

卷积神经网络在实时检测领域的研究

高新怡, 陈琦, 陈冠宇, 杨静怡, 张坤坤, 蔡华蕊

(天津商业大学, 天津 300134)

✉1275475886@qq.com; chq687002@163.com; 2920734342@qq.com;
2385333215@qq.com; 2529207862@qq.com; 1431502854@qq.com



摘要: 提出轻量模型Mini Net用于实时检测, 并保证其准确度。Mini Lower利用Group卷积与通道合并提取低阶特征, Mini Higher利用可分离的Depthwise卷积提取高阶特征。Mini模块实现的高效卷积使其大幅减少了参数量与计算量, 并且在空间维度上引入更多层次所带来的非线性, 可提升模块的提取能力。另外, 在模型中利用更精细的特征搭配多尺度预测改善小目标检测。基于一系列的消融实验验证Mini模块设计的有效性, 并透过对照实验结果证实Mini Net模型的实时性优于全卷积模型, 在参数量仅有 0.92×10^6 的情况下, 能够有效地提取目标特征。

关键词: 卷积神经网络; 轻量模型; 目标检测; 图像识别

中图分类号: TP311 **文献标识码:** A

Research on Convolutional Neural Networks in Real-time Detection

GAO Xinyi, CHEN Qi, CHEN Guanyu, YANG Jingyi, ZHANG Kunkun, CAI Huarui

(Tianjin University of Commerce, Tianjin 300134, China)

✉1275475886@qq.com; chq687002@163.com; 2920734342@qq.com;
2385333215@qq.com; 2529207862@qq.com; 1431502854@qq.com

Abstract: This paper proposes a lightweight model Mini Net for real-time detection and its accuracy is guaranteed. Mini Lower uses Group convolution and channel merging to extract low-order micros, while Mini Higher uses separable Depthwise convolutions to extract high-order micros. The efficient convolution implemented by the Mini module greatly reduces the amount of parameters and computation, and the nonlinearity brought by more layers in the space dimension is introduced, which can improve the extracting ability of the module. In addition, a combination of a finer micro and multi-scale prediction is used in the model to improve small object detection. Based on a series of ablation experiments, the effectiveness of the Mini module design is verified, and the comparative experimental results very that the real-time performance of the Mini Net model is better than that of the full convolution model. When the parameter amount is only 0.92×10^6 , the target micro can be extracted effectively.

Keywords: convolutional neural network; lightweight model; object detection; image recognition

1 引言(Introduction)

从信息化软件到电子商务, 然后到高速发展的互联网时代, 再到今天的云计算、大数据, 电子信息渗透到我们的生活、工作的方方面面。在互联网的驱动下, 人们更清晰地认识并使用数据, 不仅仅是数据统计、分析, 我们还强调数据挖掘、预测。机器学习就是对计算机一部分数据进行学习,

再对另外一些数据进行预测、判断。如今的机器视觉已逐渐成为多数学者的主要研究内容, 并且渗透到我们生活的各个领域, 如图像分类、目标定位、目标检测等。其中, 卷积神经网络(Convolutional Neural Network, CNN)的设计与研究显得尤为重要。

为了提高普通神经网络系统的效率及准确度, 本文提出

轻量模型用于实时检测，并通过一系列消融实验验证模型设计的有效性。

2 研究背景(Research background)

2.1 研究现状

经典的Le Net诞生于1998年。随后CNN的锋芒开始被SVM等手工设计的特征盖过。随着ReLU和Dropout的提出，以及GPU和大数据带来的历史机遇，CNN在2012年迎来了历史突破——Alex Net^[1]。

此后，Deep Learning不断发展，ImageNet大规模网络视觉识别挑战赛(ILSVRC)每年都会被Deep Learning刷榜。如图1所示，随着该模型被研究得越来越深入，top-5的错误率也越来越低，到2017年，降到了2.25%附近。同样，在ImageNet数据集上，人眼的辨识错误率大概为5.1%，换言之，目前的Deep Learning模型的识别能力已经超过了人眼^[2]。而如图1所示的模型代表，也是Deep Learning视觉发展的里程碑式代表。

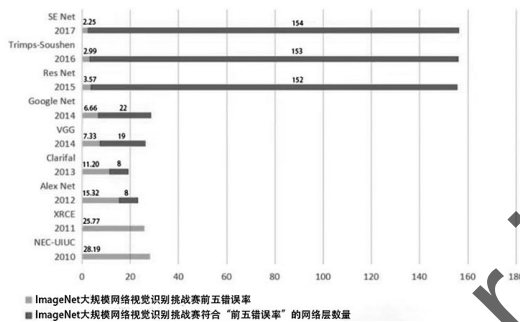


图1 卷积神经网络的发展

Fig.1 Development of convolutional neural networks

CNN主要的经典结构包括Le Net、Alex Net、ZF Net、VGG、NIN、Google Net^[3]、Res Net、SE Net等，它们是最古老的CNN模型。1985年，Rumelhart和Hinton等人提出了BP神经网络算法，使得神经网络的训练变得简单可行。目前，Deep Learning虽然还是比Cortes和Vapnic的Support-Vector Networks稍落后一点，不过其发展前景非常可观。

