

# 基于 YOLOv3 的可变时间窗自校正船只跟踪与计数<sup>①</sup>



刘 春, 栗 健

(湖北工业大学 计算机学院, 武汉 430068)

通讯作者: 刘 春, E-mail: liuchun.hust@gmail.com

**摘 要:** 能够自动识别、统计航道上的船只类型与数量, 对建设“智慧航道”、水上智能预警、通航辅助决策等具有重要意义. 通过使用 YOLOv3 预训练模型, 对船只样本图片进行训练, 调参优化得到航道中船只检测模型, 然后利用深度学习模型善于进行目标特征提取的特点, 结合目标 HSV 颜色直方特征和 LBP 局部特征来实现目标选择, 针对跟踪目标容易出现的漂移和抖动问题, 设计校正网络融合使用了基于回归的方向判断和可变时间窗的目标计数方法, 较好地实现了水上运动目标的自动检测、跟踪和自校正计数. 测试表明本文方法稳定健壮, 适用于自动分析航道视频, 提取统计数据.

**关键词:** YOLOv3; 船只识别; 目标跟踪; 自校正; 视频计数

引用格式: 刘春, 栗健. 基于 YOLOv3 的可变时间窗自校正船只跟踪与计数. 计算机系统应用, 2021, 30(11): 240-246. <http://www.c-s-a.org.cn/1003-3254/8205.html>

## Self-Correcting Ship Tracking and Counting with Variable Time Window Based on YOLOv3

LIU Chun, LI Jian

(School of Computer Science, Hubei University of Technology, Wuhan 430068, China)

**Abstract:** Automatically identifying the types and counting the numbers of ships on waterways is of great significance for the construction of “smart waterways”, intelligent early warning regarding water surface, and navigation decision support. In this study, ship sample images are trained with the YOLOv3 pre-training model, and the detection model for ships on waterways is developed after parameter adjustment and optimization. Then, considering that the deep learning model is good at extracting target features, this study combines the target HSV color histogram features and LBP local features to achieve the target selection. In view of common drift and jitter of tracking targets, a correction network is designed with the integration of regression-based direction judgment and target counting with a variable time window, which realizes the automatic detection, tracking and self-correcting counting of moving targets on water surface. The test results show that the proposed method is stable and robust, suitable for automatically analyzing channel videos and extracting statistical data.

**Key words:** YOLOv3; ship identification; target tracking; self-correcting; video count

这几年, 目标跟踪越来越成为计算机视觉研究领域中的一个焦点问题. 传统的目标跟踪方法有均值漂移 (mean shift) 算法<sup>[1]</sup> 与光流 (Lucas-Kanade) 算法<sup>[2]</sup>

等. 基于均值漂移的思想是沿与模板相似度最大的矢量方向不断迭代候选目标框, 并收敛于目标的真实位置. 但是这种方法无法很好地适应目标的尺度变化和

① 基金项目: 国家重点研发计划 (2017YFC1405401)

Foundation item: National Key Research and Development Program of China (2017YFC1405401)

收稿时间: 2021-01-20; 修改时间: 2021-02-23, 2021-03-24; 采用时间: 2021-04-02; csa 在线出版时间: 2021-10-22

快速移动. 光流法主要通过分析视频中像素灰度值随时间的变化规律来获取光流信息, 进而计算相邻序列之间的相关性来捕获目标的运动状态. 基于光流的方法假设目标在运动时亮度不变以及目标在相邻帧的位移较小, 因此这种方法只适用于背景和光照没有明显变化的场景, 然而实际应用中, 背景和光照往往处于不断变化中, 这种方法存在较大的局限性.

为了解决上述问题, 第一个基于相关滤波的方法被提出, 即平方误差最小滤波器 (Minimum Output Sum of Squared Error, MOSSE)<sup>[3]</sup>. 它根据第 1 帧的样本, 通过计算输出结果的最小均方误差, 生成一个具有判别性的滤波器, 从而有效区分背景和目标. 在跟踪过程中, 通过搜索区域经滤波器作用后的响应图像来寻找目标的位置, 响应图像的值越大, 说明该位置的图像与目标的相关性越大. 随后, 又有若干算法对此进行了大量改进, 比如基于目标的尺寸和位置, 选择相关搜索区域, 不过这些算法与搜索区域设置相关, 若过小或者过大, 可能会引起目标丢失或跟踪漂移<sup>[4]</sup>.

