

基于DRN-Faster R-CNN的复杂背景多目标鱼体检测模型

孙龙清^{1,2} 孙希蓓^{1,2} 吴雨寒^{1,2} 罗冰²

(1. 中国农业大学国家数字渔业创新中心, 北京 100083; 2. 中国农业大学信息与电气工程学院, 北京 100083)

摘要: 针对现有多目标鱼体检测大多针对受控环境进行, 泛化能力有限的问题, 提出了一种简单、有效的复杂背景下多目标鱼体检测模型。通过迁移学习构建基于DRN的特征提取方法, 对原始图像进行特征提取, 结合RPN进一步生成候选检测框; 构建基于Faster R-CNN的复杂背景多目标鱼体检测模型。在ImageNet2012数据集上的实验结果表明: 该模型对复杂背景下金鱼的检测平均精度达到89.5%, 远高于R-CNN+AlexNet模型、Faster R-CNN+VGG16模型和Faster R-CNN+ResNet101的检测精度, 表明该模型可以高效精确地实现复杂背景下的多目标鱼体检测。

关键词: 鱼体; 目标检测; 特征; 卷积神经网络; 深度残差网络

中图分类号: TP391 **文献标识码:** A **文章编号:** 1000-1298(2021)S0-0245-07

Multi-target Fish Detection Model Based on DRN-Faster R-CNN in Complex Background

SUN Longqing^{1,2} SUN Xibei^{1,2} WU Yuhang^{1,2} LUO Bing²

(1. National Innovation Center for Digital Fishery, China Agricultural University, Beijing 100083, China

2. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China)

Abstract: Target detection is the key link of fish tracking, behavior recognition and abnormal behavior detection of fish body. Therefore, fish detection has important practical significance. Due to the low imaging quality of underwater surveillance video, the complicated underwater environment, and the high visual diversity of fish bodies, multi-target fish detection in complex background is still a very challenging problem. In order to solve the problem that the existing multi-target fish detection is mostly carried out in a controlled environment and the generalization ability is limited, a simple and effective multi-target fish detection model in complex background was proposed. The feature extraction method based on DRN was constructed by transfer learning. The features were extracted from the original image, and the candidate detection frame was further generated by combining RPN. A multi-target fish detection model in complex background was constructed based on Faster R-CNN. The experimental results on the ImageNet2012 data set showed that the detection accuracy of this model for goldfish in complex background reached 89.5%, which was much higher than the detection accuracy of the R-CNN+AlexNet model, Faster R-CNN+VGG16 model and Faster R-CNN+ResNet101 model in this data set, indicating that this model can effectively and accurately realize the detection of multi-target fish in complex background.

Key words: fish; object detection; feature; convolutional neural networks; deep residual network

0 引言

基于视频的目标物体检测是计算机视觉中非常基础且重要的一个研究方向^[1]。水产养殖中鱼体的检测, 对分析鱼的行为变化、及时发现鱼异常行为、鱼病监测等有重要的作用, 对水产养殖中鱼的精

确检测, 也是实现精细化养殖的重要组成部分^[2]。

近年来, 诸多学者对鱼的检测进行了深入研究, 尤其是在特定环境和受控条件下的鱼体检测研究, 并取得了较大进展。HAN等^[3]将高斯分布特性与形态学方法相结合, 提出了一种基于粒子群聚类的鱼骨检测方法。QIAN等^[4]为了解决由于多条游动

鱼体产生的遮挡问题,提出了一种基于鱼头检测的多鱼跟踪方法。WANG 等^[5]通过鱼眼图像能量变化,快速、准确地检测鱼类的新鲜度。BOUDHANE 等^[6]基于泊松-高斯理论提出了一种水下图像预处理和鱼群定位的方法。SALMAN 等^[7]利用卷积神经网络模型解决了水下不同光强变化背景下鱼的识别。LI 等^[8]通过使用连接 ReLU、Inception 和 HyperNet 等来改变区域卷积神经网络(Region-based convolutional neural networks, RCNN)卷积层的结构,构建一种鱼类检测的神经网络,在 ImageCLEF 数据集上获得了 89.95% 的平均精度。SUN 等^[9]提出了一种利用深度学习和超分辨率的方法,实现了对低分辨率水下鱼体图像的目标识别检测。TAMOU 等^[10]通过微调 AlexNet 神经网络和迁移学习的方法,实现水下鱼体的分类检测。CHEN 等^[11]为了解决野生鱼类的分类和检测问题,提出了一种深度神经分类系统,可以利用摄像机自动对鱼进行分类。随着精确养鱼(PFF)概念^[12]的提出,提高水产养殖精准监测和控制的能力变得越来越迫切。FUKUNAGA 等^[13]通过拟合高斯混合模型,解决观赏鱼检测过程中的相互遮挡问题。上述鱼体目标检测的研究大部分都是针对受控环境进行的,如视频图像采集摄像头位置固定、背景简单、情况单一等,泛化能力有限,不适合在复杂环境中进行鱼体目标检测的需求。

