基于 YOLOv3-tiny 的视频监控目标检测算法

王均成[1,2,3],贺　超[1,2,3],赵志源[1,2,3],邹建纹[1,2,3]
(1.重庆邮电大学 通信与信息工程学院,重庆 400065;
2.先进网络与智能互联技术重庆市高校重点实验室,重庆 400065;3.泛在感知与互联重庆市重点实验室,重庆 400065)

摘　要:目标检测算法在视频监控领域有着较大的实用价值。针对当前在资源受限的视频监控系统中实现实时目标检测较为困难的情况,提出了一种基于 YOLOv3-tiny 改进的目标检测算法。该算法在 YOLOv3-tiny 架构的基础之上,通过添加特征重用来优化骨干网络结构,并提出全连接注意力混合模块来学习到更丰富的空间信息,更适合资源约束条件下的目标检测。实验数据表明,该算法相比于 YOLOv3-tiny 在模型体积降低 39.2%,参数量降低 39.8%,且在 VOC 数据集上提高了 2.7%的 mAP,在提高检测精度的同时显著降低了模型资源占用。

关键词:目标检测;视频监控;YOLOv3;特征重用;注意力机制
中图分类号:TP391.4　　**文献标识码**:A　　**DOI**:10.16157/j.issn.0258-7998.212121

中文引用格式:王均成,贺超,赵志源,等. 基于 YOLOv3-tiny 的视频监控目标检测算法[J].电子技术应用,2022,48(7):30-33,39.
英文引用格式:Wang Juncheng,He Chao,Zhao Zhiyuan,et al. Video surveillance object detection method based on YOLOv3-tiny[J]. Application of Electronic Technique,2022,48(7):30-33,39.

Video surveillance object detection method based on YOLOv3-tiny

Wang Juncheng[1,2,3],He Chao[1,2,3],Zhao Zhiyuan[1,2,3],Zou Jianwen[1,2,3]
(1.School of Communication and Information Engineering,Chongqing University of Posts and Telecommunications,Chongqing 400065,China;
2.Advanced Network and Intelligent Connection Technology Key Laboratory of Chongqing Education Commission of China,Chongqing 400065,China;
3.Chongqing Key Laboratory of Ubiquitous Sensing and Networking,Chongqing 400065,China)

Abstract:Object detection methods have great value in the application field of video surveillance. At present, it is difficult to realize real-time object detection in resource constrained video surveillance system. A object detection method based on improved YOLOv3-tiny is proposed. Based on the YOLOv3-tiny architecture, the algorithm optimizes the backbone network by adding feature reuse, and a fully-connected attention mix module is proposed to enable the network to learn more abundant spatial information, which is more suitable for object detection under resource constraints. The experimental data shows that compared with YOLOv3-tiny, the algorithm reduces the model volume by 39.2%, the amount of parameters by 39.8%, and improves the mAP of 2.7% on the VOC data set, which significantly reduces the occupation of model resources while improving the detection accuracy.
Key words:object detection;video surveillance;YOLOv3;feature reuse;attention mechanism

0 引言

近年来,目标检测算法已经广泛应用于各个视频监控场景,包括车辆检测[1]、行人检测[2]、农业检测[3]、人类异常行为检测[4]等,越来越复杂的目标检测网络展示了最先进的目标检测性能。但在实际应用中,往往需要在视频监控中一些计算能力及内存有限的设备上进行实时目标检测。例如,嵌入式平台视频监控,其可用计算资源一般仅限于低功耗嵌入式图形处理单元(Graphic Processing Unit,GPU)。这极大地限制了此类网络在相关领域的广泛应用,使得在资源受限设备上实现实时目标检测非常具有挑战。

为了实现资源有限设备上目标检测这一挑战,人们对研究和设计低复杂度的高效神经网络体系架构越来越感兴趣。而著名的 YOLO[5](You Only Look Once,YOLO)则是围绕效率设计的一阶段目标检测算法,它可以在高端图形处理器上实现视频监控目标高效检测。然而对于许多资源受限监控设备来说,这些网络架构参数量大且计算复杂度较高,使得在嵌入式等监控设备上运行时推理速度大幅下降。YOLOv3[6]是 YOLO 系列应用在各领域最普遍的算法,YOLOv3-tiny 则是在该算法的基础上简