随着出现了一些与深度学习相关的跟踪算法, 这使目标跟踪算法的研究取得了极大的进步和发展. 首次将深度学习应用于跟踪任务的方法是创造性地使用了离线训练+在线微调的深度目标跟踪框架<sup>[5]</sup>. 在这之后, Leal-Taixé等人<sup>[6]</sup>在深度学习框架下使用 tracking-by-detection 方式, 通过 two-stage 学习匹配, 研究出了一种在对行人跟踪时的数据关联新方法. 2018年, 中国科学院的何安丰等人<sup>[7]</sup>提出的 SA-Siam 目标跟踪算法在 SiamFC 基础上改进网络结构, 采用了双孪生网络的结构来提升模型在目标跟踪任务中的判别能力. 2019

年, Voigtlaender 等人<sup>[8]</sup>在 Faster R-CNN 的基础上, 引入了一种具有 Tracklet 动态规划算法的孪生两阶段全图像重检测架构, 提出的 Tracklets 的动态规划算法能够在遮挡和目标消失等挑战下具有一定效果.

但是由于航道中存在天气、水面光线变换强烈、船只种类、船只互相遮挡等一系列问题, 造成一般目标跟踪算法的检测精确度低, 船只重复计数频发. 为了解决这些问题, 本文提出了一种基于 YOLOv3 的可变时间窗自校正船只目标跟踪与计数方法, 首先通过 Darknet-53 训练网络结构, 得到一个在航道以及水运场景下的船只检测模型, 然后对目标进行全局性的 HSV 直方图与局部性的 LBP 直方图特征提取、筛选与匹配, 进行初步的目标跟踪, 再融合判断方向的多帧回归与统计计数组成的可变时间窗模型对跟踪计数进行自校正, 平滑计数时产生的抖动, 整个算法不仅可以实现航道设施的实时监测与智能预警, 还能帮助工作人员进行辅助决策且对航道内快速检修具有非常良好的效果.

## 1 YOLOv3 目标检测算法

YOLOv3 目标检测算法是一种从端到端的目标检测算法, 具有速度快、精度高的优点, 并且满足实时检测要求, 它是以 Darknet-53 网络为基础, 采用全卷积神经网络来提取图像特征, 该网络由一系列  $3 \times 3$  和  $1 \times 1$  的卷积层构成, 共有 53 层<sup>[9]</sup>. 为了解决特征提取的层数过深可能会带来的梯度消失问题, 它利用残差很大程度上减少了每次卷积的通道, 最后将输入图像的 3 个不同尺度的特征图分别进行多尺度预测并融合输出. YOLOv3 检测算法如图 1 所示.

