计算机科学与探索
Journal of Frontiers of Computer Science and Technology
ISSN 1673-9418,CN 11-5602/TP

《计算机科学与探索》网络首发论文

题目：基于深度学习的实时吸烟检测算法
作者：陈睿龙，罗磊，蔡志平，马文涛
网络首发日期：2020-10-23
引用格式：陈睿龙，罗磊，蔡志平，马文涛．基于深度学习的实时吸烟检测算法．计算机科学与探索．
https://kns.cnki.net/kcms/detail/11.5602.tp.20201022.1731.004.html

计算机科学与探索
Journal of Frontiers of Computer Science and Technology doi: 10.3778/j.issn.1673-9418.2009069

基于深度学习的实时吸烟检测算法

陈睿龙，罗　磊，蔡志平+，马文涛

国防科技大学 计算机学院，长沙 410073
+ 通信作者 E-mail: zpcai@nudt.edu.cn

摘　要: 在公共场所内吸烟，不仅对自身、他人身体健康造成潜在的危害，还存在造成火灾等现象的隐患。因此，出于健康和安全方面的考虑，本文为机场、加油站、化工仓库等严禁吸烟的场所，设计了一种基于深度学习的能快速发现和警告吸烟行为的检测模型。该模型使用了卷积神经网络对摄像头所拍摄的视频流输入帧进行处理，经过图像特征提取、特征融合、目标分类以及目标定位等过程，定位烟头的位置，进而判断出吸烟行为。常见的目标检测算法针对小目标物体检测效果不甚理想，检测速度亦有待提高。通过设计的一系列卷积神经网络模块，不但减少了模型计算量，加快了推演速度，从而能够满足实时性要求，而且提高了小目标物体(烟头)检测准确率。此外，运用了一些模型训练的技巧，提升了模型的鲁棒性。由于缺乏现有数据集，本文自制了一个与吸烟行为相关的数据集，通过对照试验，证明本文提出的算法在本数据集以及一些公开数据集上，有着更好的检测效果。

关键词: 计算机视觉；微型目标检测；实时性；吸烟检测；鲁棒性
文献标志码：A　　**中图分类号：**TP391.4

陈睿龙，罗磊，蔡志平，等. 基于深度学习的实时吸烟检测算法[J].计算机科学与探索
CHEN R L, LUO L, CAI Z P, et al. The algorithm for real-time smoking detection based on deep learning[J]. Journal of Frontiers of Computer Science and Technology

The algorithm for real-time smoking detection based on deep learning*

CHEN Ruilong, LUO Lei, CAI Zhiping+, MA Wentao
College of Computer Science, National University of Defense Technology, Changsha 410073, China

Abstract: In public, smoking behavior not only causes pathological harm to human health, but also exerts danger of fire hazards, making damage to people's safety and public property. For the health and safety considerations, we designed a real-time smoking detection model based on deep learning for stations, airports, hospitals, gas stations, chemical warehouses, and other places where smoking is strictly prohibited. Our model uses a convolutional neural network to process the input frame from the video stream captured by the webcam. Through the process of image feature extraction, feature fusion, target classification and target positioning, the coordinate of the cigarette is located,

The National Natural Science Foundation of China under Grant No. 62072465 (国家自然科学基金).

and then we can find out the smoking behaviors. Common object detection algorithms are not ideal for small target objects and the detection speed needs to be improved.We designed a series of convolutional neural network modules to reduce the amount of model parameters and pick up the inference speed to meet real-time requirements as well as improving the accuracy of small target object (cigarette) detection. We also come up with some training skills to make the model more robust. Due to the lack of relevant dataset, we produced a dataset related to smoking behaviors. Through comparative experiments, it is proved that the algorithm proposed in this article has better detection effects on our dataset and some public datasets.

Key words: computer vision; small object detection; real time; smoking detection; robustness

1 引言

随着技术不断的进步，吸烟检测的方法也随之不断的改进。传统吸烟检测的方法通常都是通过烟雾传感器、可穿戴设备等物理方式进行检测。然而，这些方法存在诸多局限，一是室外场景中烟雾浓度被极大的稀释，无法被烟雾传感器所感应；二是可穿戴设备执行检测的成本较高，需要人人拥有。此外，这种方法通过判断肢体多个部位的运动轨迹和速度，与吸烟的动作行为进行模式匹配，进而通过支持向量机(Support Vector Machine, SVM)等机器学习分类方法对匹配度进行判断，故而该类方法的检测的准确率和效率比较低[1, 2]。

除了使用物理设备的方法检测吸烟外，还有学者尝试使用传统图形学中目标检测的方法检测吸烟，这类方法主体分为三个步骤[3]：首先设定不同的大小和步长滑动窗口，然后将所有的窗口在图像上每个位置进行滑动。对于每个窗口，通过方向梯度直方图(Histogram of Oriented Gradient, HOG)或尺度不变特征变换(Scale-invariant feature transform, SIFT)方法提取出待测物体特征，最后使用分类算法对每个滑动窗口进行分类，比如 SVM、Adaboost 等方法，选取最高得分的滑窗作为检测结果。但是这类方法存在如下缺陷，首先是检测效果不理想，易受到其他物体干扰、并且定位不准确，依赖预设滑动窗口尺寸与滑动步长，其次该方法计算量也大，需要对每个滑动窗口进行特征处理与分类判断，最后手动提取特征的方式和过程比较复杂，不具备泛化性。

自从 2012 年 AlexNet 网络模型的诞生，并在当年获得了 ImageNet 图像分类比赛的冠军[4]，由此，无论是学术界还是工业界都对深度学习在计算机视觉领域中的运用给予了广泛的关注，如人脸识别、车辆检测等。在本文中，将吸烟检测问题归为目标检测问题，即通过定位行人与烟头的位置关系来判断是否存在吸烟现象。

