

基于YOLOv5s和双稳随机共振的夜间车辆检测算法

胡鹏飞, 王友国, 翟其清, 颜俊, 柏泉

引用本文

胡鹏飞, 王友国, 翟其清, 颜俊, 柏泉. [基于YOLOv5s和双稳随机共振的夜间车辆检测算法](#)[J]. 计算机科学, 2024, 51(9): 173-181.

HU Pengfei, WANG Youguo, ZHAI Qiqing, YAN Jun, BAI Quan. [Night Vehicle Detection Algorithm Based on YOLOv5s and Bistable Stochastic Resonance](#) [J]. Computer Science, 2024, 51(9): 173-181.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于MLIR的FP8量化模拟与推理内存优化](#)

FP8 Quantization and Inference Memory Optimization Based on MLIR

计算机科学, 2024, 51(9): 112-120. <https://doi.org/10.11896/jsjcx.230900143>

[基于轻量化YOLOv5s的集装箱锁孔识别算法](#)

Container Lock Hole Recognition Algorithm Based on Lightweight YOLOv5s

计算机科学, 2024, 51(6A): 230900021-6. <https://doi.org/10.11896/jsjcx.230900021>

[基于改进的YOLOv5s列车轨道线检测方法](#)

Railway Track Detection Method Based on Improved YOLOv5s

计算机科学, 2023, 50(11A): 230200101-6. <https://doi.org/10.11896/jsjcx.230200101>

[结合门控融合网络和残差傅里叶变换重建太阳斑点图](#)

Reconstruction of Solar Speckle Image Combined with Gated Fusion Network and Residual Fourier Transform

计算机科学, 2023, 50(11A): 220800229-7. <https://doi.org/10.11896/jsjcx.220800229>

[改进YOLOv5s的轻量化钢材表面缺陷检测模型](#)

Improved YOLOv5s Lightweight Steel Surface Defect Detection Model

计算机科学, 2023, 50(11A): 230900113-7. <https://doi.org/10.11896/jsjcx.230900113>

基于 YOLOv5s 和双稳随机共振的夜间车辆检测算法

胡鹏飞¹ 王友国^{1,2} 翟其清¹ 颜俊² 柏泉¹

1 南京邮电大学理学院 南京 210023

2 南京邮电大学通信与信息工程学院 南京 210023

(984253044@qq.com)

摘要 针对夜间车辆检测过程中光照不强导致漏检误检的问题,基于 YOLOv5s 和双稳随机共振提出一种改进的夜间车辆检测算法。YOLOv5s 从 4 方面进行改进:1)在 Backbone 和 Neck 中更换细小结构,提高网络对小目标的检测能力;2)加入由坐标注意力 CA 和能量注意力 SimAM 构成的双注意力机制,提高网络对目标的特征提取能力;3)采用轻量化骨干 Fasternet,减少模型参数量;4)在 Head 中采用 WIoU 损失函数,加快边界框回归损失的收敛速度。利用经典的双稳随机共振对夜间车辆数据集进行低照度图像增强,分析其有效性,并将增强后的夜间车辆图像传入改进的 YOLOv5s 网络进行训练。实验结果表明,相较于原始 YOLOv5s,融合改进的 YOLOv5s 和双稳随机共振的夜间车辆检测算法在执行远景小目标以及密集遮挡的夜间车辆检测任务时具有更高的准确率和更低的漏检率。

关键词: 双稳随机共振;低照度图像增强;YOLOv5s;双注意力机制;轻量化骨干

中图分类号 TP391

Night Vehicle Detection Algorithm Based on YOLOv5s and Bistable Stochastic Resonance

HU Pengfei¹, WANG Youguo^{1,2}, ZHAI Qiqing¹, YAN Jun² and BAI Quan¹

1 School of Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

2 School of Communication and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

Abstract Aiming at the problems of missed and false detection caused by weak illumination during night vehicle detection, an improved night vehicle detection algorithm is proposed based on bistable stochastic resonance and YOLOv5s. YOLOv5s is improved from four aspects, replacing small structures in Backbone and Neck to improve the detection ability of the network to small targets. A dual attention mechanism composed of coordinate attention CA and energy attention SimAM is added to improve the feature extraction ability of the network to the target. The lightweight backbone Fasternet is adopted to reduce the amount of model parameters. The WIoU loss function is used in Head to accelerate the convergence speed of bounding box regression loss. The effectiveness of the nighttime vehicle dataset is analyzed from quantitative and qualitative perspectives by using classical bistable stochastic resonance, and the enhanced nighttime vehicle images are passed into the improved YOLOv5s network for training. Experimental results show that, compared with the original YOLOv5s, the night vehicle detection algorithm combining improved YOLOv5s and bistable stochastic resonance has better accuracy and lower missed detection rate when performing long-range small targets and densely occluded night vehicle detection tasks.

Keywords Bistable stochastic resonance, Low-light image enhancement, YOLOv5s, Dual attention mechanism, Lightweight backbone

1 引言

车辆检测是智能驾驶过程中对周围环境感知尤为重要的一个环节,随着深度学习的发展,基于卷积神经网络的目标检测算法在自动驾驶领域取得了巨大的成就^[1],但是由于夜间光照不足等因素,误检率和漏检率较高,使得目标检测算法性能有所下降。

基于卷积神经网络的目标检测算法大致分为两阶段目标检测算法和单阶段目标检测算法。两阶段目标检测算法根据图像提取候选框,然后基于候选区域做二次修正得到检测点结果,代表算法有 RCNN^[2], Fast R-CNN^[3], Faster R-CNN^[4]。该类方法检测精度较高,但检测速度较慢。因此,大多数学者选择改进单阶段目标检测算法如 YOLOv5s 等,该类方法结构简单、计算高效,同时具备不错的检测精度。

到稿日期:2023-06-07 返修日期:2023-11-15

基金项目:国家自然科学基金(62071248)

This work was supported by the National Natural Science Foundation of China(62071248).

通信作者:王友国(wangyg@njupt.edu.cn)

Long 等^[5]引入轻量化特征增强模块(CARAFE),采用改进的多尺度特征融合网络,提出了一种轻量高效的 YOLOv5s 网络,提高了白天场景下小目标的检测精度;Zhang 等^[6]在 C3 结构中引入多头自注意力机制模块,借鉴双向特征金字塔的融合计算方式,对不同层次的特征赋予权重计算,该算法兼顾了车辆检测精度和检测速度;Deng 等^[7]提出结合级联注意力机制(通道、空间)的车辆检测算法 CAM-YOLO,能够较好地应对白天场景下较为复杂的车辆检测任务。这 3 种检测算法均侧重于白天场景,而忽略了夜间场景下的车辆检测任务。Zou 等^[8]基于 YOLOv5s 提出了一种夜间车辆检测算法,首先对传统低光照图像增强算法 HE, LIME, BM3D, RetinexNet 进行了对比,采用 LIME 算法对原始图像进行增强,将增强后的图像传入 YOLOv5s 网络进行训练;Guo 等^[9]采用 MobileNetV3 网络替换原始 YOLOv4 中的主干网络,引入 CBAM 注意力机制,利用 HDR 暗光图像增强技术进行图像预处理。这两种算法能够提高夜间远处车辆和被遮挡车辆的检测精度。

