

基于改进 YOLO 和迁移学习的 水下鱼类目标实时检测

李庆忠¹ 李宜兵¹ 牛 炯¹

摘 要 为了实现非限制环境中水下机器人基于视频图像的水下鱼类目标快速检测,提出基于改进 YOLO 和迁移学习的水下鱼类目标实时检测算法.针对 YOLO 网络的不足,设计适合水下机器人嵌入式系统计算能力的精简 YOLO 网络(Underwater-YOLO).利用迁移学习方法训练 Underwater-YOLO 网络,克服海底鱼类已知样本集有限的限制.利用基于限制对比度自适应直方图均衡化的水下图像增强预处理算法,克服水下图像的降质问题.利用基于帧间图像结构相似度的选择性网络前向计算策略,提高视频帧检测速率.实验表明,文中算法能实现在非限制环境下海底鱼类目标的实时检测.相比 YOLO,文中算法对海底鱼类小目标和重叠目标具有更好的检测性能.

关键词 卷积神经网络,深度学习,鱼类目标检测,单级式目标检测算法(YOLO),迁移学习

引用格式 李庆忠,李宜兵,牛 炯.基于改进 YOLO 和迁移学习的水下鱼类目标实时检测.模式识别与人工智能, 2019, 32(3): 193-203.

DOI 10.16451/j.cnki.issn1003-6059.201903001

中图法分类号 TP 181

Real-Time Detection of Underwater Fish Based on Improved YOLO and Transfer Learning

LI Qingzhong¹, LI Yibing¹, NIU Jiong¹

ABSTRACT To fast detect underwater fish in unrestricted underwater environment based on underwater video collected by underwater robots, a real-time detection algorithm for underwater fish based on improved you only look once(YOLO) and transfer learning is proposed. Firstly, an underwater-YOLO for the embedding computer system of underwater robots is designed to overcome the shortcomings of traditional YOLO. Then, transfer learning strategy is employed to train the underwater-YOLO network and alleviate the limitation of known underwater fish samples. A preprocessing algorithm based on contrast limited adaptive histogram equalization is proposed to overcome the problem of underwater image degradation. Finally, a video frame selection method for foreground computation of underwater-YOLO based on structure similarity between inter-frames is proposed to increase the detection frame rate. The experimental results show that the proposed algorithm achieves the goal of real-time detection of underwater fish in unconstrained underwater environment. Compared with the traditional YOLO, the proposed underwater-YOLO generates better detection performance in complex scenes with small fish and overlapped fishes.

Key Words Convolutional Neural Network, Deep Learning, Fish Target Detection, You Only Look Once(YOLO), Transfer Learning

收稿日期: 2018-11-14; 录用日期: 2019-03-13

Manuscript received November 14, 2018;

accepted March 13, 2019

国家重点研发计划项目(No.2017YFC1405202)、国家自然科学基金项目(No.61132005)、海洋公益性行业科研专项(No. 201605002)

Supported by National Key R&D Plan of China(No.2017YFC14

05202), National Natural Science Foundation of China(No. 61132005), National Marine Technology Program for Public Welfare of China(No.201605002)

本文责任编辑 叶东毅

Recommended by Associate Editor YE Dongyi

1.中国海洋大学 工程学院 青岛 266100

1.College of Engineering, Ocean University of China, Qingdao 266100

Citation LI Q Z , LI Y B , NIU J. Real-Time Detection of Underwater Fish Based on Improved YOLO and Transfer Learning. Pattern Recognition and Artificial Intelligence , 2019 , 32(3) : 193 – 203.

深海探测与作业技术是海洋技术研究的重要领域之一。水下机器人(有缆水下机器人(Remote Operated Vehicles , ROVs) /无缆水下机器人(Autonomous Underwater Vehicles , AUVs) 是目前最先进的深海探测与作业装备。要实现深海探测与作业, 水下机器人必须能快速识别理解海底环境, 并对感兴趣的目标进行准确识别与定位^[1-4]。因此, 利用装有水下摄像机的 AUVs 对水下感兴趣目标进行实时搜寻、监测具有重要研究价值和应用前景。本文目标就是研究海底非限制环境下基于视频图像鱼类目标的快速检测算法, 为海底鱼类等生物资源的监测、保护及可持续开发提供技术支持。

在地面非限制环境下, 感兴趣目标的检测算法发展迅速, 在检测精度和实时性方面已接近实用性。其主要原因是基于深度卷积神经网络(Deep Convolutional Neural Network , DCNN) 的目标检测算法可有效克服非限制环境目标检测的困难和瓶颈。其主流方法可分为如下两类。

1) 基于区域卷积神经网络(Region Convolutional Neural Network , R-CNN) 深度学习的目标检测方法。Girshick 等^[5]提出 R-CNN, 由于 R-CNN 在所有候选区域中重复进行特征提取, 影响目标检测的效率, 所以快速区域神经网络(Fast R-CNN) ^[6]采用直接在特征图像上产生候选区域的方式, 避免反复进行特征提取, 提高目标检测效率。超快区域神经网络(Faster R-CNN) ^[7]直接使用候选区域生成网络(Region Proposal Network , RPN) 产生召回率更高的候选区域, 平均准确率更高, 处理速度更快。

2) 基于单级式目标检测算法(You Only Look Once , YOLO) 深度学习目标检测算法。为了克服 R-CNN 的目标检测算法实时性差的问题, Redmon 等^[8]提出 YOLO, 摒弃候选区域生成的中间步骤, 将目标区域预测和目标类别预测整合在单个 DCNN 模型中。在 YOLO 基础上 Redmon 等^[9]提出 YOLOv2 (YOLO Version 2 , YOLOv2) , 采用诸多改进策略, 提高平均准确率和运行速度。在 YOLO 基础上, Redmon 等^[10]又提出 YOLOv3 算法(YOLO Version 3 , YOLOv3) , 在特征提取阶段, 采用浅层次特征和深层次特征的融合, 提取更具鉴别性的深层特征。对于 Microsoft COCO 数据集^[11], 准确率与 SSD 算法(Single Shot MultiBox Detector , SSD) ^[12]相当, 处理

速度为 SSD 的 3 倍。

虽然地面目标的实时检测算法已较完善, 但在水下非限制环境中, 水下感兴趣目标的实时检测仍面临许多挑战, 其主要原因是水下介质、水下光照条件、水下海底环境等都要比地面机器人面临的环境复杂。对于基于水下视觉的鱼类目标检测, 目前代表性的研究方法主要分为如下 3 类。