2.2 研究目的与意义

如今很多装置、设备都注重系统能否实时响应，相应地，系统响应及时意味着系统的有效性好。当前，许多研究好模型的重心都在建立好模型的训练上，从宏观角度看，显然花大部分时间在训练上，系统的效率就成了问题；从微观角度看，卷积本身的冗余性有待提高。本文着重研究轻量模型，从本质上分析卷积参数的有效性，进一步提高系统的效率。

3 卷积神经网络(Convolutional neural network)

3.1 神经网络

起源于1943年的M-P神经网络是广泛应用于机器学习的

人工神经网络，是按照生物神经元的结构和工作原理构造出来的一个抽象和简化的模型。其每个神经元都是一个多输入单输出的信息处理单元，且神经元输入与输出之间存在由于突触延迟所导致的固定的时滞。而现存的神经网络是由大量的神经元相互连接构成的一种具有学习能力的自适应系统。

3.1.1 感知器

在人工神经网络中，神经元具备激励特性和感知特性。与神经系统相似，基于Frank Rosenblatt提出的感知器由式(1)所确定， n 维输入 X_n 与权重 K_n 进行相乘、求和运算，再加一个可调偏置 b ，通过激励函数 $f(x)$ 映射后，得到输出 Y 。

$$Y = f\left(\sum_{i=1}^n K_i X_i + b\right) \quad (1)$$

单层感知器可视为最简单的前向神经网络，由输入层、输出层和一组可训练的权重参数组成，如图2所示。多层感知器由输入层、输出层和隐含层所构成，具有非线性特性，有效地解决了单层感知器难以处理线性不可分的问题，其结构如图3所示。

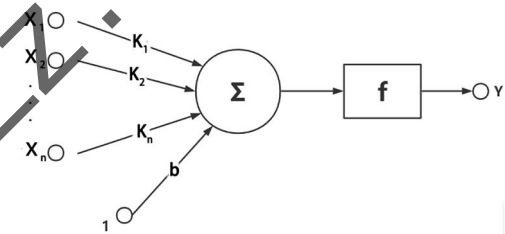


图2 单层感知器

Fig.2 Single layer perceptron

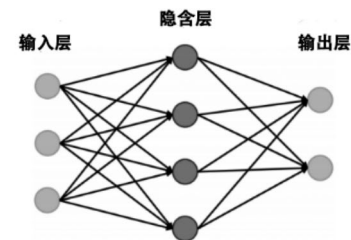


图3 多层感知器

Fig.3 Multi-layer perceptron

3.1.2 BP神经网络

BP算法是一种监督学习算法^[4]，通常情况下被用来训练多层感知机，将数据输入多层感知网络中，通过前向传播到隐含层，直至输出层后，找到实际值与理论值之间的偏差函数，根据梯度下降法进行反向传播；再由更新的权重来最小化偏差函数得到偏差的极小值，使得模型的数据尽可能地拟合真实值。算法实际的误差值取决于训练时的权重参数，训练开始前会随机分配初始权重，通过多次有效的反向传播后得到一组最小化误差的权重值。由于初始权重值是系统随机分配的，也存在一定的误差，需要多次拟合找最优。

每个神经元的输出为式(2), n 个输入 O_i , 与权重进行相乘、求和运算后得到 net_j , 通过激励函数 φ 映射得到输出 O_j 。

$$O_j = \varphi(net_j) = \varphi\left(\sum_{i=1}^n \omega_{ij} O_i\right) \quad (2)$$

权重更新为式(3), 权重调整为式(4)。

$$\omega_{ij} = \omega_{ij} + \Delta\omega_{ij} \quad (3)$$

$$\Delta\omega_{ij} = -lr \times \frac{\partial L}{\partial \omega_{ij}} \quad (4)$$

其中, L 为偏差函数。

由链式法则求得偏导数为式(5), 进而求得神经元。若神经元位于输出层, 此时输出与预测值 p 相等, 可通过直接进行求偏运算得到; 若神经元位于隐含层, 则必须进行递归运算。

$$\frac{\partial L}{\partial \omega_{ij}} = \frac{\partial L}{\partial O_j} \times \frac{\partial O_j}{\partial net_j} \times \frac{\partial net_j}{\partial \omega_{ij}} = \delta_j \times O_i \quad (5)$$

