下同样有效，还在公共数据集 Corel 5K 上进行实验，并与其他图像标注方法 FF-CNN^[11]、VSE+2PKNN-ML^[7]进行对比。各方法实验结果如表 3 所示。

表 3 各图像标注方法性能对比

Table3 Performance comparison of image annotation methods

| dataset | algorithm | $P/\%$ | $R/\%$ | $F1/\%$ | $Er/\%$ | t/s |
|-----------------------|-----------------|--------|--------|---------|---------|-------|
| violent image dataset | SVM | 58 | 43 | 49 | 59 | 0.42 |
| | C4.5 | 43 | 33 | 37 | 62 | 0.08 |
| | KNN | 58 | 30 | 39 | 56 | 0.24 |
| | CNN | 78 | 71 | 73 | 31 | 2.08 |
| | proposed method | 84 | 73 | 79 | 26 | 20.80 |
| Corel 5K | FF-CNN | 41 | 37 | 39 | - | - |
| | VSE+2PKNN-ML | 41 | 52 | 46 | - | - |
| | proposed method | 43 | 49 | 46 | - | 26.03 |

表 3 中最后一列 t 统计了每种标注方法标注一幅图像所需的时间，从侧面展示了各算法的复杂度。在 Corel 5K 数据集上，本文方法较 FF-CNN 在各项指标上均有一定的提高，相较 VSE+2PKNN-ML 在平均准确率上提高了 2 个百分点，说明本文标注方法在一般场景中同样具有有效性。

在暴恐图像数据集上，本文方法相较于机器学习算法中表现最好的 SVM，在准确率上提高了 26%，召回率提高了 30%。表明在暴恐图像数据集中，卷积神经网络提取的特征比人工选择的特征更容易被分类器学习。相较于单个 CNN，本文方法在准确率和召回率上均有所提高，表明对多个网络的集成提升了模型的标注性能。标注每幅图像用时最少的为 C4.5 算法，由于本文算法基于集成学习实现，加载多个分类子网络就需要 16 s，故本文算法标注时间较长。

图 5(a)和图 5(b)分别展示了多种算法在暴恐图像数据集中各标签类别上的准确率和召回率对比，在准确率上，本文方法表现突出，对“舰船”类别的准确率最高，达到 94%；在召回率上，本文方法和 CNN 表现相近，机器学习算法召回率均较低。

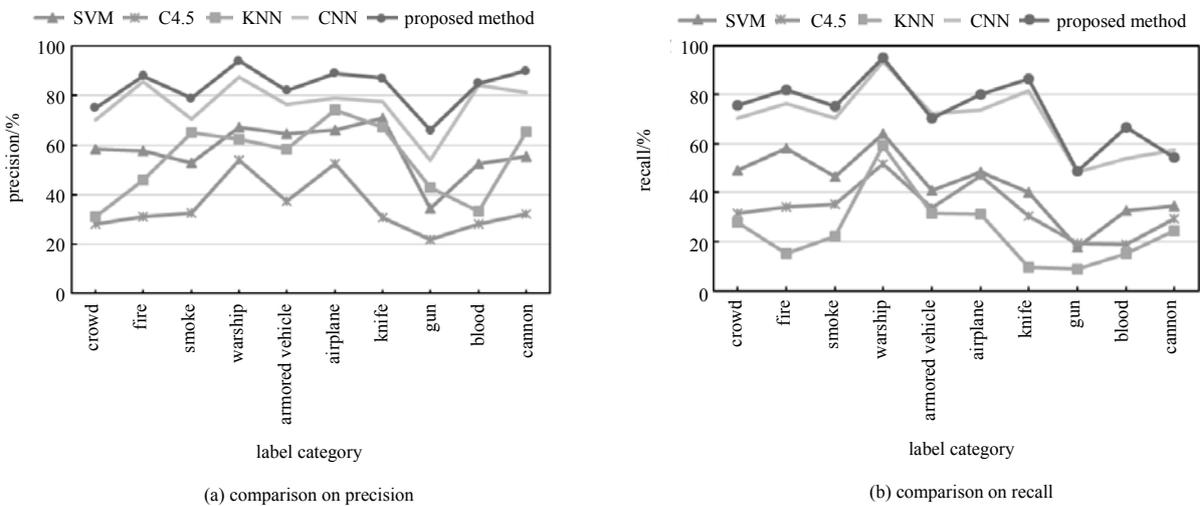


Fig.5 Performance comparison of various algorithms on different label categories
图 5 多种算法在各标签类别上的性能对比

为了验证本文对结合模块所作改进的有效性，对取均值、多数投票和本文提出的权重分配这 3 种结合方式进行对比，图 6 展示了不同结合方式在暴恐图像数据集中各标签类别上的准确率对比。

在“舰船”和“枪支”类别上，本文提出的权重分配结合方式准确率略低于其他 2 种结合方式，在其他类别上，权重分配结合方式准确率均最高。本文计算了这 3 种结合方式在各标签类别准确率上的标准方差，分别为 10.1,11.8 和 8.2。取平均值和多数投票方式在“火”、“舰船”、“血腥”这些出现频率较高、特征明显的类别上准确率较高，但在其他类别上准确率波动较大。权重分配方式在“枪支”类别上准确率稍低，在其他类别上准确率分布平稳。表明权重分配结合方式对样本不均衡会影响模型标注质量的问题有所改善。

