

DOI: 10.19650/j.cnki.cjsi.J2108848

一种远距离行人小目标检测方法*

石欣, 卢灏, 秦鹏杰, 冷正立
(重庆大学自动化学院 重庆 400044)

摘要:远距离行人小目标成像像素少、缺乏纹理信息,深度卷积神经网络难以提取小目标细粒度特征,难以准确识别与检测。本文提出一种远距离行人小目标检测方法。首先,在YOLOv4的基础上引入浅层特征改进特征金字塔,提取行人小目标细粒度特征,提出引力模型特征自适应融合方法,增加多层次语义信息之间的关联度,防止小目标特征信息流失。然后,采用增强型超分辨率生成对抗网络增加行人小目标特征数量,提高行人小目标检测准确率。最后,选取图像像素中占比范围为0.004%~0.026%的行人小目标建立试验数据集,通过与Faster RCNN、ION、YOLOv4对比实验验证。结果表明,本文方法 $mAP_{0.5}$ 提高了25.2%、26.3%、11.9%,FPS达到24,研究成果在远距离安防监测监控领域具有重要应用价值。

关键词:行人小目标;特征增强;特征自适应融合;引力模型;目标检测

中图分类号: TH701 TP183 **文献标识码:** A **国家标准学科分类代码:** 520.2040

A long-distance pedestrian small target detection method

Shi Xin, Lu Hao, Qin Pengjie, Leng Zhengli

(School of Automation, Chongqing University, Chongqing 400044, China)

Abstract: Small pedestrian targets at long distance have problems of few pixels and lack of texture information. The deep convolutional neural network is difficult to extract fine-grained features of small objects. This article proposes a long-distance pedestrian small target detection method. Firstly, based on the YOLOv4 algorithm, the shallower features are introduced to improve the feature pyramid to extract fine-grained features of pedestrian small objects. An adaptive feature fusion layer is proposed based on the gravity model to increase dependency between multiple feature layers and prevent the loss of small target feature information. Then, ESRGAN is utilized to increase pedestrian small target features number and improve pedestrian small target detection accuracy. Finally, the small pedestrian targets are selected with a proportion of 0.004%~0.026% in the image pixels to establish the self-made data set. Compared with Faster RCNN, ION, and YOLOv4, the $mAP_{0.5}$ values are increased by 25.2%, 26.3% and 11.9%. And the FPS reaches 24. The research results have important application value in the field of long-distance security monitoring.

Keywords: small pedestrian target; generative adversarial network; feature adaptive fusion; gravity model; object detection

0 引言

开阔视野、复杂场景下安防监控行人小目标识别是研究的难点^[1-3]。远距离行人目标成像像素数量少、纹理信息少^[4],根据MSCOCO数据集定义,像素的绝对值小于 32×32 即可称之为小目标^[5]。对于这一类的图像目标识别,通过深度学习计算目标的纹理和位置难以满足安防监控对行人小目标入侵检测的精度^[6-9]。因此研究行

人小目标的识别,对保证安防监控中的关键图像检测,重要区域的管控有重要意义。

目前,深度学习由于其特征提取的优势,在目标检测中得到广泛使用。Ren等^[10](2018)提出了Faster R-CNN,使用区域生成网络(RPN),根据目标大小,设置不同尺寸anchor,对行人目标的检测效果有所改进,在PASCAL VOC 2007数据集上,检测精度能达到73.2%。该算法仅使用单层特征对目标进行检测,由于其顶层感受野和小目标尺寸相差悬殊的问题导致对小目标检测的效果不佳^[11]。

收稿日期:2021-11-11 Received Date: 2021-11-11

* 基金项目:国家自然科学基金联合基金项目(U1813216)资助

Bell 等^[12]提出的 ION 网络通过感兴趣区域(ROI)从不同尺度的层裁剪特征,并使用循环神经网络(recurrent neural network, RNN)对整个图像的上下文信息进行编码,改善小目标检测,由于 ION 使用全局上下文信息,在对稀疏目标识别时误差变大。文献[13-15]提出了 YOLOv4 目标识别网络,使用 CSPDarknet53 提取目标特征,通过大量残差结构的堆叠,并使用特征金字塔结构^[16]完成不同尺度特征的融合,增强特征的表征性,以提高行人小目标的识别率^[17],由于网络下采样率过大,导致提取的小目标特征存在大量环境噪声,从而影响检测效果^[18]。

本文立足于安防监控环境下的关键帧检测,提出了一种远距离下的行人小目标检测方法。研究的数据来源为 300~400 m 视距的行人小目标,目标的绝对大小为 15 pixel×15 pixel 左右,相对大小范围为 0.004% ~ 0.026%。本文首先对 YOLOv4 进行改进,设计基于引力模型的自适应特征融合层,再引入增强型超分辨率生成

对抗网络(ESRGAN)^[19]将两个网络级联形成最终的识别检测网络。经试验分析验证了所提方法的有效性,相较于 Faster RCNN、ION、YOLOv4 等,本文方法的定位和识别精度均有较大提升。实时性有一定下降,但仍能满足实时性要求和关键帧检测的需要。

1 特征增强与改进 YOLOv4 的识别算法

1.1 增强型超分辨率生成对抗网络特征增强算法及损失函数分析

远距离处的行人目标通常占据视场万分之一以下,特征过于稀少,采用增强型超分辨率生成对抗网络(ESRGAN)^[19],将低分辨率行人样本逼近高分辨率样本,实现目标特征增强。该算法以残差块为基础,在 SRResNet 的基础上进行改动^[20],作为算法的基本框架,如图 1 所示。

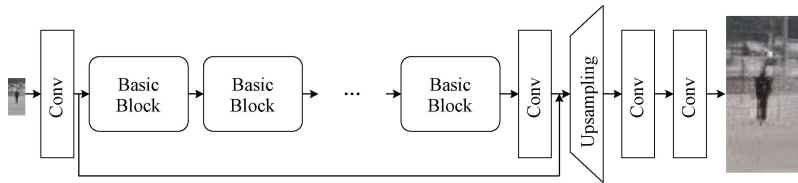


图 1 基于 SRResNet 的 ESRGAN 框架
Fig. 1 ESRGAN architecture based on SRResNet

ESRGAN 是一种在 SRGAN 上改进的超分辨率算法,在结构上分为判别器和生成器。通过借鉴相对论平均 GAN(RaGAN),完善和改进了判别器,使判别器能得出生成图像与真实图像的相似程度。判别器公式如式(1)所示。

$$\begin{cases} D_{Ra}(x_r, x_f) = \sigma(C(\text{Real})) - E[C(\text{Fake})] \rightarrow 1 \\ D_{Ra}(x_f, x_r) = \sigma(C(\text{Fake})) - E[C(\text{Real})] \rightarrow 0 \end{cases} \quad (1)$$