化的，虽然精度显著下降但具有了更少计算成本，这大大增加了在资源受限监控设备上部署目标检测算法的可行性。

本文提出了一种基于 YOLOv3-tiny 的目标检测算法 YOLOv3-SF，将改进的 ShuffleNetV2[7]网络与 YOLOv3-tiny 架构进行结合，并加入设计的注意力机制模块使神经网络能学习到更丰富的空间位置信息。实验表明，该方法能够有效优化模型的资源占用与检测精度，更适合资源有限的视频监控设备部署。

1 网络架构

YOLOv3 架构是目标检测领域中最优的算法之一。该架构通过调整骨干网结构、多尺度预测等改进方法在保持原系列优势的同时，提高了识别准确率。但是，该架构检测精度提升的同时也导致了相应计算复杂度的提高，而 YOLOv3-tiny 则是其较为轻量级的精简版本。显然，对于时效性及存储要求高的监控设备，轻量级架构才是首要的目标检测算法选择。

1.1 YOLOv3-tiny 网络架构

本文所使用的 YOLOv3-tiny 网络架构如图 1 所示。骨干网络作为目标检测任务的特征提取器，以图像作为输入，输出对应输入图像的多个不同尺度特征映射，随后通过对不同特征进行处理来获取多个预测结果，使模型更好地检测不同大小的目标物体，以此来取得更高的识别准确率。

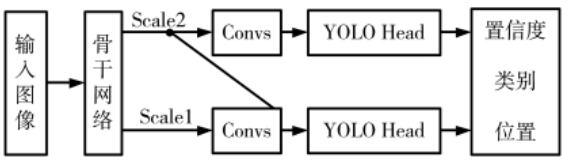


图 1 YOLOv3-tiny 网络架构

YOLOv3-tiny 网络架构参考特征金字塔网络的思想，将低级特征与高级特征拼接融合，并从不同尺度提取特征。与 YOLOv3 不同，该网络架构只保留了两种不同尺度的网络输出，目的同样是预测不同尺度的对象。预测分支不仅在骨干网络末端输出特征上独立预测，还通过将该特征图上采样到与前期特征图相同大小，然后与大特征图通道堆叠做进一步预测。每个预测框的坐标为 t_x、t_y、t_h、t_w，若预测框中心点相对于特征网格左上角坐标相对偏移量(c_x, c_y)，先验框长宽为 p_w 及 p_h，则预测框的位置有以下表示：

$$b_x=\sigma(t_x)+c_x \tag{1}$$

$$b_y=\sigma(t_y)+c_y \tag{2}$$

$$b_w=p_w e^{t_w} \tag{3}$$

$$b_h=p_h e^{t_h} \tag{4}$$

1.2 全连接注意力混合模块

对于深度卷积神经网络，高级特征包含低空间分辨率的分类信息，低级特征包含高空间分辨率的位置信息。而为了获得更准确的位置信息，许多工作将低级特征与高级特征相结合。然而这些工作一般都是直接将其相加或者拼接在一起，而不考虑通道之间的差异。受 SENet[8]的启发，本文认为卷积特征通道之间的相互依赖性很重要。因此，针对图 1 中特征融合预测分支部分，本文参考 SENet 中利用注意力机制来学习通道间的关联性，提出了将低级特征和高级特征结合的全连接注意力混合模块(Fully-connected Attention Mix Module，FCAMM)，如图 2 所示。