传统低光照图像增强算法能够提高图片的亮度,但图片色彩会出现偏差,同时会产生许多噪点,导致图片失真、细节丢失严重。然而,随机共振(SR)理论指出,在某些非线性系统中,适量的噪声能够增强信息传输,提高系统性能。20 世纪 80 年代初,Benzi 等^[10]首次提出“随机共振”这一概念。因具有独特的性能,随机共振很快被应用于弱信号检测^[11]、图像处理^[12]、生物信号处理^[13]等领域。近年来,有学者将双稳^[14]、四稳^[15]随机共振应用于低照度彩色图像增强领域。本文首先基于经典的双稳随机共振进行低照度图像增强,从定量与定性的角度验证了双稳随机共振增强有效性;然后从提升小目标检测能力、提升目标特征提取能力、使模型轻量化、修改损失函数 4 个方面改进 YOLOv5s。算法流程图如图 1 所示。

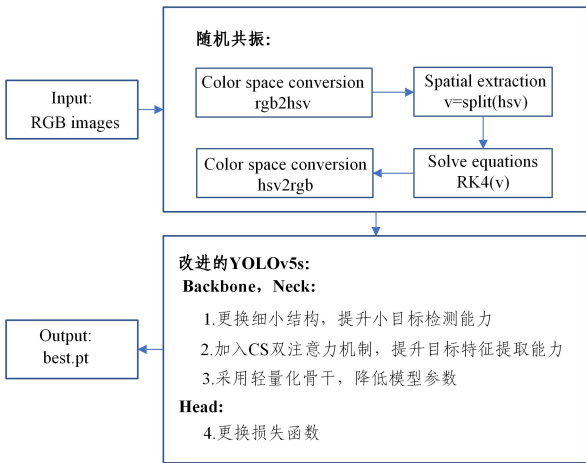


图 1 算法流程示意图

Fig. 1 Algorithm flowchart

2 低照度图像增强

采用经典的双稳随机共振进行图像增强,其朗之万方程如下:

$$\frac{dx(t)}{dt} = -\frac{\partial U(x,t)}{\partial x} + s(t) + \eta(t) \quad (1)$$

$$U(x,t) = -\frac{a}{2}x^2 + \frac{b}{4}x^4 \quad (2)$$

其中, $s(t)$ 是图像信号, $\eta(t)$ 是噪声(通常为高斯白噪声), $U(x,t)$ 为势函数。如图 2 所示,势垒高度为 $\Delta U = a^2/4b$,不稳定点 $x_0 = 0$,两个稳态点(势阱)为 $x_{\pm} = \pm \sqrt{a/b}$, a 和 b 是双稳态双阱参数。

图 3 简单描述了过阻尼布朗粒子在双稳态势阱间的跃迁过程(a)-(b)-(c)-(d)-(a)。 $s(t)$ 使双稳随机共振势函数变得不对称,当 $s(t)$ 的信号幅值较小时,很难促使粒子在两势阱之间进行周期性跃迁,此时向系统加入适当强度的噪声 $\eta(t)$ 可以促使粒子从一个势阱跃迁到另一个势阱中。随着噪声强度 $\eta(t)$ 的增大,会有一部分噪声能量向信号能量转移,促使系统的输出信号有充足的能量跃过势垒高度 ΔU ,并在两势阱之间进行跃迁,该现象即为随机共振现象。对于图像信号而言,图像中的每个像素值可以被当作一个粒子,而低照度图像大部分像素值的初始状态可以被看作是弱信号,通过加入适量的噪声促使粒子跃过势垒高度,进而向强信号过渡,从而实现低照度图像的增强。

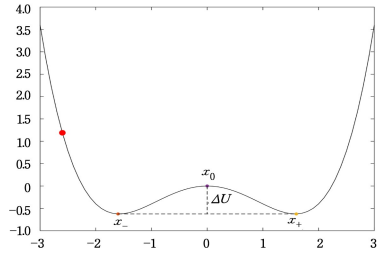


图 2 双稳随机共振势函数

Fig. 2 Function of bistable stochastic resonance

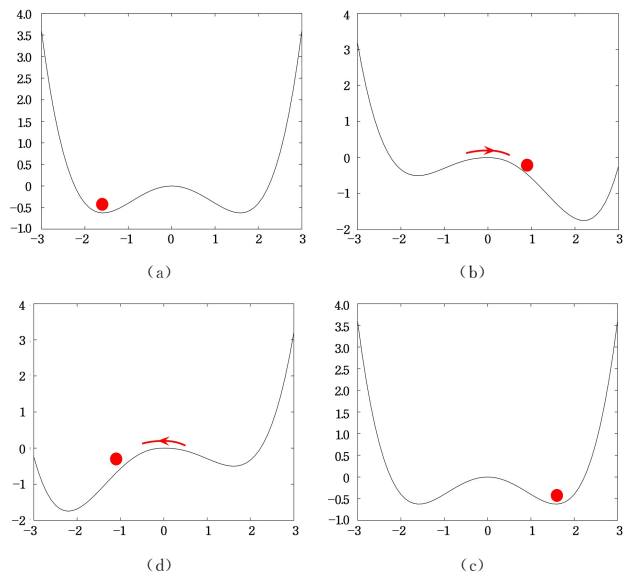


图 3 过阻尼布朗粒子在双稳态势阱间的跃迁示意图

Fig. 3 Schematic diagram of the transition of superdamping Brownian particles between bistable wells

双稳随机共振增强流程如下:

步骤 1 加载原始图像,对 RGB 图像求像素平均值 μ 。由于不同图像的明暗程度不同,因此将 μ 划分为 $[0,20)$, $[20,35)$, $[35,50)$, $[50,255]$ 4 个区间,每个区间的参数设置有所区别。

步骤 2 将 RGB 图像转为 HSV 图像,利用四阶龙格-

库塔方法(见式(3))对 V 空间求解朗之万方程,迭代步数、噪声强度等参数均由 μ 决定。

$$\begin{cases} k_1 = g_n \\ k_2 = g\left(t_n + \frac{h}{2}, u_n + \frac{h}{2}k_1\right) \\ k_3 = g\left(t_n + \frac{h}{2}, u_n + \frac{h}{2}k_2\right) \\ k_4 = g(t_{n+1}, u_n + hk_3) \\ u_{n+1} = u_n + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4) \end{cases} \quad (3)$$