其中, E 代表对 mini-batch 中的数据求均值。根据式(1)定义判别器的损失函数为:

$$L_D^{Ra} = -E_{x_r}[\log(D_{Ra}(x_r, x_f))] - E_{x_f}[\log(1 - D_{Ra}(x_f, x_r))] \quad (2)$$

根据判别器的损失函数,则可得到对称的生成器的对抗损失:

$$L_G^{Ra} = -E_{x_r}[\log(1 - D_{Ra}(x_r, x_f))] - E_{x_f}[\log(D_{Ra}(x_f, x_r))] \quad (3)$$

其中, x_f 为输入的低分辨率图像。同时,引入感知损失 L_{percep} ,其定义在预训练的深度网络的激活层之前,增强了特征的监督性,并保持了重建图像的亮度与原图像一致。

综合以上,得到 ESRGAN 的总损失函数如式(4)所示。

$$L_G = L_{percep} + \lambda L_G^{Ra} + \eta L_1 \quad (4)$$

其中, L_1 为生成图像与真实图像之间像素损失。

1.2 基于引力模型的改进 YOLOv4 算法

1) 改进的 FPN 特征融合结构

为解决小目标检测精度低下的问题, YOLOv4 采用特征金字塔(feature pyramid networks, FPN)的方式对不同尺寸的特征层进行融合,结构如图 2 所示。

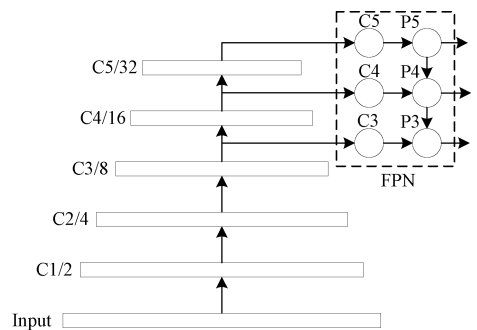


图 2 FPN 特征融合结构

Fig. 2 Feature pyramid networks architecture

以上图 2 分析,以 1 920×1 080 作为输入图像大小,行人小目标较小时可达 10 pixel×20 pixel 左右,约占比

0.01%。图像卷积至 C3 层时,下采样率为 8 倍,目标在特征图上的像素大小不足 2×3 。对于 C4、C5 层的特征,下采样率达到 16、32 倍,目标将占据不足一个像素,且每个特征层上的点不仅包括目标特征,也包含大量周围区域的特征。

本文首先针对特征提取网络中下采样率过大的问题,在 FPN 特征金字塔中引入更浅层的特征层 C2,构建如图 3 所示的改进 FPN 结构。在增加 C2 层后,目标在该层上的像素点能达到约 3×5 ,特征点数量增加为 C3 层的 4 倍,增加较多的细粒度特征。

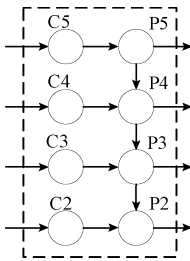


图 3 改进的 FPN

Fig. 3 The improved FPN architecture

2) 基于引力模型的自适应特征融合

特征金字塔模型通过顺序方式整合特征,会导致非相邻层次的语义信息被稀释。且考虑小目标在深层特征中包含较多环境噪声,特征融合应更偏向于使用浅层特征。本文提出基于引力模型的自适应特征融合层,通过引入 Libra RCNN 的特征平衡思想^[21],以引力模型为基础,根据自制数据集目标大小分布,自适应设定特征层的加权系数,有偏重地从各个分辨率的特征层中剥离特征信息,增加多个特征层之间的依赖关系。算法结构如图 4 所示。

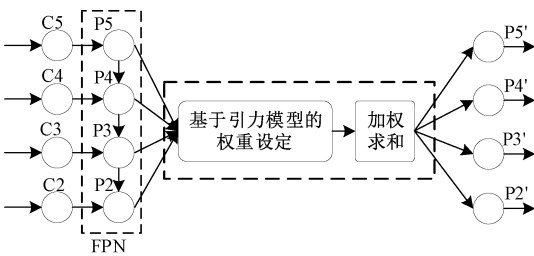


图 4 特征自适应融合算法结构

Fig. 4 Feature adaptive fusion algorithm structure

不同尺寸特征层适用于检测不同大小目标,本文将目标按大小分为 x_1, x_2, x_3, x_4 四类,类比引力^[22]来描述各特征层对不同大小目标检测效果的偏向性,类比的引力即为加权系数,万有引力公式如式(5)所示。

$$F = G \frac{M_1 M_2}{R^2} \tag{5}$$

以特征层 {P2、P3、P4、P5} 为例,图 5 为各特征层偏向性示意图,其中箭头粗细表示偏向性的强弱。

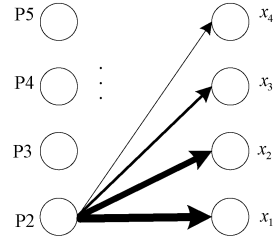


图 5 检测性能偏向性示意图

Fig. 5 Schematic diagram of detection performance bias

以 P2 层为例,P2 对 x_1, x_2, x_3, x_4 四类目标的检测性能逐步递减,类比为随着 4 类目标的在类别等级上的差级增大,P2 层的检测性能随之减弱。

由于 P2 层对应最适合检测 x_1 类目标,在量化时用 x_1 与 x_2, x_3, x_4 的分辨率倍数差作为引力间距 R ,如 x_1 与 x_2 分辨率差 2 倍, $R_{12} = 2$;与 x_3 分辨率差 4 倍, $R_{13} = 4$;与 x_4 分辨率差 8 倍, $R_{14} = 8$ 。综上,类推可得整体加权系数 $\{f_1, f_2, f_3, f_4\}$ 的关系式:

$$\begin{aligned} f_1(x_1, x_2, x_3, x_4) &= G_1 \frac{x_1 x_2}{R_{12}^2} + G_1 \frac{x_1 x_3}{R_{13}^2} + G_1 \frac{x_1 x_4}{R_{14}^2} \\ f_2(x_1, x_2, x_3, x_4) &= G_2 \frac{x_2 x_1}{R_{21}^2} + G_2 \frac{x_2 x_3}{R_{23}^2} + G_2 \frac{x_2 x_4}{R_{24}^2} \\ f_3(x_1, x_2, x_3, x_4) &= G_3 \frac{x_3 x_1}{R_{31}^2} + G_3 \frac{x_3 x_2}{R_{32}^2} + G_3 \frac{x_3 x_4}{R_{34}^2} \\ f_4(x_1, x_2, x_3, x_4) &= G_4 \frac{x_4 x_1}{R_{41}^2} + G_4 \frac{x_4 x_2}{R_{42}^2} + G_4 \frac{x_4 x_3}{R_{43}^2} \end{aligned} \tag{6}$$

