

基于深度学习的复杂天气场景交通标志检测

陆 维, 吴 锡

(成都信息工程大学 计算机学院, 四川 成都 610000)

摘 要: 针对TT100K交通标志数据集中检测目标类别数量分布不平衡、天气场景单一,以及在复杂天气场景下检测性能下降等问题,基于目标检测网络CenteNet2提出了3点改进:①提出一种基于标志替换和深度图合成图像的数据集扩展方法TT100K-sync,平衡了交通标志类别数量,新增了雾天和雾霾天两种天气场景;②基于联合训练构建一种U形网络结构U-subnet,增强特征提取主干在图像质量下降时的性能;③引入一套基于图像分块的数据预处理方法overlap-tiling,降低了目标在特征提取下采样过程中的特征损失。实验结果表明,扩展后的新数据集相较于原数据集,将交通标志类别数量的最大差距从3 176平衡到200,新增雾图和雾霾图各11 330张。采用了U-subnet和overlap-tiling后的网络,在新数据集常规天气子集TT-normal、雾天气子集TT-fog和雾霾天气子集TT-haze上的测试mAP50相比原网络分别提高了1.13%、3.26%和6.38%,显著提升了在复杂天气下的网络检测性能。

关键词: 交通标志检测;样本平衡;图像合成;联合训练;图像分块

DOI: 10.11907/rjdk.241350

开放科学(资源服务)标识码(OSID):



中图分类号: TP183; TP391.41

文献标识码: A

文章编号: 1672-7800(2025)006-0175-10

Traffic Sign Detection in Complex Weather Conditions Based on Deep Learning

LU Wei, WU Xi

(College of Computer Science, Chengdu University of Information Technology, Chengdu 610000, China)

Abstract: In response to the problems of imbalanced distribution of target categories and single weather scenes in the TT100K traffic sign dataset, as well as the decline in detection performance in complex weather scenes, three improvements are proposed based on the object detection network CenteNet2. Firstly, a dataset extension method TT100K-sync is proposed based on sign replacement and synthesis images with depth map, which balances the number of traffic sign categories and adds two new weather scenarios: foggy and hazy images. Secondly, based on joint training, a U-shaped network structure U-subnet is constructed to enhance the performance of the feature extraction backbone when image quality decreases. Thirdly, a data preprocessing method based on image tiling called overlap-tiling is introduced to reduce the feature loss of the target during feature extraction and down sampling. The experimental results show that the expanded new dataset balances the maximum difference in the number of traffic sign categories from 3 176 to 200 compared to the original dataset, and adds 11 330 fog and haze images each. The network that adopted U-subnet and overlap-tiling improved the mAP50 on the conventional weather subset TT-normal, foggy weather subset TT-fog, and haze weather subset TT-haze in the new dataset by 1.13%, 3.26%, and 6.38%, respectively, compared to the original network, significantly improving the network detection performance in complex weather conditions.

Key Words: traffic sign detection; sample balance; image synthesis; joint training; image tiling

0 引言

近年来,随着自动驾驶技术的不断发展,基于深度学

习在TT100K数据集上的交通标志检测相关研究也层出不穷^[1]。当前研究主要在原数据集基础上进行,通过改进特征提取主干、加入注意力模块、改进损失函数等方式提高检测性能。然而,这些研究存在如下局限性:①TT100K存

收稿日期:2024-05-03

扫描二维码阅读全文:



基金项目:四川省科技计划项目(重点研发项目)(2023YFG0025)

作者简介: 陆维(1983-),男,成都信息工程大学计算机学院硕士研究生,研究方向为图像处理;吴锡(1980-),男,博士,成都信息工程大学计算机学院教授、硕士生导师,研究方向为图像分析与计算成像、高性能与并行分布式计算、智慧气象与数值天气计算。本文通讯作者:吴锡。

在标志类别不平衡问题,这种不平衡会导致模型在训练过程中对高频标志过度拟合,难以正确识别低频标志,从而限制了模型在不同场景下的泛化能力^[2]。现有研究通常改进损失函数,通过权重调整以补偿少数类标志损失^[3]。虽然损失函数可以起到平衡作用,但 TT100K 标志类别不平衡问题仍然较为严重,如图 1 所示(因类别太多,x 轴每隔 10 个类别显示 1 个),头部数量和尾部数量差距超过 3 000 张,甚至末尾一些类别数量为 0,这对于损失函数设计要求较高,而且在巨大的差距下能够平衡的程度有限。②TT100K 缺乏雾、雾霾等复杂天气场景,现有研究通过传统图像增强技术给原图像添加天气特效,然而这种方式合成的图像无法模拟真实场景,因为简单地添加天气特效无法真实地模拟出复杂天气场景中的光照、遮挡、噪点等细节,导致训练出的模型在真实世界中的适应性和鲁棒性受到限制。③注意力模块虽然可以帮助提高模型性能,但在复杂天气下,由于天气条件影响目标的外观特征、边缘清晰度等因素,注意力模块的增强效果会受到一定限制^[4]。④TT100K 中目标的尺寸一般在 80 左右,属于小目标,仅靠改进网络结构无法达到较好的特征提取效果,因为特征提取的前提是目标特征的可见性,而小目标在复杂天气的影响下,其本就不明显的特征会变得更加难以识别,提取的困难程度会随着图像质量的下降而提升。

