

基于 YOLOv2 的视频火焰检测方法

杜晨锡^{1,2} 严云洋^{1,2} 刘以安¹ 高尚兵²

(江南大学物联网工程学院 江苏 无锡 214122)¹

(淮阴工学院计算机与软件工程学院 江苏 淮安 223003)²

摘要 一般火焰检测方法由于对复杂场景的应变能力较差,因此检测率较低。文中提出了一种基于改进的 YOLOv2 网络的深度学习火焰检测方法,来自动提取火焰特征;同时,针对特征提取过程中信息丢失的问题,采用聚类选取候选框,以多尺度特征融合的方法融合高层与浅层特征信息,进一步提高了模型的检测率。在 Bilkent 大学火焰视频数据集上的实验结果表明,该方法的平均正检率达到了 98.8%,检测速率达到 40 帧/s,具有较强的鲁棒性和实时性。

关键词 火焰检测, YOLOv2, 聚类, 多级特征, 特征融合

中图分类号 TP391 文献标识码 A DOI 10.11896/j.issn.1002-137X.2019.06.045

Video Fire Detection Method Based on YOLOv2

DU Chen-xi^{1,2} YAN Yun-yang^{1,2} LIU Yi-an¹ GAO Shang-bing²

(School of Internet of Things Engineering, Jiangnan University, Wuxi, Jiangsu 214122, China)¹

(Faculty of Computer & Software Engineering, Huaiyin Institute of Technology, Huaian, Jiangsu 223003, China)²

Abstract It is difficult for general flame detection methods to adapt to complex scenes, so the detection rates is low. This paper proposed a deep learning flame detection method based on an improved YOLOv2 network to extract the flame features automatically. In order to avoid the information loss in the feature extraction process, the selected anchor box by clustering is suggested and multi-scale feature fusion method is used to fuse high-level and shallow feature information, to further improve the detection rate of the model. Experimental results on the Bilkent University flame video dataset show that the average true inspection rate of the proposed method is 98.8%, and the detection rate is 40 frames/s, so its robustness and real-time performance are strong.

Keywords Fire detection, YOLOv2, Clustering, Multi-scale feature, Feature fusion

1 引言

得益于图像处理技术的进步和视频监控设备的普及,基于计算机视觉的火焰检测技术正迅猛发展。火焰图像的特征种类繁多,传统的火焰检测方法多采用手工设计的特征来构建检测模型。Dimitropoulos 等^[1]提出了一种利用火焰颜色、频闪等特征建模的检测方法,使用 SVM 分类器进行分类,但因为候选区域中运用了动态纹理分析,其计算开销较大;Celik 等^[2]提出了一种融合火焰前景和颜色信息的火焰检测器,在图像输入尺寸较低的前提下,检测速率达到 30 fps,具有较强的实时性;Poobalan 等^[3]采用 RGB 颜色模型检测火焰的颜色,并采用基于颜色的火焰分割技术提取感兴趣区域,总体检测精度达到 90%,具有一定的实用性;耿庆田等^[4]利用 RGB 模型和 YCbCr 模型开发新的颜色识别规则来建立火焰检测模型,减少了照明背景变化带来的干扰,在不利的照明背景条件下展现出了较强的鲁棒性。手工设计特征的检测模型

在应用场景、检测精度和检测速率方面都存在一定的局限性。近年来,深度学习理论在目标检测等领域取得了突破性进展,因而利用特征学习算法强大的表征和建模能力的火焰检测技术,具有重要的研究价值和应用前景。Frizzi 等^[5]提出了一种基于经典卷积神经网络的火焰检测模型(其网络结构包含 8 层),并直接对特征图进行火焰检测,取得了一定的检测效果,但对于不同焰色的火焰,检测精度差距较大。Muhammad 等^[6]提出了一种基于 GoogleNet 网络的火焰检测架构,并采用迁移学习的手段进行微调,该方法在检测精度和效率两方面取得了一定的平衡,但是仍具有较高的误报率。近年来,随着硬件计算能力的提高,更多的深层卷积神经网络模型被应用到目标检测领域,主要分为以 R-CNN 系列网络为主的双步目标检测算法和以 YOLO、SSD 网络为主单步目标检测算法。前者由区域建议方法产生一系列样本候选框,再通过卷积神经网络进行样本分类;后者将提取候选框和分类两个任务融合到一个网络中,直接将目标边框定位问题转化为回归问题

到稿日期:2018-06-01 返修日期:2018-08-26 本文受国家自然科学基金项目(61402192),江苏省“六大人才高峰”项目(2013DZXX-023),江苏省“青蓝工程”,江苏省高等学校自然科学研究重大项目(18KJA520001),淮安市“533 英才工程”资助。