1) 基于水下图像底层特征的鱼类目标检测算法。Hsiao 等^[13]提出真实水下非限制环境中鱼类目标的检测识别算法, 通过提取水下图像的各种底层特征, 并利用稀疏表达技术设计分类器进行鱼类目标的辨识, 识别率可达 81.8% , 但实时性较差。Cutter 等^[14]利用 Haar-like 特征和多个级联分类器实现海底非限制环境下岩石鱼类目标的检测, 检测率可达 89% , 但对于鱼类非正常姿态、背景复杂、光照条件恶劣、噪声过大、分辨率较低等非限制环境的检测结果并不理想, 误检率较高。

2) 基于 R-CNN 深度学习的鱼类目标检测算法。Seese 等^[15]提出鱼类目标检测分类算法, 首先利用融合技术估计背景图像并实现鱼类前景区域的分割, 然后对于分割的鱼类前景区域, 再利用 DCNN 自动提取特征与分类。Li 等^[16]改进 Faster R-CNN 的结构, 提出适合水下鱼类目标检测的轻型 R-CNN , 虽然检测率可达 89.95% , 但实时性仍欠佳, 仅为 11 帧/秒。

3) 基于 YOLO 深度学习的鱼类目标检测算法。Sung 等^[17]把一般地面目标实时检测的 YOLO 网络结构模型移植到水下鱼类目标的检测, 利用特定的水下视频作为训练、测试数据集, 取得较好的分类精度和实时性, 但缺乏考虑水下机器人的实际硬件限制和水下光照环境的复杂多变性。

通过上述分析可知, 目前对水下非限制环境中基于视觉的鱼类目标的检测主要存在如下三方面问题。1) 由于光在水体介质中存在严重的衰减和散射效应, 造成水下图像质量下降, 为水下鱼类目标的特征学习、识别理解造成严重障碍。2) 由于海底鱼类目标大多具有未知性, 鱼类目标在自动识别学习时, 已知标签的样本数量非常有限, 传统的训练学习方法不再适用。3) 水下机器人研制采用嵌入式计算机系统, 相比一般台式计算机, 硬件运算能力具有较大的限制性。因此, 基于 YOLO 等深度学习的地面目标实

时检测算法不再适合水下机器人的嵌入式系统。

为了克服上述问题,本文提出基于改进 YOLO 和迁移学习的水下鱼类目标实时检测算法。针对 YOLO 网络结构进行改进,提出适合水下机器人嵌入式系统的精简 YOLO 结构(Underwater-YOLO),可以胜任水下鱼类目标的实时检测要求。针对网络训练时样本不足的问题,采用迁移学习的方法训练 Underwater-YOLO 网络。针对水下海底图像散射模糊、对比度较低、光照不均问题,提出基于限制对比度自适应直方图均衡化(Contrast Limited Adaptive Histogram Equalization, CLAHE)的水下图像增强预处理算法。此外,提出基于图像结构相似度(Structural Similarity, SSIM)的视频帧选择性网络前向计算策略,有效提高检测速率。

1 YOLO 实时目标检测算法

YOLO 目标检测算法将目标检测问题转化为一个回归问题,仅使用一个深度卷积神经网络模型进行目标检测,能以较高的准确率实现快速目标检测与识别。

YOLO 的网络结构如图 1 所示。由图 1 可知, YOLO 网络结构主要包括两部分。1) 特征提取。利用一个具有 20 个卷积层的基础网络提取输入图像的特征,可从大小为 $448 \times 448 \times 3$ 的彩色图像中提取 $14 \times 14 \times 1024$ 维特征。2) 回归预测。使用随机初始化权值的 4 层卷积进一步运算特征,产生维度为 $7 \times 7 \times 1024$ 维包含特征,然后使用 4096 个尺寸为 7×7 的卷积核将特征矩阵进行二维卷积操作,产生 4096×1 维全连接特征,通过全连接层产生 539×1 维输出特征,将输出特征重排为 $7 \times 7 \times 11$ 维,作为回归的输出矩阵。

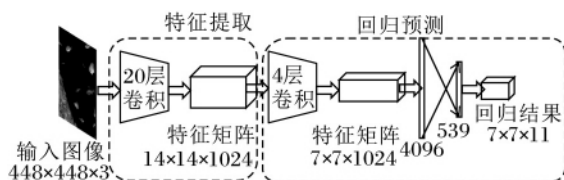


图1 YOLO 网络结构

Fig.1 Network structure of YOLO

回归输出矩阵包含目标边界框的尺寸和坐标、边界框内包含目标的概率及目标所属类别。这些信

息的编码方式如图 2 所示。

图 2 为输出矩阵第 1、2 维度取 1 时,第 3 维度的编码信息,即一个图像块的编码信息。目标检测结果以边界框选中目标的方式呈现,其中, x, y 为边界框中心点相对于网格左上方的偏移量, w, h 为边界框的宽度和高度, C 为边界框内包含目标的置信度,编码结果给出两个边界框的预测值,根据置信度 C 选择合理的边界框, $P(c)$ 为被检测目标属于某一类别的概率,为目标检测结果确定类别。

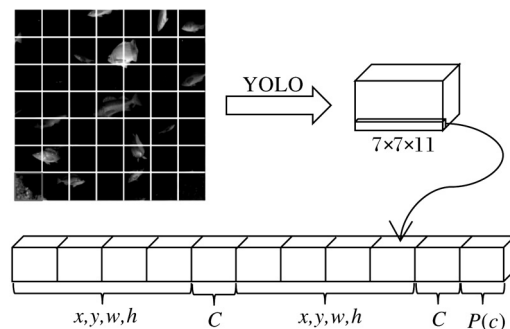


图2 YOLO 输出矩阵信息编码

Fig.2 Coding mode of output matrix information of YOLO

将训练样本的标注信息处理成为输出矩阵的编码顺序,可以建立理想输出与实际输出的损失函数。

坐标预测的损失函数如下:

$$loss_{coord} = \sum_{i=0}^{72} \sum_{j=0}^2 1_{ij}^{obj} \left[\left(x_i - \hat{x}_i \right)^2 + \left(y_i - \hat{y}_i \right)^2 \right] + \sum_{i=0}^{72} \sum_{j=0}^2 1_{ij}^{obj} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right]. \quad (1)$$

预测网格框包含目标时的损失函数:

$$loss_{obj} = \sum_{i=0}^{72} \sum_{j=0}^2 1_{ij}^{obj} \left[\left(C_i - \hat{C}_i \right)^2 \right].$$

预测网格框不包含目标时的目标函数:

$$loss_{noobj} = \sum_{i=0}^{72} \sum_{j=0}^2 1_{ij}^{noobj} \left[\left(C_i - \hat{C}_i \right)^2 \right].$$

预测目标类别的损失函数:

$$loss_{class} = \sum_{i=0}^{72} 1_i^{obj} \sum_{c \in class} \left[p_i(c) - \hat{p}_i(c) \right]^2.$$

综合部分的损失函数求得总损失函数:

$$loss = \lambda_{coord} \cdot loss_{coord} + loss_{obj} + \lambda_{noobj} \cdot loss_{noobj} + loss_{class}.$$

以损失函数作为卷积神经网络的优化准则,利用随机梯度下降法进行优化,即可完成 YOLO 网络的训练。

2 基于改进 YOLO 的水下鱼类目标检测算法

YOLO 目标检测算法不能直接应用于水下图像鱼类目标的检测,主要原因体现在如下 3 方面.