本文通过借鉴 YOLO(You Only Look Once)[5]高性能检测算法，设计了一种轻量级吸烟检测网络模型，该模型通过融合多层次不同的特征图向量、增加注意力机制模块、残差模块以及 SPP(Spatial Pyramid Pooling)模块等改良了原网络结构，提高了小型目标的检测准确率；同时，减少了模型的卷积核参数，进而减少了模型计算量，加快了模型最终的推演速度，达到满足检测实时性要求。针对模型鲁棒性问题，通过训练数据增强、改变损失函数与激活函数、增加正则化方法、利用上下文信息等方法提高了本文模型的鲁棒性。

本文第 2 节介绍目标检测的相关工作，包括单阶段和双阶段的典型检测算法。第 3 节介绍吸烟检测中的难点，主要包括细粒度目标难以检测的问题、实时性问题以及模型鲁棒性。通过对这些难点的分析，提出了具体的改进与解决方法。第 4 节将详细介绍本文模型的结构。第 5 节将介绍本文自己制作的数据集，以及所设计模型在该数据集以及其他数据集上的性能表现。

2 相关工作

传统目标检测算法包括 VJ(Viola and Jones)级

联检测器[6]、HOG 检测器[7]以及 DPM(Deformable Parts Model)模型[8]等，存在着计算量大、手工提取特征复杂、特征表征性能较弱以及模型的泛化能力较差等问题，很难解决不同场景中的吸烟检测问题。而卷积神经网络中的卷积核作为“天然”的滤波器，具有优越的特征提取能力，这也正是其在计算机视觉领域中取得颠覆性突破的主要因素之一。除此之外，通过使用多种场景的数据集进行训练，使得卷积神经网络模型具有很强的泛化能力，因此，目前深度学习已成为目标检测领域的首选解决方案。

通常目标检测所使用的经典特征提取网络如 VGG(Visual Geometry Group)[9]、GoogLeNet[10]、ResNet[11]等，之所以使用在图像分类领域里面取得显著成效的预训练网络结构，是因为其具有强大的特征抽取能力，通过提取到的大量特征完成高难度的多图像分类任务。而目标检测同样需要大量的图像特征，因此，检测模型的骨干网络(backbone)通常会使用 GoogLeNet 的 Inception 结构、ResNet 的残差结构等，不仅可以避免神经网络反向传播更新权重时的梯度消失等问题，还能够加速模型收敛。

目标检测算法发展至今，出现了两大流派，双阶段和单阶段检测算法。前者的代表算法主要包括 Faster RCNN[12]、FPN(Feature Pyramid Networks)[13]、RFCN(Region Fully Convolutional Networks)[14]以及 Cascade RCNN[15]等，以 Faster RCNN 为例，这类算法首先通过基础卷积神经网络提取图像特征，输出特征图。再使用 RPN(Region Proposal Network)网络，对输入的特征图中的每个位置，使用 softmax 预测 $2\times k$ 个分数，k 为本文预设的锚框(anchor)个数，2 表示前景和背景两个类别的得分，同时利用边框回归，对每个特征图的位置预测 $4\times k$ 个坐标回归特征矩阵，使得前景样本的锚框通过变换更加接近真值，之后通过候选网络层(proposal layer)通过非极大值抑制(Non Maximum Suppression, NMS)[16]和得分排序，筛选生成候选区域(region proposals)。综合候选区域与之前得到的特征图信息，经过 ROI (Region of Interest)池化生成候选特征图(proposal feature maps)，将其传输至全连接层完成最终的物体分类和边框回归定位。后者的代表算法则包括 YOLO、SSD[17]、RetinaNet[18]以及 EfficientDet[19]等，以 YOLO 为例，这类算法将分类问题转换为回归问题，不需要经过提取候选区域步骤，而是直接通过卷积神经网络得出目标的位置与类别。通过基础卷积神经网络提取图像特征后，直接在每个特征图上执行目标分类与边框回归定位，同样借助锚框加速边框回归，输出向量再经过非极大值抑制得出最终预测结果。两类算法各有优劣，单阶段的算法长于速度，具有较快的模型推演速度，但在预测精度方面则稍逊一筹；相比之下，双阶段的检测算法的目标检测准确率较高，但模型推演速度较慢。

基于视觉的吸烟检测易受图像噪声干扰进而产生误检，并且烟头目标较小，难以发现与识别，因此，目前学术界中，基于目标检测的吸烟检测方法较少，相关工作和理论并不完善。本文借助目标检测理论的基本思路，将烟头视为待检目标，通过设计卷积神经网络的结构，在本人制作的数据集中进行训练，与经典的深度学习检测器 YOLO、SSD、FasterRCNN 等相比，对吸烟行为的检测，同时具有更高的检测准确率和检测速度。此外，在一些公开数据集中，本文的算法模型也有着更好的表现。

3 吸烟检测难点以及解决方案

通过深度学习目标检测来完成吸烟检测的方法，存在着如下难点：

3.1 小目标检测

首先，需要对小目标物体给出相关定义。在微软 COCO 数据集中，存在着对小物体目标的描述，指的是目标面积小于 32×32 的物体，单位为像素。小目标物体由于分辨率较低、图片拍摄过程中较之于大物体更容易出现模糊、抖动等现象的干扰，并且抗噪能力较弱，即使是常见的图形学噪声如椒盐噪声、高斯噪声等也容易对目标物体造成较大程度的干扰，通过去噪手段也很难完全恢复小物体目标