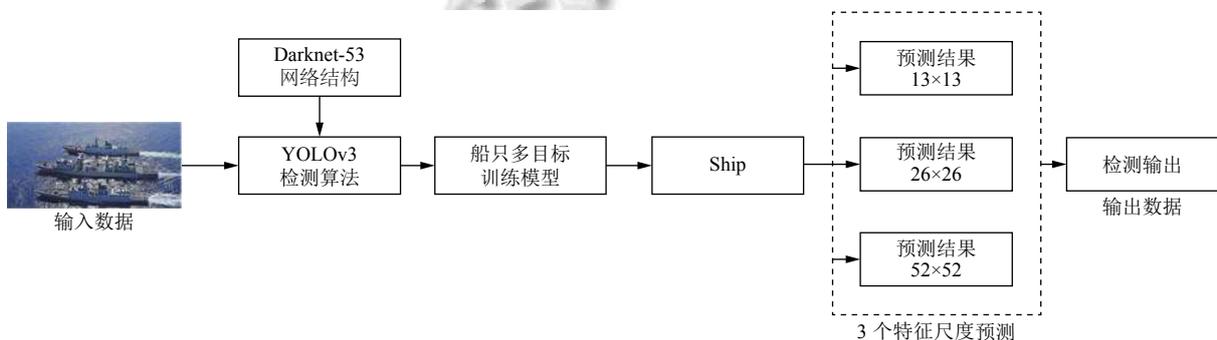


图 1 YOLOv3 检测算法流程

本文针对 YOLOv3 的 LeakyReLU 激活函数进行了更适合本项目的优化, 数学表达式为  $f(x)=\max(ax,x)$ ,  $a$  为训练过程中自学习的参数, 使神经网络节点在每一

次训练的反向传播过程中都能快速准确地保持更新状态, 速度更快, 精度更佳, 在项目中使用更稳定. 而更高版本的 YOLO 使用了 Mish 激活函数, 无法适用此优

化,因此选择YOLOv3目标检测算法.后续跟踪时所有候选目标均来自于此YOLOv3目标检测算法检测到的目标.

## 2 目标跟踪算法

目标跟踪是指给定视频序列初始帧中的目标位置,在后续帧中预测每一帧中目标的位置,得到目标完整的运动方向和轨迹的过程<sup>[10]</sup>.通常情况下,典型的目标跟踪算法流程是由运动模型、外观模型、观测模型以及在线更新机制构成.虽然在目标跟踪领域已经有很多的研究成果,但是仍有许多关键技术有待突破.其中一个比较重要的问题是跟踪时如果物体被阻挡,那么物体重新出现时既可能会对它进行恢复识别与恢复计数,又有可能因目标检测识别率原因把目标识别为新物体,从而重复识别与重复计数.为此本文在跟踪过程中融合了判别与校正.在一个可变时间窗内进行计数抑制和恢复.此跟踪算法如算法1.

算法1.跟踪算法

- 1) 在第一帧图像中确定跟踪目标,将该目标设为模板,并获得其HSV与LBP特征向量;
- 2) 继续进行目标检测,得到若干检测待选目标;
- 3) 获得每个待选目标的HSV和LBP特征,依次计算它们与跟踪目标的相似度得分,确定出一个相似度得分最高的待选目标;
- 4) 使用可变时间窗自校正模型以避免出现计数重复以及方向误判和抖动.

可变时间窗自校正模型分为两个部分,分别是基于可变时间窗的目标计数和基于多帧回归的运动方向判别.前者可以在船只发生遮挡并重复计数以后实现计数自校正,而后者会根据船只的运动趋势来实现船只运动方向的自校正判断.

### 2.1 特征提取与目标选择

在目标跟踪时,可能会出现多个待跟踪目标,为了能够在视频画面中匹配一个与最初确定的模板最相近的目标来跟踪,因此需要对它们进行选择<sup>[11]</sup>.故通过HSV颜色直方图的特点结合LBP直方图的特点一起对待跟踪的物体来选择.

HSV空间特征是全局性的,它可以较好地描述物体形状的整体信息,尤其适合描述一些形变较大的目标<sup>[12]</sup>.LBP特征恰恰相反,它比较适合描述一些局部性的特征,比如目标的局部纹理,以及目标的空间边缘信息.非常重要的是,计算LBP特征比较轻松,检测速度将会大大缩短<sup>[13]</sup>.式(1)为LBP特征的提取过程:

$$P(x_c, y_c) = \sum_{p=0}^{P-1} 2^p s(i_p - i_c) \quad (1)$$

其中,  $(x_c, y_c)$ 表示中心像素;  $i_c$ 是灰度值;  $i_p$ 是相邻像素的灰度值;  $s(\cdot)$ 是一个符号函数,即:

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (2)$$

HSV直方图特征相似度计算公式为:

$$s(H_1, H_2) = \sqrt{1 - \frac{1}{\sqrt{H_1 H_2 N^2}} \sum_I \sqrt{H_1(I) H_2(I)}} \quad (3)$$