1) 在结构上, YOLO 网络结构过于复杂, 前向推理计算时需要较高的计算能力. 而水下机器人搭载的嵌入式处理系统计算能力有限, 所以无法在嵌入式系统上运行 YOLO. 此外, 由于 YOLO 仅在单一尺度下进行目标检测, 难以检测位置重叠、体态较小的鱼类目标.

2) 海底鱼类样本相对匮乏, 已知标签的样本数量非常有限, 传统训练方法无法有效充分训练泛化性能较高的深度神经网络, 必须根据水下图像的特点提出能使用小样本充分训练网络的方法.

3) 由于海底视频图像呈现对比度较低、纹理细节模糊及光照不均匀等现象, 所以 YOLO 无法适应海底观测视频图像.

为了克服上述问题, 本文提出基于改进 YOLO 和迁移学习的鱼类目标实时检测算法, 算法总体框架如图 3 所示.

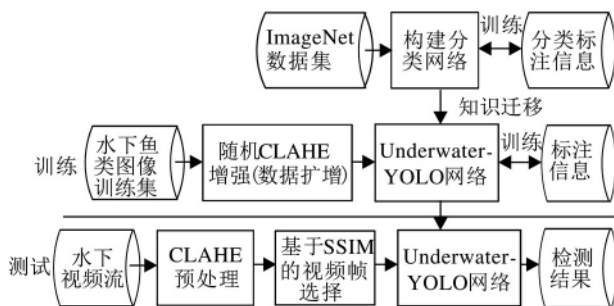


图 3 本文算法总体框架

Fig.3 Framework of the proposed method.

本文算法主要包括如下 4 个模块: Underwater-YOLO 网络结构设计、Underwater-YOLO 网络的迁移学习、基于 CLAHE 的水下图像增强预处理、基于 SSIM 的视频帧选择性前向推理计算. 下面具体介绍各模块的实现过程.

2.1 Underwater-YOLO 网络结构设计

为了既能适合水下机器人的嵌入式系统的计算能力, 又能实现对水下鱼类小目标、重叠目标的准确检测, 本文设计适合水下机器人进行鱼类目标检测的网络结构(Underwater-YOLO), 如图 4 所示.

由图 4 可知, Underwater-YOLO 主要包括特征提取和预测回归两部分. 特征提取部分包括 7 个卷

积层, 每个卷积层框图中的数据分别表示卷积核数量、卷积核尺寸、卷积步长(省略者默认步长为 1). Underwater-YOLO 的特征提取部分仅使用 7 层卷积, 并以步长为 2 的卷积运算进行下采样, 如此设计的目的是为了减少水下机器人嵌入式系统进行目标检测时的运算量. 而在回归预测部分, 本文舍弃全连接网络结构, 整个网络使用全卷积神经网络(Fully Convolutional Network, FCN) 的结构, 克服在回归过程中将空间特征转化为二维特征丢失空间信息的弊端^[18].

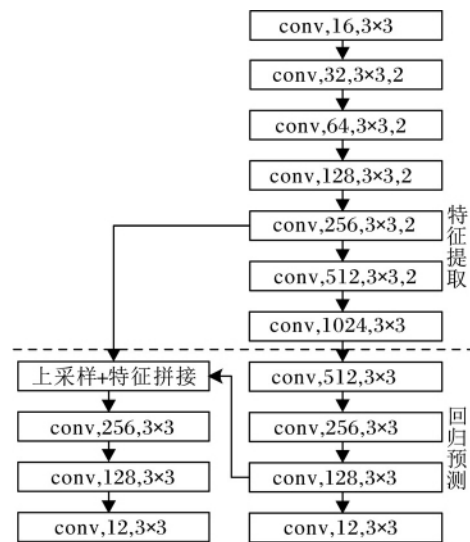


图 4 Underwater-YOLO 的网络结构

Fig.4 Network structure of Underwater-YOLO

为了实现水下鱼类小目标的检测, 在回归预测部分采用特征拼接实现多尺度预测方式. 对于深层特征, 由于下采样使特征尺寸减小, 导致越深层次的特征尺寸越小, 因此深层特征中小目标的信息会因为下采样而丢失. 为了解决仅使用深层特征进行目标检测导致微小目标被忽略的缺陷, 采用多尺度的检测方式, 即联合浅层特征和深层特征, 提高小目标的检测率. 图像特征在 Underwater-YOLO 网络结构中的传递流程如图 5 所示, 该过程可直观理解为多尺度回归. 图中每个子框中的数据分别表示为(特征宽度 特征高度 特征通道数).

从图 5 可看出, 输入图像在 Underwater-YOLO 的特征提取部分逐层提炼, 形成尺寸小、维数大的深层抽象特征. 深层特征在回归预测部分继续进行卷积操作, 最终产生粗尺度预测矩阵. 由于特征提取部分伴随下采样, 一些特征的细节信息丢失. 在预测回归阶段, 将深层特征进行上采样, 使之与浅层特征尺寸匹配, 然后拼接不同层的尺寸, 生成融合特征. 使

用融合特征继续进行回归预测,最终产生细尺度预测矩阵。

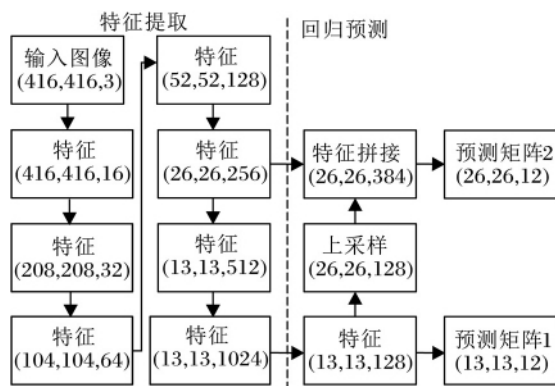


图5 Underwater-YOLO 的特征流程图

Fig.5 Feature flow chart of underwater-YOLO

Underwater-YOLO 的输入为 3 个通道的彩色图像,通过 Underwater-YOLO 网络后产生 $13 \times 13 \times 12$ 和 $26 \times 26 \times 12$ 两个尺度的预测矩阵。其中 $13 \times 13 \times 12$ 的预测矩阵编码将图像划分为 13×13 个网格,每个网格内包含鱼类目标的置信度及边界框的坐标值。而 $26 \times 26 \times 12$ 的预测矩阵编码将图像划分为 26×26 个网格,每个网格内包含鱼类目标的置信度和边界框的坐标信息。