杜晨锡(1991-),男,硕士生,主要研究方向为机器学习、模式识别, E-mail: tencyrush@163.com; 严云洋(1967-),男,博士,教授, CCF 会员,主要研究方向为数字图像处理、模式识别, E-mail: yunyang@hyit.edu.cn(通信作者); 刘以安(1963-),男,博士,教授,主要研究方向为模式识别、数据融合; 高尚兵(1981-),男,博士,副教授,主要研究方向为数字图像处理、模式识别。

进行处理。两种检测方法在候选框生成策略上存在差异,导致其在性能上也有不同,前者在检测准确率和定位精度上占优,后者则在算法速度上占优^[7]。YOLOv2 模型^[8]不仅继承了单步检测算法在速率上的优势,在 VOC2012 数据集上达到了最高的检测速率,即 67 帧/s,并且具有最佳的检测精度。如表 1 所列。

表 1 检测框架性能表

Table 1 Performance of detection framework

Frame Work	Fast R-CNN	Faster R-CNN VGG-16	Faster R-CNN ResNet	YOLO	SSD 300	SSD 500	YOLOv2 416×416
mAP	70.0	73.2	76.4	63.4	74.3	76.8	76.8
FPS	0.5	7	5	45	46	19	67

视频火焰检测对时效性具有较高的要求。因此在优先考虑检测速率的前提下,本文综合网络模型的检测精度,提出一种基于 YOLOv2 网络的改进的火焰检测模型,先通过聚类选取候选框,再应用多级尺度特征融合方法提升模型的检测精度。实验结果表明,该模型具有较强的鲁棒性和实时性。

2 YOLOv2 网络

YOLOv2 模型在 YOLO 的基础上,借鉴先进的目标检测研究成果,使得检测精度得到了较大的提升,主要体现在区域建议手段和特征提取网络两个方面。

2.1 候选框提取

YOLOv2 在区域建议手段中引入了 anchor 机制^[9],即选定 k 个固定宽高比的候选框,用于预测最终的边界框,以使模型更容易学习。以 416×416 的输入尺寸为例,最终输出的特征图尺寸为 13×13 ,即为 13×13 个网格。对于每个落入网格的目标,记 $\Pr(\text{object})$ 为 1,否则为 0;再分别计算 k 个候选框与真实框的交并比 IOU (Intersection-Over-Union):

$$IOU_{pred}^{truth} = \frac{\text{area}(\text{truth}) \cap \text{area}(\text{pred})}{\text{area}(\text{truth}) \cup \text{area}(\text{pred})}$$

随后,通过 IOU 值计算得到一组类别置信度 $Conf_i$:

$$Conf_i = \Pr(\text{class}_i) \times IOU_{pred}^{truth}$$

最后,选定类别置信度为最大值的候选框来预测真实框。

2.2 网络结构

YOLOv2 以 Darknet-19 框架为基础,包含 19 个卷积层和 5 个最大池化层。神经网络的高层信息体现了更强的语义信息,对于识别问题较为有效;而浅层的特征由于分辨率较高,对于目标定位有天然的优势。因此,将多种尺度特征融合即可获得高分辨率、强语义的特征。YOLOv2 正是利用这一点,采用了类似于 ResNet 的 skip connection^[10]的模块。首先使用 route 层将浅层特征传递下来,然后通过 reorg 层降低输入特征图的维度,以便融合高层特征和浅层细粒度特征,对浅层特征进行二次利用,提高了模型的检测能力。

3 算法改进

3.1 维度聚类候选框

火焰作为一种非刚体,在燃烧过程中其形态变幻无常,具有不同的长宽比。并且 YOLOv2 的 anchor 维度是基于 VOC2007 和 VOC2012 数据集确定的,并不具有普遍性。因

而选定一组适应火焰数据集的候选框就显得尤为重要。针对这一问题,提出利用 k-means 聚类的方式在自制火焰数据集上确定 anchor 参数,以获得更高的 IOU 分数,加快模型的收敛速度。

k-means 是一种常见的聚类分析算法,采用距离作为相似性的评价指标,即认为两个对象的距离越近则相似度越大,距离靠近的对象组成簇,因此首先要确定聚类簇的个数 k 。如图 1 所示,使用 k-means 聚类算法绘制出不同 k 值及其对应的 IOU 曲线图。