LBP直方图特征相似度计算公式为:

$$s(L_1, L_2) = \sqrt{1 - \frac{1}{\sqrt{L_1 L_2 M^2}} \sum_I \sqrt{L_1(I) L_2(I)}} \quad (4)$$

其中,  $H_1$ 表示候选目标的HSV特征直方图向量,  $H_2$ 表示模板目标的HSV特征直方图向量;  $N$ 表示直方图bin数目;  $\bar{H}_k = \frac{1}{N} \sum_J H_k(J)$ ,  $H_k(J)$ 表示 $H_k$ 中序号为 $J$ 的bin的HSV颜色向量统计值;  $L_1, L_2$ 分别表示候选目标和模板目标的LBP特征直方图向量;  $M$ 表示直方图bin数目;  $\bar{L}_k = \frac{1}{M} \sum_J L_k(J)$ ,  $L_k(J)$ 表示 $L_k$ 中序号为 $J$ 的bin的LBP颜色向量统计值.

通过上面的计算,我们可以非常轻松地得到待跟踪目标的HSV和LBP的直方特征,以及候选船只与模板的特征相似度,设置HSV特征的权重为1,LBP特征的权重为2,就计算出了所有待跟踪船只的相似度得分.最后通过相似度得分的多少来选择船只.

### 2.2 可变时间窗自校正跟踪计数

#### 2.2.1 可变时间窗计数

在计算HSV直方特征相似度和LBP直方特征相似度后,每一艘被跟踪的船只在画面中出现后从始至终都应被认定为同一船只.因此,当每艘船只被检测出时,通过计算HSV直方特征和LBP直方特征,对每个不同的特征都会被赋予一个独一无二的flag值,并且使计数加一.这个flag值也会随检测线框一同标出.当后续跟踪时,因为跟踪的船只具有和初始时非常近似的特征相似度,所以为它们赋予了和初始时同一个flag值,表明是同一船只,不会使计数加一.如图2所示.

但是,在实际的跟踪过程中,可能会出现2种情况对计数结果进行误判:(1)在两船有遮挡现象时,可能会出现“融为一体”的情况,两船整体可能会展现出与

之前船只较大的相似度差别; (2) 当船只被遮挡并重新出现时, 可能由于角度, 光线等问题, 被遮挡的船只在

遮挡前后也显示出了较大的相似度差别. 这两种情况都可能会被认为出现了一艘新的船只, 使计数错误加一.



图2 船只计数原理

为此针对第1种情况设计了一种可变时间窗的计数方法. 思想是如果被跟踪的某些船只突然“消失”, 那么说明船只可能被遮挡. 此时间窗长度会与遮挡船与被遮挡船的大小差成正比, 速度差成反比. 因为在视频图像中, 越小以及越慢的船只通过同一遮挡物所需的时间往往越长.

在实际的航行过程中, 因为每艘船只的速度变化较小, 所以可以将被跟踪船只出现到消失这段时间的平均速度作为船只的速度. 而两船的大小差可以根据跟踪船只的线框长度相减得到.

因此, 时间窗的长度根据船只大小与速度进行变化, 即:

$$L = a \cdot \left| \frac{x_1 - x_2}{v_1 - v_2} \right| \quad (5)$$

其中,  $L$ 表示时间窗大小,  $a$ 为待定参数,  $x_1, v_1$ 分别为被遮挡船的长度与速度,  $x_2, v_2$ 分别为遮挡船的长度与速度.

但是上述第2种情况属于自然条件不可控, 因此可以对通过式(5)计算得到的时间窗长度进行适当延长, 提高容错率.

最后判断在上述计算出的时间窗口内, 这艘“新”的船只是否消失且原船只再次出现. 如果是, 那么就对计数进行校正, 减去这些出现的“新”船只的数量.