对于重叠目标的检测问题,YOLO 在图像的每个子块中最多只能检测到一个目标,若在一个子块中同时出现多个目标,会造成重叠目标的漏检。为此,对输出的预测结果矩阵进行重新编码。借鉴 Faster R-CNN 中锚点的思想,为一个图像子块设置多个锚点,每个锚点可编码目标的边界框的坐标值、存在目标的置信度及类别。由于锚点被初始化为不同尺寸,发生重叠的目标可以被不同的锚点捕获。预测矩阵中信息的编码方式如图 6 所示。

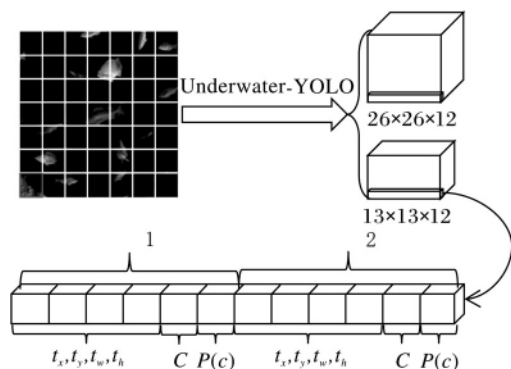


图6 Underwater-YOLO 预测矩阵编码方式

Fig.6 Encoding mode of prediction matrix information of Underwater-YOLO

2 个预测矩阵的第 3 维长度一致,编码相同的信息。 t_x, t_y 为边界框偏离图像子块左上顶点的位置信息。 t_w, t_h 为确定边界框宽度和高度的参数。上述 4 个参数可解决边界框在图像中的定位问题。 C 为边界框中是否包含目标的置信度。 $P(c)$ 为边界框内目标的类别判断,判断目标属于鱼类的概率。上述所有参数可使用一个锚点进行描述,通过引入锚点,增加目标检测的预设候选区域,有利于解决重叠目标的漏检问题。图像中图像块、锚点与边界框分布及尺寸关系如图 7 所示。

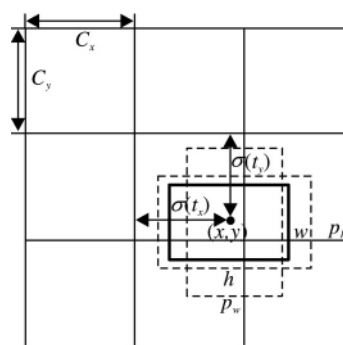


图7 锚点框与边界框的位置关系

Fig.7 Positional relationship between anchor box and bounding box

图 7 中 c_x, c_y 为图像子块对于图像左上角的偏移量。 $\sigma(t_x), \sigma(t_y)$ 为边界框中心点相对于图像子块的偏移量按照图像子块的尺寸归一化的结果。 p_w, p_h 为锚点框的宽度和高度。 (x, y) 为边界框的中心点坐标。 w, h 为边界框的宽度和高度。由 x, y, w, h 可确定边界框在图中的位置,可通过预测矩阵中的偏移参数计算求得:

$$x = \sigma(t_x) + C_x, y = \sigma(t_y) + C_y,$$

$$w = p_w \exp(t_w), h = p_h \exp(t_h).$$

锚点框的初始尺寸与指数函数相乘求得边界框的尺寸,当为锚点指定不同的初始尺寸时会获得不同尺寸的边界框。因此,可通过设置多个不同尺寸的锚点框的方式改善目标检测的图像子块中检测多个目标的效果。

通过使用多尺度预测与多锚点的编码方式,一幅图像能产生更多的待检测区域,在该网络结构中共产生

$$(13 \times 13 + 26 \times 26) \times 2 = 5070 \text{ 个}$$

检测区域,包含 4 种不同尺寸的锚点框。通过这一改进能极大增加小目标检测的召回率,并且能检测重叠的目标。

2.2 Underwater-YOLO 的迁移学习

为了训练 Underwater-YOLO,通过采集水下鱼类图像,以人工标注的方式建立水下鱼类目标检测数据集.数据集共有 1 500 幅 ROV 在海底环境下的实拍图像,其中 500 幅图像挑选自 labeled fishes in the wild 数据集^[14],1 000 幅图像来源于实拍水下视频中的关键帧.将图像数据集样本分为训练集和验证集两部分,训练集共有 1 000 个样本,验证集共有 500 个样本.

由于已标注的水下鱼类检测数据较少,直接用于训练 Underwater-YOLO 难以获得具有泛化性的训练结果.特别是特征提取部分的卷积层需要充分训练,才能从图像提取出具有鉴别性的特征.为此,本文采用迁移学习^[19]的方法训练卷积神经网络,Yosinski 等^[20]论证在迁移学习中特征迁移的有效性.在迁移学习中,将本文标注的 1 000 幅水下图像样本集作为目标域,将 ImageNet 分类数据集^[21]作为源域,使用微调(Finetune)的方式迁移学习通过分类网络训练的模型.

本文的迁移学习分为两部分:分类网络训练;将分类网络的知识迁移到检测模型,在此基础上进行 Underwater-YOLO 的训练.

对于分类网络部分,首先根据 Underwater-YOLO 网络结构建立分类网络结构.如图 8 所示,以 Underwater-YOLO 特征提取部分的卷积层为基础,通过添加一层卷积核数量为 1 000 的 13×13 的卷积层,将特征矩阵转化为 $1\,000 \times 1$ 维特征向量,然后再加一层 softmax 层将特征向量转化为概率输出.

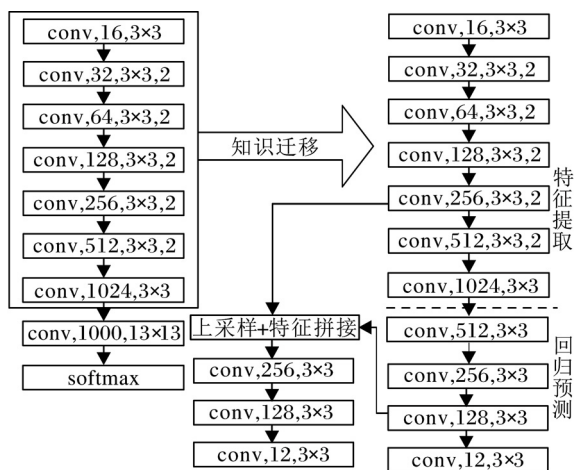


图 8 Underwater-YOLO 的知识迁移过程

Fig.8 Knowledge transferring of Underwater-YOLO

建立分类网络后,使用 ImageNet 数据集进行训

练.网络训练超参数选择如下:批次大小(Batch Size)为 64,输入图像高度为 416,输入图像宽度为 416,动量为 0.9,权值衰减为 0.000 1,最大迭代次数为 100 000.学习率调整策略如下:第 1 ~ 10 000 次迭代设置为 0.1,第 10 000 ~ 80 000 次迭代设置为 0.01,第 80 000 ~ 100 000 次迭代设置为 0.001,微调,由此减小权值在极小值附近的震荡.