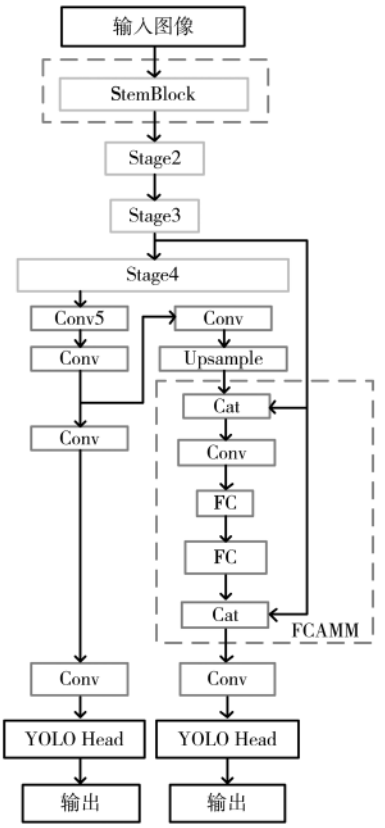


图 2 YOLOv3-SF 网络架构

通过向预测分支中添加全连接注意力混合模块可以更自适应地融合多层次特征，得到具有更高识别能力的特征。具体来说，FCAMM 先将低级特征和高级特征拼接在一起后使用 3×3 卷积层来进行通道缩减；然后利用两个全连接层学习不同通道间的非静态、非线性相互内在联系，通道间乘法为通道重新加权生成调制权重；最后，向全连接层输出添加低级特征来执行特征融合。

与文献[8]中标准的 SE 注意力模块相比，全连接注意力混合模块则仅保留两个全连接层来学习不同通道特征重要程度，在拥有相似性能的同时进一步降低运算量，能更好利用有限模型容量。另外，注意力模块输出的高级特征再次与骨干网络阶段的低级特征混合，形成类残差结构，能更好地保护特征信息完整性。简而言之，本文所提出的全连接注意力混合模块在计算成本和特征表达之间实现了更有力的平衡。

2 骨干网络

从 YOLOv1[5]到 YOLOv3，每一代性能的提升都与骨干网络的改进密切相关。本文所采用的骨干网络基于 ShuffleNetV2 网络模型，是由旷视在 2018 年提出的轻量级卷积神经网络。本文将改进后的 ShuffleNetV2 网络作为骨干网络与 YOLOV3-tiny 架构进行结合，来用于视频监控系统资源受限设备上的目标检测任务。

2.1 ShuffleNetV2 网络模型

ShuffleNetV2 网络模型根据衡量模型复杂度的指标，并联系理论与实际得到了相关适用的改进策略：(1)使用 1×1 卷积核均衡输入输出通道大小；(2)减少组卷积使用；(3)减少网络分支；(4)减少元素级运算。

根据这些策略，引入了图 3 所示单元进行网络的构建，其中的 S1 即表示步幅为 1。即，在开始时先将特征映射在通道维度分离为两个分支，一分支做同等映射，另一分支使用 1×1 卷积来使输入输出通道相同，满足策略(1)；而且其中的两个 1×1 卷积不再是组卷积，满足策略(2)；只有两个分支的输出进行拼接及通道混合，符合策略(3)；同时通道混合可以和下一个通道分离合并成一个元素级运算，符合策略(4)。

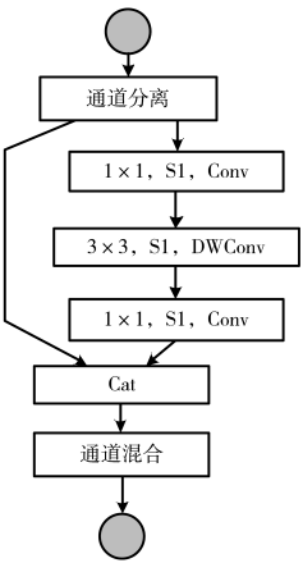


图 3 ShuffleNetUnit 结构

ShuffleNetV2 的结构主要是由图 3 中的单元堆叠而成的 3 个不同的阶段组成，实现了高效卷积。ShuffleNetV2 同样可以设定每个模块的通道数，如 1.5×、1×，进而调整网络的复杂程度。

2.2 ShuffleNetV2-S 网络结构

本文所使用的骨干网络 ShuffleNetV2-S 是 ShuffleNetV2 的改进版。ShuffleNetV2 在网络的前期结构开始时就利用步幅为 2 的卷积及最大池化层来缩小特征图的大小，虽然有效降低了计算成本，但早阶段的低级特征对视觉任务非常重要，过早减小特征图的大小会损害特征表征力，不利于目标检测。受到密集连接[9]和并行结构[10]的启发，重新设计了早期结构 StemBlock，如图 4 所示。在原有结构的基础上，添加分支以用于加强后期结构前的低级特征传播，鼓励特征重用，然后与原结构使用 Add 操作连接来满足后续阶段的输入要求。