深度学习方法^[14]的产生极大地提高了语音识别、视觉目标识别、目标检测、植物病虫害、农产品品质检测^[15-18]等领域的技术水平。REN 等^[19]在

RCNN 中,将区域生成网络(Region proposal network, RPN)经过端到端的训练生成高质量区域用于目标检测,使得检测系统在 GPU 上的帧率为 5 f/s。HE 等^[20]设计了 152 层的深度残差网络(Deep residual network, DRN),在 ImageNet 测试集上的误差降低到 3.57%。MANDAL 等^[21]在鱼体变形、低分辨率等复杂背景条件下,将引入的 RPN 与 3 个分类网络相结合,建立鱼体检测系统,实验准确率为 82.4%,远高于先前提出的方法。

针对现有多目标鱼体检测大多针对受控环境进行,泛化能力有限的问题,本文以金鱼为例,提出基于 DRN-Faster R-CNN 的复杂背景多目标鱼体检测模型,利用现有开放模型预训练初始化权重,通过迁移学习构建基于 DRN 的特征提取方法,对原始图像进行特征提取,结合 RPN 进一步生成候选检测框;利用 Faster R-CNN 网络对新模型进行训练,构建基于 Faster R-CNN 的复杂背景多目标鱼体检测模型,以实现复杂背景下的多目标鱼体检测,为进一步监测鱼的行为提供良好基础。

1 材料和方法

1.1 样本数据集

本文采用 ImageNe2012 数据集下编号为 n01443537 的不同背景下不带标注的 1 400 幅金鱼图像作为样本数据集。根据图像的特点将图像按照背景复杂度不同分为河流、养殖池塘、鱼缸、观赏水族箱和其他背景等 5 大类。分类后的图像数据如图 1 所示。

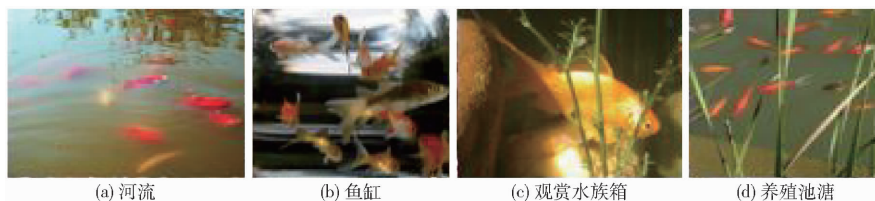


图 1 样本数据集

Fig. 1 Data set of samples

1.2 图像预处理

由于分类后的数据集中,各个背景图像的数量不同,需要对样本数据集中的图像进行相应的处理来扩充数据集,保证不同背景下图像数据的均匀性。图像预处理主要包括:

(1)针对标注样本集不足的问题,采用裁剪、旋转、翻转、随机缩放及色度空间变换等方法,对已有的数据集进行扩充。最终保证每种背景下的图像达到 1 500 幅,确保数据集的均匀性,提高模型的鲁棒性。

(2)使用主流的 LabelImage 进行人工标注,以

便于生成适合 TensorFlow 框架下的数据格式。

(3)将图像进行等比例缩放,缩放后的最大边长为 1 024 像素,最小边长为 600 像素,确保模型训练的速度和质量。

(4)经过预处理和扩充后,最终训练集图像总数为 4 800 幅,随机地从每一类数据集中划分出 80% 作为训练集、20% 作为验证集。测试集图像总数为 1 200 幅。

预处理后标注的实例图像如图 2 所示。

1.3 实验方法和设计

计算机配置 8 GB 内存, CPU 为 Ubuntu 16.04



图 2 标注后的图像
Fig. 2 Image after annotation

操作系统中的 Intel(R) Core (TM) i7 - 6700。算法采用 TensorFlow^[22] 深度学习框架和 Python 3.6.5 实现。实验方法设计包括模型训练、测试、对比验证。

(1)模型训练分为 3 部分,采用预训练的 DRN 模型提取输入图像的卷积特征图;输入 RPN 网络层后输出所有可能的候选检测框;输入模型的第 5 个残差块进行候选检测框的分类和位置修正。

(2)模型测试使用的图像数量为 1 200 幅,在每次迭代结束后测试误差反向传播,达到更新训练参数的目的。

(3)对比验证:为验证本模型的识别效果,选择 R - CNN + AlexNet^[23] (简称 RCNN)、Faster R - CNN + VGG16(简称 FRCNN + VGG16)、Faster R - CNN + ResNet101(简称 FRCNN + ResNet)主流目标检测方法和本文方法进行对比。