步骤 3 将增强后的 HSV 图像转换为 RGB 图像,得到增强后的结果。

将双稳随机共振增强算法与直方图均衡化(HE)、具有色彩恢复的多尺度 Retinex 算法(MSRCR)、四稳随机共振增强算法进行比较,对低照度图像增强后的图像如图 4 所示。原图(a)属于均匀低光照图像,(b)属于非均匀低光照图像,(c)属于非常不均匀的低光照图像。HE 算法增强后的图像产生了很多噪点,曝光严重,同时很多纹理特征被丢失。MSRCR 算法增强后的图像色彩出现明显偏差,图像失真严重。双稳随机共振算法增强后的图像整体亮度均有所提升,可以看出,(a)列色彩更加鲜艳饱满,灯光没有产生曝光,图片的纹理特征也较为清晰,同时左侧边缘的公交车不容易被忽略;(b)列整体亮度提升最为明显,色彩更加真实;(c)列局部亮度提升明显,树枝、边缘车辆等局部物体纹理特征较原图更加清晰。相较于双稳随机共振,四稳随机共振算法增强后的图像整体亮度更高,但同时图片局部产生了微弱曝光,会有小部分细节丢失,如(a)列中灯光周围、(c)列中左侧面包车背部,同时四稳随机共振增强算法时间复杂度明显高于双稳随机共振增强算法。整体而言,双稳随机共振通过加入适量噪声,能有效增强低光照图像,同时能够保留图像的纹理特征。

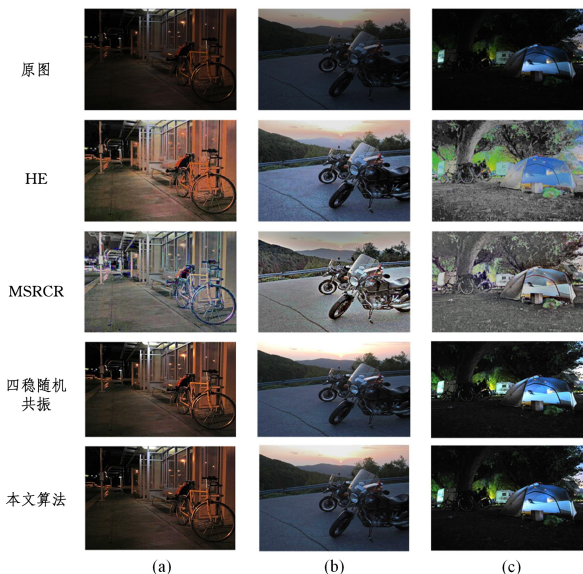


图 4 原始图像与增强图像对比图

Fig. 4 Comparison of primitive images and enhanced images

另外,采用 PSNR、SSIM、互信息(Mutual Information)3 个指标对增强后的图像进行定量分析,结果如表 1 所列,同时对

增强后图像的噪声方差进行估计,如表 2 所列。

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (4)$$

其中, μ_x 和 μ_y 分别代表 x 和 y 的平均值, σ_x 和 σ_y 分别代表 x 和 y 的标准差, σ_{xy} 代表 x 和 y 的协方差, c_1 和 c_2 分别为常数,防止分母为 0。

$$MSE = \frac{1}{3 \times H \times W} \sum_{R,G,B} \sum_{i=1}^H \sum_{j=1}^W (X(i, j) - Y(i, j))^2 \quad (5)$$

$$PSNR = 20 \log_{10} \left(\frac{2^n - 1}{\sqrt{MSE}} \right) \quad (6)$$

其中, $X(i, j)$, $Y(i, j)$ 分别代表图像 X, Y 在坐标 (i, j) 处的像素值; H, W 分别为图像的高度和宽度; n 为每像素的比特数,一般取 8。

$$H(Y) = -\sum P(y) \log(P(y)) \quad (7)$$

$$I(X; Y) = H(Y) - H(Y|X) \quad (8)$$

表 1 评价指标结果

Table 1 Results of evaluation indicators

	Algorithms	PSNR	SSIM	MI
(a)	HE	11.8384	0.2894	3.6899
	MSRCR	9.8394	0.2179	1.1340
	四稳随机共振	20.5779	0.6941	4.2708
	双稳随机共振	20.6974	0.6920	4.3012
(b)	HE	12.7490	0.4172	4.3285
	MSRCR	9.8876	0.1974	1.3013
	四稳随机共振	15.2099	0.6928	4.8698
	双稳随机共振	16.3547	0.6889	4.8158
(c)	HE	6.8151	0.0789	2.4254
	MSRCR	9.2024	0.1014	0.9975
	四稳随机共振	17.7433	0.5674	3.6544
	双稳随机共振	20.6092	0.6042	3.7711

由表 1 可知, HE 算法、MSRCR 算法的 PSNR 和 SSIM 值都很小,说明增强后的图像质量很差,与原始图像结构相似性程度很低;而随机共振增强后的 PSNR 和 SSIM 值相对较大,体现了随机共振增强算法的有效性。互信息 MI 表示不确定度的减少量,也就是 X 为 Y 提供的信息量, MI 值越大说明 X 为 Y 提供的信息量越大,说明增强后的图片细节信息越丰富。表 1 中, HE 和 MSRCR 增强后的图像 MI 值相对较小,说明增强后的图像细节信息丢失较为严重。对于随机共振而言,双稳、四稳随机共振增强后的图像 MI 值相对较大,当图 4 中(a)列和(c)列图像整体亮度较低时,双稳随机共振的 MI 值略大于四稳随机共振,当(b)列图像整体亮度较高时,四稳随机共振的 MI 值略大于双稳随机共振。

表 2 噪声方差估计结果

Table 2 Results of evaluation indicators

	(a)	(b)	(c)
HE	0.00297	0.00389	0.00261
MSRCR	0.00315	0.00331	0.00244
四稳共振	0.00319	0.00239	0.00234
本文算法	0.00255	0.00217	0.00229

由表 2 可知, 4 种增强算法的噪声方差估计值均保持在 0.002~0.003 之间,其中 HE 和 MSRCR 估计值同样相对较大,双稳随机共振噪声方差估计值略小于四稳随机共振。

3 改进 YOLOv5s 算法

3.1 YOLOv5s 理论算法

YOLOv5s v6.0 版本包括 input 数据输入端、Backbone 骨干网络、Neck 特征金字塔和 Head 目标检测头 4 部分。在数据输入端,采用 Mosaic 数据增强算法,随机选取 4 张图片,随机缩放、裁剪后拼接。在 Backbone 中,Conv 模块由标准卷积、BN 和 SiLU 激活函数构成;C3 模块是一种轻量级的卷积神经网络模块,由 3 个标准卷积、多个 Bottleneck 模块以及

Concat 模块构成;SPPF 模块是将输入的特征图串行通过 3 个 5×5 的最大池化层,作用是增大感受野。在 Neck 中,采用 FPN 和 PAN 的特征金字塔网络,通过上采样与下采样操作将不同层的特征图融合起来,生成了具有多尺度的特征图。Head 目标检测头主要由 Bounding box 损失函数(CIOU)以非极大值抑制(Soft-NMS)构成。