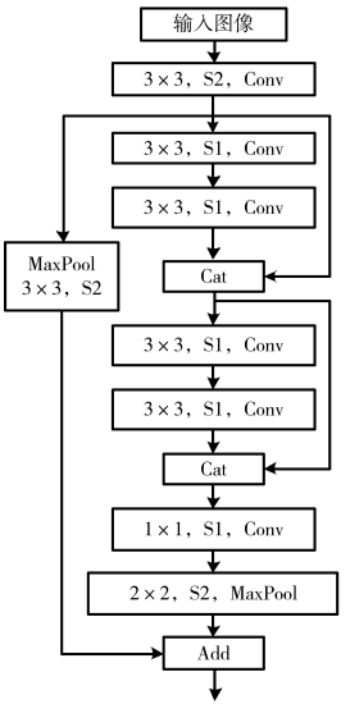


图 4 StemBlock 结构

其中，相较于文献[9]网络中的每个层都会与前面所有层在通道维度上连接，本文所设计的 StemBlock 中的每个层则仅与前面的两层跳连。这样的设计不仅同样加强了特征传播，而且在一定程度上保证了计算量。然后，并行连接原有结构来保证原网络完整性，并与特征复用分支进行元素级相加实现特征融合。本文所改进的骨干网络与 ShuffleNetV2 网络相比，不仅获得了更丰富的低级特征信息，而且维持了网络计算复杂度。

3 实验

本实验所使用的计算机配置为：Intel-Xeon E5-2678 v3 的 CPU，GTX1060 的 GPU，Windows 10 操作系统，程序在 PyTorch 框架下运行。

3.1 数据集

本实验使用来自 PASCAL 可视化对象分类挑战 2007 的数据集，该数据集中有 20 个分类，图片大小不一，共包含 9 963 张用于训练和验证的图像，其中训练集图片 5 011 张，测试集图片 4 952 张。在本实验中，所有图片被转换成 416×416 大小作为网络输入。

3.2 超参数设置及性能指标

本文中神经网络的主要超参数设置如表 1 所示。本文主要使用每秒传输帧数(Frames Per Second，FPS)、

表 1 超参数设置

实验参数	取值
epoch/个	800
batch 大小	32
batch 数量/个	500 200
学习率	0.001
优化方法	SGD
动量	0.97
权重衰减	0.000 456

均值平均精度(mean Average Precision,mAP)、参数量(parameters,Param)以及模型大小 4 个评价指标评判所提出 YOLOv3-SF 算法的检测效果。

3.3 实验结果

为了将目标检测运用在资源有限的视频监控设备上,本文所提出的方法是基于 YOLOv3-tiny 模型架构,因此本文中的网络模型实验结果主要与 YOLOv3-tiny 进行了比较。而为了体现 YOLOv3-SF 的改进提升效果,同样加入了与 1.5×、1×不同复杂度的 ShuffleNetV2 的性能对比,如表 2 所示。

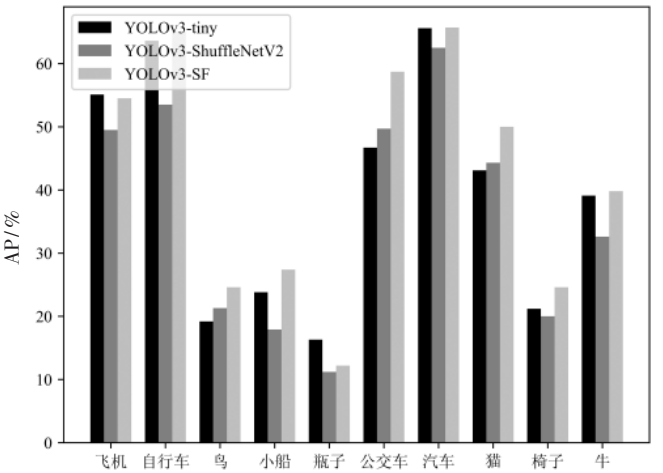
表 2 不同网络模型的性能对比

网络模型	模型/MB	Param	mAP	FPS
v3-tiny	33.4	8.71M	0.431	33.9
v3-ShuffleNetV2-1×	14.1	3.62M	0.381	40.2
v3-ShuffleNetV2-1.5×	19.8	5.12M	0.403	37.4
v3-SF-1×	14.5	3.73M	0.435	33.9
v3-SF-1.5×(本文)	20.3	5.24M	0.458	33.1