2 模型构建与结果分析

2.1 算法原理

本文检测模型主要包括图像特征提取、候选检测框生成、候选检测框分类和定位 3 部分,完整的算法流程如图 3 所示。

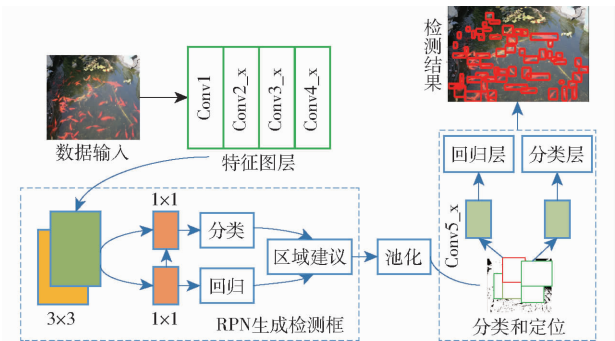


图 3 多目标鱼体检测算法流程图

Fig. 3 Flow chart of multi-target fish detection algorithm

图像特征提取部分采用预训练的 DRN 模型,利用迁移学习的方法对图像信息进行特征提取,4 层残差网络分别为:Conv2_x、Conv3_x、Conv4_x、Conv5_x,

每一部分分别有 3、4、23、3 个残差块。候选检测框生成部分采用 RPN,它可以接受一幅任意尺寸的图像作为输入,然后输出 1 组长方形的感兴趣区域,每个感兴趣区域都会获得一个是否包含物体的打分值。

候选检测框分类和定位部分是最终的候选检测框的分类和候选框位置的修正,这个过程将使用图像特征提取部分提供的感兴趣区域框的位置,在同一幅卷积特征图上找到相应的特征,这些特征将作为输入使神经网络最后判断出物体检测位置和可信度。

2.2 模型构建

2.2.1 迁移学习

多层的深度神经网络模型通常具有海量的模型参数需要进行训练和学习,占用巨大的计算机资源且对训练样本的数量要求高。而且现有大规模图像数据库,如 ImageNet、VOC 等,只包含行人、汽车、飞机等常见目标,只能检测数据库中所含有的特定类型的目标。但经大型数据库训练完成的模型参数已具备较强的目标提取能力,可作为迁移学习的初始化参数,显著提高学习任务的效率^[24]。迁移学习是对在大规模的图像数据集上预训练好的网络模型的模型参数,通过少量新目标数据集进行监督训练和微调,使迁移学习后得到的神经网络模型具备对该新目标的检测能力。

本文采用迁移学习优化网络参数,进行模型的训练。在 ResNet - 101 预训练模型的基础上进行迁移学习,利用预训练模型提取的权重信息,作为新的特征提取权重,对样本数据集进行特征提取。如图 4 所示,基于 DRN 结构开展迁移学习,采用 Faster R - CNN 算法进行训练。整体结构由 2 部分组成:RPN 生成检测框信息;使用 Faster R - CNN 检测候选框并进行识别、分类和边框回归。其中,模型的部分超参数设置如表 1 所示。

2.2.2 图像特征提取

目标检测效果取决于图像特征提取,然而,传统的卷积层或全连接层在信息传递时,通常都会存在信息丢失、损耗等问题。残差网络通过直接将输入信息绕道传到输出,网络模型只需要学习输入和输出相差的部分,一定程度上保护了信息的完整性,简化了学习的复杂度。

(1)残差块优化选择

在残差网络中,直接使用多个卷积层学习输入和输出之间的非线性映射比较困难,通过用多层卷积去学习输入和输出之间的残差,可以使网络更容易训练。此时每个残差模块的输出数据都由输入数