3.2 YOLOv5s 算法改进

本文从 4 方面改进 YOLOv5s,改进后的 YOLOv5s 模型整体结构如图 5 所示。

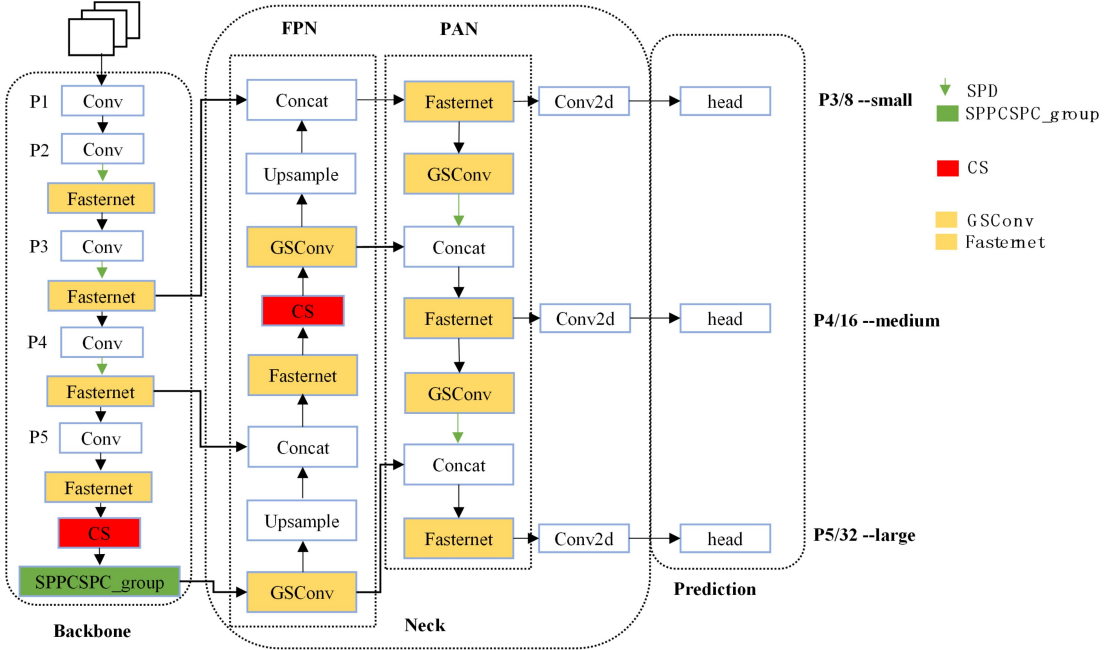


图 5 改进后的 YOLOv5s 模型结构图

Fig. 5 Diagram of improved YOLOv5s model structure

3.2.1 提升小目标检测能力

采用 SPD 结构^[16]替换下采样操作,如图 6 所示,当 $scale=2$ 时,输入的原始特征图的维度是 $s \times s \times c$,我们分别沿着 x 与 y 轴方向,间隔选取,即蓝、金、黑、白,会得到 4 个大小为 $s/2 \times s/2 \times c$ 的特征图,进而沿着通道方向进行 Concat 拼接操作,从而得到大小为 $s/2 \times s/2 \times 4c$ 的特征图。当 $scale=2$ 时,SPD 结构等价于 $stride$ 为 2 的下采样操作,该方法不丢失可学习信息,会增加少量的参数量。在改进后的 YOLOv5s 模型中,将骨干网络 Backbone 中 P2, P3, P4 之后的下采样、Neck 中的下采样替换为 SPD 结构。

采用 SPPCSPC^[17]模块替换 SPPF 模块,如图 7 所示,SP-PCSPC 模块首先将特征图分为两部分,其中一部分通过 4 条最大池化操作分支,kernel size 分别为 1, 5, 9, 13, 这样可以获得 4 种感受野,从而可以更好地区别大目标和小目标;另一部分进行常规的卷积操作,然后将两部分结果进行拼接。在实验过程中,最终采用 SPPCSPC_group 模块,该模块与 SPPCSPC 模块结构完全相同,前者 Conv 模块中参数 $groups$ 设为 4,后者 $groups$ 为 1。分组卷积可以把原来的卷积操作分成 $groups$ 个小组分别处理,这样可以减少卷积操作的参数量和计算量。

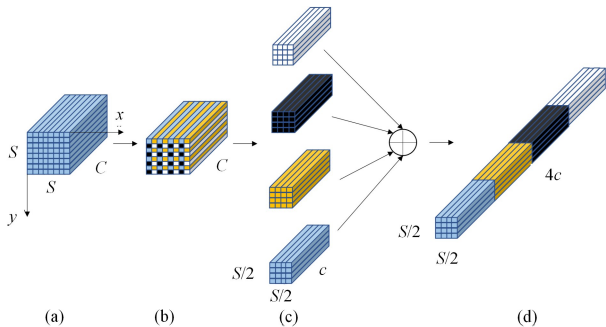


图 6 $scale=2$ 的 SPD 结构图

Fig. 6 SPD structure diagram with $scale=2$

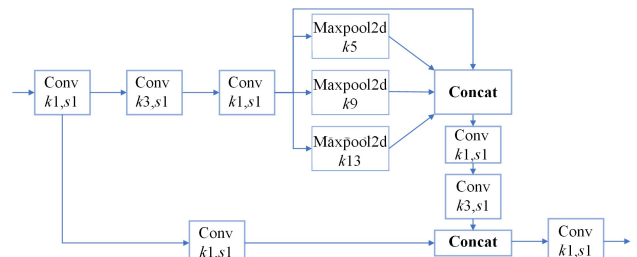


图 7 SPPCSPC 结构图

Fig. 7 SPPCSPC structure diagram

3.2.2 基于 CS 双注意力机制特征提取模块

双注意力机制由坐标注意力 CA^[18] (Coordinate Atten-

tion)和能量注意力 SimAM^[19] 串行构成,记为 CS 注意力机制,结构如图 8 所示。

CA 模块可以看作是一个用来增强特征表示能力的计算模块,CA 使用尺寸为 $(H, 1)$ 和 $(1, W)$ 的池化核沿着 X 坐标和 Y 坐标对输入特征图的每个通道进行全局平均池化,然后将两个方向的特征图拼接在一起,使用共享的 1×1 卷积进行卷积操作,得到大小为 $1 \times (W + H) \times C/r$ 的中间特征 f ,然后将特征图 f 切分成单独的张量 $f^h \in R^{C/r \times H}$ 和 $f^w \in R^{C/r \times W}$,分别与两个大小为 1×1 的卷积核 F_h 和 F_w 进行卷积操作,得到与 X 具有相同通道数的注意力权重 g^h 和 g^w ,最后与原始特征图进行加权计算,得到最终输出 Y 。