的特征。其次，小目标物体受制于其本身尺寸，携带的图形学信息较少，因此在提取特征的过程中，能提取到的特征非常少。

针对小目标物体难以发现的问题，可以利用模型浅层的特征图(feature map)向量。特征图是与目标检测相关的基本概念，输入图像经过一个卷积核或者池化核后，输出的二维矩阵向量就是特征图，与特征图息息相关的另一个概念为感受野(Receptive Field)。所谓感受野，指的是卷积神经网络每一层输出的特征图上每个位置的输出向量在输入图片上映射的区域大小。感受野的大小与原输入图像所经过的卷积层或者池化层中核的尺寸呈平方正相关关系，感受野越大的特征图，它的每个位置的向量所能“感受”的区域越大，越能够捕捉深层次的高维隐藏特征以及大型物体目标的特征，相反，小感受野的特征图通常捉浅层细节特征以及小型物体目标的特征。对于卷积神经网络来说，浅层网络的特征图所具有的感受野较小，可以发现较小的目标物体。同理，如果模型的卷积神经网络层数较多，深层的特征图感受野越大，特征图中的一个向量代表了原图较大区域的特征，则更多的将小物体周围背景或者其他物体的特征纳入特征表示，从而使模型丧失发现小物体的能力。浅层网络通常提取的特征偏向于细节特征，主要包括边缘、形状等方面；而深层网络则提取较为抽象的特征，如图 1 所示。对此，可以借鉴 FPN 网络的思想，将深浅层次的特征图进行融合，既尽可能多的保留小物体的特征，又可以很好地检测不同尺度下的待测物体。这里的融合指的不是特征图对应二维空间位置向量值的相加，而是在维度(channel)层面的扩张，可以理解为多张特征图的“堆叠”。

Fig.1　Feature maps of different scales

图 1　不同尺度特征图

(左图为原图，从左至右特征图所在的卷积层深度依次增加)

注意力机制同样能够增强模型检测小物体目标能力。注意力模型最初被应用于机器翻译任务，2017 年，SENet[20](Squeeze-and-Excitation Networks)通过设计的注意力模块，取得了最后一届 ImageNet 图像分类比赛的冠军，标志着注意力机制在计算机视觉领域中的成功运用。2018 年，注意力模型 CBAM[21](Convolutional Block Attention Module)在 SENet 的基础上，设计了结合空间位置和特征通道两个维度的注意力模块，取得了更好的效果，如图 2 所示。

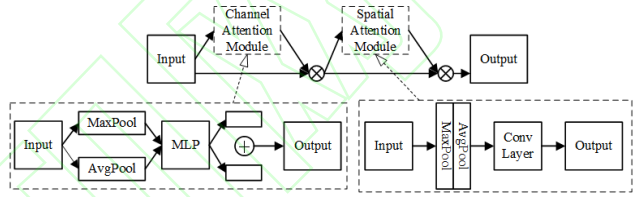


Fig. 2　Convolutional Block Attention Module

图 2　卷积注意力模块

特征通道表示某层特征图的个数，与该层卷积核的个数相等。假设特征图的输入为 c(特征通道数) × h(特征图高度) × w(特征图宽度)，通过将特征图的 h 和 w 进行平均池化和最大池化压缩成一维，之后将这两个 $c \times 1 \times 1$ 的输出向量加和，得出 $c \times 1 \times 1$ 的输出向量，将其作用于原输入特征图执行卷积相乘运算，来增强特征通道维度的注意力。空间位置的特征注意力机制则分别将平均池化和最大池化作用于输入特征图，再将二者在特征通道维度进行拼接，此时特征图变为 $2c \times h \times w$，然后将其输入至一个卷积核个数为 c 的卷积层，最后再与原特征图进行卷积相乘，完成空间位置的注意力增强。实验证明[20, 21]，注意力模块能够帮助卷积神经网络提取到更加鲁棒的特征，本文设计了注意力模块来解决小目标物体难以捕捉的问题，如图 4 (c) 所示，它的设计借鉴了 CBAM 的设计思想，模块的上半部分通过融合经过最大池化层和平均池化层的特征图向量，增强了空间位置的注意力，下半部分通过加和经过最大池化层和平均池化层的特征图向量，增强了特征通道维度的注意力。

3.2　轻量级特征提取网络

吸烟检测需要关注实时性问题，需要及时发现并警示吸烟行为，由于吸烟动作和过程较为短暂，若不实时地执行检测，及时作出响应，则容易出现漏检的

情况。目标检测是两个子任务的结合，即图像分类和边框定位，而两者都需要大量的待检物体的特征。因此，检测模型通常由两个部分组成，骨干网络(backbone)和执行检测的头部(heads)。骨干网络通常使用大量的卷积层提取特征，头部则使用所提取的特征完成目标定位和分类过程，模型推演的时间损耗主要在于骨干网络。骨干网络的规模之所以越来越庞大，是为了学习更为复杂的非线性映射关系，提取更多潜在的特征，但随着层数和参数量增多，推演计算量也不可避免地增加，因此，需要在保证能够提取到充分的目标特征的基础上，设计轻量的特征提取网络。

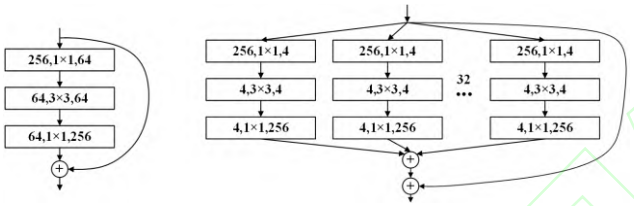


Fig.3 ResNet (left) module and ResNeXt module (right)

图 3 ResNet 模块(左)与 ResNeXt 模块(右)

实际上，在一些场景中部署检测模型，需要考虑硬件资源的计算能力，这也是设计轻量级特征提取网络的另一重要因素。比如将模型部署到 Nvidia TX2 等开发板上，由于开发板的算力远远达不到深度学习专用显卡的算力，模型的推演速度将进一步被减缓，因此，设计轻量型特征提取网络更加具有研究意义。本文设计了残差模块和 SPP 模块，并将其与普通卷积层组建形成骨干特征提取网络。与常规的卷积神经网络层组成的骨干特征提取网络相比，本文的骨干网络在不增加参数量的情况下，提高了特征提取能力。