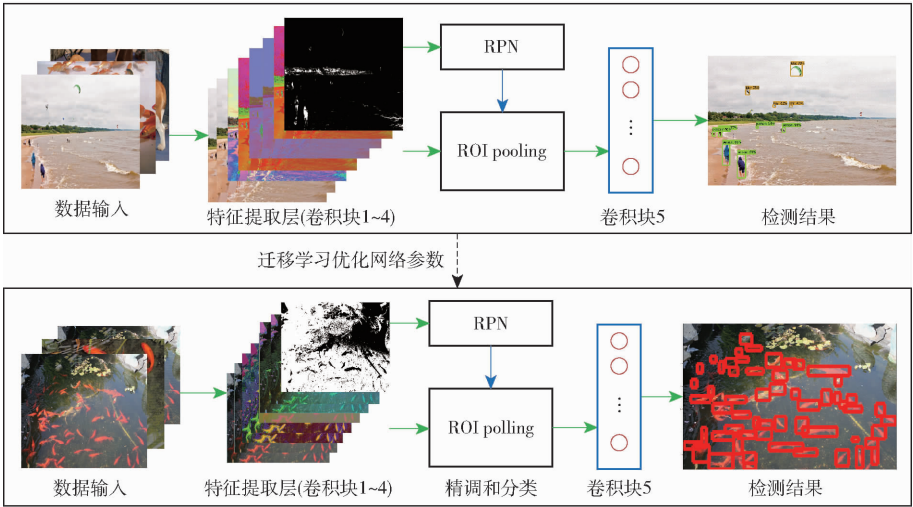


图 4 迁移学习结构图

Fig.4 Structure diagram of transfer learning

表 1 模型超参数设置

Tab.1 Hyperparameter settings of model

参数	数值
训练轮数	2 000
批处理量	12
池化卷积核尺寸	2 × 2
池化步长	2
学习率	0.000 03
测试数量	800

据和残差共同构成。由于每个模块中的输入和输出之间出现了“直连通道”，梯度传播会更容易，从而有利于构建更深的模型。

图 5 为 2 种不同的残差模块。如图 5 所示,这 2 种 ResNet 结构分为 ResNet34 残差块(图 5a)和 ResNet50/101/152 残差块(图 5b)。其中图 5b 为本文采用的残差块示意图,采用了批归一化设计,降低了参数数量,通过第 1 个 1 × 1 的卷积把 256 维通道数降到 64 维后再通过 1 × 1 卷积恢复,总参数数量为 1 179 648,是不使用批量归一化层的参数数量(69 632)的 16.94 倍。

本文对残差块进行计算优化,即将图 5a 中的两个 3 × 3 的卷积层替换为图 5b 中的 1 × 1、3 × 3 和 1 × 1 的卷积层。图 5b 结构中的 3 × 3 卷积层首先在一个 1 × 1 卷积层下降维,减少了计算量,后又通过另一 1 × 1 的卷积层进行了还原,既保持了模型的精度又减少了计算量。

(2) DRN 提取图像特征

采用 101 层的 DRN 结构用于提取原始数据的图像特征,具体步骤如下:网络模型以任意尺寸的图像作为输入,经过卷积核为 7 × 7、步长为 2 的卷积后输入最大池化层进行池化;将输出结果依次送入 4 层残差网络,各层残差网络的残差块数量分别为

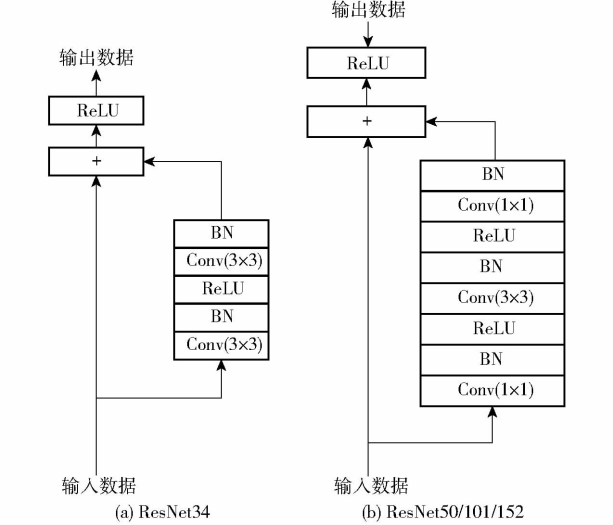


图 5 2 种 ResNet 结构对比

Fig.5 Comparison of two kinds of ResNet structures

3、4、23、3,每个残差块包含 3 个卷积层(第 1 层和第 3 层尺寸为 1 × 1,第 2 层尺寸为 3 × 3)。考虑到计算速度,采用预训练的 DRN 模型 ResNet101,利用迁移学习的方法对图像信息进行特征提取。第 4 层残差网络 Conv4_x 的输出为区域建议网络和感兴趣区域(Region of interest,ROI)池化共享的部分,第 5 层残差网络 Conv5_x 作用于 ROI 池化之后的特征图,此时特征图的维度符合 ResNet101 中 Conv5_x 的输入,如图 6 所示。

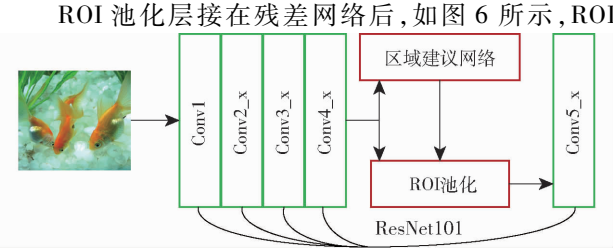


图 6 DRN 模型

Fig.6 Model diagram of DRN

池化层的输入是任意大小的卷积,输出是固定维数的向量,卷积层的通道数不变,即 ROI 池化层可以把任意宽度、高度的卷积特征转换为固定长度的向量。通过 ROI 池化层将卷积特征中不同形状的区域对应到同样长度的向量特征,将原始图像中的不同长宽区域都对应到一个固定长度的向量特征,完成各个区域的特征提取工作。

引入残差网络,通过添加跳跃连接将对函数的拟合变为每层对某一残差的拟合,缓解了深层网络难以收敛的问题,提高了模型对鱼类这种小而密集分布的目标的检测性能。

2.2.3 候选检测框生成

在 R - CNN、SPPNet^[25] 和 Fast R - CNN^[26] 中,目标物体的候选框提取算法均为选择性搜索(Selective search, SS)方法。在该方法中,候选框是图像经过分割之后的超像素区域不断相互融合所形成的,非常耗时。为了减少计算量、提高检测的速度和精度,引入 RPN 结构代替已产生的候选框,把物体检测整个流程融入一个神经网络中,专门用来提取候选框。