其中,

$$\delta_j = \frac{\partial L}{\partial O_j} \times \frac{\partial O_j}{\partial net_j} = \begin{cases} \left(\frac{\partial L}{\partial O_j}\right) \times \frac{\partial}{\partial net_j} \varphi(net_j), & j = \text{output neuron} \\ \left(\sum_l \delta_l \omega_{jl}\right) \times \frac{\partial}{\partial net_j} \varphi(net_j), & j = \text{hidden neuron} \end{cases}$$

3.2 卷积

随着神经网络的不断发展, 在自适应学习系统的基础上, 使用梯度下降法实现的多层次神经网络能够有效地解决系统处理非线性问题。神经元早期所采用的是全连接方式对数据进行拟合, 在处理高像素图像时, 模型容易出现过拟合的状况。

3.2.1 卷积神经网络

卷积神经网络利用局部感知视野、权值共享与空间或者时间的下采样实现平移、缩放和形变的不变性^[5], 进一步改善全连接网络在图像识别领域的缺陷。

为了进行高层次的特徵提取, 利用输入图像拓扑结构的方式, 使得卷积核提取到局部特徵, 再通过逐步滤波的结合, 得到高层次特徵。特徵图中的神经元是由上一层的一组局部神经元与单一卷积核进行卷积所得到的。

将单个卷积核设置为一组权重和一个可选择的偏置, 卷积核可以在不同的区域内检测到相同的特徵, 再进行相乘、求和运算后, 得到一个平面特徵图。而特徵图的所有神经元共享权重, 进而降低特徵图的复杂度。

计算出的特徵图通过卷积层与下采样层进行交替搭建, 进而降低特徵图的空间分辨率, 最后在网络末端结合全连接层与分频器输出预测结果。

3.2.2 串联式与并联式

(1) 串联式

由于传统的卷积神经网络在拟合复杂度较高的非线性数据时会加大下一层的计算负担, 故提出串联式系统简化计算, 即串联多的 3×3 卷积与最大池化层组合来推展模型层数, 进而大幅度提升模型的识别能力。

(2) 并联式

如果数据集的分布可由相对稀疏的网络进行拟合, 可分析某些激活值的相关性, 将相关性高的神经元聚合连接在一起, 从而减轻过拟合和降低卷积参数的计算量。

在高层特徵空间中, 彼此之间的距离相对较远, 所以使用到大尺寸的卷积核数量相对较多, 难以避免计算量的增多。因此, 在 3×3 、 5×5 卷积前与 3×3 池化, 在其后加入能进行通道交互且减少数据计算量的 1×1 卷积。此模型在实现提取高层特徵的同时, 还能够控制其空间、时间复杂度处于合理的范围并具有一定的准确性。模型如图4所示。

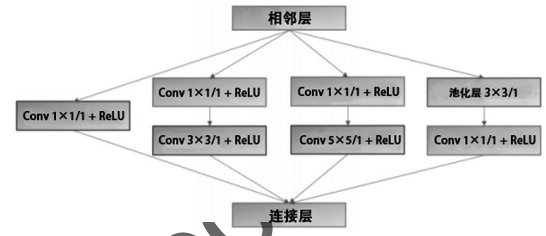


图4 并联式卷积模型

Fig.4 Parallel convolution model

3.2.3 残差

残差模块如图5所示, 通过捷径方式连接到原始输入层 x , 得到残差函数表达式 $F(x) = H(x) - x$, 再通过元素层级的加法得到 $H(x)$, 并经过激励函数映射得到相应的输出值。引入残差函数映射能够在分支突出微小变化, 使得权重对分支变化更加敏感, 从而降低模型的训练难度。

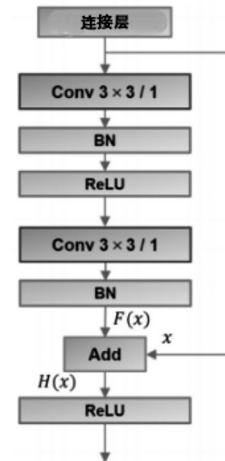


图5 残差模块

Fig.5 Residual module

3.3 轻量卷积

为了解决系统的效率问题, 引入轻量卷积。轻量模型主要是模型的卷积层进行组合与设计。通常情况下, 在卷积层引入与传统卷积不同的Group卷积和Depthwise卷积。

3.3.1 Group卷积

传统的卷积是对上一层所有特徵通道进行卷积处理, 如图6所示。而Group卷积是先将特徵通道进行分组, 使其在不同的GPU上进行运算, 不同的卷积核对各自上一层分组后的

