

数字信号与图像处理

课程论文

班 级： 计算2114

姓 名： 庄佳强

学 号： 202121331104 ；

日 期： 2024.01 ；

成 绩： ；

评语：

图像识别—基于yolo的图像检测识别

庄佳强

（集美大学计算机工程学院，福建 厦门 361021）

**摘要：**数字信号处理（DSP）和目标检测在计算机科学领域中都扮演着重要的角色。数字信号处理通过采样、量化和离散化等技术，将连续时间信号转换为数字形式，为图像处理提供了基础。而目标检测则是计算机视觉领域的一个核心任务，旨在识别图像或视频中的特定目标。本文聚焦于数字信号处理与目标检测的融合，特别关注于You Only Look Once（YOLO）系列算法，探讨了这一领域的研究意义、先前的研究成果以及相关的算法介绍,并且通过实验来对比验证yolo的图像检测能力。

**关键词：**图像识别;YOLO;数字信号处理;图像检测;

Image recognition - YOLO based image detection and recognition

Zhuangjiaqiang

（ School of Computer Engigeer，Jimei University，Xiamen，361021，China）

**Abstract:** Digital signal processing (DSP) and object detection both play important roles in the field of computer science. Digital signal processing converts continuous time signals into digital form through techniques such as sampling, quantization, and discretization, providing a foundation for image processing. Object detection is a core task in the field of computer vision, aimed at identifying specific targets in images or videos. This article focuses on the integration of digital signal processing and object detection, with a particular focus on the You Only Look Once (YOLO) series of algorithms. It explores the research significance, previous research results, and related algorithm introductions in this field, and compares and verifies YOLO's image detection ability through experiments.

**Key words:** Image recognition;YOLO;digital signal processing;image detection;

###### 0、引言

数字信号处理（DSP）在图像处理领域扮演着关键的角色。数字信号处理涉及采样、量化、滤波等技术，允许对图像进行数字化处理。这种处理方式使得图像能够以离散的形式被计算机理解和处理，为图像识别任务提供了基础。

在传统方法中，通常需要多次处理图像，对图像进行多次扫描，这导致了计算复杂度的增加和实时性能的下降。这种多步骤的处理方式对于需要实时响应的应用场景，如自动驾驶和视频监控，显然是不理想的。

为解决这一问题，YOLO的创始人Joseph Redmon于2016年提出了YOLO，将目标检测任务转化为回归问题。这一思想的核心是通过单一的前向传播即可同时预测目标的位置和类别，从而实现对整个图像的一次全面观察。YOLO的设计理念在一定程度上解决了实时性和准确性的平衡问题，成为目标检测领域的一项创新[6]。

YOLO算法的不断发展与数字信号处理的进步有着紧密的关联。随着数字信号处理技术的不断演进，YOLO算法在处理图像特征、优化计算等方面得到了更强大的支持，从而推动了目标检测领域的发展。其继任版本YOLO v2、v3、v4等在网络结构、损失函数、骨干网络等方面的改进，也使得YOLO算法更好地适应了不同应用场景和需求[1]。

###### 1、研究意义

YOLO的研究推动了实时目标检测的发展，通过将目标检测任务转化为回归问题，实现了对整个图像的一次全面观察。这种思想通过数学信号处理中的优化算法和回归分析等方法得以实现，提高了目标检测的效率。对实时性能的强调也与数学信号处理中的实时算法和优化有关。

YOLO通过深度学习技术进行端到端的图像处理，利用卷积神经网络等结构来提取图像特征。在数学信号处理中，这种端到端的处理方式也与信号处理的全局优化和特征提取相关。YOLO的研究推动了这种综合性方法的发展，加强了深度学习在数字信号处理中的应用。

本文中将在YOLO v4的基础上，进行研究，对应不同的训练出来模型进行对比测试实验，比较出较优的模型进行后续的使用。

###### 2、之前的研究成果

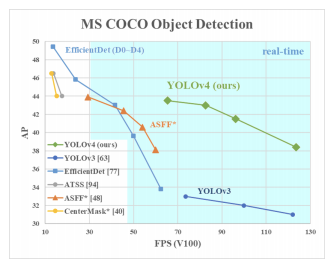
YOLO v1: 实现了端到端的目标检测，较好地平衡了速度和准确性。然而，在小目标检测和处理重叠目标方面仍存在一定的挑战[2]。