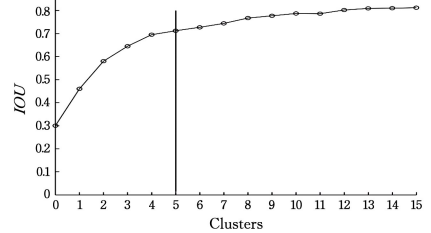


图 1 聚类候选框维度

Fig. 1 Clustering anchor box dimensions

随着 k 值的逐渐增大, IOU 值的变化趋于平缓。增加候选框个数能获得更高的 IOU 分数,但是也势必会增加模型的计算量。因此,综合考虑模型的复杂度和召回率两个因素,可选取 $k=5$ 作为聚类最优个数。

3.2 多级特征融合

火灾发生初期是火焰检测的关键时期,此时的火焰形态多为小型火焰,这就要求模型对小目标具有较高的检测能力,更多地利用高分辨率的浅层特征来构建高级语义特征图。针对这一问题,本文引入 SharpMask 方法^[11]中的自上而下侧连接结构用以扩增前馈网络。自上而下的路径用以构建高分辨率强语义特征,侧连接则为了使定位更加准确。利用该模型将更浅层的 52×52 尺度融入到最终检测特征图当中,并采用 refinement 模块取代原有的较为粗犷的 reorg 层。具体来说,每一个 refinement 模块 R_i 将低维度上采样而成的特征 M_i 和横向传递来的高维度特征 F_i 融合起来,再通过上采样向下传递。在结构上, M_i 和 F_i 具有相同的维度, R_i 模块的目的就是使 $M_{(i+1)}$ 具有 M_i 和 F_i 两倍的维度,以减少信息丢失。模型结构如图 2 所示。

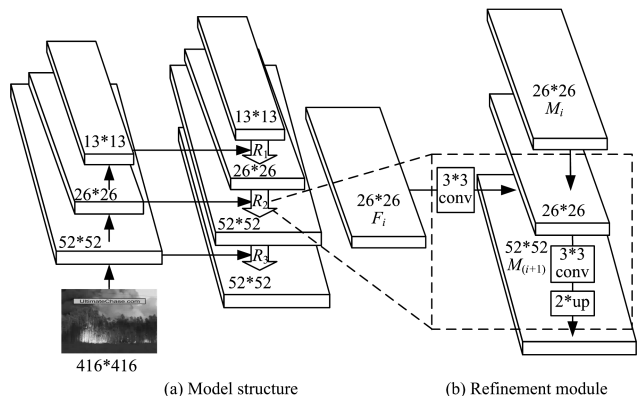


图 2 模型结构图

Fig. 2 Architectures for model

4 实验结果及分析

4.1 实验环境

实验环境:CPU 为 AMD Ryzen 7 1700 3.0GHz,GPU 为 NVIDIA GeForce GTX 1070 Ti,内存为 8 GB,操作系统为 Ubuntu 16.04。

4.2 实验数据集

数据集分为两部分:1)通过网络和拍摄等手段收集的火焰视频构建的训练集 FireData,场景包括室内、室外、白天、夜晚等;2)通过 Bilkent 大学火灾数据库 VisiFire 构建的测试视频集。将全部视频转换成帧图片后,按一定次序抽取,并手动按 VOC 格式进行打标。数据集的内容描述如表 2 所列。

表 2 数据集描述

Table 2 Dataset description

Item	FireData	VisiFire	Total
Pictures	5 624	2 090	7 714
Objects	10 920	3 130	14 050

4.3 验证聚类候选框

本文通过 IOU 分数来评价 YOLOv2 和本文提出的两种候选框方案。

采用 VisiFire 数据集,通过 k-means 算法对数据集中的目标维度进行聚类,并以最终得到的聚类维度作为候选框 anchors 的参数,最终得到的 anchors 参数分别为 (5.93, 7.00), (8.13, 17.32), (12.06, 9.26), (16.81, 20.61), (27.37, 26.65)。由于本文改进模型的最终输出尺度为 YOLOv2 模型的 4 倍,因此须将 YOLOv2 中的 anchors 参数分别乘 4 作为对比。如表 3 所列,经过聚类产生的 anchors 参数与火焰形态有着更高的重叠率。

表 3 候选框的对比

Table 3 Comparison of anchor box generation

Box Generation	k	AvgIOU
YOLOv2	5	0.63
Proposed	5	0.71

4.4 验证多级特征融合

为了验证本文提出的多级特征融合模型的有效性,首先使用 ImageNet 数据集,分别对 YOLOv2 和本文模型进行预训练,随后在 FireData 上微调检测器,最终在 VisiFire 数据集上进行测试,并根据测试结果绘制 PRC (Precision-Recall Curve) 曲线,且以 mAP (mean Average Precision) 作为评价标准。从图 3 所示的结果可以看出,本文提出的改进模型在大部分情况下具有更高的查准率和查全率,并且 mAP 提升了 8.0,改进效果明显。