RPN 的本质是基于滑窗的无类别目标检测器,输入图像经过 DRN 的前 4 层残差网络进行特征提取后,送入候选检测框生成模块,由 RPN 输出感兴趣的目标区域。此时 RPN 并不能判断出感兴趣区域是不是金鱼,而是把可能含有目标的区域选取出来,即感兴趣区域。同时,RPN 也会在提取的特征映射图上框定这些 ROI 对应的大致位置,即输出检测的候选矩形框。

在 RPN 的输入特征图中,假设以每个像素点为中心都存在不同面积和长宽比的参考框。参考框在对预测的候选框中心位置和信息做参数回归时,起到了中间桥梁的作用。

在 RPN 中,特征图先经过一层卷积计算,再分别经过 2 个并行的卷积层得到输出张量。其中一个卷积层的输出张量表示每个参考框内的物体是目标还是背景,另一个卷积层的输出张量表示候选框的中心位置和尺寸信息。

Faster R - CNN 损失函数定义为

$$L(p_i, \mathbf{t}_i) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(\mathbf{t}_i, \mathbf{t}_i^*) \quad (1)$$

式中 N_{cls} ——每次迭代所用到的批处理量
 N_{reg} ——RPN 输入特征图中参考框的总数
 p_i ——第 i 个参考框中包含物体的概率
($0 \leq i < N_{cls}$)

λ ——权重平衡参数

$\mathbf{t}_i$ ——预测参考框的坐标向量

p_i^* ——标注信息

$\mathbf{t}_i^*$ ——真实值参考框的坐标向量

当第 i 个参考框与任意一个标注的包围框的交并比(IoU)大于阈值 0.7,或者与某个标注的包围框有最大的 IoU 时, $p_i^* = 1$,否则, $p_i^* = 0$ 。 $\mathbf{t}_i = [t_x, t_y, t_w, t_h]$ 和 $\mathbf{t}_i^* = [t_x^*, t_y^*, t_w^*, t_h^*]$ 分别表示预测的候选框和标注的包围框的位置、尺寸信息(即中心坐标值 x 、 y 和宽度 w 、高度 h)经过归一化之后的值,计算公式为

$$\begin{cases} t_x = (x - x_a) / w_a \\ t_y = (y - y_a) / h_a \\ t_w = \lg(w / w_a) \\ t_h = \lg(h / h_a) \end{cases} \quad (2)$$

$$\begin{cases} t_x^* = (x^* - x_a) / w_a \\ t_y^* = (y^* - y_a) / h_a \\ t_w^* = \lg(w^* / w_a) \\ t_h^* = \lg(h^* / h_a) \end{cases} \quad (3)$$

其中,下角标为 a 的变量,表示该变量用于表征第 i 个参考框的相关参数;上角标为 $*$ 的变量,表示该变量用于表征标注的包围框的相关参数;没有上角标和下角标的变量用于表征 RPN 预测的候选框的相关参数。

输入的原始图像经过 DRN 进行特征提取后得到的特征图,首先输入 RPN 网络产生多个候选框,计算它们在给定的真值框上交、并集上的交点面积比来表示检测框的定位精度,即交并比重叠度,计算公式为

$$\varphi(B_{gt_i}, B_{dt_i}) = \frac{\text{Area}(B_{gt_i} \cap B_{dt_i})}{\text{Area}(B_{gt_i} \cup B_{dt_i})} \quad (4)$$

式中 B_{gt_i}, B_{dt_i} ——第 i 个标注的真值框和候选检测框

2.2.4 候选检测框分类和定位

为了精确地检测多目标鱼体,需要对候选检测框进行分类和定位。

假设每个候选检测框都是独立的,定义每个候选检测框面积为 a_i ,经过实验后设置交并比重叠度阈值为 $\theta = 0.5$,筛选后得到合适的候选检测框面积 b_i 为

$$b_i = \begin{cases} 1 & (a_i > \theta) \\ 0 & (a_i \leq \theta) \end{cases} \quad (5)$$

将筛选后的候选检测框输入 DPN 最后一层残差网络中进行目标分类框回归,得到最终的分类和目标位置、尺寸检测结果。

2.3 结果与分析

2.3.1 模型检测性能分析

(1)检测精度

模型的性能由平均精度均值 (mAP) 来评价。RCNN、FRCNN + VGG16、FRCNN + ResNet 与本文方法的平均精度均值对比如表 2 所示。

表 2 不同模型检测精度对比

Tab.2 Comparison of mean average precision of different models		%
检测模型	mAP	
RCNN	58.0	
FRCNN + VGG16	71.0	
FRCNN + ResNet	76.0	
本文方法	89.5	

(2)检测时间

为了验证本文方法的检测速度,对比了其他目标检测模型。使用不同目标检测模型 (RCNN、FRCNN + VGG16、FRCNN + ResNet) 对同一批图像进行检测,比较该批图像检测时间的平均值,如表 3 所示。