### 2.2.2 多帧回归方向判断

理论上, 跟踪目标时, 根据在相邻两帧同一目标的区域坐标大小, 即可计算出目标的运动方向. 如第一帧目标的坐标为(12, 22), 第二帧坐标为(14, 22), 则目标在向画面的右侧移动. 但是在实际的应用过程中, 可能会随着画面与目标的同步移动使目标物体的坐标出现

抖动, 如果按上面的方法判断, 向同一个方向运动的物体极有可能出现(12, 22), (11, 22), (13, 22)这样的坐标形式而判断为先向左后向右的移动顺序. 采用多帧回归判断算法可抑制此问题: 通过对当前帧前后的若干帧中目标物体的坐标作回归分析, 得出物体的运动趋势, 再结合相邻两帧的坐标对比做出综合判断, 以此来确定目标物体的运动方向.

在判断船只运动方向时, 可以建立下面的线性关系:

$$Y = A + BX + \varepsilon \quad (6)$$

其中,  $A$ 和 $B$ 为待定参数,  $\varepsilon$ 为随机误差项,  $X$ 表示时间,  $Y$ 表示船只的横坐标.

当我们判断某一时刻船只的运动方向时, 将这一时刻的帧与它之前若干相邻帧的船只坐标提取出来, 通过回归方程计算可以得出 $B$ 的正负值, 当 $B$ 为正值时, 可初步判断船只运动方向趋势为向画面的右侧移动, 反之则向左侧移动. 当船只向左侧移动时, 在跟踪线框之上标注L, 反之标注R.

随机选取一帧, 记录与之相邻的之前若干帧坐标数据(此时时刻为0), 如表1. 若以传统方式判断前后两帧坐标, 则会发现横坐标变小, 会发生船只向画面左侧移动的误判. 而通过计算回归方程可以得出 $Y=2.4978+0.0383X$ , 因 $B=0.0383$ 为正值, 可判断此时船只向画面右侧移动.

常识和实验验证发现, 判断船只方向应该以船只运动趋势来作为基本判断依据, 而不是前后两帧的变化, 因此本方法可以有效地减少抖动, 提高判断正确率.

## 2.3 融合自校正模型的改进YOLOv3神经网络

可变时间窗自校正跟踪计数模型可以与YOLOv3

神经网络进行融合,设计一个校正网络来实现本文自校正模型的功能。

该校正网络为一个链式结构,包含2个模块.第1个模块包含3个卷积单元,第2个模块包含3个全连接层.然后,修改YOLOv3模型的网络结构,使校正网络融入到基于YOLOv3的跟踪网络结构中.它可以处理跟踪网络的输出结果,对其进行校正,随后根据这个结果对跟踪网络及其自身进行网络更新.如图3所示.

利用校正网络的结果同时更新跟踪网络与校正网络,并将其作为下一帧学习的网络.本研究采用校正网

络确定的最佳样本的尺度作为追踪目标的尺度<sup>[14]</sup>,计算速度大大加快,也节省了计算开销.