残差模块 ResNet 在图像分类任务中取得了显著成效，其具体的结构如图 3 所示，ResNeXt[22]借鉴继承了 ResNet 的残差思想，并且其通过实验证明了其提出的分组多基数路径结构可以在不增加参数复杂度的前提下学习更多的特征，提高分类准确率，同时还减少了模型超参数的数量。因此，本文借鉴了 ResNeXt 模块，如图 4 (a) 所示，本文在检测模型的特征提取的骨干网络中设计了残差模块，代替了传统的简单卷积网络层。该模块首先通过卷积步长为 2 的卷积操作完成特征图的下采样过程，之后参照了 ResNeXt 结构，通过建立的 32 组多路径结构，完成分组卷积，同时使用了两个残差组卷积的堆叠，通过双层残差进一步增强网络的特征学习能力，避免出现梯度爆炸、梯度消失等问题。

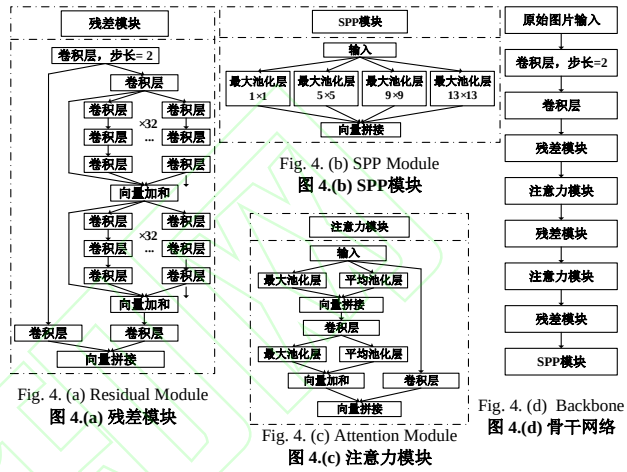


Fig. 4. (a) Residual Module
图 4.(a) 残差模块

Fig. 4. (b) SPP Module
图 4.(b) SPP模块

Fig. 4. (c) Attention Module
图 4.(c) 注意力模块

Fig. 4. (d) Backbone
图 4.(d) 骨干网络

Fig.4 Basic network module of detection model

图 4 检测模型的基础网络模块

SPPNet[23](空间金字塔池化)原本是为了解决卷积层和全连接层相接时，不同尺寸的输入特征图无法产生固定长度的特征表示，从而导致卷积层输出与全连接层输入连接失败的问题。本文则是利用了 SPPNet 的潜在特征融合特性，它利用不同步长的最大池化核作用于输入特征图，得到不同大小感受野的输出特征图，然后通过特征通道维度的向量拼接，融合了多级特征，加强网络对抽象特征、深层语义特征的学习能力。相比于执行多次卷积操作后，再进行向量拼接，取得相同效果的同时，显著减少模型运行复杂度和参数计算量。因此，本文借鉴了 SPPNet 模块，在骨干网络的末尾设计了 SPP 模块，如图 4 (b) 所示，它包含了四路不同尺寸的池化核模块，输入的特征图分别通过四个池化层后会产生四类感受野面积不同的特征图，其中通过 1×1 池化核的特征图感受野最小，通过 13×13 池化核的特征图感受野最大感受野范围正比于池化核的尺寸。SPP 模块通过融合四种不同感受野的特征图向量，弥补了整体网络下采样次数不足的缺陷，进而有利于发现目标物体整体特征以及深层语义特征。

3.3 模型鲁棒性

基于深度学习的目标检测算法模型除了主体是

由卷积神经网络构成，还包括了常规的组件，如激活函数、损失函数、正则化方法等。激活函数为模型提供了非线性建模的能力，卷积和池化操作仅仅是矩阵运算，仅仅是高维空间的线性建模与计算，而激活函数通过输入输出的非线性映射，可使神经网络模型去学习并拟合非线性函数的函数。目标检测中子问题可归结为目标分类和坐标回归定位，无论是分类还是回归函数都不可能仅仅是线性函数，因此，需要选择性能更强的激活函数 Mish[24]代替原有的激活函数。Mish 激活函数如式(1)所示：

$$f(x)=x\cdot\tanh(\ln(1+e^{x})) \tag{1}$$

损失函数衡量模型预测的好坏，度量预测值与实值之间的差异，也就是模型训练的目标。损失函数通过公式的方式描述了本文要解决的问题，而目标检测的损失函数可分为两个部分，分类损失和边框回归损失。分类损失函数选择 YOLO 算法中的二分类交叉熵(Binary Cross Entropy, BCE)损失函数，该部分的损失函数包括两个部分，置信度损失和类别损失如式 (2) 所示：

$$\begin{aligned}L_{class}=&-\lambda_{obj}\sum_{i=0}^{S\times S}\sum_{j=0}^{anchor-1}I_{ij}^{obj}[t_{conf}\log(pred_{conf})+(1-t_{conf})\log(1-pred_{conf})]\\&-\lambda_{noobj}\sum_{i=0}^{S\times S}\sum_{j=0}^{anchor-1}I_{ij}^{noobj}[t_{conf}\log(pred_{conf})+(1-t_{conf})\log(1-pred_{conf})]\\&-\sum_{i=0}^{S\times S}\sum_{j=0}^{anchor-1}\sum_{c\in classes}I_{ij}^{obj}[t_{cls}(c)\log(pred_{cls})+(1-t_{cls})\log(1-pred_{cls})]\end{aligned} \tag{2}$$