表 3 不同模型的单幅图像检测时间

Tab.3 Detection time of different models for single image		s
模型	检测时间	
RCNN	1.20	
FRCNN + VGG16	0.75	
FRCNN + ResNet	0.35	
本文方法	0.40	

实验结果表明,本文方法在目标图像的检测速度上高于传统的 RCNN 和 FRCNN 方法。

2.3.2 本文方法检测效果分析

图 7 为 ImageNet2012 数据集上对图像进行检测的部分检测结果。从图 7 可以看出,本文方法能够在不同的复杂背景下同时对多目标进行检测,且在自然河流中目标尺寸较小、分布密集、背景比较复杂;在水族箱、鱼缸等环境下目标尺寸较大,本文方法在这些复杂的情形下均能正确地对多目标进行准确的检测,证明了本文方法的有效性和复杂背景下的可迁移性。

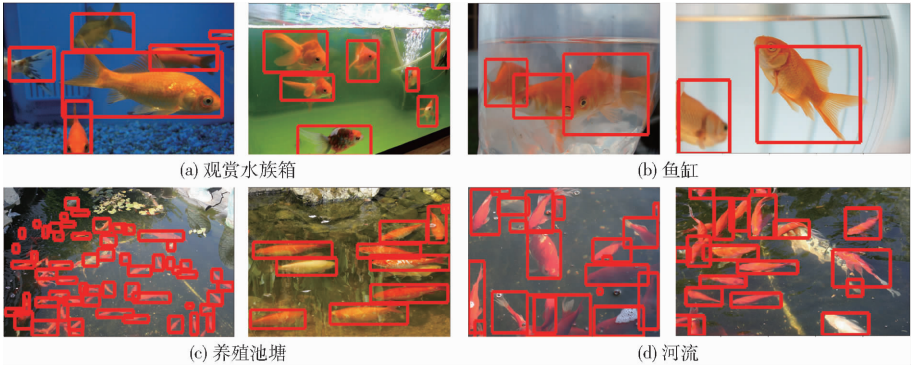


图 7 本文方法在 ImageNet2012 数据集上部分检测效果

Fig.7 Detection effect of algorithm proposed on ImageNet2012 data set

为了验证算法的性能,对比不同方法的检测效果。选取了 4 种具有代表性的测试图像:河流背景、小溪背景、水族箱背景、养殖池塘背景。其中养殖池塘的图像鱼体目标小而密集、水族箱中的目标尺寸较大、池塘背景复杂且鱼体目标较小、河流中鱼体目标尺寸较大,目标距离较为紧凑。检测对比效果如图 8 所示。

由图 8 可以看出 RCNN 的检测效果较差,其原因在于它首先在原始图像上产生候选框,缩放到统一尺寸后,才开始对每幅图像进行特征提取,可能会对不同尺寸图像的特征信息造成很大的损失,采用 AlexNet 结构进行特征提取不充分,对复杂背景多目标的适应性较差。FRCNN + VGG16 检测效果一般,尤其在尺度背景变化较大的情况下,检测效果不佳,其原因在于采用了 VGG16 对特征的提取较为粗糙,很难满足变化多尺度的检测要求。FRCNN + ResNet

方法采用了更深的残差网络模型提取图像特征,具有更加优异的特征表达能力,取得了较好的检测效果,但是仍然存在一定数量的漏检和检测框错位的现象。在此基础上,本文方法采用优化了特征提取过程的残差模块,使得对不同尺度目标图像的特征提取更加准确,因此取得了最好的检测性能,有效减少了漏检和错检。

3 结论

(1)针对现有多目标鱼体检测大多针对受控环境进行,泛化能力有限的问题,利用预训练模型初始化权重,通过迁移学习构建基于 DRN 的特征提取方法,对残差块进行优化选择后对原始图像进行特征提取,结合 RPN 进一步生成候选检测框;利用 Faster R-CNN 网络对新模型进行训练,构建了基于 Faster R-CNN 的复杂背景多目标鱼体检测模型。