表 4 展示了不同方法在暴恐图像标注任务中的实际效果。当图像中的暴恐元素有遮挡时，本文方法会有错误标注或缺漏标注的情况，但总体标注准确率较高。

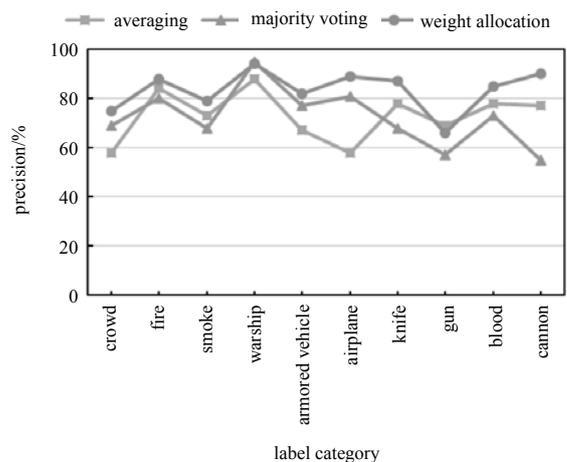


Fig.6 Precision comparison of different ensemble strategies on different label categories
图 6 不同结合方式在各标签类别上的准确率对比

表 4 暴恐图像标注示例
Table 4 Examples of violent image annotation

| examples |  |  |  |  |
|-----------------|---|---|--|---|
| ground truth | knife | warship airplane | gun | fire airplane |
| SVM | none | warship | none | crowd |
| C4.5 | airplane gun | none | crowd | crowd |
| KNN | crowd | warship | crowd | crowd |
| CNN | knife gun | warship | knife gun | fire smoke |
| proposed method | knife | warship airplane | gun | fire |

4 结论

本文提出了一种基于集成分类的暴恐图像标注方法，通过多标签学习建立图像特征与语义关键词之间的映射函数，并提出一种利用权重分配矩阵对各子网络的输出结果进行融合的集成学习结合方式，改善了样本不平衡会导致标注模型在各标签类别上标注精确度差异性大的问题。改善了样本不平衡会导致标注模型在各标签类别上标注精确度差异性大的问题。暴恐图像标注模型的准确率和召回率分别达到 84% 和 73%，相较于传统机器学习算法有较大提升，并且在各标签类别上的标注准确率分布稳定。

参考文献：

- [1] 徐勇,张慧. 图像自动标注方法研究综述[J]. 现代情报, 2016,36(3):144-150. (XU Yong,ZHANG Hui. Summary of automatic image annotation method[J]. Journal of Modern Information, 2016,36(3):144-150.)
- [2] 刘凯,张立民,孙永威,等. 利用深度玻尔兹曼机与典型相关分析的自动图像标注算法[J]. 西安交通大学学报, 2015, 49(6):33-38. (LIU Kai,ZHANG Limin,SUN Yongwei,et al. An automatic image annotation algorithm using deep Boltzmann machine and canonical correlation analysis[J]. Journal of Xi'an Jiaotong University, 2015,49(6):33-38.)
- [3] MAJIDPOUR J,KHEZRI E,HASSANZADE H,et al. Interactive tool to improve the automatic image annotation using MPEG-7 and multi-class SVM[C]// IEEE Conference on Information and Knowledge Technology. Urmia,Iran:IEEE, 2015:1-7.
- [4] MEHMOOD Z,MAHMOOD T,JAVID M A. Content-based image retrieval and semantic automatic image annotation based on the weighted average of triangular histograms using support vector machine[J]. Applied Intelligence, 2018,48(1):166-181.
- [5] JI Q,ZHANG L,LI Z. KNN-based image annotation by collectively mining visual and semantic similarities[J]. KSII Transactions on Internet & Information Systems, 2017,11(9):4476-4490.
- [6] SU F,XUE L. Graph learning on K nearest neighbors for automatic image annotation[C]// Proceedings of the 5th ACM on International Conference on Multimedia Retrieval. Shanghai,China:ACM, 2015:403-410.
- [7] ZHANG W,HU H,HU H. Training visual-semantic embedding network for boosting automatic image annotation[J]. Neural Processing Letters, 2018,48(3):1503-1519.
- [8] 张华忠,侯进. 基于决策树 C4.5 集成算法的图像自动标注[J]. 计算机应用研究, 2018,321(7):308-310. (ZHANG Huazhong,HOU Jin. Image annotation based on decision tree C4.5 ensemble algorithm[J]. Application Research of Computers, 2018,321(7):308-310.)
- [9] FAKHARI A,MOGHADAM A M E. Combination of classification and regression in decision tree for multi-labeling image annotation and retrieval[J]. Applied Soft Computing, 2013,13(2):1292-1302.
- [10] RAJCHL M,LEE M,OKTAY O,et al. Deep cut: object segmentation from bounding box annotations using convolutional neural networks[J]. IEEE Transactions on Medical Imaging, 2017,36(2):674-683.
- [11] LIN Y,ZHANG H. Automatic image annotation via combining low-level colour feature with features learned from convolutional neural networks[J]. Neuro Quantology, 2018,16(6):679-685.
- [12] LIU F,XIANG T,HOSPEDALES T M,et al. Semantic regularization for recurrent image annotation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu,HI,USA:IEEE, 2017:2872-2880.
- [13] BREIMAN L. Bagging predictors[J]. Machine Learning, 1996,24(2):123-140.
- [14] KRIZHEVSKY A,SUTSKEVER I,HINTON G E. ImageNet classification with deep convolutional neural networks[C]// Advances in Neural Information Processing Systems. Lake Tahoe,Nevada,USA:[s.n.], 2012:1097-1105.
- [15] DENG J,DONG W,SOCHER R,et al. ImageNet: a large-scale hierarchical image database[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami,FL,USA:IEEE, 2009:248-255.