其中 S 代表输出特征图大小，anchor 表示每个特征图向量负责预测的锚框个数，λ_{obj} 和 λ_{noobj} 表示惩罚项因子，t_{conf} 和 $pred_{conf}$ 分别表示存在物体的置信度真值和预测值，t_{cls} 和 $pred_{cls}$ 则分别表示物体的类别真值与预测值，I_{ij}^{obj} 表示第 i 个位置的第 j 个锚框是否存在待测物体，若存在物体其值为 1，不存在则为 0。

早期的目标检测算法 YOLO 使用 MSE 损失函数、Faster RCNN 使用 L1-smooth 损失函数，作为边框回归损失函数，然而以 Ln 范数衡量回归损失并不准确。2019 年，DIoU[25]损失函数的提出，使得目标检测的边框回归过程较之以往的损失函数更加快速与准确，该损失函数如式 (3) 所示：

$$L_{regression}=1-\frac{|B\cap B^{gt}|}{|B\cup B^{gt}|}+\frac{\rho^{2}(b,b^{gt})}{c^{2}} \tag{3}$$

其中 B 和 B^{gt} 分别表示预测的边框和边框真值，$B\cap B^{gt}$ 与 $B\cup B^{gt}$ 表示二者的交集面积和并集面积，b 和 b^{gt} 表示 B 和 B^{gt} 的中心点 $\rho^{2}(\ )$表示欧式距离，c 为覆盖 B 和 B^{gt} 最小矩形框的对角线长度。由此，得出检测模型的总体损失函数为式 (4) ：

$$L_{detection}=L_{class}+L_{regression} \tag{4}$$

机器学习中常用的正则化方法为 Dropout[4]，在神经网络前向传播的过程中以一定的概率忽略输入神经元，通过模拟人类的遗忘现象，防止模型出现过拟合现象，使模型更加鲁棒。Dropout 通常被广泛用于全连接层的正则化，但它作用于卷积层的效果并不明显。卷积层中的激活单元是空间关联的，即便是使用 Dropout 随即丢弃特征图某些位置上的向量，物体的信息仍然能够通过卷积网络传输到下一层。因此，本文使用 DropBlock 方法完成对特征图的正则化约束，以一定的概率忽略图中的相邻区域块而非某一个点，如图 5 所示。

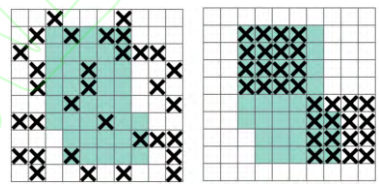


Fig. 5 Dropout regularization (left) and DropBlock regularization (right)

图 5 Dropout 正则化(左)与 DropBlock 正则化(右)

对于深度学习算法来说，数据集的重要性不言而喻，数据集的质量决定了检测模型的质量。通常来说，数据量越大、包含的待测场景越多，检测模型的泛化性能越强，检测准确率也会相应的提高。通过训练数据增强技术如 Mixup、CutMix[26]、多尺度放缩、平移、旋转以及对称等，进一步提高训练数据的多样性，防止模型训练数据太过单一。最后，利用上下文信息，在检测时设置两个待检测类，行人和烟。输出的最终预测结果向量经过非极大值抑制后，通过先选取出置信度较大的行人检测框，反向排除吸烟检测框与行人检测框 IoU (Intersection over Union) 小于 0 的预测向量，减少烟头误检率，提高检测准确率。

4 模型结构设计

本节将介绍本文提出的总体吸烟检测模型结构以及检测流程。该模型借鉴了 YOLO 算法的单阶段检测思想，直接通过深度学习模型对待测物体进行目标分类与边框回归定位，得到物体的位置与类别。

4.1 整体网络结构

首先是负责特征提取的骨干网络部分，如图 4 (d)所示，它由卷积层、残差模块、SPP 模块和注意力模块组合而成，学习和提取待测物体特征。图 6 所示的是本文提出的吸烟检测模型的整体网络架构，本文借鉴了谷歌 EfficientDet 模型中的 BiFPN 结构，融合了 3 种不同感受野尺度的特征图输出向量。经过下采样的特征图尺寸相应缩减为原特征的一半，通过双线性插值的上采样方法，扩大特征图尺寸，以便与原特征图完成特征通道维度的向量拼接。最终获得三个尺度的输出结果，尽可能的覆盖不同尺寸的香烟目标，使用非极大值抑制算法、上下文信息计算处理以及预测置信度排序等方法处理汇总三层的输出向量，排除干扰项，得出最终的待测目标类别以及在图像中的位置。

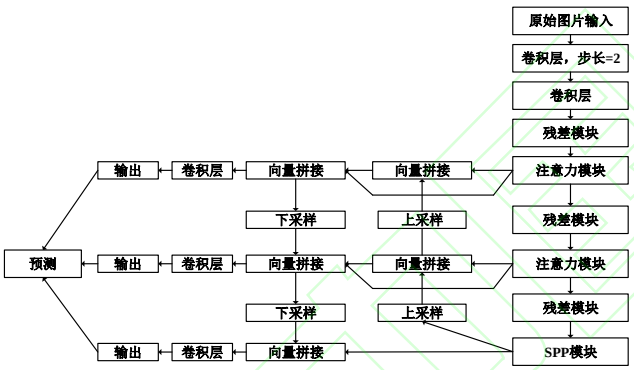


Fig. 6 Overall network structure of the detection model

图 6 检测模型的整体网络结构

4.2 吸烟检测流程

本文通过实时视频流协议(Real Time Streaming Protocol, RTSP)按照一定的帧数间隔读取视频帧，将帧进行简单预处理后输入到本文设计的吸烟检测模型中，经过模型计算出置信度较高的烟头以及行人所在位置，首先使用 NMS 排除冗余检测框，之后通过上下文信息关联算法，将每个烟头预测坐标与行人预测坐标进行 IOU 计算，排除计算值小于某个阈值的预测框，表示烟头与行人毫无关联，不存在吸烟现象。最后将捕捉到的吸烟现象视频帧进行标注并保存至本地，并向有关人员发出警示信息，具体的检测流程如图 7 所示。