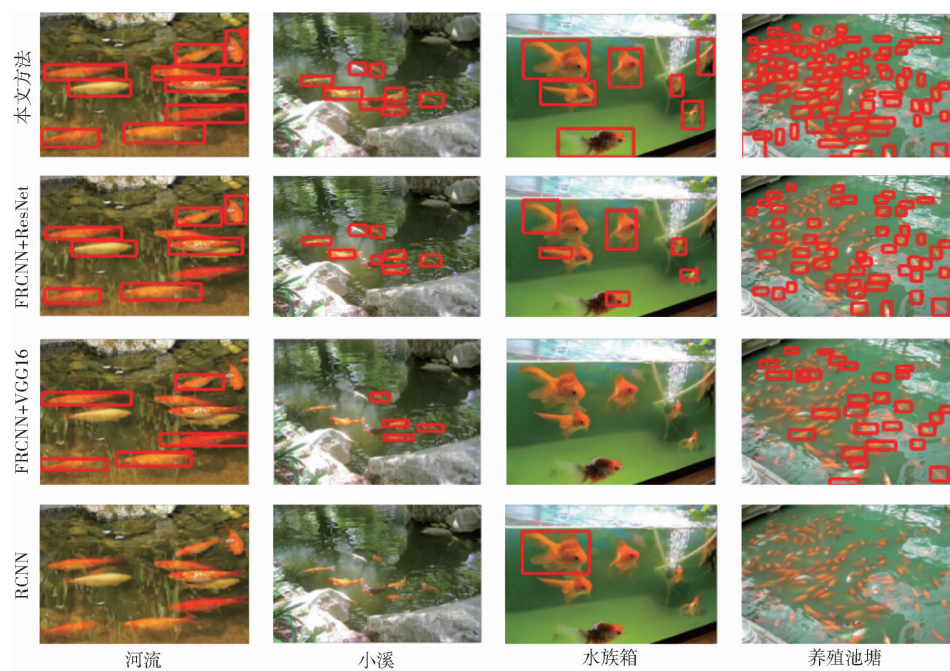


图 8 不同模型在河流、小溪、水族箱、养殖池塘背景下的检测效果

Fig.8 Detection effects of different models in background of rivers , streams , aquariums and breeding ponds

(2)实验结果表明,本文提出的模型对复杂背景下金鱼的检测平均精度均值达到了 89.5%,远高于 RCNN + AlexNet 模型、Faster R - CNN + VGG16 模型和 Faster R - CNN + ResNet101 的检测精度,表明该模型在复杂背景下的多目标鱼体检测上具有更优越的性能。

参 考 文 献

[1] SZEGEDY C, LIU N W, JIA N Y, et al. Going deeper with convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2015.

[2] QIN H, LI X, LIANG J, et al. DeepFish: accurate underwater live fish recognition with a deep architecture [J]. Neurocomputing, 2016, 187:49 - 58.

[3] HAN Y, SHI P. An efficient approach for fish bone detection based on image preprocessing and particle swarm clustering[C]//International Conference on Intelligent Computing. Springer, Berlin, Heidelberg, 2007.

[4] QIAN Z M, WANG S H, CHENG X E, et al. An effective and robust method for tracking multiple fish in video image based on fish head detection[J]. BMC Bioinformatics, 2016, 17(1):251.

[5] WANG F, ZANG Y, WO Q, et al. Fish freshness rapid detection based on fish-eye image[C]//PIAGENG 2013: Image Processing and Photonics for Agricultural Engineering. International Society for Optics and Photonics, 2013.

[6] BOUDHANE M, NSIRI B. Underwater image processing method for fish localization and detection in submarine environment [J]. Journal of Visual Communication and Image Representation, 2016,39(8):226 - 238.

[7] SALMAN A, JALAL A, SHAFAIT F, et al. Fish species classification in unconstrained underwater environments based on deep learning[J]. Limnology and Oceanography: Methods, 2016,14(9):570 - 585.

[8] LI X, TANG Y, GAO T. Deep but lightweight neural networks for fish detection[C]//OCEANS 2017-Aberdeen. IEEE, 2017.

[9] SUN X, SHI J, DONG J, et al. Fish recognition from low-resolution underwater images[C]//International Congress on Image & Signal Processing. IEEE, 2017.

[10] TAMOU A B, BENZINOUE A, NASREDDINE K, et al. Underwater live fish recognition by deep learning[C]//International Conference on Image & Signal Processing. Springer, Cham, 2018.

[11] CHEN G, SUN P, SHANG Y. Automatic fish classification system using deep learning[C]//IEEE International Conference on Tools with Artificial Intelligence. IEEE Computer Society, 2017.

[12] MARTIN F, FRANK K, NORTON T, et al. Precision fish farming: a new framework to improve production in aquaculture [J]. Biosystems Engineering, 2018, 173:176 - 193.

[13] FUKUNAGA T, KUBOTA S, ODA S, et al. GroupTracker: video tracking system for multiple animals under severe occlusion [J]. Computational Biology and Chemistry,2015,57:39 - 45.

[14] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553):436.