表1 实验中测得的数据坐标

时刻	横坐标
-8	2.1
-7	2.2
-6	2.4
-5	2.3
-4	2.3
-3	2.5
-2	2.4
-1	2.5
0	2.4

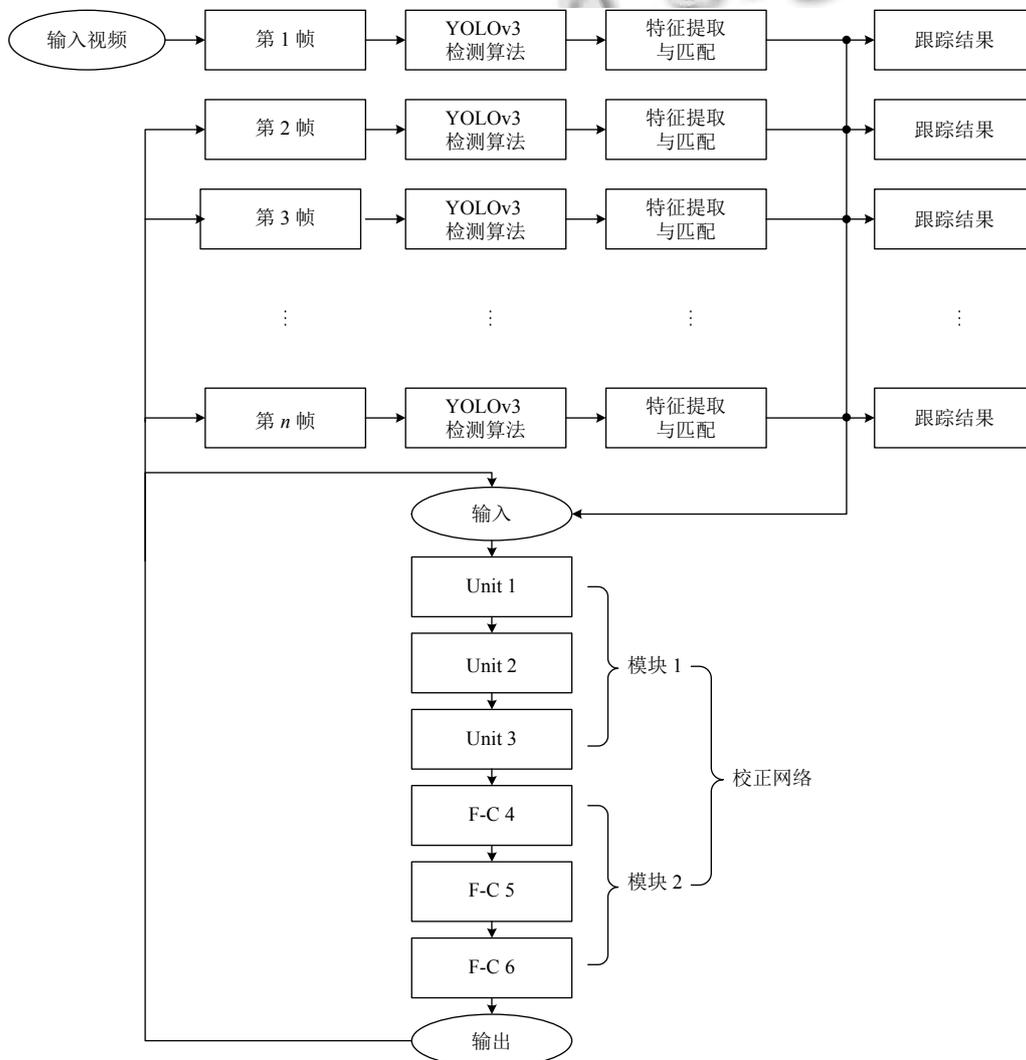


图3 融合网络结构图

### 3 测试结果

本实验采用了 NVIDIA 2060 的 GPU 与 PyCharm

编辑器的开发环境.通过长江实地拍摄以及上网搜索得到 1565 张船只图片,制作 VOC 数据集时将图片按

6:2:2 的比例分为训练集, 验证集与测试集, 使训练集为 939 张图片, 测试集为 313 张图片, 验证集为 313 张图片. 使用 Darknet-53 训练网络结构, 根据实际情况对样本分类与算法修改, 将自校正模型添加到网络结构中去. 开始训练之前, 根据 GPU 性能表现, 修改 batch=64, subdivisions=8, learning\_rate=0.001, momentum 动量系数设置为 0.9, policy 学习率调整的策略为 steps. 然后对数据集进行训练, 同时系统可以自动保存所有的记录和训练好的模型. 其中有一个自动覆盖的最优训练模型以供使用.

不仅如此, 系统还生成了训练时的当前平均损失以供研究者实时观察. 因为它是评价目标检测效果的重要指标<sup>[15]</sup>, 所以我们可以非常直观地看到训练完成后的效果. 当我们进行了大约 400 次的训练以后, 此时平均损失变化极小. 这时说明训练已经基本完成, 可以达到一个比较好的效果.

当前平均损失收敛图如图 4 所示.