YOLO v2: 对小目标和不同尺寸目标的检测性能有所提升，同时支持大规模类别的目标检测。

YOLO v3:在保持实时性的基础上，显著提升了目标检测的准确性，成为YOLO系列的一个重要版本[4]。

YOLO v4:在各个方面进行了创新，进一步提高了目标检测的性能，适应了更广泛的场景和应用[5]。

通过图一可以得出在YOLO v4在相同FPS下AP较YOLO v3大约有10%提升，在相同AP下，YOLO v4的FPS 较YOLO v3 有较大幅度的提升。



图一

###### 3、典型算法介绍

本文重点分析介绍YOLO v4的模型框架。

BN算法:

反向传播时经过该层的梯度是要乘以该层的参数。

当正向传播为：

那么反向传播是就为

考虑到有多个卷积层

因此当从l层到k层时就有

其中当>1时会导致梯度消失现象而当 <1时会导致梯度爆炸现象。而BN算法是用来解决这样的传播问题。BN的作用就是抹去w的scale的影响。

//小batch结果

//小batch类别

//正则化

//扫描和运动

最后的 就是 归一化(BN)后的值,归一化后所有值满足均值为0方差为1

接着又进行了scale加上shift操作(y=scale\*x+shift)，每个神经元增加了两个参数scale和shift参数，这两个参数是通过训练学习到的，意思是通过scale和shift把这个值从标准正态分布左移或者右移一点并长胖一点或者变瘦一点，每个实例挪动的程度不一样，这样等价于非线性函数的值从正中心周围的线性区往非线性区移动。

在训练中使用BN是要计算均值和方差的，而这两个统计量是随着样本不同而变化的。如果在测试中依然遵循这样的方式，那么无疑同一个样本在不同的batch中预测会得到不一样的概率值，这显然是不对的。

在测试中，BN根据训练过程中计算的均值和方差，使用滑动平均去记录这些值。在测试的时候统一使用记录下来的滑动平均值，这一点可以从源码中看出来。所以在TensorFlow或者Pytorch中，BN的代码分别有is\_training 和 self.training字段，就是为了区别使用行为的。

CBN算法：

batch比较小时，BN在batch维度统计不准确，导致准确率下降。在无法扩大batch训练的前提下(显卡性能不足)，CBN通过收集最近几次迭代信息来更新当前迭代时刻的均值和方差，这样就变向实现了扩大batch目的。我们知道在当前迭代时刻，参数已经更新了N次，存储的前几个迭代的参数肯定无法直接和当前迭代次数进行合并计算，也就是由于网络权重的变化，不同迭代产生的网络激活无法相互比较。故需要找到一种解决办法。一种可行的方法就是利用当前和先前迭代的均值和方差[5]。

由于梯度迭代属性,网络权重在迭代过程中缓慢改变，我们可以用泰勒展开式，根据。

=(1)

=(2)

和是网络的权重值的统计量。

是展开式的高阶无穷小项,较小时可以忽视。

对于某一层L，假设当前迭代时刻为t,则时刻中的统计量和,在第t时刻的参数,对于时刻的统计值，可以在处展开，为权重参数。

局部梯度中和会急剧减少。在浅层上网络的权重的变化对深层上的激活分布减少的影响。可以通过减少BN层的协变量来解释。

因此我们把上述的式子转化为:

(3)

(4)

(5)