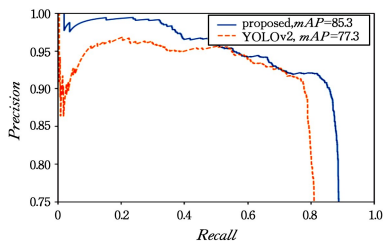


图 3 模型性能的对比

Fig. 3 Comparison of model performance

4.5 对比实验

将本文算法与文献[5]、文献[9]及文献[12-13]中的算法进行对比,实验采用从 Bilkent 大学火灾数据库 VisiFire 中选取的 9 段视频。视频示例如图 4 所示,对比检测结果如表 4 和表 5 所列。其中 TP 代表火焰检测率, FP 代表漏检率, TN 代表非火焰视频正确率, FN 代表误检率。

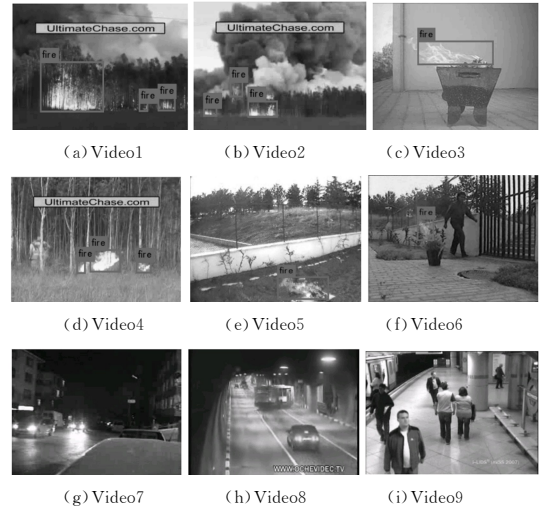


图 4 视频示例

Fig. 4 Video examples

表 4 火焰视频的检测结果

Table 4 Detection results on flame videos

视频	总帧数	文献[5]		文献[9]		文献[12]		文献[13]		本文	
		TP	FP	TP	FP	TP	FP	TP	FP	TP	FP
Video1	200	89.5	10.5	100.0	0.0	98.0	2.0	96.0	4.0	98.5	1.5
Video2	216	90.3	9.7	99.1	0.9	97.7	0.9	95.8	4.2	98.6	1.4
Video3	439	90.9	9.1	97.5	2.5	92.7	7.3	90.7	9.3	99.3	0.7
Video4	170	87.6	12.4	100.0	0.0	97.1	2.9	94.7	5.3	98.2	1.8
Video5	595	91.8	8.2	99.3	0.7	94.3	5.7	93.0	7.0	99.0	1.0
Video6	470	91.3	8.7	96.4	3.6	93.6	6.4	92.1	7.9	98.9	1.1
平均值	348.3	90.2	9.8	98.7	1.3	98.0	2.0	93.7	6.3	98.8	1.2

表 5 非火焰检测视频结果

Table 5 Detection results on non-flame videos

视频	总帧数	文献[5]		文献[9]		文献[12]		文献[13]		本文	
		TP	FP	TP	FP	TP	FP	TP	FP	TP	FP
Video7	144	82.6	17.4	97.9	2.1	94.4	5.6	93.1	6.9	98.6	1.4
Video8	120	83.3	16.7	97.5	2.5	96.6	3.4	94.2	5.8	98.3	1.7
Video9	228	88.6	11.4	100.0	0.0	97.8	2.2	96.9	3.1	100.0	0.0
平均值	164	84.8	15.2	98.5	1.5	96.3	3.7	94.7	5.3	99.0	1.0

文献[5]基于经典卷积神经网络,针对整个视频帧训练模型,并在最后一层特征图应用滑动窗口来检测火焰。该模型结构较为粗糙,对于帧数较高(如 Video3, Video5, Video6, Video9)的视频,检测率有所提高,不过整体检测率仍然偏低。文献[9]、文献[12]及文献[13]的模型为当下具有代表性的 3 类检测模型。文献[9]中的 Faster-RCNN 为典型的双步检测器,在 Fast-RCNN 的基础上融入了 RPN 作为区域建议手段,在 Video1, Video2 和 Video4 的稳定场景下展现出了极佳的检测性能,不过对于 Video3 和 Video6 这种火焰形态变化较大的场景,检测率则有所降低,同时因为采用了 RPN 和全连接层,整体计算量较大,帧率仅为 14 帧/s,无法满足实时性要求。文献[12]给出的 SSD 单步检测器采用 VGG-16 进行特