对于上述模型,需依据 VOC2012 作为标准数据集,求解引力系数 $\{G_1, G_2, G_3, G_4\}$ 。此时 $f_1 = f_2 = f_3 = f_4 = 0.25$,归一化后的 x 均为 0.25。求解后得到最终的关系模型为:

$$\begin{aligned} f_1(x_1, x_2, x_3, x_4) &= 21 \times \left(\frac{x_1 x_2}{R_{12}^2} + \frac{x_1 x_3}{R_{13}^2} + \frac{x_1 x_4}{R_{14}^2} \right) \\ f_2(x_1, x_2, x_3, x_4) &= 9 \times \left(\frac{x_2 x_1}{R_{21}^2} + \frac{x_2 x_3}{R_{23}^2} + \frac{x_2 x_4}{R_{24}^2} \right) \\ f_3(x_1, x_2, x_3, x_4) &= 21 \times \left(\frac{x_3 x_1}{R_{31}^2} + \frac{x_3 x_2}{R_{32}^2} + \frac{x_3 x_4}{R_{34}^2} \right) \\ f_4(x_1, x_2, x_3, x_4) &= 9 \times \left(\frac{x_4 x_1}{R_{41}^2} + \frac{x_4 x_2}{R_{42}^2} + \frac{x_4 x_3}{R_{43}^2} \right) \end{aligned} \tag{7}$$

在得到加权系数的关系模型后,通过对自制数据集的目标进行分类统计,求出 4 个特征层对 4 类目标的引力分量,求和并归一化后即可得到各个特征层的权重,根

据以上流程,以自制数据集为例,自适应权重求解过程如下。

(1) 对数据集目标进行分类统计

以 VOC2012 作为标准数据集,以目标相对大小进行从小到大排序,对数量进行累加,在数量四等分的情况下对目标进行分段分类。得到的大小分类区间为 $[0, 0.0119]$, $[0.0119, 0.0499]$, $[0.0499, 0.2296]$, $[0.2296, 1]$ 。统计自制数据集汇总目标大小分布情况如图 6 所示。

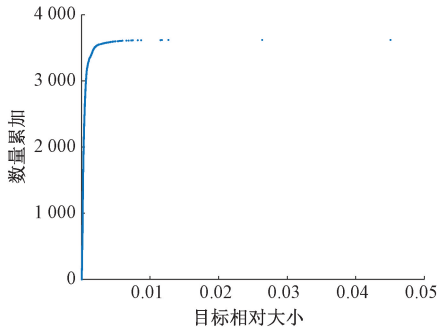


图 6 自制数据集目标大小分布

Fig. 6 Our own dataset targets size distribution

根据分类区间,统计目标占比,并进行归一化,结果如表 1 所示。

表 1 目标类别占比
Table 1 Target category ratio

目标类别	x_1	x_2	x_3	x_4
数量占比	0.830 4	0.120 7	0.038 4	0.010 5

(2) 根据引力模型求解权重分量

将得到的占比代入式(4),得到基于数据集的自适应权重,如式(8)所示。

$$\begin{aligned}
 f_1 &= 0.571 0 \\
 f_2 &= 0.236 7 \\
 f_3 &= 0.029 3 \\
 f_4 &= 0.006 6
 \end{aligned}
 \tag{8}$$

即得 P2、P3、P4、P5 层的权重对应为 0.571 0、0.236 7、0.029 6、0.006 6。结果显示,自适应后的权重更偏向于下采样率最低的 P2 层。

1.3 级联的远距离行人小目标识别方法

本文通过将 1.2 中的特征增强和改进 YOLOv4 进行级联,构建小目标检测网络,提高小目标识别精度。

1) 疑似小目标判断

为了对小目标进行特征增强,首先需要对视场中的疑似小目标进行筛选、判断。首先统计自制数据库中目标,目标长宽比在 1.7 : 1 ~ 3.5 : 1 范围时符合

MSCOCO 数据集小目标定义^[5]的目标,占小目标总数的 89%,小目标的长宽比分布如图 7 所示。

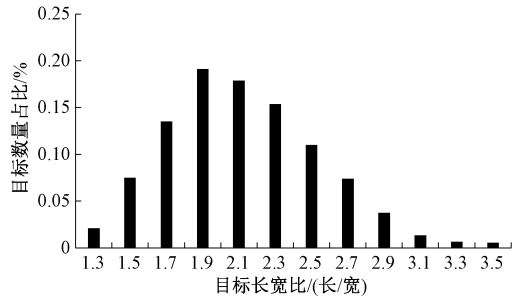


图 7 行人目标大小分布情况

Fig. 7 Size distribution of small target

其次,在不进行任何参数调整的前提下,使用 Faster RCNN 对自制数据集进行检测,经统计,测试目标在置信度小于 0.6 时满足小目标定义的数量占所有小目标的 87%,图 8 为所有小目标在测试时的置信度分布情况。