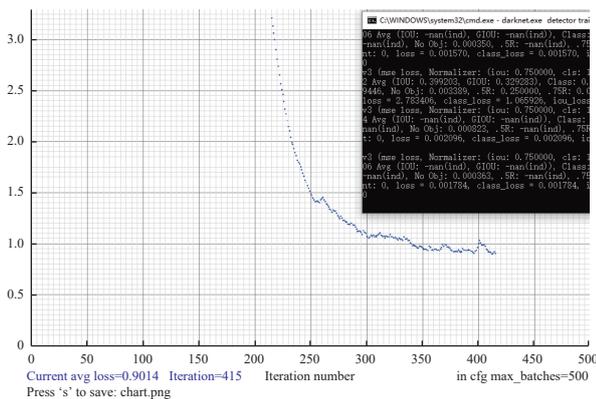


图 4 当前平均损失收敛图

随机挑选一张船只图片进行检测, 可以看出检测时间为 35.723 ms, 船只检测 ship 匹配度为 90%, 如图 5 所示.

对水运以及航道场景下的视频测试主要应对船只种类的多样性、太阳或灯光等不同光线的照射、拍摄视角不同等问题. 本文采用了多种类别监控与视频素材, 测试的场景丰富, 涵盖了各种天气条件下的多种场景, 以保证该目标跟踪与计数方法在视频或监控等多种场景下都可以稳定运行. 使用如上的船只检测模型, 通过上传不同场景下船只的视频, 即可实现目标检测并跟踪计数, 并展示出较高的跟踪成功率以及计数准确率. 跟踪效果如图 6 所示.

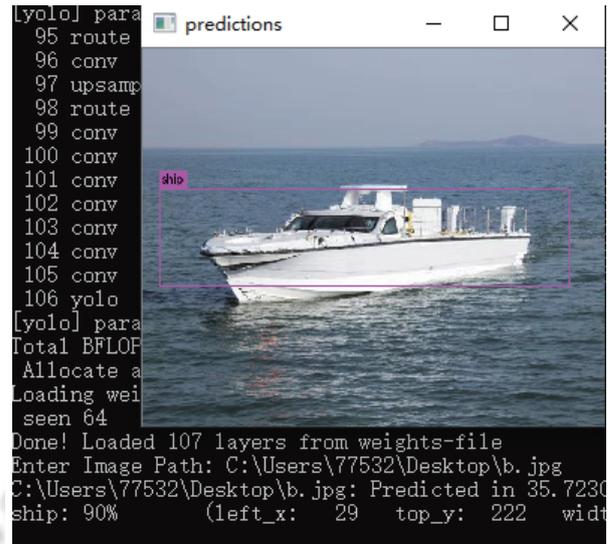


图 5 船只目标检测效果

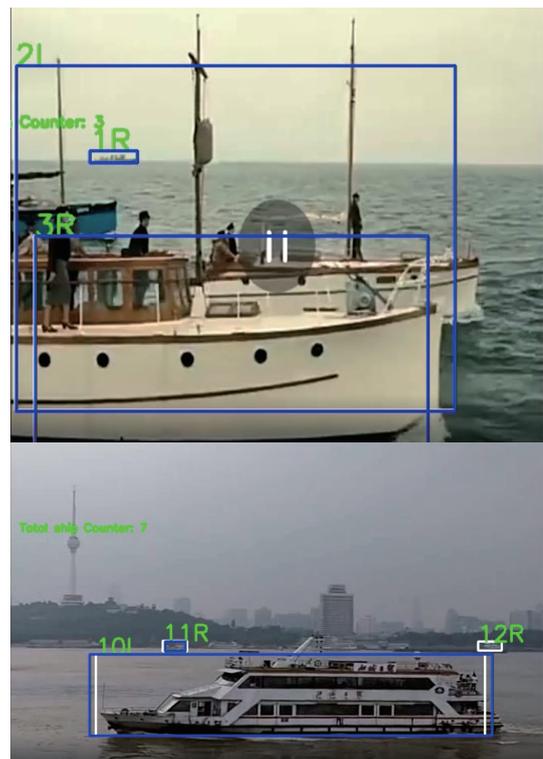


图 6 船只目标跟踪效果

为了验证本文提出的自校正计数模型是否可以通过使用可变时间窗对通航的船只进行计数的自校正, 进行了对照实验, 实验组为融合了自校正模型的改进网络, 对照组为未融合的普通 YOLOv3 网络. 通过对比是否应用可变时间窗自校正跟踪计数模型, 对计数结果进行了统计. 实验中使用了 9 个不同的视频, 分别统