训练完分类网络后,可将分类网络训练的知识迁移到 Underwater-YOLO 特征提取部分,迁移过程如图 8 所示.

使用 ImageNet 数据集训练完分类网络后,卷积核的权值经过充分训练,能提取泛化特征.这些卷积核就可作为分类网络学习到的知识.将这些知识迁移到 Underwater-YOLO 的特征提取部分,设置特征提取部分的学习率为 0,锁定特征提取部分的卷积核.回归预测部分的卷积核通过随机初始化的方式赋值,然后使用水下图像样本集训练回归预测部分的卷积核.

由于水下图像训练集数量较少,本文提出结合限制对比度自适应直方图均衡化(CLAHE)^[21]的迁移学习训练方法.在训练回归检测网络阶段,同时使用数据扩增的手段,使小样本训练集发挥最大效果.采用的数据扩增手段包括:1) 随机旋转图像.将原始图像随机旋转 $-15^\circ \sim 15^\circ$.2) 调整曝光度和饱和度.将图像从 RGB 颜色空间转化到 HSV 空间,对曝光度和饱和度随机调整至原先的 $1/1.5 \sim 1.5$ 倍.3) 随机调整亮度分量,使亮度增加或减少原先的 $-0.1 \sim 0.1$ 倍.4) 随机调节限制对比度自适应直方图均衡化的裁剪阈值,获得不同直方图分布规律的训练样本,提升训练结果的泛化性.

此外,由于锚点框的尺寸也会影响目标检测的效果,若能为锚点框初始化合理的尺寸,则能有效提高目标检测的指标.本文使用聚类方法,以训练样本中已标注的边界框的尺寸为参考,确定锚点框的尺寸.选择 k 个聚类中心,以边界框和聚类尺寸的 IOU 指标确定距离函数:

$$d(\text{bbox } k\text{means}) = 1 - \text{IOU}(\text{bbox } k\text{means}).$$

以该距离函数作为优化准则进行聚类,针对水下鱼类目标检测数据集,获得与训练样本中边界框重合率最大的锚点尺寸.按照聚类结果初始化锚点尺寸,提高预测边界框与实际目标的重合率.

2.3 基于 CLAHE 的水下图像增强预处理

对于水下图像,由于光线的严重衰减,图像整体对比度较低,噪声较大,存在光照不均匀的现象.为此,本文采用限制对比度直方图均衡化(CLAHE)

方法进行水下图像的增强处理。

限制对比度自适应直方图均衡化图像增强预处理算法步骤如下。

算法 CLAHE

step 1 将输入图像从 RGB 空间转换到 HSV 颜色空间。

step 2 将亮度分量 V 图像划分为 k 个大小为 $m \times n$ 的子块, 每个子块相互连续, 互不重叠。

step 3 计算子区域中每个灰度级可以平均分配的像素个数

$$N_{ave} = \frac{N_{CR-x} N_{CR-y}}{N_{gray}},$$

其中 N_{gray} 为子区域中灰度级的数量, N_{CR-x} 为子区域 x 轴方向的像素数, N_{CR-y} 为 y 轴方向的像素数。限制每个灰度级包含的像素数不超过平均值 N_{ave} 的 N_{clip} 倍, 则实际剪切极限值

$$N_{CL} = N_{clip} \cdot N_{ave},$$

其中 N_{clip} 为截取限制系数, 含义是限制子区域每个灰度级包含的像素数不超过平均像素数的 N_{clip} 倍。

step 4 对每个子块, 按照裁剪极限值 N_{CL} 剪切灰度直方图, 多余的像素数量重新分配到各个灰度级中, 设已被剪切的像素总数为 N_{sum_clip} , 可得每个灰度级均分的剪切像素数 N_{add} ,

$$N_{add} = \frac{N_{sum_clip}}{N_{gray}},$$

重新分配的过程如下:

$$H(i) = \begin{cases} N_{CL} + N_{add} & H(i) > N_{CL} \\ H(i) + N_{add} & H(i) \leq N_{CL} \end{cases}$$

其中 $H(i)$ 为原始区域中第 i 个灰度级的像素数。

step 5 重复 step 3 的分配过程, 剩余的像素数为 $N_{sum_clip} \leq N_{gray}$, 被分配像素步长值

$$S = \frac{N_{gray}}{N_{sum_clip}},$$

平均分配到每个灰度级。

step 6 对每个子块剪切后的灰度直方图进行直方图均衡化。

step 7 使用双线性插值的方法去除图像分块处理后产生的块状效应。

step 8 将增强预处理后的 V 分量图像和原先的 H 、 S 分量一起转换到 RGB 颜色空间, 得到增强后水下图像。

2.4 基于结构相似度的视频帧选择性前向推理计算

在利用水下机器人视觉系统进行水下鱼类目标检测时, 若利用 Underwater-YOLO 网络对视频的每

帧都进行前向计算检测, 由于帧速一般为 30 帧/秒, 则水下机器人的嵌入式系统的计算能力难以承受。考虑到视频相邻帧间存在大量信息冗余, 本文提出基于帧间图像结构相似度 (SSIM) 的视频帧选择方法, 即只对关键帧进行 Underwater-YOLO 网络的前向检测计算, 相似的多余帧不需要重复进行前向检测计算。

SSIM 为一种基于结构信息衡量图像相似程度的评价准则。结构相似度将图像的相似度建模为亮度 (l)、对比度 (c) 和结构 (s) 的组合指标。令 x 、 y 表示相邻的帧间图像, 两者之间的亮度相似指标使用均值 (μ_x 、 μ_y) 进行估计, 对比度相似指标使用标准差 (σ_x 、 σ_y) 进行估计, 结构相似指标使用协方差 (σ_{xy}) 进行估计。

SSIM 可使用上述 3 个指标的组合, 定义如下:

$$SSIM(x, y) = [l(x, y)]^\alpha [c(x, y)]^\beta [s(x, y)]^\gamma,$$

其中

$$l(x, y) = \frac{2\mu_x \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1},$$

$$c(x, y) = \frac{2\sigma_x \sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2},$$

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3},$$

$\alpha \geq 0$ $\beta \geq 0$ $\gamma \geq 0$, 分别用于调节亮度、对比度、结构相似度的权重。当取 $\alpha = \beta = \gamma = 1$, $C_3 = C_2/2$, 可得

$$SSIM(x, y) = \frac{(2\mu_x \mu_y + C_1)(2\sigma_x \sigma_y + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}.$$