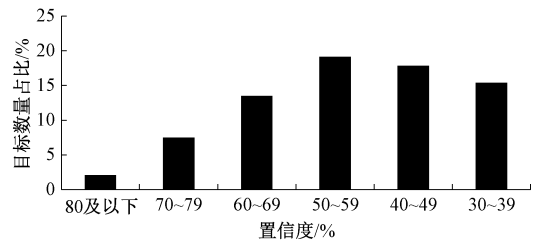


图 8 小目标测试置信度分布情况

Fig. 8 Confidence distribution of small target

根据以上分析,得出小目标判据如下:目标长宽比在 1.7 : 1 ~ 3.5 : 1 范围内;识别置信度低于 0.6。

整体判断流程如图 9 所示。

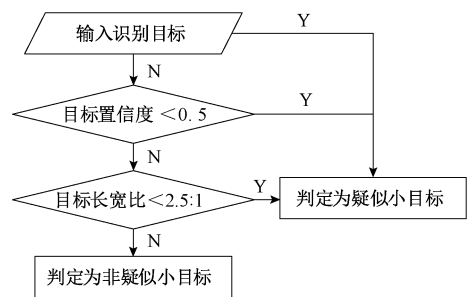


图 9 疑似小目标判断过程

Fig. 9 Suspecious small target judgment process

2) 方法框架

通过上述疑似小目标判据,将超分辨率生成对抗网络与改进的 YOLOv4 算法结合,形成最终的远距离行人小目标识别方法。方法框架如图 10 所示。

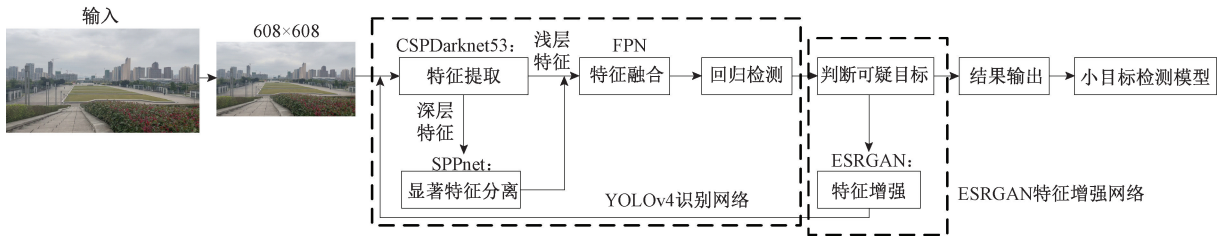


图 10 远距离小目标方法框架

Fig. 10 Framework of the long-distance small target recognition algorithm

具体步骤如下。

- (1) 输入任意尺寸的照片,并统一缩放到 608×608 大小;
- (2) 将统一尺寸的图像输入到改进后的 CSPDarknet53 主干特征提取层;
- (3) 经过 SPPnet 和特征金字塔把拥有不同语义信息的特征进行拼接处理,得到 $(19, 19, 75)$, $(38, 38, 75)$, $(76, 76, 75)$ 3 个尺度特征;

(4) 对于识别结果,使用,则将其判断为疑似小目标,并输入到超分辨率算法中,进行特征增强;

(5) 将增强后的图像覆盖于原图位置,根据目标相邻的最小距离调整增强后的目标位置,保证处理后的目标不与其他目标重叠,并重新进行识别;

(6) 将第 2 次识别目标的置信度赋给第 1 次的识别结果,以保证原始图像完整。

综上,整个方法实现流程如图 11 所示。

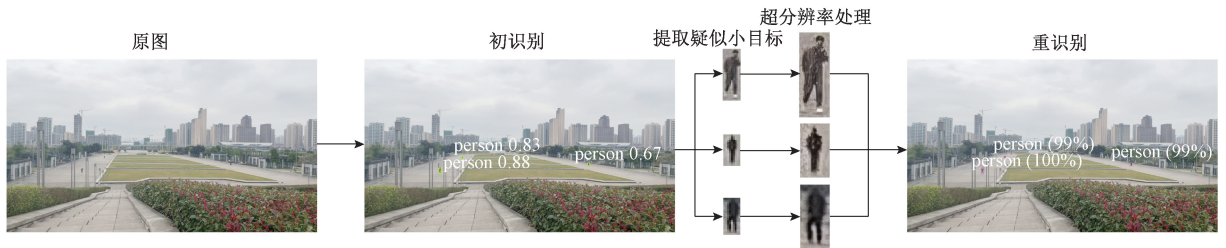


图 11 方法整体识别过程展示

Fig. 11 Analysis of the overall recognition process of the algorithm

2 分析与讨论

2.1 行人小目标数据集获取

1) 数据采集与处理平台

本文使用海康威视 DS-2ZMN3007 摄像头,配合飞越 FY-SP1510 二自由度转台,移动电源 SK6,稳压电源 LPV-300-24,搭建室外远距离监控环境。海康威视 DS-2ZMN3007 摄像头拥有较快的快门捕捉速度和 1080P 的高清分辨率,可视距离远,能捕捉较远视距下的行人小目标,本文通过采用 RSTP 网络通信协议与摄像头进行通信,获取视频信息,通过数据处理服务器对视频信息进行标定等操作。采集设备实物图如图 12 所示。

2) 实验数据集

(1) 对于深度学习识别算法,通过摄像头拍摄制作小目标行人数据集对算法训练和测试。图像分辨率 1920×1080 ,场景设定于地形起伏变化较小、无明显遮挡物的空旷场景,目标距离 $300 \sim 400$ m,行人目标在视场

占比约万分之二,绝对像素值约 15×15 ,符合 MSCOCO 数据集小目标定义。识别样本集共 1964 张,按照 6:2:2 的比例设置训练集、验证集和测试集。

(2) 对于特征增强算法,使用自制行人数据集进行模型训练,图像分辨率 1920×1080 ,场景设定于无明显遮挡物的空旷场景,目标距离 $10 \sim 50$ m,占据视场比例约 20%,通过裁剪制作高分辨率行人图像共 632 张,按 6:2:2 的比例设置训练集、验证集和测试集。使用 Set5、

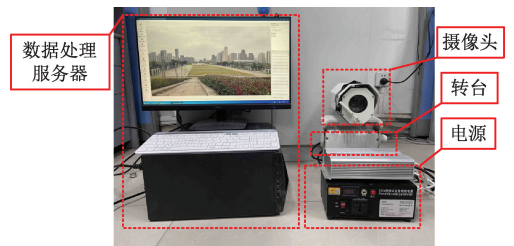


图 12 数据采集与处理平台

Fig. 12 Data acquisition and processing platform

Set14、BSD100 数据集和自制行人数据集对算法完成验证和测试。

3) 训练参数

训练网络硬件配置: Intel Core i7-8700K CPU 3.7 GHz 处理器, 32 GB 运行内存, GPU GeForce RTX 2070 8 G。算法试验环境为 Ubuntu16. 04, 使用 Python 和 C++语言编写网络完成训练和测试。

特征增强算法训练参数: 设置 ESRGAN 训练迭代次数 400 000, 输出与输入图像的放大长宽比例均为 4, 训练输入的图像块数量 128, 初始学习率 2×10^{-4} , 权重衰减周期为 250 000, 最小学习率 10^{-7} 。