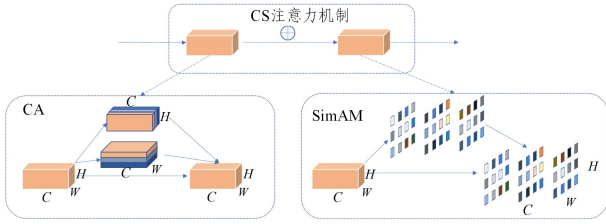


图 8 CS 注意力机制结构图

Fig. 8 CS attention mechanism structure diagram

SimAM 注意力模块受启发于人脑注意力机制,其采用能量函数计算注意力权重,是一种无需额外参数的轻量型 3D 注意力模块。能量函数如式(9)所示:

$$e_i(\omega_i, b_i, y, x_i) = \frac{1}{M-1} \sum_{i=1}^{M-1} (-1 - (\omega_i x_i + b_i))^2 + (1 - (\omega_i t + b_i))^2 + \lambda \omega_i^2 \quad (9)$$

其中, $M = H \times W$ 是每个通道的能量个数, t 和 x_i 分别是输入特征图中单个通道的目标神经元和其他神经元, ω_i 和 b_i 是变换的权重和偏置, λ 是正则化系数。式(9)存在关于 ω_i 和 b_i 的解析解:

$$\omega_i = -\frac{2(t - \mu_i)}{(t - \mu_i)^2 + 2\sigma_i^2 + 2\lambda}, b_i = -\frac{1}{2}(t + \mu_i)\omega_i \quad (10)$$

其中, $\mu_i = \frac{1}{M-1} \sum_{i=1}^{M-1} x_i$, $\sigma_i^2 = \frac{1}{M-1} \sum_{i=1}^{M-1} (x_i - \mu_i)^2$ 是该通道中除 t 之外的所有神经元的均值和方差, ω_i 和 b_i 是在单个通道上获得的,假设每个通道中的所有神经元都遵循相同的分布,进而得到最小能量解:

$$e_i^* = \frac{4(\hat{\sigma}_i^2 + \lambda)}{(t - \hat{\mu})^2 + 2\hat{\sigma}_i^2 + 2\lambda} \quad (11)$$

其中, $\hat{\mu} = \frac{1}{M} \sum_{i=1}^M x_i$, $\hat{\sigma}^2 = \frac{1}{M} \sum_{i=1}^M (x_i - \hat{\mu})^2$, 每个神经元的重要性可由 $1/e_i^*$ 得到,最终结果如式(12)所示:

$$\tilde{X} = \text{sigmoid}\left(\frac{1}{E}\right) \odot X \quad (12)$$

CA 模块在考虑通道维度的同时,将空间维度拆分成 X 方向和 Y 方向,可以通过学习得到自适应通道权重。在此基础上,SimAM 将通道维度和空间维度统一,每个通道内所有神经元均与其他神经元存在线性可分性,通过评估每个神经元的重要性,找出信息丰富的神经元,使得模型关注更加有用的信息。本文将 CS 双注意力机制模块放在 Backbone 中 SP-PCSPC_group 模块之前以及 Neck 中进行一次上采样之后。

3.2.3 基于轻量化卷积模块

1) GSConv 模块

如图 9 所示,GSConv 模块^[20] 首先对输入通道数为 $C1$ 的特征图进行 Conv2d、BN、激活函数操作,得到通道数为 $C2/2$ 的特征图 X ,对该部分特征图进行深度可分离卷积(DSC)操作得到特征图 X' ,将 X 和 X' 进行拼接得到通道数为 $C2$ 的特征图,然后进行通道混洗得到输出特征图,这样可以将标准卷积的信息完全混合到 DSC 的输出中。GSConv 模块将标准卷积和深度可分离卷积结合,在特征提取过程中,可以使输出尽可能接近标准卷积,同时降低计算成本,更好地平衡模型的准确性和速度。在改进后的网络中,仅将 Neck 中的 Conv 模块替换为 GSConv 模块。

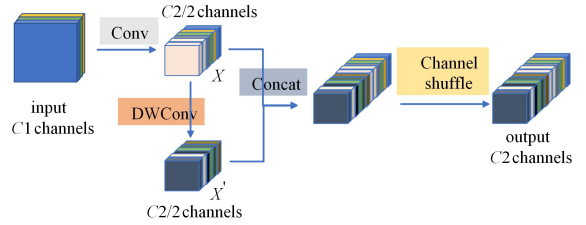


图 9 GSConv 结构图

Fig. 9 GSConv structure diagram

2) Fasternet 卷积神经网络

采用轻量化骨干 Fasternet^[21] 替换 YOLOv5s 中所有 C3 模块,如图 10 所示。Fasternet 是在 Partial Convolution (PConv)的基础上提出的,PConv 只在输入特征图的一部分上(本文选取 1/4 channel)进行常规卷积操作,并保持其余通道不变,目的是减少冗余计算和内存访问,从而能够更加高效地提取特征。Fasternet 首先对输入的特征图进行一次 PConv 操作,卷积核大小为 3×3 ,然后经过一次 1×1 卷积、BN、ReLU 激活函数,再进行一次 1×1 卷积操作,将得到的特征图与输入特征图进行拼接,得到最终的输出特征图。

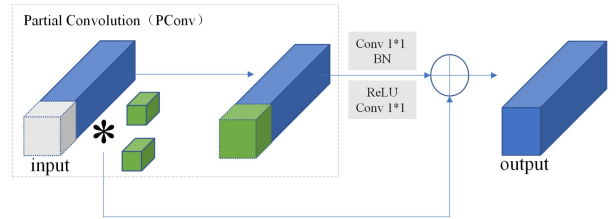


图 10 Fasternet 结构图

Fig. 10 Fasternet structure diagram

3.2.4 基于 WIoU 损失函数设计

在 head 目标检测头中,边界框回归 BBR (Bounding Box Regression) 的损失函数对目标检测至关重要,定义良好的 BBR 损失函数将能明显提升模型的性能。Tong 等^[22] 提出了具有动态非单调聚焦机制的损失函数 Wise-IoU (WIoU),使用“离群度”替代 IoU 对锚框产生估计,离群度小意味着锚框质量高,为其分配一个小的梯度增益,以便使边界框回归聚焦到普通质量的锚框上。对离群度较大的锚框分配较小的梯度增益,能有效防止低质量样本产生较大的有害梯度。WIoU v3 版本离群度、损失函数公式如下:

$$\beta = \frac{\mathcal{L}_{IoU}^*}{\bar{\mathcal{L}}_{IoU}} \in [0, +\infty), r = \frac{\beta}{\delta \alpha^{\beta - \delta}} \quad (13)$$

$$\mathcal{L}_{WIoUv3} = r \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*}\right) \bar{\mathcal{L}}_{IoU} \quad (14)$$

其中, \mathcal{L}_{IoU}^* 为梯度增益, 随 \mathcal{L}_{IoU} 的减小而减小; $\bar{\mathcal{L}}_{IoU}$ 为动量 m 的滑动平滑值; α, β 为超参数。由于 $\bar{\mathcal{L}}_{IoU}$ 是动态的, 锚框的质量划分也是动态的, 因此 WIoU 在每一时刻都能作出最符合当前情况的梯度增益分配策略。