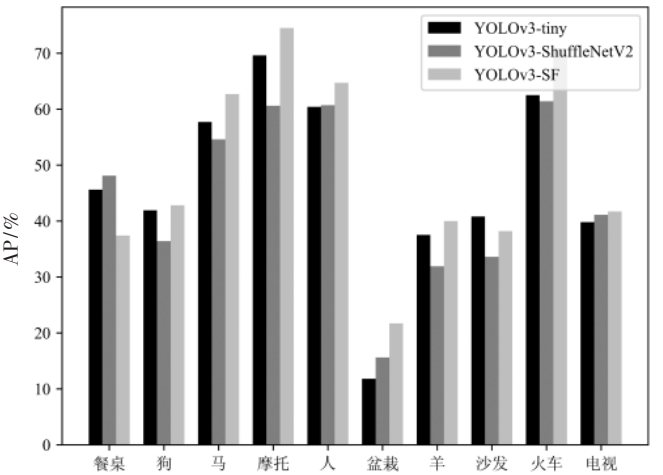
本文所提算法与 YOLOv3-tiny 相比参数量降低 39.8%的同时,mAP 提高了 2.7%,模型体积降低 39.2%,而且 FPS 基本不变;与对应复杂度的 ShuffleNetV2 直接作为骨干网相比,mAP 则有效提升了 5.5%。

本文主要采用改进的复杂度为 1.5 的 ShuffleNetV2 作为特征提取网络,以下对比实验均基于 ShuffleNetV2-1.5×网络。图 5 中展示了 VOC 数据集中每个类别目标的具体 AP 测试结果。可以看出,本文所改进提出的 YOLOv3-SF 算法的检测能力明显高于 YOLOv3-tiny 及 YOLOv3-ShuffleNetV2。实验结果在共 20 个目标类别中仅有瓶子、沙发两类的 AP 结果略低于 YOLOv3-tiny,但全部高于直接将 ShuffleNetV2 作为骨干网络。

为了验证所提改进之处的有效性,本文通过删除部分网络结构来研究网络的性能。通过表 3 所示的消融实验结果可以看出,本文改进后的 ShuffleNetV2-S 网络在保持低模型复杂度的同时,mAP 比原始的骨干网络提高了 3.7%。再结合 FCAMM 融合多层次特征后,总体比 ShuffleNetV2 提高了 5.5%。因此,本文所改进的骨干网络及提出的全连接注意力混合模块都能有效提升性能,同时保证了轻量级模型规模。



(a)前 10 个类别目标



(b)后 10 个类别目标

图 5 每类别目标的测试 AP 值

表 3 消融实验

网络模型	StemBlock	FCAMM	Param	mAP
ShuffleNetV2	no	no	5.12M	0.403
ShuffleNetV2+FCAMM	no	yes	5.16M	0.408
ShuffleNetV2-S	yes	no	5.21M	0.440
YOLOv3-SF	yes	yes	5.24M	0.458

本文还测试了在 YOLOv3-tiny 架构下不同骨干网络的性能,测试结果如表 4 所示。结果表明,ShuffleNetV2-S 骨干网络不仅更轻量级,而且同时满足了检测精度及实

表 4 不同骨干网络的性能对比

骨干网络	模型/MB	Param	mAP	FPS
tiny	33.4	8.71M	0.431	33.9
ShuffleNetV2[7]	19.8	5.12M	0.403	37.4
ShuffleNetV2-S	20.2	5.21M	0.440	40.7
DarkNet-19[11]	95.6	23.85M	0.511	42.9
VGG-16[12]	70.5	17.59M	0.479	25.7
DenseNet-121[9]	38.2	9.89M	0.510	25.9

(下转第 39 页)