识别算法训练参数: 设置 YOLOv4 训练进行迭代 12 000 次, 权重权值衰减速率为 0.000 5, 训练批量为 32, 动量 0.9, 初始学习率为 10^{-3} 。

2.2 模型评估

1) 特征增强算法指标

采用峰值信噪比 (PSNR) 和结构相似性 (SSIM) 两个指标作为衡量超分辨率结果好坏的量化指标。PSNR 公式如下:

$$PSNR = 10 \times \lg \left(\frac{MAX_I^2}{MSE} \right) \quad (9)$$

其中, MAX_I^2 表示图片像素点可能达到的最大像素值; MSE 表示真实图像 I 与虚拟图像 K 之间的均方误差。

SSIM 的公式如下:

$$SSIM(x, y) = l(x, y) \times c(x, y) \times s(x, y) \quad (10)$$

其中, $l(x, y)$ 为像素亮度, $c(x, y)$ 为像素对比度, $s(x, y)$ 为图像结构。

PSNR 和 SSIM 越高, 算法生成的虚假图像与真实图像的相似程度越高。

2) 识别算法指标

选用平均检测精度 (mAP), 交并比 (IOU) 作为算法性能比较的最终指标。由于实验对象仅为行人目标, 算法的平均检测精度 mAP 等效于行人类别的识别精度 AP, 公式如下。

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \quad (11)$$

式中: n 为类别总数, 本文 $n=1$; i 为类别序号; AP_i 为平均精度函数, 为该算法识别第 i 类目标的精确率。

通常情况下, 将 IoU = 0.5 时的 mAP 记作 $mAP_{0.5}$; 以 0.05 为间隔, 将 IoU 从 0.5 依次递增到 0.95, 再将在各 IoU 情况下求得的 mAP 求平均值, 记作 $mAP_{0.5:0.95}$ 。通过以上两个不同标准下的 mAP, 来衡量模型的检测精度。

2.3 目标特征增强结果分析

通过横向对比 Bicubic、SRCNN^[23]、SRGAN 三种超分辨率算法, 说明 ESRGAN 算法对于小目标增强的优点。图 13 所示为 ESRGAN 四种损失函数随时间变化的曲线。

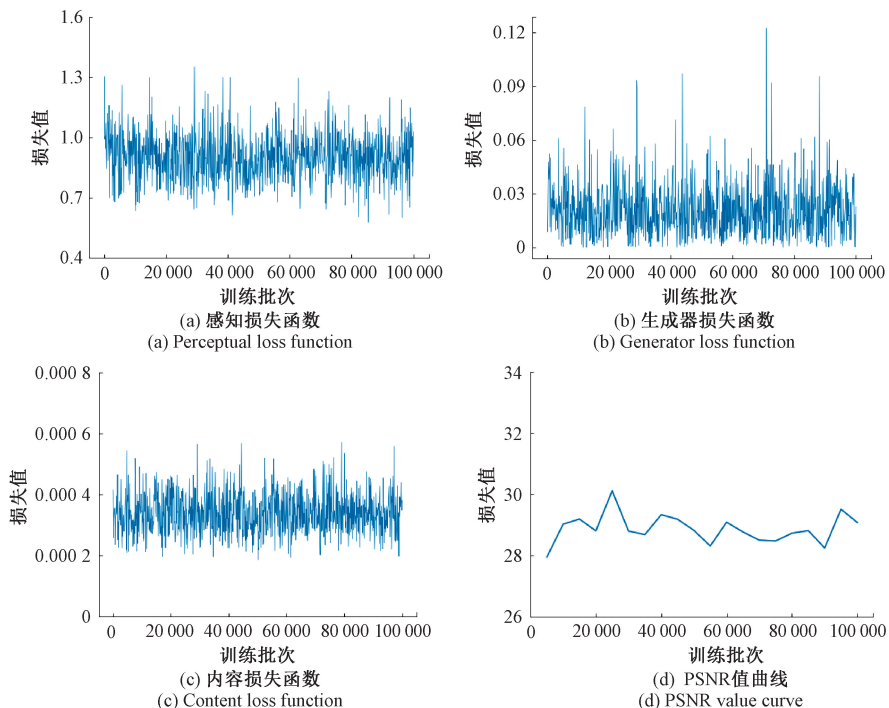


图 13 ESRGAN 损失函数。

Fig. 13 ESRGAN loss function

图 13 表明 ESRGAN 在迭代次数约 20 000 时, 各损失值已达到比较稳定的状态。实验分别记录各种算法在

3 种公开数据集和自制行人目标数据集下样本重建的平均 PSNR 和 SSIM 值, 放大比例设置为 4 倍, 即长宽各扩大为

4倍。表2、3所示为 Bicubic、SRCNN、SRGAN、ESRGAN 四种算法在放大比例为4倍时重建图像的 PSNR 和 SSIM 值。

表2 不同超分辨率算法的平均 PSNR

Table 2 Average PSNR of different super-resolution algorithms

数据集	放大倍数	PSNR/dB			
		Bicubic	SRCNN	SRGAN	ESRGAN
Set5	4	26.7	29.4	28.2	30.8
Set14	4	24.5	27.7	26.5	27.1
BSD100	4	23.7	26.1	26.8	26.2
自制数据集	4	24.4	27.1	27.5	27.3

表3 不同超分辨率算法的平均 SSIM

Table 3 Average SSIM of different super-resolution algorithms

数据集	放大倍数	SSIM			
		Bicubic	SRCNN	SRGAN	ESRGAN
Set5	4	0.94	0.94	0.89	0.87
Set14	4	0.84	0.91	0.85	0.75
BSD100	4	0.93	0.94	0.86	0.82
自制数据集	4	0.86	0.88	0.82	0.79

图14所示为各算法在自制数据集上的测试展示图,图14(a)~(d)下方小图从左至右依次为 Bicubic 生成图、SRCNN 生成图、SRGAN 生成图、ESRGAN 生成图、原分辨率图像。