4 实验结果及分析

4.1 实验环境

实验租用云服务器, 采用 ubuntu20.04 操作系统, CPU 为 15 vCPU Intel(R) Xeon(R) Platinum 8358P CPU @ 2.60GHz, GPU 为 RTX 3090(24GB), 内存为 80GB, 所用的深度学习框架为 PyTorch 1.11.0, Python 环境为 3.8, Cuda 为 11.3。

4.2 实验数据集

本次实验从公共数据集 ExDark、BDD100K、UA-DETRAC、华为 SODA10M 中挑选部分夜间车辆图片, 同时自行拍摄夜间照片视频等作为私有数据集, 总计 4893 张图片, 以 8:1:1 的比例划分训练集、验证集、测试集, 数据集中检测对象共有 5 类, 包括 car, bus, bicycle, motorcycle, truck。利用双稳随机共振算法对原始数据集(Dark)进行数据增强, 将增强后的数据集(Light)传入 YOLOv5s 网络进行训练。

4.3 实验过程与结果分析

本文选定精确率 P 、召回率 R 、准确率 AP 、 $AP[0.5:0.95]$ 、参数量 Params、边界框损失 box_loss 、分类损失 cls_loss 作为评价指标, 如表 3 所列。

表 3 各评价指标

Table 3 Explanation of each evaluation indicator

评价指标	说明
P (精确率)	$P = \frac{TP}{FP + TP}$
R (召回率)	$R = \frac{TP}{FN + TP}$
AP (准确率)	$AP = \int_0^1 P(R) d$
$AP[0.5:0.95]$	选定 $IoU = 0.50, 0.05, 0.95$ 共 10 个阈值 每个阈值计算 AP 值, 最后取平均
Params	模型的参数量
box_loss, cls_loss	预测锚框与 GT 框的误差、预测锚框的分类与真实标注分类是否一致

将本文模型与上述模型进行对比实验, 表 4 列出了模型整体的实验结果, 图 11 是各个模型在各个类别上的 AP 值结果图。模型 D 是基线算法, 模型 L 的 P 和 AP 值与模型 D 非常接近, 但是召回率 R 值增加了 1.7%, 说明原始数据集经过双稳随机共振增强后, 可使检测更加全面, 原来容易被忽略的车辆被检测到的概率更大; LW 是将 CIoU 损失函数替换为 WIoU, 可以看出 P, R, AP 值均提升了 0.5%, 由图 12 可知, LW 的 box_loss 收敛要比 L 快得多, 说明了 WIoU 损失函数的动态非单调聚焦机制能够为锚框合理地分配梯度增益, 有效地平衡了高低质量样本; LW2CS 采用了 SPD 和 SPPCSPC_group 结构, 加入了 CS 双注意力机制, 使得 R 值提升了 3%,

AP 值提升了 2%, 模型检测更加全面, 同时提升了对小目标的检测精度, 但是参数量增加了 29%; LW2CSFG 采用轻量化模块和骨干, 参数量减少了 25%, 同时最终的 AP 和 P 值均提高了 1.3%。由图 11 可知, car, bus, bicycle, motorcycle, truck 的 AP 均值分别为 0.968, 0.912, 0.817, 0.845, 0.876, 每一类在 6 个模型检测下的 AP 值整体呈上升趋势。

表 4 各模型的对比实验结果

Table 4 Results of each model comparison test

序号	模型	P	R	AP	参数量
1	D(基线算法)	0.926	0.809	0.876	7033114
2	L	0.923	0.826	0.877	7033114
3	LW	0.931	0.814	0.882	7033114
4	LW2	0.933	0.824	0.89	8902426
5	LW2CS	0.923	0.838	0.895	8934754
6	LW2CSFG	0.939	0.807	0.889	6686458

注: D 是未改进的 YOLOv5s 在原始低照度数据集上训练; L 是未改进的 YOLOv5s 在增强后的数据集上训练; LW 是更换损失函数为 WIoU; LW2 是采用 SPD 和 SPPCSPC_group 结构; LW2CS 是加入 CS 双注意力机制; LW2CSFG 是采用轻量化模块。

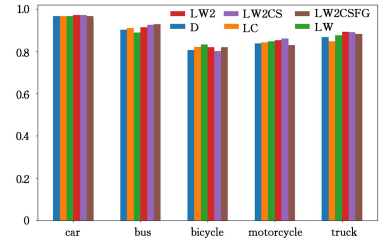


图 11 模型在各类别上的 AP 值

Fig. 11 AP of model on various categories

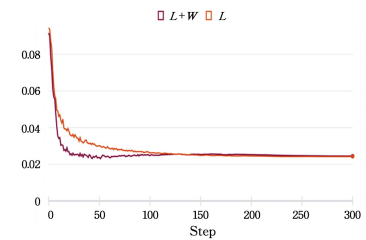


图 12 LW 与 L 对比图

Fig. 12 LW and L comparison

针对表 4 中模型 4(LW2)和模型 6(LW2CSFG), 从参数量角度进行消融实验, 结果如图 13 所示。模型 LW2 将 Backbone 中 P2, P3, P4 之后的下采样操作替换为 SPD 结构, 而 LW2-s 仅替换 P2 和 P3 之后的下采样操作, 参数量由 8902426 减少至 8738146, 减少了 2.8%, 但是准确率 AP 值、召回率 R 值均下降了 2%, 所以本文最终选择 LW2 模型。图 13(b) 中, 模型 LW2CSFallG 即本文提出的模型, 将 YOLOv5s 的 Backbone 和 Neck 中所有 C3 模块均替换为 Fasternet, 仅将 Neck 中的 Conv 模块替换为 GSConv, 而 LW2CSFallG 在此基础上将 Backbone 中的 Conv 模块也替换为 GSConv, 参数量由 6686458 减少至 5915098, 参数量降低了 12%, 但是准确率 AP 值在训练 220 个 epoch 之前下降明显, 高达 2.5%, 并且分类损失 cls_loss 较 LW2CSFallG 有所提升; 另外, LW2CSFallG 完整训练 300 epoch 的时长为 18.3h, 而 LW2CSFallG 仅需 5h。综合考虑, 本文最终选取模型 LW2CSFallG, 即模型 6(LW2CSFG)。