- reflectance spectroscopy(VNIRS) in the field[J]. *Geoderma*, 2016, 261(2): 151 – 159.
- [10] 李广信, 王超, 冯美臣, 等. 冬小麦冠层光谱与土壤供氮状况相关性研究[J]. *农业机械学报*, 2017, 48(5): 275 – 281.
LI Guangxin, WANG Chao, FENG Meichen, et al. Correlation between soil nitrogen situation and canopy spectra of winter wheat[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2017, 48(5): 275 – 281. (in Chinese)
- [11] 马永杰, 郭俊先, 郭志明, 等. 基于近红外透射光谱及多种数据降维方法的红富士苹果产地溯源[J]. *现代食品科技*, 2020, 36(6): 303 – 309
MA Yongjie, GUO Junxian, GUO Zhiming, et al. Origin tracing of red Fuji apple based on near infrared transmission spectrum and various dimension reduction methods [J]. *Modern Food Science and Technology*, 2020, 36(6): 303 – 309. (in Chinese)
- [12] 张瑶, 李民赞, 郑立华, 等. 基于近红外光谱分析的土壤分层氮素含量预测[J]. *农业工程学报*, 2015, 31(9): 121 – 126.
ZHANG Yao, LI Minzan, ZHENG Lihua, et al. Prediction of soil total nitrogen content in different layers based on near infrared spectral analysis[J]. *Transactions of the CSAE*, 2015, 31(9): 121 – 126. (in Chinese)
- [13] LIU X, TOSUN D, WEINER M W, et al. Locally linear embedding (LLE) for MRI based Alzheimer's disease classification [J]. *Neuro Image*, 2013, 83: 148 – 157.
- [14] YUAN X, GE Z, YE L, et al. Supervised neighborhood preserving embedding for feature extraction and its application for soft sensor modeling: supervised neighborhood preserving embedding [J]. *Journal of Chemometrics*, 2016, 30(8): 430 – 441.
- [15] MAATEN L V D, HINTON G. Visualizing data using t-SNE[J]. *Journal of Machine Learning Research*, 2008, 9: 2579 – 2605.
- [16] BELKIN M, NIYOGI P. Laplacian eigenmaps for dimensionality reduction and data representation [J]. *Neural Computation*, 2003, 15(6): 1373 – 1396.
- [17] BREIMAN L. Random forests[J]. *Machine Learning*, 2001, 45(1): 5 – 32.
- [18] 李民赞. 光谱分析技术及其应用[M]. 北京: 科学出版社, 2006.
- [19] 吕杰, 郝宁燕, 崔晓临. 利用可见光近红外的尾矿区农田土壤 Cu 含量反演[J]. *农业工程学报*, 2015, 31(9): 265 – 270.
LÜ Jie, HAO Ningyan, CUI Xiaolin. Inversion model for copper content in farmland of tailing area based on visible-near infrared reflectance spectroscopy[J]. *Transactions of the CSAE*, 2015, 31(9): 265 – 270. (in Chinese)
- [20] 汪成龙, 李小昱, 武振中, 等. 基于流形学习算法的马铃薯机械损伤机器视觉检测方法[J]. *农业工程学报*, 2014, 30(1): 245 – 252.
WANG Chenglong, LI Xiaoyu, WU Zhenzhong, et al. Machine vision detecting potato mechanical damage based on manifold learning algorithm[J]. *Transactions of the CSAE*, 2014, 30(1): 245 – 252. (in Chinese)
- [21] 贾渊, 李振江, 彭增起. 结合 LLE 流形学习和支持向量机的猪肉颜色分级[J]. *农业工程学报*, 2012, 28(9): 147 – 152.
JIA Yuan, LI Zhenjiang, PENG Zengqi. Pork color grading based on LLE manifold learning and support vector machine[J]. *Transactions of the CSAE*, 2012, 28(9): 147 – 152. (in Chinese)

(上接第 251 页)

- [15] LI C, PENG X, JING T, et al. A learning-based approach for leaf detection in traffic surveillance video[J]. *Multidimensional Systems & Signal Processing*, 2017(11): 1 – 10.
- [16] ALVARO F, SOOK Y, SANG K, et al. A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition[J]. *Sensors*, 2017, 17(9): 2022.
- [17] DENG L, WANG Y, HAN Z, et al. Research on insect pest image detection and recognition based on bio-inspired methods [J]. *Biosystems Engineering*, 2018, 169: 139 – 148.
- [18] INKYU S, ZONGYUAN G, FERAS D, et al. DeepFruits: a fruit detection system using deep neural networks[J]. *Sensors*, 2016, 16(8): 1222.
- [19] 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 & Machine Intelligence*, 2015, 39(6): 1137 – 1149.
- [20] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016.
- [21] MANDAL R, CONNOLLY R M, SCHLACHERZ T A, et al. Assessing fish abundance from underwater video using deep neural networks[C] // 2018 International Joint Conference on Neural Networks (IJCNN), 2018: 1 – 6.
- [22] ABADI M. TensorFlow: learning functions at scale[J]. *Acm Sigplan Notices*, 2016, 51(9): 1.
- [23] GUPTA S, ARBELAEZ P, GIRSHICK R, et al. Indoor scene understanding with RGB-D images: bottom-up segmentation, object detection and semantic segmentation[J]. *International Journal of Computer Vision*, 2015, 112(2): 133 – 149.
- [24] LONG M, WANG J, DING G, et al. Transfer feature learning with joint distribution adaptation[C] // 2013 IEEE International Conference on Computer Vision (ICCV). IEEE Computer Society, 2013.
- [25] 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): 346 – 361.
- [26] GIRSHICK R. Fast R-CNN[C] // 2015 IEEE International Conference on Computer Vision (ICCV). IEEE Computer Society, 2015.