通过计算视频相邻帧之间的结构相似度, 可判断视频内容的变化幅度。当画面几乎不发生变化时, 相邻帧保持较高的结构相似度, 即 $SSIM(x, y) \approx 1$, 此时视频内容基本固定, 不需要进行深度学习的前向推理。当画面变化累积到一定程度, 视频帧的结构相似度变小, 此时需要进行深度学习前向推理, 更新目标检测的结果。

视频帧选择性前向推理计算过程如图 9 所示。

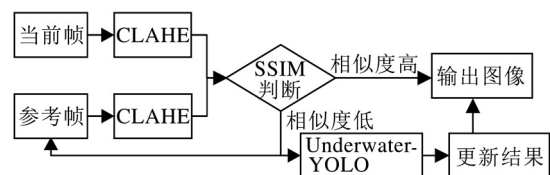


图9 视频帧选择性前向推理计算过程

Fig.9 Forward inference computation based on key frame selection

以视频的第一帧图像进行卷积神经网络的前向运算并作为参考帧,与视频的下一帧进行对比,计算两帧之间的结构相似度.设定结构相似度阈值,若两帧结构相似度大于阈值,不进行前向推理,直接输出图像并附加前一帧的检测结果.若两帧的结构相似度小于等于规定的阈值,对当前帧进行 Underwater-YOLO 网络的前向推理,并将当前帧赋值为新的参考帧.

3 实验及结果分析

为了验证 Underwater-YOLO 网络的有效性,基于 Underwater-YOLO 网络进行水下鱼类目标检测的可行性,将 YOLO 网络和 Underwater-YOLO 应用于水下鱼类检测.实验采用的数据集为手工标注的 ROV 实拍水下图像.服务器的配置参数如下: intel i5CPU、64 GB 内存、GTX1080ti GPU(11 GB 显存)、Ubuntu16.4 操作系统.

实验主要分为 3 部分: 1) 测试 Underwater-YOLO 与 YOLO 基础网络进行鱼类目标检测的性能,验证 Underwater-YOLO 网络结构的有效性.2) 使用结合限制对比度自适应直方图均衡化的迁移学习方法及锚点维度聚类的方法训练深度神经网络,讨论本文提出的训练策略的有效性.3) 使用本文提出的基于 SSIM 的选择性前向推理的方法进行视频目标检测,验证选择性前向推理方法在实时鱼类检测过程中的有效性.

3.1 Underwater-YOLO 检测结果

本部分实验旨在验证 Underwater-YOLO 网络结构的有效性,因此在训练阶段仅采用迁移学习加数据扩增的策略以相同的超参数训练 YOLO、YOLOv2、YOLOv3 及 Underwater-YOLO 网络.将 500 幅测试图像输入训练好的网络进行鱼类目标的位置回归.当神经网络预测的目标边界框与手工标注的边界框交并化(Intersection Over Union, IOU) 大于等于 0.5 时,认为成功检测目标;否则,目标漏检.选择准确率(P)、召回率(R)、平均交并比(Mean Intersection over Union, mIOU)、每秒检测帧数(Frames per Second, FPS) 作为评价准则.

各种目标检测算法在水下目标检测的性能如表 1 所示.

文献[14]为基于 Haar-like 特征的 Adaboost 水下图像鱼类检测方法,使用相似的鱼类检测数据集,检测准确率为 66%,对于复杂的水下环境有较差的

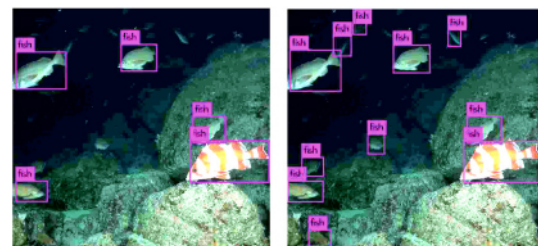
鲁棒性,因此使用传统方式的鱼类检测准确率受到限制.相比 Adaboost 目标检测方法,基于 YOLO 的目标检测算法具有较好的性能指标.使用最基本的 YOLO 网络结构,目标检测的准确率比 Adaboost 目标检测方法提高 10%.而本文建立的 Underwater-YOLO 在准确率和召回率上都高于 YOLO 基础网络,与 YOLOv2 相当.由于 YOLOv3 具有深层的网络结构,使用多种改进策略,目标检测的准确率和召回率偏高.但 Underwater-YOLO 仅使用 7 层卷积神经网络进行特征提取,大幅降低目标检测的运算量,检测速度可达 122 帧/秒,远高于 YOLO、YOLOv2、YOLOv3.

表 1 水下目标检测实验结果

Table 1 Experimental results of underwater object detection

方法	准确率 /%	召回率 /%	平均 交并比	检测速度 /帧/秒
Adaboost	66	—	—	—
YOLO	76	62	0.62	32
YOLOv2	83	74	0.71	41
YOLOv3	92	79	0.78	40
Underwater-YOLO	89	73	0.66	122

使用基本的 YOLO 网络和 Underwater-YOLO 进行水下图像测试,典型结果对比如图 10 所示.



(a) YOLO

(b) Underwater-YOLO

图 10 2 种网络结构检测结果对比

Fig.10 Comparison of detection results of 2 networks

图 10(a) 为 YOLO 网络使用迁移学习的策略进行训练后检测鱼类的结果,(b) 为 Underwater-YOLO 网络经过相同的训练过程后的检测结果.对比两者可以看出,使用 YOLO 网络结构进行鱼类目标检测时,仅能检测较明显的目标,漏检体态较小、光线较暗处的目标.而使用 Underwater-YOLO 网络进行目标检测时,能够检测一些不明显的鱼类,大幅提升目标检测的召回率.

实验结果表明,在水下鱼类检测中,Underwater-

YOLO 能以较少的卷积层实现目标检测,不仅具有较好的准确率和召回率,还大幅提升目标检测的运算速度.网络中的特征融合、多尺度检测及使用锚点框编码输出信息可以有效提升网络性能.

3.2 训练策略对神经网络性能的影响

在训练 Underwater-YOLO 过程中本文采用多种训练策略(A+B+C),本节实验验证各种训练策略对目标检测性能指标的影响.对 Underwater-YOLO 使用不同的训练策略进行训练,通过测试集可获得不同训练策略下目标检测性能指标,如表 2 所示.表 2 中, A 表示迁移学习, B 表示基于 CLAHE 的图像预处理, C 表示使用维度聚类方法确定锚点的初始尺寸.