通道进行处理,如图7所示。

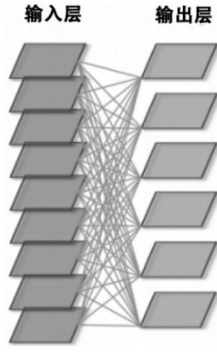


图6 传统卷积

Fig.6 Traditional convolution

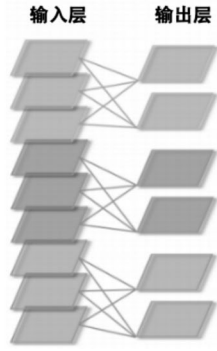


图7 Group卷积

Fig.7 Group convolution

此处以上一层特微通道数为 C_{in} ,经过本层卷积核操作后,输出通道为 C_{out} 为例。使用宽为 K_w 、高为 K_h 的一般卷积核,那么,它的单一卷积核尺寸为: $K_w \times K_h \times C_{in}$;忽略偏置参数,该层参数量为: $K_w \times K_h \times C_{in} \times C_{out}$;若使用等高等宽的Group卷积核,将上一层通道分为 G 组,参数量为: $(K_w \times K_h \times \frac{C_{in}}{G} \times \frac{C_{out}}{G}) \times G = K_w \times K_h \times \frac{C_{in}}{G} \times C_{out}$,相比于一般的卷积核,Group卷积的参数量为原来的 $\frac{1}{G}$ 倍。

3.3.2 Depthwise 卷积

Depthwise卷积模型如图8所示,若上一层特微通道数为 C_{in} ,且用等高等宽的Depthwise卷积核,那么其单一卷积核尺寸仅为 $K_w \times K_h \times 1$,该层参数量也仅为 $K_w \times K_h \times 1 \times C_{in}$,参数量大幅度降低。

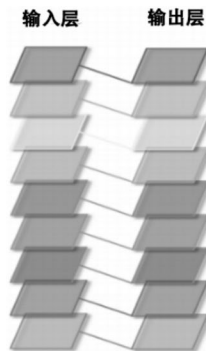


图8 Depthwise卷积

Fig.8 Depthwise convolution

3.4 目标检测

3.4.1 Two Stages算法

Two Stages的主要算法为R-CNN,将检测问题转化为分类问题,使用选择性阶层分组方式对候选区域进行提取,通过图像分割算法得到多区域,根据相似度逐层合并得到多个候选框,再对每个候选框缩放到固定尺寸,输入卷积神经网络进行特微提取,再送入SVM进行分类,得到准确位置。

FAST-R-CNN^[6]改进了R-CNN的缺点,将原始图像一次性输入卷积神经网络,并将最后获得的特微送入池化层提取到相应的特微区域,并将候选框实现最大池化,输出固定尺寸的特微图,解决了全连接层需要固定输入,缩放特微区域导致失真的问题。

使用选择性搜索提取候选区域会占用很多检测时间,将候选区域提取到卷积神经网络中,引入区域生成网络RPN,将卷积层输出的特微图进行类别和背景判断,根据所获得的候选框对应之前卷积网络输出的特微图,将其输入池化层中,再分别送入softmax分类器和校正边界回归器中,获得最终的预测结果。

3.4.2 One Stage算法

目标检测^[7]中的One Stage算法是直接回归物体的类别概率和位置坐标值,比阶层分组提取特微方法预先提取候选框的控制更加快速,可以实现即时检测。One Stage算法中有代表性的是YOLO系列算法,整体为单一管道,直接从单张图像回归出边界框的类别和准确位置。其优点是检测速度快,便于训练,且准确度高于R-CNN系列。

YOLOv3将原始图片缩放为 $H \times W$,并将其输入单一网络中,经过卷积层与池化层的处理,将特微图分成 $S \times S$ 的单元格,且每个单元格预测 B 个边界框,每个边界框预测 $(5+C)$ 个值,包含框的相对中心坐标 (d_{cx}, d_{cy}) 、相对偏移宽高 (d_w, d_h) 、置信度分数为 c 以及 C 个类别条件概率。 d 表示相对单元格偏移。

在测试时,如式(6)所示,将预测框的类别条件概率 C 与置信度分数 c 相乘后,得到特定类别置信度分数 box_score ,并根据所有预测框的特定类别置信度分数进行过滤和非极大值抑制消除,从而得到最终预测结果。

$$Pr(class_i | object) \times Pr(object) \times (IOU_{pred}^{truth}) = Pr(class_i) \times (IOU_{pred}^{truth}) \quad (6)$$

式(6)中,含有目标也就是其中心点落入该单元格, $Pr(object)=1$;不含目标, $Pr(object)=0$ 。 IOU_{pred}^{truth} 表示真实框与预测框的面积比。