[12] 解维坤,陈龙,张凯虹,等.一种 FPGA 的在线编程测试方法:202010898140.0[P].2020.
[13] 肖驰,关炀.一种基于 TCL 语言的数字电路快速测试方法:201911262773[P].2019.
[14] 钱宏文,刘继祥,刘会,等.一种基于 LabVIEW 调用 Vivado Tcl 脚本自动化测试方法:202110763875.7[P].2021.
[15] 周元甲.基于 10GE 网络的 USB2.0 HUB 研究[D].北京:北京邮电大学,2018.
[16] 曹俊文.高速 USB 控制器 CY68014A 的应用[J].物探装备,2015,25(5):329-333.
[17] 杨少博.USB3.0 高速数据传输技术研究及应用[D].太原:中北大学,2016.

(收稿日期:2021-12-16)

作者简介:
刘继祥(1986-),男,工程师,主要研究方向:自动化控制与测试、电路应用验证。

扫码下载电子文档

(上接第 33 页)

时性要求。本文算法中改进后的骨干网络相较其他卷积神经网络在降低计算复杂性以及提高模型表达性之间实现更有力的平衡。

4 结论

本文针对资源受限视频监控系统中传统目标检测算法复杂度高、资源占用大的问题,提出了一种基于 YOLOv3-tiny 架构和 ShuffleNetV2 网络的轻量级算法。通过对特征传播方式进行研究,本文对架构中的 ShuffleNetV2 骨干网络结构进行了重新设计以加强低级特征复用;并提出了一种基于注意力机制的多层次特征融合模块来丰富目标空间位置信息,最后将该算法在 VOC 数据集上进行训练、测试。实验结果表明,本文设计的目标检测算法较 YOLOv3-tiny 的检测精度有一定提升,且对于低算力低存储的监控平台,本文方法具有更强的适用性。本文目标检测任务是在资源极其有限的监控设备上运行,检测速度也会更有限,因此下一步还可通过稀疏、量化等修剪方法对网络模型进行优化,提高检测速度。

参考文献

[1] HU H N,CAI Q Z,WANG D,et al.Joint monocular 3D vehicle detection and tracking[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision,2019:5390-5399.
[2] 刘欣,李卫龙,张灿明.基于边窗滤波和扩张卷积的矿井行人检测[J].电子技术应用,2020,46(10):42-46,50.
[3] ZHANG K,CHEN X,WANG H.Research on external quality inspection technology of tropical fruits based on computer vision[M].Recent Developments in Data Science and Business Analytics.Springer,Cham,2018:165-174.
[4] 陈纪铭,陈利平.一种优化 FCN 的视频异常行为检测定位方法[J].重庆邮电大学学报(自然科学版),2021,33(1):126-134.
[5] REDMON J,DIVVALA S,GIRSHICK R,et al.You only look once:unified,real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2016:779-788.
[6] REDMON J,FARHADI A.Yolov3:an incremental improvement[J].arXiv preprint arXiv:1804.02767,2018.
[7] MA N,ZHANG X,ZHENG H T,et al.ShuffleNet V2:practical guidelines for efficient CNN architecture design[C]//Proceedings of the European Conference on Computer Vision,2018:116-131.
[8] HU J,SHEN L,SUN G.Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2018:7132-7141.
[9] HUANG G,LIU Z,VAN DER MAATEN L,et al.Densely connected convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2017:4700-4708.
[10] SZEGEDY C,IOFFE S,VANHOUCKE V,et al.Inception-v4,inception-resnet and the impact of residual connections on learning[C]//Proceedings of the AAAI Conference on Artificial Intelligence,2017.
[11] REDMON J,FARHADI A.YOLO9000:better,faster,stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2017:7263-7271.
[12] SIMONYAN K,ZISSERMAN A.Very deep convolutional networks for large-scale image recognition[J].arXiv preprint arXiv:1409.1556,2014.

(收稿日期:2021-09-03)

作者简介:
王均成(1996-),通信作者,男,硕士研究生,主要研究方向:深度学习与计算机视觉,E-mail:wangjc5736@163.com。
贺超(1990-),男,博士研究生,主要研究方向:光纤无线通信网络。
赵志源(1998-),男,硕士研究生,主要研究方向:深度学习与计算机视觉。

扫码下载电子文档

版权声明