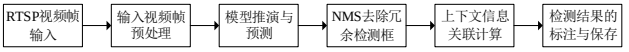


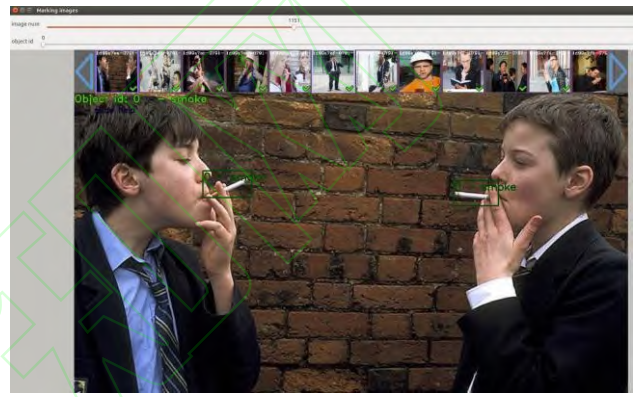
Fig. 7 Flow chart of smoking detection

图 7 吸烟检测流程示意图

5 实验分析

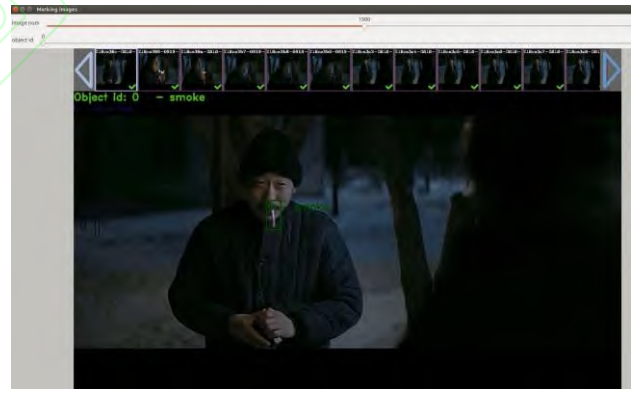
由于缺乏相应的吸烟检测数据集，首先完成相关数据集的制作，使用该数据集训练检测模型，深度学习训练框架为 Darknet，训练所使用的显卡型号为 NVIDIA GeForce RTX 2080Ti，操作系统为 64 位 Ubuntu kylin 16，配置 CUDA10.0，CUDNN7.5。

5.1 数据集的制作



(a) Data annotation (part one)

(a) 数据标注（一）



(b) Data annotation (part two)

(b) 数据标注（二）

Fig.8 Data annotation

图 8 数据标注

本文制作的数据集中的数据来源于爬取的谷歌和百度图片、HMDB 人类行为数据库中吸烟片段以及自己录制的吸烟视频。数据标注使用的是 Github 开源标注工具 YOLO Mark。目标标注的内容为五元组(class, x, y, w, h)，其中 class 表示物体类别，为 Int 类型的整数；0<=x, y, w, h<=1 表示目标物体的中心点坐标以及标注的真值框的高度和宽度在原输入图像中的比例，通过归一化方法处理标注

值，方便后续模型的推演与计算。数据标注的可视化如图 8 所示，数据集的划分如表 1 所示，干扰项指的是没有任何吸烟目标的样本。

Table 1　Settings of our Dataset

表 1　数据集的划分

	训练集	验证集	测试集
吸烟样本	3100	200	300
干扰项	500	20	50

5.2　实验结果

在使用 Darknet 训练框架搭建好本文的检测模型后，首先在本文自制的数据集上展开了实验。

首先进行数据预处理过程，数据增强手段包括多尺度放缩、平移旋转、对称、随机擦除以及 CutMix。整个训练过程中的迭代次数为 30000，批量大小为 64，使用 Adam 梯度优化器，起始学习率为 0.001，权重衰减系数为 0.0005，正则化方法采用 Dropblock，损失函数则为本文中的式(4)。图 9 所示为本模型在自制数据集上训练时的损失值的变化以及训练集和验证集的 mAP 的变化，发现模型在验证集的检测效果非常好。训练完成的模型在真实场景中的检测效果图如图 10 所示。

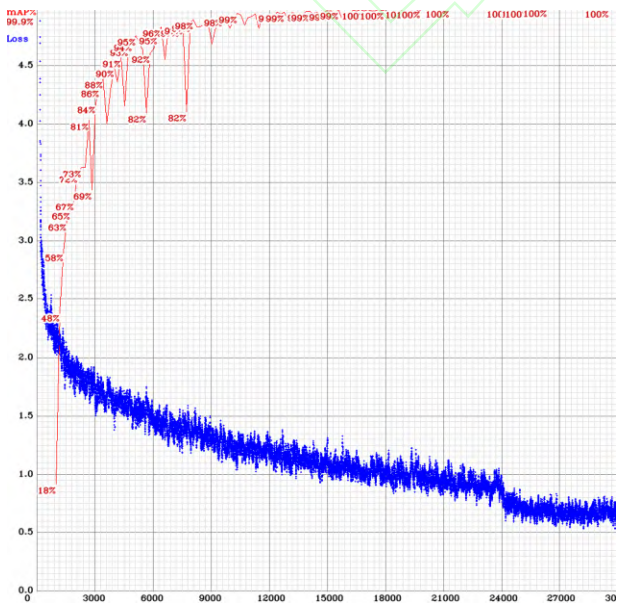


Fig.9　Training log

图 9　训练日志

(a) Detection results (part one)

(a) 检测效果图（一）

(b) Detection results (part two)