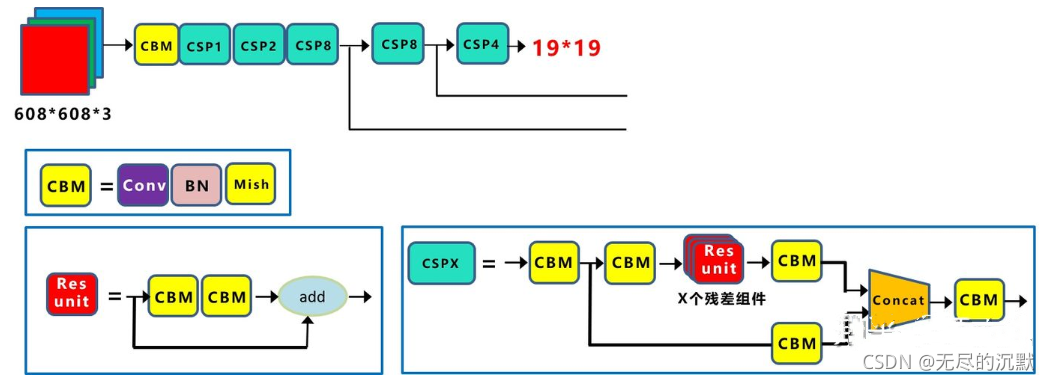
=(6)

(7)

最后得出：

(8)

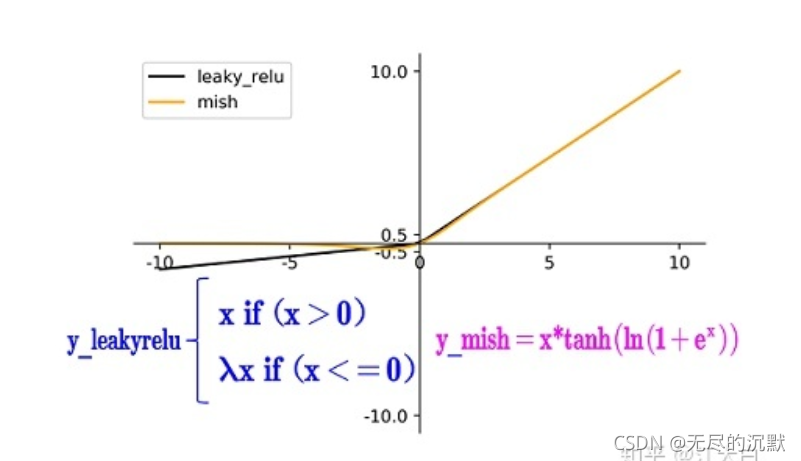
BackBone:



图二

见图2，YOLO v4中采用了主干网络为 CSPDarknet53，其中包含了5个CSP模块，所以输入的图像为608\*608,图像变化为608->304->152->76->38->19 。而且YOLO作者在BackBone采用了Mish激活函数[9]。

Mish激活函数：



图三

见图3 ,mish在负值时并不是完全截断，而允许比较小的负梯度流入从而保证了信息的流动。mish激活函数无边界，这让他避免了饱和（有下界，无上界）且每一点连续平滑且非单调性，从而使得梯度下降更好。

Dropblock:

Yolov4中使用的Dropblock，其实和常见网络中的Dropout功能类似，也是缓解过拟合的一种正则化方式。传统的Dropout很简单，一句话就可以说的清：随机删除减少神经元的数量，使网络变得更简单。

Neck:

在目标检测领域，为了更好的提取融合特征，通常在Backbone和输出层，会插入一些层，这个部分称为Neck。相当于目标检测网络的颈部，也是非常关键的。Yolov4的Neck结构主要采用了SPP模块、FPN+PAN的方式。

在SPP模块中，使用k={1\*1,5\*5,9\*9,13\*13}的最大池化的方式，再将不同尺度的特征图进行凹形操作。spp模块是YOLOv4中在YOLOv3的基础上加了的模块，而PAN则也是YOLOv4的创新模块[10]。

实验中使用算法：

实验中图像识别任务为检测出人物在smoke。

整体：

先检测通过person检测模型，检测出帧中的人。之后通过脸部检测模型在人的基础检测出人脸。之后在图像中扣出人脸，放到本次实验对比的模型中检测人脸。

对识别出的特征进行NMS算法：

IOU(交并比)

IOU=Area(A∩B)/Area(A∪B)(A为框a的面积，B为框b的面积)

NMS算法:

（1）将所有框的得分排序，选中最高分及其对应的框

（2）遍历其余的框，如果同特征下和当前最高分框的重叠面积(IOU)大于一定阈值，我们就将框删除。

（3）从未处理的框中继续选一个得分最高的，重复上述过程。

对得出的特征进行选择算法:

对应检测出5个特征的模型，检测出smoke\_hand进行选择，然后遍历其他特征，找出smoke特征,如果二者的交集和smoke的面积的差大约一个值，表示，当前人正在抽烟。

对应检测出3个特征的模型，只需检测出的smoke就进行选择。

###### 4、实验和结论

本次实验中，使用两个yolo v4 的检测模型，其中一个模型检测3个特征，另外一个模型检测5个特征，通过验证yolo v4在检测多特征下的检测能力和检测速度。

a. 数据集

实验的数据集来自老师，包括1500张无批注照片，其中包含smoke，smoke\_hand，backgroud,people特征。

b. 实验设计

在相同的检测环境下，前面的people检测，face检测使用相同的模型。

只在最后的smoke检测阶段使用不同的检测模型。

评估标准为：准确率和FPS。

检测率:可以合理的看出模型的检测能力。

FPS:可以看出模型的检测速度。

c.实验设置

硬件配置：GPU GTX3050TI。

参数配置: 实验中使用二者使用模型的IoU值为0.85，置信度值域为0.3。

d. 结果分析

同一图片的不同模型：

图4 5特征模型 图5 3特征模型

|  |  |  |
| --- | --- | --- |
| 模型 | 检测率 | FPS |
| 3特征模型 | 0.26 | 112 |
| 5特征模型 | 0.2 | 110 |
| 5特征模型(严格模式) | 0.13 | 109 |

其中5特征模型(严格模式)和5特征模型区别为5特征模型只要有smoke or smoke\_hand 就算检测成功。而5特征模型(严格模式)只有同时检测出smoke and Smoke hand 才算检测成功。

e. 结果解释

5特征模型中和3特征模型进行比较，看出特征增加后，检测率有着较显著的降低，这可能是因为特征过多，模型有些欠拟合，导致检测率下降。

而5特征模型和5特征模型(严格模式),主要是因为同时检测出smoke and smoke\_hand 的检测率过低，导致检测率过低。

f. 局限性和未来工作

局限性

在我拿到这个5特征模型时,模型训练迭代不高。实验数据不够全面，可能无法完整的反应出检出率。

改进

等新的模型训练好还需要再次实验。

扩大数据集范围特征范围。

g.结论

通过实验，验证了YOLO v4的图像识别检测能力。而且通过对比实验，检验出了当多特征时的模型的检测能力会有所下降。

###### 总结

总的来说，YOLO在图像识别领域具有重要的作用，在数字信号处理中对图像的检测处理提供了新的方法。基于YOLO的图像检测与识别在实时性、端到端设计以及多类别目标检测等方面取得了显著的成果，不断推动着计算机视觉领域的发展。未来的研究方向将集中在提高算法在复杂场景中的性能、处理小目标的能力以及更广泛的应用领域。

而YOLO现在面临的挑战与未来发展方向包括小目标检测，复杂场景处理，实时性能与准确性平衡：不断追求实时性能与准确性的平衡。

**参考文献**

1. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). "You Only Look Once: Unified, Real-Time Object Detection." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
2. Redmon, J., & Farhadi, A. (2017). "YOLO9000: Better, Faster, Stronger." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
3. Redmon, J., & Farhadi, A. (2018). "YOLOv3: An Incremental Improvement." arXiv preprint arXiv:1804.02767.
4. Bochkovskiy, A., Bochkovskiy, A. (2020). "YOLOv4: Optimal Speed and Accuracy of Object Detection." arXiv preprint arXiv:2004.10934.
5. Bochkovskiy, A. (2020). "YOLOv4-CSP: Exposing Spurious Convolutional Scaling." arXiv preprint arXiv:2006.08224.
6. Chen, Y., & Pock, T. (2017). Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), 1256-1272.
7. Oppenheim, A. V., & Schafer, R. W. (2010). Discrete-Time Signal Processing. Pearson.
8. Chen, Y., & Pock, T. (2017). Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), 1256-1272.
9. ultana F., Sufian A., Dutta P.A review of object detection models based on convolutional neural networkIntelligent Computing: Image Processing Based Applications (2020)
10. Zhiqiang, W., & Jun, L. (2017, July). A review of object detection based on convolutional neural network. In 2017 36th Chinese Control Conference (CCC) (pp. 11104-11109). IEEE.