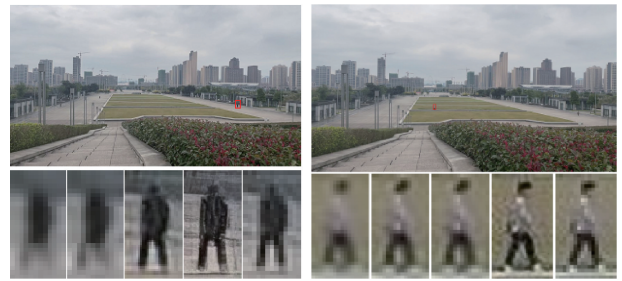
通过实验可以得出,Bicubic 的 PSNR 和 SSIM 值低于 SRCNN、SRGAN、ESRGAN 算法;SRCNN 在各个数据集的 PSNR 值位于第一或第二,SSIM 均位于第一;SRGAN 和 ESRGAN 的 PSNR、SSIM 值相差不大,低于 SRCNN。从最终重建效果的角度,Bicubic 观感最差;SRCNN、SRGAN 还原了部分细节,总体比较模糊;ESRGAN 增强细节最多,视觉效果更符合实际原目标,优于前3种算法。

综上,ESRGAN 在保持较高水平的 PSNR 的同时,相比于其他算法能生成更多的细节,且能拥有更好的重建观感,更符合原行人目标。

2.4 目标识别结果分析

1) 识别结果数据分析

由于本文提出的改进小目标检测算法具有较强的可移植性,可嵌入到大多数多尺度检测算法中,因此本文的对比重点为原始的 YOLOv4 算法,再选择主流小目标检测算法 ION 和 Faster RCNN 进行横向对比,总体说明算法的有效性。



(a) 人物1的重建对比
(a) Super-resolution reconstruction comparison of the first person

(b) 人物2的重建对比
(b) Super-resolution reconstruction comparison of the second person



(c) 人物3的重建对比
(c) Super-resolution reconstruction comparison of the third person

(d) 人物4的重建对比
(d) Super-resolution reconstruction comparison of the fourth person

图14 不同超分辨率算法的重建效果展示图

Fig. 14 Comparison results of reconstruction effects of super-resolution methods

首先对改进前后识别算法训练过程进行分析,两种模型对应的训练损失曲线如图15所示。

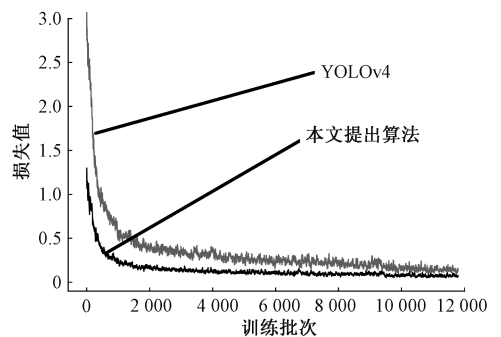


图15 训练损失曲线

Fig. 15 Training loss curves

由图15可得,原 YOLOv4 在训练迭代次数达到 10 000 次时基本收敛,达到平稳状态,损失值在 0.15 左右,最终损失值为 0.129 9;本文改进后的算法在训练迭代次数达到 2 000 次时基本收敛,损失值在 0.1 左右,最终损失值为 0.072。试验得出,改进后算法的收敛速度更快,最终损失值更小。

在相同的自制数据集下,本文通过比较 $mAP_{0.5}$ 、 $mAP_{0.5:0.95}$ 和 FPS,对 4 种算法的检测精度和实时性进行对比,相应的检测结果如表 4 所示。

表 4 各模型识别性能对比

Table 4 Comparison of recognition performance of each model

模型	特征提取网络	mAP _{0.5} /%	mAP _{0.5:0.95} /%	FPS
Faster-RCNN	Resnet101	61.5	31.1	56
ION	VGG16	60.4	36.2	83
YOLOv4	CSPDarknet53	74.9	40.8	61
本文算法	CSPDarknet53	86.7	51.4	24

结合检测结果,在检测精度上,本文提出的级联小目标检测算法在 mAP_{0.5} 上比 Faster-RCNN、ION 和原 YOLOv4 分别提高 25.2%、26.3%、11.9%;对于定位精度要求更高的 mAP_{0.5:0.95},本文算法分别提高 20.3%、25.2%、10.6%。数据显示,本文改进 YOLOv4 引入更浅层特征,构建自适应特征融合层,并级联超分辨率模型,对极小行人目标的特征进行增强,在行人小目标识别方面有较好的改善,在不同定位要求下的检测精度均高于前 3 种算法。

在算法运行速度上,ION 速度最快达到 83 FPS,本文提出的小目标检测算法平均处理每张图片的时间达 41.7 ms,仅 24 FPS,为 4 种算法中最低,勉强满足实时性,经统计,算法各部分消耗时间如表 5 所示。

表 5 本文级联算法各步骤消耗时间

Table 5 Time consumed by each step of the cascade algorithm in this article

步骤	特征生成	检测	结果刷新
耗时时间/ms	20.8	19.7	1.2
总计		41.7	

根据结果分析,特征生成消耗时间占总时间近 1/2,从而影响实时性。从关键帧检测的角度分析,本文算法在勉强满足实时性的情况下,检测精度有提大提升,能很好满足监控背景下的远距离关键帧检测,具有一定的可靠性。