为了避免训练不稳定引起所需收敛时间过长的问题,YOLOv2和YOLOv3将中心坐标采用对应于单元格左上角的

相对位置, 求出Sigmoid函数 φ , 使得中心坐标能够落在单元格内。如图9所示, 特微图的宽为 B_w 、高为 B_h , 边界框的宽为 A_w 、高为 A_h , 边界框中心坐标为 (d_{cx}, d_{cy}) , 偏移宽高为 (d_w, d_h) , 单元格距离左上角特微图的距离为 (o_x, o_y) , 可以通过式(7)和式(8)定下 (B_x, B_y) , 进而得到预测框的中心坐标在特微上的准确位置。

$$B_x = \varphi(d_{cx}) + o_x \quad (7)$$

$$B_y = \varphi(d_{cy}) + o_y \quad (8)$$

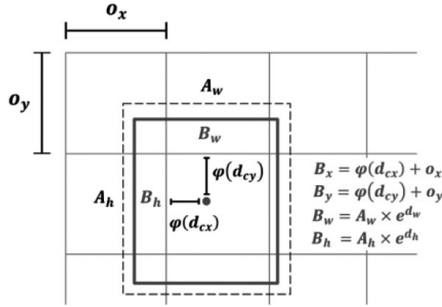


图9 边界框在特微图中的准确位置

Fig.9 Exact position of bounding box in micrograph

4 Mini Net 模型(Mini net model)

本研究基于Mini卷积模块的设计, 在高层与低层分别采用不同性质的卷积模块进行特微的提取, 通过减少模型的参数数量和计算量实现即时检测, 并保证一定的准确性。

4.1 检测系统

对Mini Net模型的检测流程图如图10所示, 将原始图片经缩放转换为固定尺寸输入系统中, 经过Mini Net模型处理后, 直接输出目标的确切位置, 采用单一管道的系统模式提升检测效率。此处输入的图片均为彩色RGB数据, 并不对原始图像进行灰度处理降维。

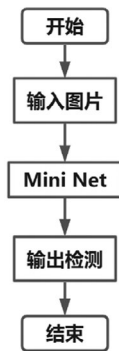


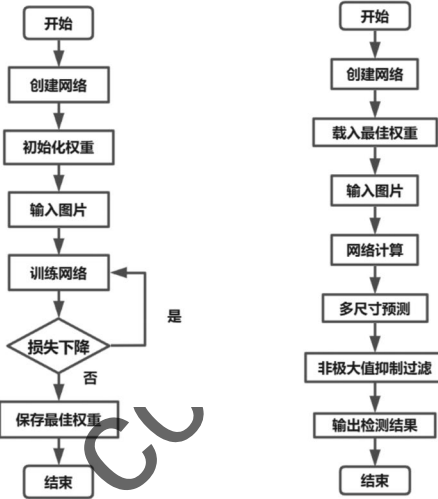
图10 检测流程图

Fig.10 Detection flow chart

4.1.1 系统流程

Mini Net检测系统分为训练阶段和测试阶段, 训练阶段如图11(a)所示, 建立Mini Net网络并对权重进行初始化, 将训练集缩放完的图片输入网络进行前向传播, 计算出偏差

函数, 再通过梯度下降法进行反向传播调整权重值, 经过多次训练, 最终得到特微图的权重参数。测试阶段如图11(b)所示, 建立Mini Net网络, 将最佳权重输入网络中, 将缩放后的图像输入网络进行前向传播进而实现多尺度预测, 再利用NMS过滤得到最后的检测结果。



(a)训练阶段

(b)测试阶段

图11 Mini Net检测系统

Fig.11 Mini Net detection system

4.1.2 输入前处理

首先, 计算出模型输入尺寸与原始尺寸的宽高比, 并取最小值作为缩放比例, 为了避免缩放后的边界大于原图边界裁剪到原始图像, 再将原始图像乘以缩放比例得到新的尺寸。由于输入图像的宽和高皆为416, 但是原始图像的宽高比不一定是1:1, 故采用等比例缩放图像, 放置到宽高皆为416且RGB规定为(128,128,128)的灰度底片上。为了在训练过程中不出现重叠效果, 选择中间值128作为底色进行训练。

4.1.3 模型架构

Mini Net模型是在高低阶段分别采用不同的卷积模块组合成的, 在模型低层阶段采用Mini Lower模块来提取低阶特微, 在高层阶段采用Mini Higher模块来提取高阶特微, 最后搭配两个尺寸输出预测, 整体框架如图12所示。

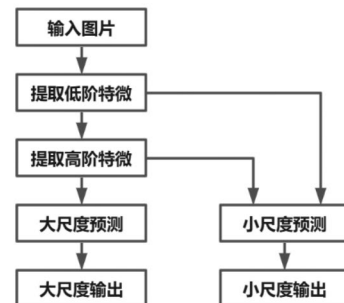


图12 Mini Net架构

Fig.12 Mini Net architecture

4.2 运作模式

Mini Net模型的运作模式是基于One Stage算法的YOLOv3, 将整个任务视为回归问题, 不需要预先提取候选框, 而是直接将图片输入神经网络中进行处理, 将特微图分割成单元格的形式, 在特微空间上进行全局的目标检测。考虑到轻量模型的计算量问题, 此处将采用YOLOv3-tiny的两个尺度进行输出预测, 并且借鉴YOLOv3先验框的偏置回归控制, 将其分割在两个尺度上进行训练和测试。