计了应用算法前后的跟踪计数效果,实验结果如表2所示。

表2 应用自校正算法前后计数效果对比

视频序号	船只实际数目	原计数	改进后计数
1	12	21	12
2	52	70	54
3	24	30	26
4	5	6	5
5	17	20	18
6	8	10	8
7	30	36	32
8	22	30	23
9	20	25	20

由表2可以看出,在没有使用本文提出的自校正计数模型时,因为画面抖动,遮挡导致计数非常不准确,尤其是2号视频序列船只计数为70,船只被重复计数了18次。而在使用了自校正计数模型后,2号视频序列的船只计数为54次。参考其他视频序列实验数据,表明已基本满足目标实时跟踪与计数的性能要求。

#### 4 结语

本文介绍了基于YOLOv3目标检测算法的航道水运场景下船只目标跟踪与自校正计数方法,包括系统的架构、数据集制作、模型训练、特征提取、目标跟踪、方向判别与计数、检测结果和结果分析。通过对大量视频素材的测试结果分析,本文所讲述的船只目标跟踪与计数系统在多种场景、多种天气条件下均能抵抗抖动稳定计数。但是在出现两条相似船只且有交错重合的条件下,仍然会造成计数出错,在后续进一步的研究中将针对这种情况做进一步的优化。

#### 参考文献

- Du K, Ju YF, Jin YL, *et al.* Object tracking based on improved MeanShift and SIFT. Proceedings of the 2012 2nd International Conference on Consumer Electronics, Communications and Networks. Yichang: IEEE, 2012. 2716–2719.
- Shi JB, Tomasi. Good features to track. 1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 1994. 593–600.

- Bolme DS, Beveridge JR, Draper BA, *et al.* Visual object tracking using adaptive correlation filters. Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco: IEEE, 2010. 2544–2550.
- 汤一明,刘玉菲,黄鸿.视觉单目标跟踪算法综述.测控技术,2020,39(8):21–34.
- Wang NY, Yeung DY. Learning a deep compact image representation for visual tracking. Proceedings of the 26th International Conference on Neural Information Processing Systems. Lake Tahoe: Curran Associates, 2013. 809–817.
- Leal-Taixé L, Canton-Ferrer C, Schindler K. Learning by tracking: Siamese CNN for robust target association. 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Las Vegas: IEEE, 2016. 418–425.
- He AF, Luo C, Tian XM, *et al.* A twofold siamese network for real-time object tracking. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 4834–4843.
- Voigtlaender P, Luiten J, Torr PHS, *et al.* Siam R-CNN: Visual tracking by re-detection. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020. 6577–6587.
- 郑志强,刘妍妍,潘长城,等.改进YOLO V3遥感图像飞机识别应用.电光与控制,2019,26(4):28–32.
- 孟球,杨旭.目标跟踪算法综述.自动化学报,2019,45(7):1244–1260.
- 田健,王开军.自适应紧致特征的超像素目标跟踪.中国图象图形学报,2017,22(10):1409–1417. [doi: 10.11834/jig.160619]
- 毕威,黄伟国,张永萍,等.基于图像显著轮廓的目标检测.电子学报,2017,45(8):1902–1910. [doi: 10.3969/j.issn.0372-2112.2017.08.014]
- 吉培培,陈恳,刘哲,等.结合HOG-LBP特征及多实例在线学习的随机藏目标跟踪方法.宁波大学学报(理工版),2015,28(4):42–47.
- 张涛,张乐.一种基于多尺度特征融合的目标检测算法.激光与光电子学进展:1–11. <http://kns.cnki.net/kcms/detail/31.1690.TN.20200811.1650.062.html>. (2020-08-12) [2020-11-20].
- 高铭.基于深度学习的复杂交通环境下目标跟踪与轨迹预测研究[博士学位论文].长春:吉林大学,2020.