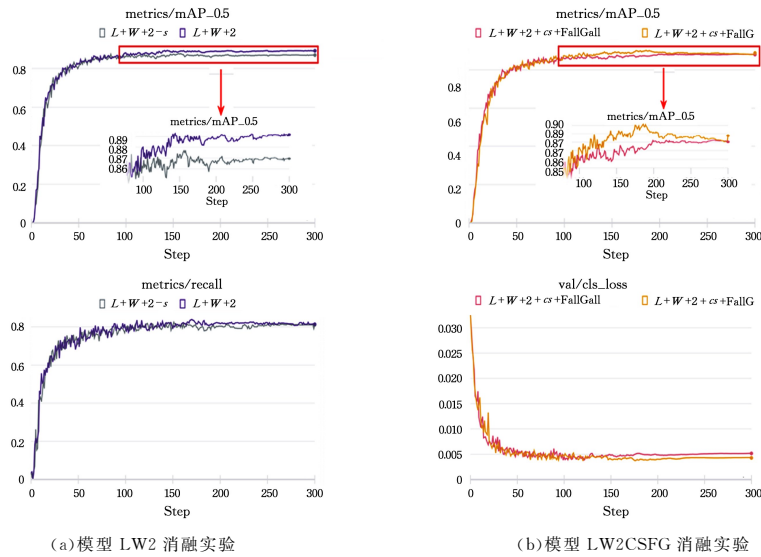


图 13 参数量消融实验结果对比图

Fig. 13 Comparison chart of parameter ablation experimental results

为了验证 CS 双注意力机制在特征提取过程中能够更加关注有用的信息,我们将 CS 双注意力机制与单一的 CA、SimAM 注意力机制分别进行对比。如图 14 所示,将表 4 中模型 6 基于这 3 种注意力机制生成的特征图进行可视化,并将模型 6 的注意力机制分别替换为 CA 和 SimAM 进行模型训练,训练结果如图 15 所示。图 14 表明,相较于单一的 CA 或者 SimAM 注意力机制,CS 双注意力机制能够更加精确地定位感兴趣的物体,且能较好地避开部分具有干扰性的背景信息。由图 15 可知,分别基于 3 种注意力机制进行模型训练,在 220 个 epoch 之前,CS 双注意力机制准确率明显高于单一的 CA 或者 SimAM 注意力机制。

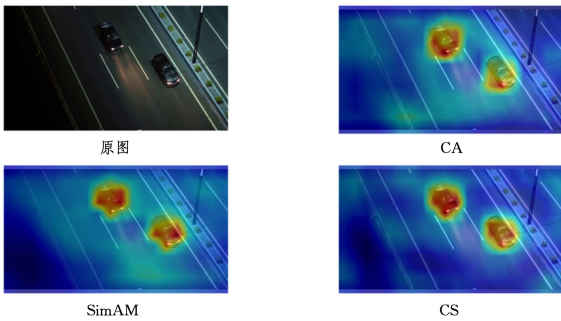


图 14 模型基于 3 种注意力机制生成的特征图可视化结果

Fig. 14 Visualization of feature map generated by model based on three attention mechanisms

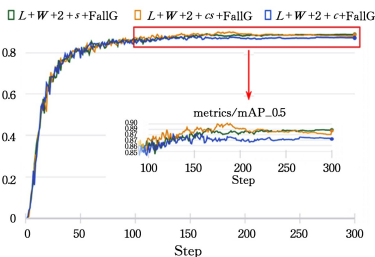


图 15 3 种注意力机制训练结果对比图

Fig. 15 Comparison chart of training results of three kinds of attention mechanisms

从时间复杂度上来看,SimAM 注意力机制模块只使用了一个全局池化层(GAP),时间复杂度约为 $O(C_{in} \times C_{out})$,其中 C_{in} 为输入通道数, C_{out} 是输出通道数。CA 注意力机制模块使用了平均池化、多个卷积层、BN 层等,整体时间复杂度由卷积层决定,约为 $O(M^2 \times K^2 \times C_{in} \times C_{out})$,其中 M^2 是输出特征图面积, K^2 是卷积核面积。CS 双注意力机制由 CA 和 SimAM 串行构成,整体时间复杂度同 CA 模块一致,但 CS 模块先将空间维度拆分,通过学习得到自适应通道权重,再将通道和空间维度统一,从而能够使模型更加关注有用的信息,整体性能优于单一的 CA 或者 SimAM 注意力机制。

为了验证双稳随机共振低照度图像增强的有效性,对模型 D(基线算法)、模型 DW2CSFG(仅改进 YOLOv5s)、模型 LW2CSFG 进行了一组对比实验,如图 16 所示。图 16(a)显示,模型 DW2CSFG 和模型 D 的准确率 AP 值曲线基本重合,仅分类损失 cls_loss 收敛速度略快一点;图 16(b)显示,低照度图像经过双稳随机共振增强后,传入改进的 YOLOv5s 训练得到的模型 LW2CSFG 的 AP 值比 DW2CSFG 高出 2.8%,分类损失 cls_loss 也明显低于 DW2CSFG。通过这组对比实验可以发现,如果仅在原夜间车辆数据集上进行 YOLOv5s 改进,检测的准确率几乎保持不变,而经过双稳随机共振增强后传入改进的 YOLOv5 网络,准确率提升明显,分类损失明显下降。这也说明了双稳随机共振进行低照度图像增强的有效性。

联合表 4 以及图 16 发现,对原始夜间车辆数据集仅进行双稳随机共振增强或仅进行 YOLOv5s 改进,车辆检测的准确率均无明显提升,只有先进行双稳随机共振增强后,再传入改进的 YOLOv5s 网络,车辆检测的准确率才会显著提升。

此外,针对测试集 476 张图片进行了误检、漏检的统计,结果如表 5 所列。测试集共有 1439 个标注,基线算法模型 D 误检 4 辆车,漏检 161 辆车,漏检率达到 11.2%;模型 LW2CSFG 误检为 0,漏检 57 辆车,漏检率为 4%,相比 D 模型降低了 7.2%。