表 2 Underwater-YOLO 使用不同训练策略的网络性能对比
Table 2 Performance comparison of Underwater-YOLO using different training strategies

策略	准确率/%	召回率/%	平均交并比
Underwater-YOLO	79	66	0.58
策略 A	89	73	0.66
策略 B	81	80	0.64
策略 C	80	69	0.71
策略 A+B	93	83	0.68
策略 A+B+C	93	84	0.72

由表 2 可得,直接使用水下图像训练集进行网络训练,获得的结果缺乏泛化性,当使用测试集进行测试时,准确率和召回率很低.策略 A 充分训练卷积神经网络部分的卷积核,在测试过程中准确率得到明显提升.策略 B 能克服光照不均匀情况下暗区目标难以检测的缺陷,召回率明显提升.策略 C 使预测边界框更接近目标位置, mIOU 得到明显提升.而策略 A+B 为本文提出的结合 CLAHE 的迁移学习方法,由表可知,准确率和召回率都得到提升,高于单独使用一种策略.综上所述,综合使用 A+B+C 训练策略,准确率可达到 93%,召回率可达到 84%, mIOU 可达到 0.72.相比直接进行目标检测,准确率提升 14%,召回率提升 18%, mIOU 提升 14%.

由实验结果可知,通过使用多种策略的组合,提升最大的性能指标是召回率.目标检测中的召回率是指被正确检测的鱼类目标占图像中所有应被检测出鱼类的比率.若直接训练 Underwater-YOLO 而不使用本文的改进措施,召回率仅有 66%,这意味着有大量的目标被漏检.而引入本文的改进策略,召回率提升至 84%.

图 11 为使用普通迁移学习训练 Underwater-

YOLO 与使用本文训练策略进行网络训练的典型检测结果对比.

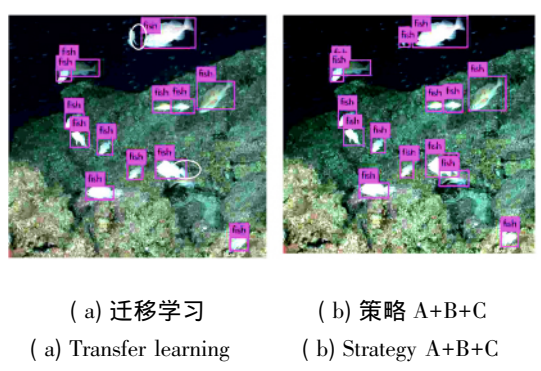


图 11 2 种训练策略检测结果对比

Fig.11 Comparison of detection results using different training strategies

图 11(a) 为 Underwater-YOLO 网络使用普通迁移学习策略训练后的检测结果,(b) 为 Underwater-YOLO 网络使用本文的结合 CLAHE 的迁移学习的策略及使用锚点框维度聚类方法进行训练的检测结果.由检测结果可看出,两者都具有检测重叠目标的能力,这是由于网络结构中使用锚点的编码方式与多尺度预测的结果.但(a) 中仍存在目标漏检的情况,漏检的目标使用椭圆边框标注.上方未被检测的原因是鱼类目标体态特殊,目标较小;下方未被检测的原因是目标发生严重重叠,不易观察.使用本文的训练策略,由(b) 的结果可看出,本文方法可克服在鱼类体态不正常和严重重叠状态下目标漏检的问题,正如表 2 所示,由于漏检率下降从而有效提升本文方法的召回率.

综上所述,迁移学习可提升目标检测的准确率,基于 CLAHE 的图像预处理可提升目标检测的召回率,基于维度聚类的锚点框初始化可提升目标检测的 mIOU 及召回率,综合各种策略可发挥其优势,提高整体性能指标.

3.3 选择性前向推理对目标检测速度的影响

为了验证基于 SSIM 的选择性前向推理视频加速检测的有效性,在 GTX1080ti GPU 上统计视频目标检测的平均速度,结果如表 3 所示.

由表 3 可知,选择性前向推理的视频加速算法在 GPU 平台上检测速度有所提升,但提升幅度并不理想,主要原因是在 GPU 上运行卷积神经网络单次前向推理的耗时较少,并且引入 SSIM 的计算过程仍由 CPU 完成,检测速度难以得到很大提升.

然而水下机器人难以配备高能耗的 GPU,主要运算平台为嵌入式处理器.由于电脑 CPU 与嵌入式

处理器的性能相当,本文统计在 Intel i5CPU 平台上选择性前向推理视频检测的处理速度,平均处理速度如表 4 所示。

表 3 在 GTX1080ti GPU 平台上视频目标检测处理速度

Table 3 Speed of video object detection on GTX1080ti GPU platform

方法	帧/秒	
	全帧运算	选择帧运算
Faster R-CNN ^[22]	9.8	—
YOLOv3	16	22
Underwater-YOLO	122	140

表 4 在 Intel i5CPU 平台上视频目标检测处理速度

Table 4 Speed of video object detection on Intel i5 CPU platform

方法	帧/秒	
	全帧运算	选择帧运算
YOLOv3	0.2	2
Underwater-YOLO	4	28

由表 4 可得,在 CPU 平台上选择性前向推理过程可显著提升目标检测的速度。对于 YOLOv3 网络,由于网络结构较复杂,进行全帧运算仅能达到 0.2 帧/秒。综上所述,通过使用选择性前向推理的视频检测方法可有效减少深度神经网络的推理次数,提高目标检测效率。Underwater-YOLO 网络通过使用选择性前向推理视频加速方法,可将目标检测的速度提升到 28 帧/秒,实现在 CPU 平台上的实时检测。

4 结 束 语

本文利用深度学习的方法,提出基于改进 YOLO 与迁移学习的水下鱼类目标实时检测算法。构建 Underwater-YOLO 网络结构,使用特征融合的方法进行多尺度的目标检测,可检测出位置重叠和体态较小的鱼类目标。在算法的实时性方面,通过减少网络结构的卷积层数与卷积核的数量,并且通过计算帧间结构相似度,选择性地更新深度网络的前向推理,实现在水下机器人嵌入式系统上的水下鱼类目标实时检测。提出的迁移学习方法能够获得训练出具有较强泛化性能的网络模型。提出的基于限制对比度自适应直方图均衡化预处理算法可以去除水下图像的散射模糊现象,并且克服光照不均匀问题。实验表明,在水下视频测试集上,鱼类目标检测

速度在 CPU 平台上达 28 帧/秒,准确率达 93%,召回率达 84%。深海生物资源的识别与统计是一个迫切需要解决的问题,通过应用本文的水下鱼类目标实时检测算法,可以使水下机器人更好地服务于深海生物资源的探索与开发。