4.2.1 训练阶段

Mini Net模型采用监督学习进行模式训练, 包含两个输入 I_1 和 I_2 , 计算偏差函数的输出端 O , 第一个输入 I_1 为图片中目标对应的真实框 G 的标记值, I_2 为缩放后图片的RGB值。

(1) 输入端 I_1

训练集的每一个目标对应的真实框都包含五个标记值, 分别为: 边界左上角坐标 $(rx_{\min}, ry_{\min})$ 、右下角坐标 $(rx_{\max}, ry_{\max})$ 、所属类别 C_{id} 。先将真实框标记值进行转换:
 $G_{cx} = \frac{rx_{\min} + rx_{\max}}{2}$, $G_{cy} = \frac{ry_{\min} + ry_{\max}}{2}$, $G_w = rx_{\max} - rx_{\min}$, $G_h = ry_{\max} - ry_{\min}$, 再将真实框转化后的标记值进行归一化处理, 如式(9)所示:

$$(G_{ncx}, G_{ncy}, G_{n\omega}, G_{nh}) = \left(\frac{G_{cx}}{\omega}, \frac{G_{cy}}{h}, \frac{G_w}{\omega}, \frac{G_h}{h} \right) \quad (9)$$

格式转换完成后, 将每个真实框 G 对应到先验框 A 上, 从而决定先验框的训练顺序。

(2) 输入端 I_2

缩放后的原始图像在模型训练前, 先将RGB归一化为 $[0, 1]$, 有助于训练的稳定, 再将归一化的数据输入Mini Net网络中, 经前向传播后输出 y_{pred} , 最终两个卷积层分别输出不同尺寸的特微图。小尺度输出的张量形状为: $[bs, 13, 13, 18]$, 大尺度输出的张量形状为: $[bs, 26, 26, 18]$, 18表示输出三个预测框, 且每个预测框预测六个值。

(3) 输出端 O

整个系统的输出端为损失层, 损失层接收两个输入, 分别为实际标记值 y_{true} 和模型预测值 y_{pred} , 再通过偏差函数计算出两者的误差值, 经过多次训练, 寻找最合适的权重参数。在损失函数计算前, 需进行维度转换, 如表1所示。

表1 维度转换

Tab.1 Dimension transformation

预测值与标记值	小尺度张量形状	大尺度张量形状	张量的项目值
y_{pred}	转换前 $[bs, 13, 13, 18]$	$[bs, 26, 26, 18]$	—
	转换后 $[bs, 13, 13, 3, 6]$	$[bs, 26, 26, 3, 6]$	$[d_{cx}, d_{cy}, d_w, d_h, c_{po}, c_{po}]$
y_{true}	$[bs, 13, 13, 3, 6]$	$[bs, 26, 26, 3, 6]$	$[G_{ncx}, G_{ncy}, G_{nw}, G_{nh}, c_i, C_i]$

此处采用交叉熵作为偏差函数, 用概率形式表示交叉熵的值, 因为损失层的真实值和预测值皆为六个项目, 故采用

多任务训练模式。

4.2.2 测试阶段

Mini Net测试阶段由训练阶段取得的权重进行运算, 先将原始图片缩放到模型所需要的尺寸, 将RGB进行归一化处理并进行前向传播, 输出不同尺寸的预测结果, 将置信度分数 c_p 和条件概率 C_p 相乘, 得到特定置信度分数 box_{score} , 将置信度分数 $box_{\text{score}} < 0.2$ 的预测框过滤掉, 再对过滤后的图片进行NMS处理消除重叠框, 直至处理完所有预测框, 得到预测结果。

4.3 Mini 模块设计

此处根据不同阶段提取的特微特性设计不同形式的轻量模型, 分别为基于低层次提取的基本特微Mini Lower和基于高层次提取的高阶特微Mini Higher, 再将二者进行组合得到轻量模型Mini Net。

4.3.1 Mini Lower模块

Mini Lower模块主要利用Group卷积, 其不同组的卷积核分别作用于分组后的特微图上。首先, 对输入模块的特微采用 1×1 卷积进行信息融合, 并将卷积数量定为输入特微通道数量的一半, 从而可以实现特微交互, 并能降低参数计算量; 接着对处理完的特微进行Group卷积运算, 考虑到过多分组会导致特微破碎化, 所以只分为两组, 并进行 3×3 卷积操作, 卷积核的数量取决于合并后的输出通道数。此处采用通道层级的合并策略, 一方面可以大幅减少参数量, 另一方面也减少了卷积带来的不必要的参数计算量。