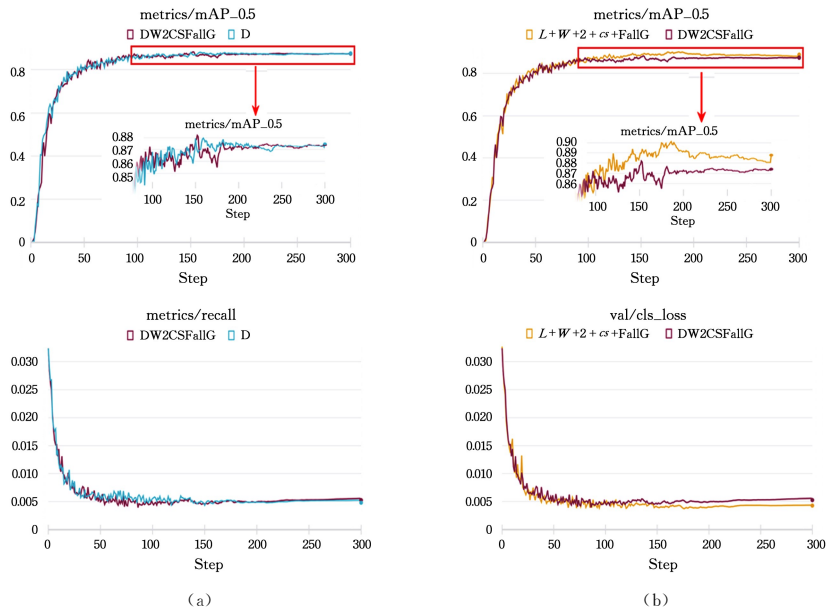


图 16 随机共振消融实验结果对比图

Fig. 16 Comparison chart of stochastic resonance ablation experimental results

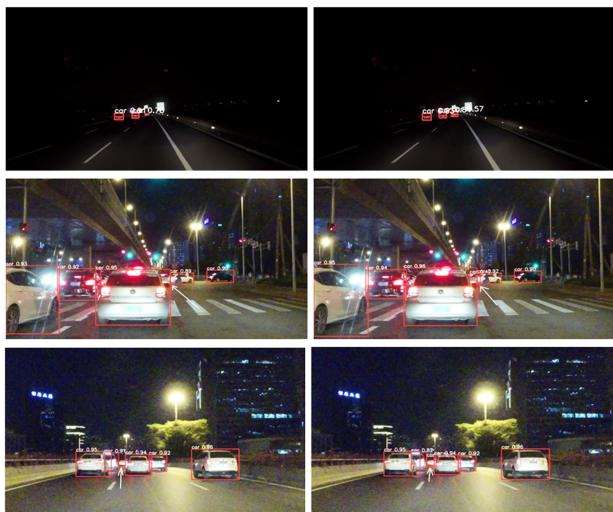
表 5 漏检误检统计结果

Table 5 Results of missed and false detection

测试集	模型	误检	漏检	漏检率
共 1439 个	D	4	161	0.112
标注	LW2CSFG	0	57	0.040

4.4 检测结果对比

为了进一步评价模型 D 和 LW2CSFG 在处理夜间车辆检测任务时的性能差异,在测试集中随机选取部分图片进行检测,结果如图 17 所示。左侧均以模型 D 的 best.pt 模型进行检测,远景小目标车辆因为密集遮挡而存在漏检现象;右侧均以模型 LW2CSFG 的 best.pt 模型进行检测,都能够更为准确地检测出,减少了漏检现象的发生。整体而言,对于夜间车辆检测任务,尤其是远景小目标、密集遮挡任务,LW2CSFG 比基线算法 D 更有优势,能够有效地减少漏检误检的现象发生。



(a) YOLOv5s(D) (b) 改进的 YOLOv5s(LW2CSFG)

图 17 检测结果对比图

Fig. 17 Comparison chart of detection results

结束语

本文提出了一种改进的 YOLOv5s 模型 (LW2CSFG),首先利用双稳随机共振对夜间车辆数据集进行低照度图像增强,使得检测结果更加全面;采用具有动态非单调聚焦机制的损失函数 $WIoU$,加快了边界框回归损失的收敛速度,同时平衡了训练过程中的高低质量样本,提高了模型的泛化能力;用 SPD 和 SPPCSPC_group 结构替换部分下采样、SPPF 结构,引入 CS 双注意力机制,提高了网络对目标的特征提取能力,最终提升了小目标的检测精度;最后采用 Fasternet 和 GSCov 轻量化模块,使得参数量减少 25%,准确率提升 1.3%,漏检率降低 7.2%。本文改进的网络能够降低夜间车辆的漏检率,较好地应对远景小目标、密集遮挡的目标检测任务。

在接下来的工作中,我们将从模型剪枝的角度对训练好的 best.pt 模型进行通道剪枝、模型微调,在不降低检测精度的条件下,满足夜间车辆检测实时性要求。另外也会考虑利用 C++ 结合 TensorRT 将最终模型在 Nvidia Jetson Nano 和 Jetson Xavier NX 等开发板上进行部署。

参考文献

[1] ZHANG X Y, GAO H B, ZHAO J H, et al. Overview of deep learning intelligent driving methods [J]. Journal of Tsinghua University (Science and Technology), 2018, 58(4): 438-444.

[2] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich Feature-Hierarchies for Accurate Object Detection and Semantic Segmentation [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2014: 580-587.

[3] GIRSHICK R. Fast R-CNN [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2015: 1440-1448.

[4] 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.
- [5] LONG S, SONG X F, ZHANG S, et al. Vehicle detection research on improved aerial images of YOLOv5s[J]. Laser Journal, 2022, 43(10): 22-29.
- [6] ZHANG H D, ZHANG R Q, TONG L, et al. Vehicle target detection method based on improved YOLOv5s[J]. Journal of Chongqing University of Technology(Natural Science), 2023, 37(7): 80-89.
- [7] DENG T M, LIU X H, WANG L, et al. Vehicle detection algorithm combined with cascading attention mechanism[J]. Computer Engineering and Applications, 2023, 59(21): 141-150.
- [8] ZOU Y, LONG W, LI Y Y, et al. Research on vehicle detection algorithm in low-illumination environment[J]. Machine, 2022, 49(7): 66-74.
- [9] GUO K Y, WANG S D, LI X, et al. Multi-target detection of vehicles in dim scenes based on Dim env-YOLO algorithm[J]. Computer Engineering, 2023, 49(3): 312-320.
- [10] BENZI R, SUTERA A, VULPIANI A. The mechanism of stochastic resonance[J]. Journal of Physics A: Mathematical and General, 1981, 14(11): L453-L457.
- [11] LIU J, WANG Y G. Performance investigation of stochastic resonance in bistable systems with time-delayed feedback and three types of asymmetries[J]. Physica A: Statistical Mechanics and Its Applications, 2018, 493: 359-369.
- [12] LIU J, HU B, WANG Y G. Optimum adaptive array stochastic resonance in noisy grayscale image restoration[J]. Physics Letters A, 2019, 383(13): 1457-1465.
- [13] WUEHR M, BOERNER J C, PRADHAN C, et al. Stochastic resonance in the human vestibular system-Noise-induced facilitation of vestibulospinal reflexes[J]. Brain Stimulation, 2018, 11(2): 261-263.
- [14] GAO W, XIAO H F. Low-light color image enhancement based on dynamic bistable stochastic resonance[J]. Chinese Journal of Liquid Crystals and Displays, 2021, 36(6): 861-868.
- [15] WEI M, HU X F, LIN M. Low-illumination image enhancement method based on tetrastable stochastic resonance[J]. Chinese Journal of Liquid Crystals and Displays, 2022, 37(7): 871-879.
- [16] SUNKARA R, LUO T. No more strided convolutions or pooling: A new CNN building block for low-resolution images and small objects[C]// Machine Learning and Knowledge Discovery in Databases, France, 2023: 443-459.
- [17] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[J]. arXiv: 2207. 02696, 2022.
- [18] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2021: 13713-13722.
- [19] YANG L X, ZHANG R Y, LI L, et al. SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks[C]// Proceedings of the 38th International Conference on Machine Learning, PMLR, 2021: 11863-11874.
- [20] LI H, LI J, WEI H, et al. Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicles[J]. arXiv: 2206. 02424, 2022.
- [21] CHEN J, KAO S, HE H, et al. Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks[J]. arXiv: 2303. 03667, 2023.
- [22] TONG Z, CHEN Y, XU Z, et al. Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism[J]. arXiv: 2301. 10051, 2023.



HU Pengfei, born in 1999, postgraduate. His main research interests include stochastic resonance and deep learning-based vehicle detection.



WANG Youguo, born in 1968, Ph.D., professor, Ph.D supervisor. His main research interests include signal and information processing, social network information dissemination and control, stochastic resonance theory and application.

(责任编辑:何杨)