(b) 检测效果图（二）

Fig.10　Detection results

图 10　检测效果图

本文制作的测试数据集在 YOLOv3、SSD、RetinaNet 以及本文提出的目标检测算法模型的检测结果如表 2 所示，通过在配有英伟达 GeForce RTX 2080Ti 显卡的主机上进行测试实验，可以看到，本文提出的检测算法在检测准确率(mAP)，以及模型推演速度(FPS)上要高于本文所借鉴的 YOLOv3 算法以及其他代表性的单阶段检测算法。表明了本文的算法在总体吸烟目标目标检测准确率以及检测速度方面有了相应的提升。

Table 2　Detection results on our dataset

表 2　不同算法在本文数据集上的检测结果

	Backbone	mAP(%)	FPS
YOLOv3	Darknet-53	78.7	94
SSD	VGG-16	75.4	62
Retina Net	Resnet-50	84.6	47
Ours	Proposed	86.3	103

之后，使用本文提出的模型在公开数据集 PASCAL VOC 数据集上进行训练测试，训练模型时增

添了一块 TITAN 显卡，使用双显卡进行训练。表 3 则描述了本文算法在 PASCAL VOC 数据集上的训练检测结果，推演过程使用的是 GeForce RTX 2080Ti 显卡，与 YOLOv3、SSD、Faster RCNN 等经典检测算法相比，本文算法仍然具有一定的优势。

Table 3 Detection results on PASCAL VOC

表 3 PASCAL VOC 数据集上的检测结果

	Backbone	mAP(%)	FPS
YOLOv3	Darknet-53	76.8	91
SSD	VGG-16	74.3	67
Faster RCNN	VGG-16	73.2	17
Ours	Proposed	79.3	98

为了验证本算法在小目标检测方面的提升，本文在公开数据集 Tsinghua-Tencent 100K 交通标志牌检测数据集上进行了模型训练与测试，该数据集包括了三类交通标志牌，即禁令类标志、警示类标志以及指示类标识，因此训练数据的标注类别，设置为 3。该数据集的训练集有 6107 张图片，测试集有 3073 张图片，之后选择了 YOLOv3 模型(本文模型的主要参考和对比模型)、未加 AttentionBlock（该部分替换为普通卷积层）的本文模型、未使用多层特征图双向融合的本文模型（该部分替换为普通的单向 FPN 网络结构）以及完整的本文模型进行实验对比，最终的检测结果如表 4 所示：

Table 4 Detection results on TT100k

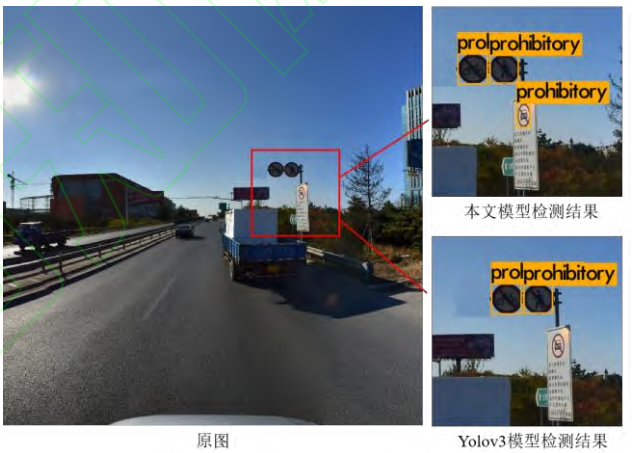
表 4 在 TT100k 数据集上的检测结果

	注意力模块	多尺度双向特征图融合	mAP(%)	FPS
YOLOv3	-	-	87.2	91
Ours	√	×	90.5	104
	×	√	89.7	98
	√	√	94.3	103

检测的推演过程使用的是 GeForce RTX 2080Ti 显卡，经过实验对比，本文算法利用了 Attention Block 模块以及多尺度双向特征图融合预测等结构，加强了对小目标的识别与检测能力。图 11 则是本文算法与 YOLOv3 算法检测交通标志牌的对比实例图。

(a) Detection results (part one)

(a) 检测效果（一）



(b) Detection results (part two)

(b) 检测效果(二)

Fig.11 Comparison of detection results on traffic sign

图 11 交通标志检测效果对比图

6 结束语

本文根据吸烟行为的实际应用场景，提出了一种基于深度学习的能快速发现和警告吸烟行为的检测模型模型，该模型对细粒度的小目标具有较好的检测效果。一方面为了解决模型实时性检测的问题，对提取图像特征的卷积神经网络骨干网络的结构进行了优化，不仅减少了模型参数量与计算量、加快模型推演速度，进而提高检测速度，还能将该结构用于计算资源受限的场景中。另一方面为了改善模型的鲁棒性，将新型激活函数 Mish 引入到本文模型的卷积层中，同时在检测模型中增加正则化 DropBlock 模块，防止模型过拟合；其次，选择 DIoU 边框回归损失函数替换常规的均方根误差损失，提高目标物体定位的准确

率；最后，利用上下文信息，减少目标物体的误检率。通过本文自制的数据集训练本文提出的模型，验证了模型具有较好的检测效果。受限于数据集的多样性，本文的模型在实际生产应用中效果不理想，未来将进一步扩充数据集，完成模型增量训练，进一步减少误检与漏检率，提升模型在实际工业运用中的检测效果。

References:

[1] Senyurek V, Imtiaz M, et al. Cigarette Smoking Detection with An Inertial Sensor and A Smart Lighter[J]. Sensors, 2019, 19(3): 570-588.

[2] Senyurek V Y , Imtiaz M H , Belsare P , et al. Smoking detection based on regularity analysis of hand to mouth gestures[J]. Biomedical Signal Processing and Control, 2019, 51: 106-112.