征提取,并融合了多级特征图用于检测,检测尺度从 38×38 到 1×1 不等,然而由于整体检测尺度偏小,且没有采用侧连接结构,因此对于小型目标(如 Video5 和 Video6)的检测不理想。文献[13]中的 MobileNet 为轻型检测器的代表,在复现该模型时采用 SSD 作为检测器,该模型为了减少整体参数,使下采样的位置整体靠前,因此位于模型深层的检测尺度相对于文献[12]更小,不过对计算能力要求较小,检测速度可达 140 帧/s。

本文基于改进的 YOLOv2 构建网络模型,采用自上而下的传递结构加大了模型的深度,并且通过融合低层的特征加强了检测特征的语义信息,因而对小型火焰目标(如 Video1, Video2, Video5, Video6)的检测效果较好。同时,由于模型本身采用中心点预测的方法,减少了回归过程中的震荡幅度,并且本文采用的侧连接结构提供了更精确的位置信息,因而对于火焰形态随外界因素影响变化较大的场景(如 Video3, Video6),该模型亦能够有效地提供更精确的定位。实验结果表明,在火焰视频中,本文模型对不同场景下不同长宽比的火焰都展现出了较高的检测率,平均检测率优于其他模型。在非火焰视频中,本文模型对于灯光等类火物体也有着最低的误检率,且计算参数较少,检测速率快,达到了 40 帧/s,具有很强的实时性。

结束语 本文以 YOLOv2 模型为基础,提出通过聚类候选框的方法提升定位的精度,又提出通过自上而下侧连接结构改进模型从而加强特征信息,与同类算法相比,本文提出的模型具有最高的正检率和最低的误检率,并且具有较强的鲁棒性。当然改进模型的结构更为复杂,计算量更大,检测速度有所下降,但仍能满足实时性的要求。下一步的工作将着力扩充数据集,加强模型对复杂场景的应变能力。同时,在满足模型检测精度的前提下,进一步改进模型,以减少计算量,增强该模型的实用性。

参 考 文 献

- [1] DIMITROPOULOS K, BARMPOUTIS P, GRAMMALIDIS N. Spatio-Temporal Flame Modeling and Dynamic Texture Analysis for Automatic Video-Based Fire Detection[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2015, 25(2): 339-351.
- [2] CELIK T, DEMIREL H, OZKARAMANLI H. Automatic fire detection in video sequences[C]// European Signal Processing Conference. IEEE, 2015: 1-5.
- [3] POOBALAN K, LIEW S C. Fire detection algorithm using image processing techniques[C]// International Conference on Artificial Intelligence and Computer Science. 2015.
- [4] GENG Q T, YU F H, ZHAO H W, et al. New algorithm of flame detection based on color features[J]. Journal of Jilin University (Engineering and Technology Edition), 2014, 44(6): 1787-1792. (in Chinese)
耿庆田, 于繁华, 赵宏伟, 等. 基于颜色特征的火焰检测新算法[J]. 吉林大学学报(工学版), 2014, 44(6): 1787-1792.
- [5] FRIZZI S, KAABI R, BOUCHOUICHA M, et al. Convolutional neural network for video fire and smoke detection[C]// Conference of the IEEE Industrial Electronics Society. IEEE, 2016: 877-882.
- [6] MUHAMMAD K, AHMAD J, MEHMOOD I, et al. Convolutional Neural Networks based Fire Detection in Surveillance Videos[J]. IEEE Access, 2018, PP(99): 1.
- [7] ZHANG H, WANG K F, WANG F Y. Advances and Perspectives on Applications of Deep Learning in Visual Object Detection[J]. Acta Automatica Sinica, 2017, 43(8): 1289-1305. (in Chinese)
张慧, 王坤峰, 王飞跃. 深度学习在目标视觉检测中的应用进展与展望[J]. 自动化学报, 2017, 43(8): 1289-1305.
- [8] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger [C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Washington: IEEE Computer Society, 2017: 6517-6525.
- [9] REN S, GIRSHICK R, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.
- [10] HE K, ZHANG X, REN S, et al. Identity Mappings in Deep Residual Networks[C]// Computer Vision - ECCV 2016. Springer International Publishing, 2016: 630-645.
- [11] PINHEIRO P O, LIN T Y, COLLOBERT R, et al. Learning to refine object segments[C]// Lecture Notes in Computer Science. Heidelberg. Springer, 2016: 75-91.
- [12] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector[C]// European Conference on Computer Vision. Springer, Cham, 2016: 21-37.
- [13] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications [J]. arXiv preprint arXiv:1704.04861, 2017.