4.3.2 Mini Higher模块

Mini Higher模块主要利用Depthwise卷积, 每个卷积核各自操作所对应的单一特微通道, 并分别使用单一卷积核进行运算处理。使用Depthwise卷积能够大大减少计算量。基于Mini Lower在模块前端引入 1×1 卷积搭配池化层与激励函数所带来的效率, 故此处也使用 1×1 的卷积组合, 并将卷积核数量设为 $0.5C_m$ 来降低参数的计算量, 接着进行Depthwise卷积, 最后 1×1 卷积相当于是对Depthwise卷积输出的特微进行融合, 以此拟合目标特微所需的位置。

4.4 检测模型

4.4.1 YOLOv3-tiny模型

图片输入模型经 3×3 卷积核处理后, 利用 2×2 最大池化降低特微图尺寸搭配 3×3 卷积增加通道数, 经过五次基本特微后, 特微通道数增加至512; 接着采用 3×3 卷积搭配 1×1 卷积提取高阶特微, 此阶段 3×3 卷积核的数目较多, 用以提取更多特性的高阶特微, 其后利用 1×1 卷积来降低通道数量; 输出端采用 3×3 卷积提取表达性特微, 搭配 1×1 卷积输出结果。

4.4.2 Mini Net模型

图片输入模型后, 先对输入图片的上边界和左边界填0处

理,使得特微图的宽和高均降至原来的一半;接着使用五个Mini Lower模块提取基本特微,并加倍特微通道,穿插四个池化层来降低特微图的尺寸,该阶段操作完之后,特微图的尺寸降为 13×13 ,通道数增加至512;对于高阶特微采用Mini Higher进行提取;最终两个尺度的输出端采用 1×1 卷积进行预测。

5 实验结果(Experimental result)

5.1 开发环境介绍

此研究在中央处理器上执行所有操作,没有使用具有大量平行运算能力的圆形处理器,用Python完成即时检测系统的设计,神经网络的搭建基于TensorFlow和聚类算法Keras,数据集均采用WIDER FACE对模型进行监督学习评估,没有使用另外的数据集。

5.2 数据的预处理

5.2.1 WIDER FACE

数据集为WIDER的子集,所有图片通过Google和Bing等搜索得到^[8],进行类别处理,删除相似度较高的图片保证样本的丰富性。如图13所示为携带多样属性的大规模数据,可充分保证正负样本,不需要额外的数据集。



图13 WIDER FACE数据集

Fig.13 WIDER FACE data set

5.2.2 数据集预处理

训练集本身的特性将影响模型的泛化能力,需要对数据集进行过滤和筛选,保证数据集内的数据真实有效。

5.2.3 过滤和筛选

首先,针对有问题的数据进行筛选,删除十个标记值为0的不符图片,并剔除宽 $w \leq 0$ 或高 $h \leq 0$ 的边界框。如表2所示为对数据集与验证集进行初步处理。

表2 数据集与验证集的初步处理

Tab.2 Preliminary processing of data set and verification set

WIDER FACE		原始数量	删除数量	有效数量
训练集	图片	12,880	4	12,876
	边界框	159,420	27	159,393
验证集	图片	3,226	14	3,212
	边界框	39,422	7	39,415

5.2.4 统计与聚类

先统计训练集的边界框,并聚类出实验所需要的六个先验框,通过使用K-means的欧式距离函数计算出所有数据点与各个群集中心的距离。为了减小统计和聚类引起的误差,对初始值的选取是从所有数据点中挑选六个点,而非任意的任意值。

5.3 训练方法

基于Mini卷积模块的有效设计,使得模型在整体训练的过程中更加稳定,因此在检测数据集上训练检测模型。在数据集上采用多阶段训练策略,在特定阶段搭配超参数调整,从而提高模型的训练效率和检测的准确性。

5.3.1 超参数与优化器

此研究对于周期的确定不采用预先固定,而是在遍历整个训练集的每个周期之后,利用周期更新权重值,在验证集上计算平均误差,从而判断训练效果。常见的批次数量 bs 的设定方式有BGD、SGD和MBGD。BGD方式是将所有样本输入网络中,将样本全部遍历一遍得到更新后的权重,此方法的计算量过于庞大,收敛速度非常慢;SGD方式的每一次训练仅选取一个样本输入网络中,避免了大量的计算,模型通常遍历少数样本就可以收敛;MBGD方式每次训练选取一个批次的 m 个样本输入网络,把整体数据分为若干批次,再决定该批次权重更新的梯度方向,这样的数据既稳定,又不会造成计算量大的问题。权重更新的有效性将影响模型的数据拟合能力,非线性高的权重将会调整更多来拟合每一个数据点,但是这样容易出现过拟合的情况,如图14所示。