2) 识别效果展示与分析

本文以 IoU=0.5 为定位标准,选取不同相邻情况和不同尺寸大小的小目标检测结果对算法进行分析,结果如图 16 所示,对应目标的置信度如表 6 所示。

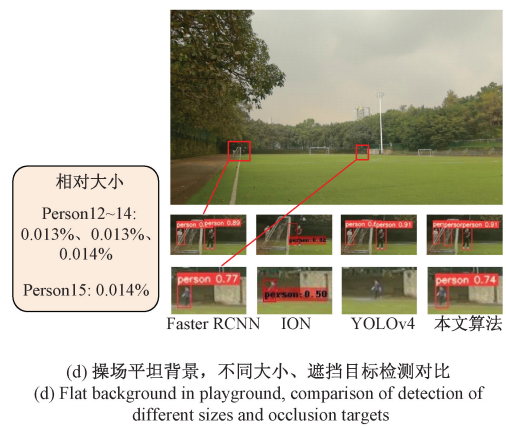
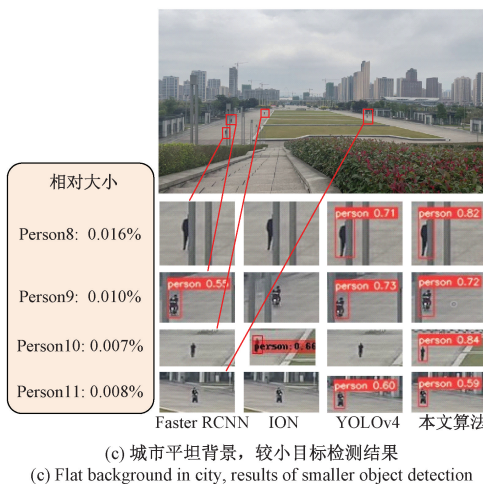
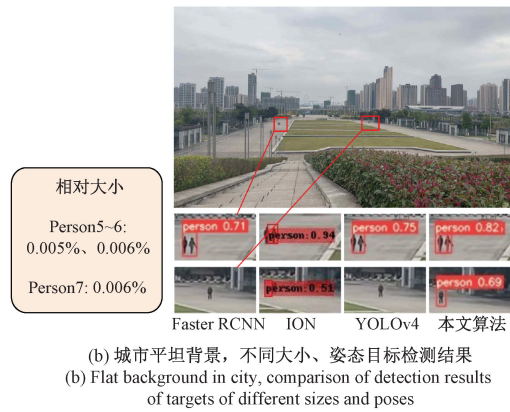
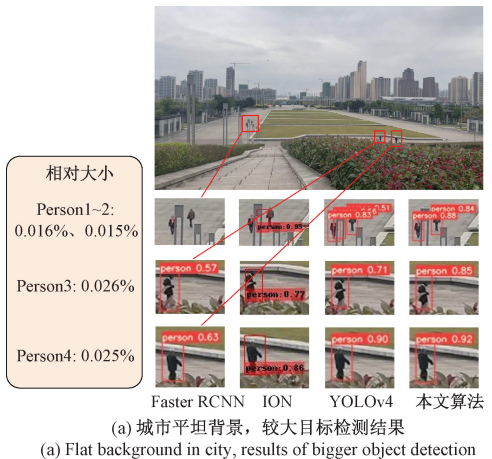


图 16 本文级联算法横向对比结果

Fig. 16 The results of the horizontal comparison between the cascade algorithm in this article and other algorithms

将图 16 中的目标按从左至右的顺序统计其相对大小和检测置信度,目标置信度结果如表 6 所示。

表 6 各算法识别第 331 张图像的目标置信度
Table 6 Four methods to identify the confidence of the 331th image

目标(相对大小)	Faster RCNN	ION	YOLOv4	本文算法
Person 1(0.016%)	Lost	Lost	83%	88%
Person 2(0.015%)	Lost	95%	51%	84%
Person (误检)	-	-	56%	no
Person 3(0.026%)	57%	77%	71%	85%
Person 4(0.025%)	63%	86%	90%	92%
Person 5(0.005%)	71%	94%	75%	82%
Person 6(0.006%)	Lost	Lost	Lost	55%
Person 7(0.006%)	Lost	51%	Lost	69%
Person 8(0.016%)	Lost	Lost	71%	82%
Person 9(0.010%)	55%	Lost	73%	72%
Person 10(0.007%)	Lost	66%	Lost	84%
Person 11(0.008%)	Lost	Lost	60%	59%
Person 12(0.013%)	59%	Lost	64%	75%
Person 13(0.013%)	Lost	Lost	Lost	54%
Person 14(0.014%)	89%	82%	91%	91%
Person 15(0.014%)	77%	50%	Lost	74%

图 16 和表 6 所示为识别结果与所有目标检测的置信度数据。本文将从目标大小、不同姿态、遮挡和误检情况对结果进行分析。

(1) 不同大小

试验结果中目标的相对大小可分为小于 0.001%、0.001%~0.002%、大于 0.002% 三类。目标 5、6、7、10、11 为大小最小的 0.001% 以下,为微小目标;目标 1、2、8、9、12~15 大小在 0.001%~0.002% 内,为一般小目标;目标 3、4 大小大于 0.002%,为较大小目标。

对于不同尺度的小目标,本文提出算法由于对 YOLOv4 改进,能提取更细粒度的特征,且使用超分辨率进行特征生成,对所有目标均无漏检,且置信度最高。其他 3 种算法存在不同程度的漏检,且检测置信度较低,容易形成误检,如原 YOLOv4 算法出现误检目标。

(2) 不同姿态情况

展示图中,目标 9、11 为骑行行人目标,区别于一般站立行人,此时 3 种算法均出现一定的漏检情况。本文提出算法能较好适应不同姿态的行人小目标检测,无明显漏检现象。

(3) 不同遮挡情况

实验展示如目标 1、8、13 中的遮挡情况,此时 Faster RCNN 和 ION 的漏检现象严重,本文算法无漏检且置信度最高。

经对比,Faster RCNN 作为经典二阶段算法,对小目标有一定检测能力,ION 和 YOLOv4 作为近年较好的小目标检测算法也有较好效果,,但对于极小目标和在不同姿态、遮挡情况下,仍然会出现漏检。本文提出算法在 YOLOv4 的基础上进行改进并级联特征生成算法,对于小目标检测有明显改善效果,对不同尺寸、不同姿态、不同遮挡情况的行人小目标,识别精度均有较大提升,具有较好的可靠性。

2.5 结果分析

1) 定性分析:对于不同尺寸大小、不同姿态和遮挡情况下的行人小目标,主流小目标检测算法 Faster RCNN、ION 和 YOLOv4 存在漏检、置信度低、定位精度差的问题。本文提出的小目标算法通过设计自适应的特征融合层,更好的提取和融合了细粒度的特征,并用超分辨率增大了目标的相对大小,对占比 0.002% 以下的小目标有较好的改善,识别精度有一定的提升。2) 定量分析:通过自制行人小目标数据集,选择 3 种主流小目标检测算法与本文算法进行比较。相比之下,本文算法对行人小目标检测的 $mAP_{0.5}$ 分别提高了 25.2%、26.3%、11.9%, $mAP_{0.5;0.95}$ 提高了 20.3%、15.2%、10.6%。本文提出的算法运算速度在 RTX2070 8 G 的环境下达到 24 FPS,勉强满足实时性要求。

3 结 论

通过对远距离下的行人小目标识别算法进行研究,本文设计了一种针对远距离行人小目标的改进 YOLOv4 与 ESRGAN 级联检测算法。通过引入更浅层特征对原 YOLOv4 的 FPN 进行改进,并设计了基于引力模型的自适应特征融合层,更充分的提取和融合细粒度的行人特征。经改进 YOLOv4 初识别后,根据疑似小目标判据提取疑似目标,并使用 ESRGAN 算法进行超分辨率特征增强,再使用改进 YOLOv4 对增强后的行人目标进行二次识别。