参 考 文 献

- [1] QIAO X, BAO J H, ZENG L H, *et al.* An Automatic Active Contour Method for Sea Cucumber Segmentation in Natural Underwater Environments. *Computers and Electronics in Agriculture*, 2017, 135: 134-142.
- [2] LI Q Z, ZHANG Y, ZANG F N. Fast Multicamera Video Stitching for Underwater Wide Field-of-View Observation. *Journal of Electronic Imaging*, 2014, 23(2): 367-368.
- [3] BONIN-FONT F, OLIVER G, WIRTH S, *et al.* Visual Sensing for Autonomous Underwater Exploration and Intervention Tasks. *Ocean Engineering*, 2015, 93(1): 25-44.
- [4] MAHMOOD A, BENNAMOUN M, AN S J, *et al.* Deep Image Representations for Coral Image Classification. *IEEE Journal of Oceanic Engineering*, 2019, 44(1): 121-131.
- [5] GIRSHICK R, DONAHUE J, DARRELL T, *et al.* Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation // *Proc of the IEEE Conference on Computer Vision and Pattern Recognition*. Washington, USA: IEEE, 2014: 580-587.
- [6] GIRSHICK R. Fast R-CNN // *Proc of the IEEE International Conference on Computer Vision*. Washington, USA: IEEE, 2015: 1440-1448.
- [7] REN S Q, HE K M, GIRSHICK R, *et al.* Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 6(1): 1137-1149.
- [8] REDMON J, DIVVALA S, GIRSHICK R, *et al.* You Only Look Once: Unified, Real-Time Object Detection // *Proc of the IEEE Conference on Computer Vision and Pattern Recognition*. Washington, USA: IEEE, 2016: 779-788.
- [9] REDMON J, FARHADI A. YOLO9000: Better, Faster, Stronger // *Proc of the IEEE Conference on Computer Vision and Pattern Recognition*. Washington, USA: IEEE, 2017: 6517-6525.
- [10] REDMON J, FARHADI A. YOLOv3: An Incremental Improvement [C/OL]. [2018-10-11]. <https://arxiv.org/pdf/1804.02767.pdf>.
- [11] LIN T Y, MAIRE M, BELONGIE S, *et al.* Microsoft COCO: Common Objects in Context // *Proc of the European Conference on Computer Vision*. Berlin, Germany: Springer, 2014: 740-755.
- [12] LIU W, ANGUELOV D, ERHAN D, *et al.* SSD: Single Shot Multibox Detector // *Proc of the European Conference on Computer Vision*. Berlin, Germany: Springer, 2016: 21-37.
- [13] HSIAO Y H, CHEN C C, LIN S I, *et al.* Real-World Underwater Fish Recognition and Identification, Using Sparse Representation. *Ecological Informatics*, 2014, 23: 13-21.
- [14] CUTTER G, STIERHOFF K, ZENG J M. Automated Detection of Rockfish in Unconstrained Underwater Videos Using Haar Cascades

- and a New Image Dataset: Labeled Fishes in the Wild // Proc of the IEEE Winter Applications and Computer Vision Workshops. Washington, USA: IEEE, 2015: 57-62.
- [15] SEESE N, MYERS A, SMITH K, *et al.* Adaptive Foreground Extraction for Deep Fish Classification // Proc of the 2nd Workshop on Computer Vision for Analysis of Underwater Imagery. Washington, USA: IEEE, 2016: 19-24.
- [16] LI X, TANG Y H, GAO T W. Deep But Lightweight Neural Networks for Fish Detection // Proc of the OCEANS 2017. Washington, USA: IEEE, 2017. DOI: 10.1109/OCEANSE.2017.8084961.
- [17] SUNG M, YU S C, GIRDHAR Y. Vision Based Real-Time Fish Detection Using Convolutional Neural Network // Proc of the OCEANS 2017. Washington, USA: IEEE, 2017. DOI: 10.1109/OCEANSE.2017.8084889.
- [18] LONG J, SHELHAMER E, DARRELL T. Fully Convolutional Networks for Semantic Segmentation // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2015: 3431-3440.
- [19] PAN S J, YANG Q. A Survey on Transfer Learning. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345-1359.
- [20] YOSINSKI J, CLUNE J, BENGIO Y, *et al.* How Transferable Are Features in Deep Neural Networks? // Proc of the 27th International Conference on Neural Information Processing Systems. Cambridge, USA: The MIT Press, 2014: 3320-3328.
- [21] RUSSAKOVSKY O, DENG J, SU H, *et al.* Imagenet Large Scale Visual Recognition Challenge. International Journal of Computer Vision, 2015, 115(3): 211-252.

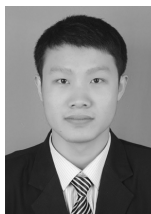
- [22] LI X, SHANG M, HAO J, *et al.* Accelerating Fish Detection and Recognition by Sharing CNNs with Objectness Learning // Proc of the OCEANS 2016. Washington, USA: IEEE, 2016. DOI: 10.1109/OCEANSAP.2016.7485476.

作者简介



李庆忠 博士 教授, 主要研究方向为图像处理、信号处理、模式识别. E-mail: liqingzhong@ouc.edu.cn.

(LI Qingzhong, Ph.D., professor. His research interests include image processing, signal processing and pattern recognition.)



李宜兵(通讯作者) 硕士研究生, 主要研究方向为智能信息处理、模式识别. E-mail: liyibing@stu.ouc.edu.cn.

(LI Yibing (Corresponding author), master student. His research interests include intelligent information processing and pattern recognition.)



牛炯 博士研究生, 工程师, 主要研究方向为高频雷达信号处理、高频雷达海洋环境监测技术. E-mail: 459258810@qq.com.

(NIU Jiong, Ph.D. candidate, engineer. His research interests include high frequency radar signal processing and high frequency radar marine environment monitoring technology.)

关于召开 2019 国家机器人发展论坛暨 Robocup 机器人世界杯中国赛、2019 年中国自动化产业年会的通知

一、活动背景

为贯彻落实好《中国制造 2025》将机器人作为重点发展领域的总体部署,推进我国机器人产业快速健康可持续发展,中国自动化学会在过往四届论坛的基础上,将于 2019 年 4 月 17 日至 21 日在绍兴组织召开 2019 国家机器人发展论坛暨 Robocup 机器人世界杯中国赛、2019 年中国自动化产业年会,同期将开展科技成果展示,聚焦多种类智能机器人的发展、应用与创新,壮大中国机器人事业,为实现中国制造强国梦而作出贡献。

二、组织机构

主办单位 中国自动化学会、中共绍兴市委组织部、中共绍兴市柯桥区委、绍兴市柯桥区人民政府

联合主办单位 中国人工智能产业发展联盟(AIIA)

承办单位 中共绍兴市柯桥区委组织部、浙江千人计划绍兴产业园(绍兴金柯桥科技城)建设管理委员会

协办单位 教育部高等学校自动化类专业教指委

支持单位 中国自动化学会智能自动化专业委员会