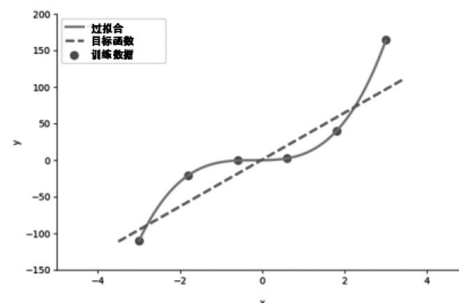


图14 过拟合现象

Fig.14 Over fitting phenomenon

优化器采用MBGD搭配L2进行权重更新，再搭配具有Momentum动量特性与RMS自适应监督学习的Adam作为优化器。

5.3.2 数据增强

此处对于不同训练阶段采用不同的处理策略。数据增强相当于增加更多样化的数据量，因此可以避免过拟合，提高检测准确度。多尺度缩放分为三步：第一步，在有效的范围内随机生成新的宽高比，这样不会导致检测图像与原始图像的比例相差太大；第二步，获取随机缩放值 s ，以避免缩放比例过大或过小导致输入尺寸的边界裁剪掉过多的原始数据；第三步，设置RGB=(128,128,128)的信箱模式，找到底片最合适的位置。

5.4 实验结果分析

首先，为了说明Mini Lower的有效性，将Mini Lower与其他模块进行对照实验，并探讨模型在合并更加精细的特徵后，是否能够改善整体的检测效果，最后对Mini Net与YOLOv3-tiny进行评估比较。此策略将带来更好的检测性能。

采用Mini Lower模块设计的Mini Net能够更加准确及时地进行检测。以下按照表3来设计对照实验，Mini Net-A组无残差分支，故卷积核的数量增加；Mini Net-B组将Mini Lower模块后的通道合并改为残差模块中的元素相加形式；Mini Net-C组保证合并后的通道数为384，而非Mini Net中的416通道。实验结果如表4所示。

表3 多阶段训练

Tab.3 Multistage training

训练阶段	训练集	批次数量	初始学习率/%	数据增强
第一阶段	第一子集	32	0.1	—
第二阶段	第二子集	16	0.01	—
第三阶段	第三子集	16	0.01	✓

表4 模型比较

Tab.4 Model comparison

模型	参数量	验证误差/%
Mini Net	923,420	26.92
Mini Net-A	1,021,876	27.65
Mini Net-B	2,334,292	26.79
Mini Net-C	914,876	28.75

6 结论(Conclusion)

对于轻量模型Mini Net，在参数量仅有 0.92×10^6 的情况下，能够有效地提取目标特徵。由于卷积本身有冗余问题，相较于全卷积使用大量的参数学习特徵，根据低阶特徵和高阶特徵分别设计的Mini Lower和Mini Higher能够更准确地

学习特徵。在卷积模块中增加任何操作都会增加模型的计算量，进而影响检测速度。轻量模型将批标准化层和激励函数都设计在模块前端，在数据堆叠的过程中，特徵的交互作用使得数据在各个部分间共享，从而减少了不必要的数据计算。

本文提出多阶段的训练策略，第一阶段使得收敛快速而稳定，第二阶段和第三阶段改善系统出现过拟合的问题，利用更精细的特徵来改善小目标的检测。

参考文献(References)

[1] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 6(60):84-90.

[2] RUSSAKOVSKY O, DENG J, SU H, et al. TimageNet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(03):221-252.

[3] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]// CVPR. 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE, 2015:1-9.

[4] 刘品.BP神经网络结构优化研究及应用[D].北京:中国地质大学,2016.

[5] 陈灏然.基于卷积神经网络的小目标检测算法研究[D].无锡:江南大学,2021.

[6] GIRSHICK R. Fast R-CNN[C]// ICCV. Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015:1440-1448.

[7] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]// CVPR. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu HI, USA: IEEE, 2017:936-944.

[8] YANG S, LUO P, LOY C C, et al. WIDER FACE: A face detection benchmark[C]// CVPR. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. LasVegas, USA: IEEE, 2016:5525-5533.

作者简介:

高新怡(2001-), 女, 本科生.研究领域: 自动化.

陈琦(1970-), 女, 博士, 副教授.研究领域: 控制理论与应用.

陈冠宇(2001-), 男, 本科生.研究领域: 计算机科学与技术.

杨静怡(2001-), 女, 本科生.研究领域: 自动化.

张坤坤(2001-), 女, 本科生.研究领域: 自动化.

蔡华蕊(2000-), 女, 本科生.研究领域: 通信工程.