试验表明,本文提出的方法相比于原 YOLOv4,对 0.002% 以下的小目标和相邻、遮挡小目标的识别精度有较好的提升,相比原 YOLOv4, $mAP_{0.5}$ 提高了 11.9%, $mAP_{0.5;0.95}$ 提高了 10.6%。FPS 从 61 下降为 24,在满足实时性的同时,提高小目标检测精度,对远距离安防监控领域有重要应用价值。

下一步研究工作致力在不降低检测性能的前提下简化算法结构、精简运算流程,提高算法实时性,检测距离更远、更小的行人目标。

参考文献

- [1] 刘颖,刘红燕,范九伦,等. 基于深度学习的小目标检测研究与应用综述[J]. 电子学报, 2020, 48(3): 590-624.
- LIU Y, LIU H Y, FAN J L, et al. A survey of research and application of small object detection based on deep learning[J]. Acta Electronica Sinica, 2020, 48(3): 590-624.
- [2] SUN C, AI Y, WANG S, et al. Mask-guided SSD for small-object detection[J]. Applied Intelligence, 2021, 51(20):1-12.
- [3] 孙丹,王莉莉. 基于空-频域映射与虚警抑制的弱小目标检测算法[J]. 电子测量与仪器学报, 2019, 33(1): 31-39.
- SUN D, WANG L L. Dim target detection algorithm based on spatial-frequency domain mapping and false alarm suppression[J]. Journal of Electronic Measurement and Instrumentation, 2019, 33(1): 31-39.
- [4] KE X, LIN X, QIN L. Lightweight convolutional neural network-based pedestrian detection and re-identification in multiple scenarios [J]. Machine Vision and Applications, 2021, 32(2):1-23.
- [5] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft coco: Common objects in context [C]. ComputerVision-ECCV 2014, Cham, 2014: 740-755.
- [6] LEE G, HONG S, CHO D. Self-supervised feature enhancement networks for small object detection in noisy images[J]. Ieee Signal Processing Letters, 2021, 28: 1026-1030.
- [7] 鞠默然, 罗海波, 王仲博, 等. 改进的 YOLO v3 算法及其在小目标检测中的应用[J]. 光学学报, 2019, 39(7):245-252.
- JU M R, LUO H B, WANG ZH B, et al. Improved YOLO v3 algorithm and its application in small target detection [J]. Acta Optica Sinica, 2019, 39(7): 245-252.
- [8] 宁强, 秦鹏杰, 石欣, 等. 复杂地面背景下的红外目标检测算法[J]. 光子学报, 2019, 48(4):188-200.
- NING Q, QIN P J, SHI X, et al. Infrared target detection algorithm under complex ground background [J]. Acta Photonica Sinica, 2019, 48(4):188-200.
- [9] LIAN J, YIN Y, LI L, et al. Small object detection in traffic scenes based on attention feature fusion [J]. Sensors, 2021, 21(9):3031.
- [10] 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, 2017, 39(6):1137-1149.
- [11] 王瑞, 史天运, 包云. 一种基于视频的铁路周界入侵检测智能综合识别技术研究[J]. 仪器仪表学报, 2020, 41(9):190-197.
- WANG R, SHI T Y, BAO Y. Research on an intelligent comprehensive recognition technology in railway perimeter intrusion detection based on videos [J]. Chinese Journal of Scientific Instrument, 2020, 41(9):190-197.
- [12] BELL S, ZITNICK C L, BALA K, et al. Inside-outside net: Detecting objects in context with skip pooling and recurrent neural networks [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2874-2883.
- [13] YANG D, J ZHANG, XU S, et al. Real-time pedestrian detection via hierarchical convolutional feature [J]. Multimedia Tools and Applications, 2018, 77(19): 25841-25860.
- [14] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection[J]. ArXiv Preprint, 2020, ArXiv: 2004.10934.
- [15] REDMON J, FARHADI A. YOLOv3: An incremental improvement [J]. ArXiv Preprint, 2018, ArXiv: 1804.02767.
- [16] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.
- [17] 崔艳鹏, 王元皓, 胡建伟. 一种改进 YOLOv3 的动态小目标检测方法[J]. 西安电子科技大学学报, 2020, 47(3):1-7.
- CUI Y P, WANG Y H, HU J W. Detection method for a dynamic small target using the improved YOLOv3 [J]. Journal of Xidian University, 2020, 47(3):1-7.
- [18] 郑太雄, 江明哲, 冯明驰. 基于视觉的采摘机器人目标识别与定位方法研究综述[J]. 仪器仪表学报, 2021, 42(9):28-51.
- ZHENG T X, JIANG M ZH, FENG M CH. Vision based target recognition and location for picking robot: A

- review[J]. Chinese Journal of Scientific Instrument, 2021, 42(9): 28-51.
- [19] WANG X, YU K, WU S, et al. Esrgan: Enhanced super-resolution generative adversarial networks [C]. Proceedings of the European Conference on Computer Vision (ECCV) Workshops, 2018.
- [20] LEDIG C, THEIS L, HUSZÁR F, et al. Photo-realistic single image super-resolution using a generative adversarial network [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4681-4690.
- [21] PANG J, CHEN K, SHI J, et al. Libra r-cnn: Towards balanced learning for object detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 821-830.
- [22] 杜隆胤. 基于万有引力定律的分类方法研究[J]. 计算机应用与软件, 2013, 30(2): 205-207.
DU L Y. An exploration of classification based on the law of universal gravitation [J]. Computer Applications and Software, 2013, 30(2): 205-207.
- [23] DONG C, LOY C C, HE K, et al. Learning a deep convolutional network for image super-resolution [C].

European Conference on Computer Vision, Springer, Cham, 2014: 184-199.

作者简介



石欣(通信作者), 分别于 2003 和 2010 年于重庆大学分别获得硕士学位和博士学位, 现为重庆大学教授、博士生导师, 主要研究方向为智能控制、模式识别。

E-mail: meetshixin@gmail.com

Shi Xin (Corresponding author) received his M. Sc. degree and Ph. D. degree both from Chongqing University in 2003 and 2010, respectively. He is currently a professor and a Ph. D advisor at Chongqing University. His main research interests include intelligent control and pattern recognition.



卢灏, 2019 年于西南科技大学获得学士学位, 现为重庆大学硕士研究生, 主要研究方向为模式识别与计算智能。

E-mail: haolu145@foxmail.com

Lu Hao received his B. Sc. degree from Southwest University of Science and Technology in 2019. He is currently a master student at Chongqing University. His main research interests include pattern recognition and intelligent computing.