[3] Wu P, Heieh J W, Cheng J C, et al. Human Smoking Event Detection Using Visual Interaction Clues[C]// Proceedings of the 20th International Conference on Pattern Recognition, Istanbul, August 23-26, 2010. Washington: IEEE Computer Society, 2010, 56: 4344-4347

[4] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]// Proceedings of the 25th International Conference on Neural Information Processing Systems. 2012, 25(2): 1097-1105.

[5] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, June 27-30, 2016. Washington: IEEE Computer Society, 2016: 779-788.

[6] Viola P A , Jones M J. Rapid Object Detection using a Boosted Cascade of Simple Features[C]// Proceedings of the 2001 IEEE Conference on Computer Vision and Pattern Recognition, Kauai, Dec 8-14, 2001. Washington: IEEE Computer Society, 2001, I-I

[7] Dalal N and Triggs B. Histograms of oriented gradients for human detection[C]// Proceedings of the 2005 IEEE Conference on Computer Vision and Pattern Recognition, San Diego, June 20-25, 2005. Washington: IEEE Computer Society, 2005: 886-893.

[8] Felzenszwalb P, McAllester D, Ramanan D. A discriminatively trained, multiscale, deformable part model[C]// Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, June, 23-28, 2008. Washington: IEEE Computer Society, 2008: 1-8.

[9] Simonyan K , Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. arXiv:1409.1556, 2014.

[10] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, June 27-30, 2016. Washington: IEEE Computer Society, 2016: 2818-2826.

[11] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, June 27-30, 2016. Washington: IEEE Computer Society, 2016: 770-778.

[12] Ren S, He K , Girshick R , et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(6): 1137-1149.

[13] Lin T Y, Dollar P, Girshick R B, et al. Feature pyramid networks for object detection[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, July 21-26, 2017. Washington: IEEE Computer Society, 2017,1(2): 4.

[14] Dai J, Li Y, He K, et al. R-FCN: Object detection via region-based fully convolutional networks[C]// Proceedings of the 30th International Conference on Neural Information Processing Systems, Red Hook, May 20, 2016. Curran Associates Inc. 2016, 379-387.

[15] Cai Z , Vasconcelos N. Cascade R-CNN: Delving into High Quality Object Detection[J]. arXiv:1712.00726, 2017.

[16] Neubeck A , Gool L J V. Efficient Non-Maximum Suppression[C]// Proceedings of the 18th International Conference on Pattern Recognition, Hong Kong, August 20-24, 2006. Washington: IEEE Computer Society, 2006, 850-855.

[17] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[C]// LNCS 9905: Proceedings of the 14th European Conference on Computer Vision, Amsterdam, October 8-16, 2016. Heidelberg: Springer, 2016, 21-37.

[18] Lin T Y , Goyal P , Girshick R , et al. Focal Loss for Dense Object Detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 99: 2999-3007.

[19] Tan M X, Pang R M, Le Q V. EfficientDet: Scalable and efficient object detection[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Seattle, 2020. Washington: IEEE Computer Society, 2020: 10778-10787.

[20] Hu J , Shen L , Albanie S , et al. Squeeze-and-Excitation Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.

[21] Woo S, Park J C, Lee J Y, et al. CBAM: Convolutional block attention module[C]// LNCS 11211: Proceedings of the 15th European Conference on Computer Vision, Munich, September 8-14, 2018. Heidelberg: Springer, 2018, 3-19.

[22] Xie S N, Girshick R, Dollar P, et al. Aggregated residual transformations for deep neural networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, July 21-26, 2017. Washington: IEEE Computer Society, 2017, 1492-1500.

[23] He K , Zhang X , Ren S , et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37(9): 1904-1916.

[24] Misra D. Mish: A Self Regularized Non-Monotonic Neural Activation Function[J]. arXiv:1908.08681, 2019.
[25] Zheng Z H, Wang P, Liu W, et al. Distance-IoU Loss: Faster and better learning for bounding box regression.[J] Proceedings of the 34th AAAI Conference on Artifificial Intelligence, 2020, 34(7): 12993-13000
[26] Yun S, Han D, Chun S, et al. CutMix: Regularization strategy to train strong classifiers with localizable features.[C]// Proceedings of the IEEE International Conference on Computer Vision, Seoul, October 27- November 2, 2019. Washington: IEEE Computer Society, 2019, 6023–6032.

CHEN Ruilong was born in 1996. He is a M.S. Candidate at National University of Defense Technology. His research interests include computer vision, etc.
陈睿龙(1996-)，男，新疆奎屯人，国防科技大学硕士研究生，主要研究领域为计算机视觉。

LUO Lei was born in 1984. He received the Ph.D. degree in computer science from National University of Defense Technology in 2013. He is a lecturer at National University of Defense Technology. His research initerests inlcude computer vision and system software.
罗磊(1984-)，男，湖南醴陵人。2013 年于国防科技大学获博士学位，目前为国防科技大学讲师，主要研究领域为计算机视觉、系统软件。

CAI Zhiping was born in 1975. He received the Ph.D degree in computer science from National University of Defense Technology in 2005. He is a professor and Ph.D. supervisor at National University of Defense Technology. His research interests include artificial intelligence, network measurement and edge computing.
蔡志平(1975-)，男，湖南益阳人，2005 年于国防科技大学获得博士学位，目前为国防科技大学教授，主要研究领域为人工智能、网络测量、边缘计算。

MA Wentao was born in 1993. He received the M.S. degree in information and Communication Engineering at Central South University of Forestry and Technology in 2020. He is a Ph.D. Candidate at National University of Defense Technology. His research initerests inlcude computer vision, image mosaic and image retrieval.
马文涛(1993-)，男，安徽临泉人，于 2020 年 6 月获中南林业科技大学工学硕士学位，目前在国防科技大学攻读博士学位，主要研究领域为计算机视觉、图像拼接、图像检索。