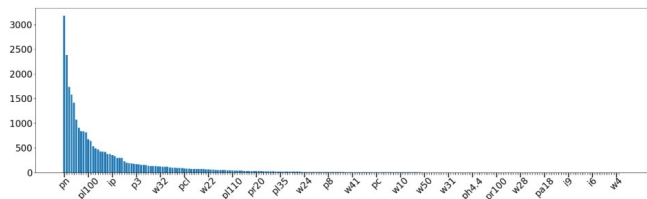


Fig. 1 The distribution of traffic sign category quantities in TT100K

图 1 TT100K 交通标志类别数量分布图

本文首先通过扩展数据集从源头解决问题,构建一个更大、更多样化的数据集,让模型学习到更广泛的数据表示,并作为一种自然的正则化手段直接提高模型对未见数据的预测能力,确保研究成果能真正走向实际应用^[5,6]。具体而言,本文提出了一种数据集扩展流程 TT100K-sync,平衡了各交通标志类别的数量,新增了雾和雾霾 2 种复杂天气场景,更好地模拟了真实世界的交通场景。其次,构建一个联合训练的 U 形网络结构 U-subnet。该结构包含了 1 个检测子网和 1 个恢复子网,可以端到端地进行训练并共同学习 3 个任务:能见度增强、物体分类和物体定位,通过多任务共同学习来获得特征提取主干的性能改进。再次,引入一种高级图像预处理技术 overlap-tiling,通过图像重叠分块以降低小尺寸交通标志在特征提取中的特征损失,以增强交通标志小目标的检测能力。

总结而言,目前基于 TT100K 交通标志检测的研究都是在原始数据集上进行,无论算法如何改进,都无法回避原数据集本身的问题。而本文研究从根源上解决数据集问题,再针对更贴近真实场景的合成图像以改进网络结构

和检测流程,在国内交通标志检测模型的训练、测试、评估、应用上具有现实意义。

1 相关工作

1.1 样本类别平衡

现阶段针对交通标志类别数量不平衡的处理可以分为 3 种:①Meng 等^[7]筛选出类别数量大于 100 的 45 个类别进行训练和测试,其他数量不足 100 的类别则不参与检测。这种处理的优点是实施简单,但缺点比较明显,忽略其他大部分的标志导致模型在实际应用中识别范围有限,只能作为一种实验手段。②Wu 等^[8]通过改进损失函数以平衡高频类别和低频类别在分类器上的差距,这种做法虽然可以在一定程度上解决类别不平衡的问题,但当类别繁多且类别之间数量差距巨大时,损失函数平衡的效果就会下降。③Nanni 等^[9]通过替换标志的方式直接更改各类别的数量分布情况,从根源上解决标志类别数量不平衡问题。需要特别指出的是,这种直接改变数据集的做法除能够有效解决不平衡问题外,还可以灵活地根据具体情况不断迭代数据集,比如针对测试结果中某些表现不佳的类别或者某些特定场景(比如高速路)相关的类别进行数量调整。但同时,这种方法的成本也较高,替换所需的各种图片要准备齐全,替换过程也要精心设计,以避免在替换后对模型产生不必要的干扰。

1.2 雾图与雾霾图合成

关于合成天气,长期以来主流的做法是 Feng 等^[10]依靠传统的计算机图像方法,在真实图像上进行特效合成,其最大的缺点是逼真程度不够。以雾图为例,随机数生成函数无法准确地模拟真实世界中复杂的雾气分布,生成的数据对真实场景检测提供的帮助有限。

近几年,随着深度学习的快速发展,出现了基于生成模型的合成算法。一项引人注目的研究成果是由 Lin 等^[11]利用生成对抗网络(GANs)实现的动态天气效果合成,其介绍了一种新的深度学习框架,称为 Weather-GAN,专门用于生成具有各种天气条件(如晴天、雨天、雪天等)的真实感图像。该方法能够从给定的晴朗天气图像出发,通过学习不同天气条件下的图像分布特征,合成逼真的特定天气效果图像。通过这种方法,成功模拟了多种复杂且动态变化的天气效果,显著提升了天气效果合成的真实感和多样性。但这种做法存在 2 点不足:①生成对抗网络需要大量的真实数据进行训练,如果输入的数据质量不高或代表性不足,会影响生成图像的质量和多样性,而大量真实数据的收集需要极大的时间成本;②生成前无法为天气效果指定天气参数,比如生成能见度 100 m 的雾图,这为后续研究和实验带来了不便。

还有一种方式是基于深度图的合成算法,其中有代表性的是 Sakaridis 等^[12]基于双目深度图的合成算法,它所生成

的 Cityscapes-Foggy 数据集受到了 Cityscapes 官方的认可。其原理是利用精准的双目深度图估计出场景中每个像素点到相机的距离,从而推断出不同位置的雾的密度,再利用雾气的光学模型模拟出更真实的雾气分布效果。需要特别注意的是:①虽然上文只提到雾图的合成,但雾霾图的合成也类似,对于具体处理上的不同,本文在后面章节再作阐述;②该方式的前提是深度图, Cityscapes 官方提供了精准的双目深度图但 TT100K 没有,在后面章节将详细说明如何处理该挑战。

1.3 复杂天气场景的交通标志检测

对于常规天气场景的交通标志检测,目前主流的做法是改进网络结构,比如引言中提到的改进损失函数,由陈德海等^[13]于2022年提出。其在 YOLO-v3 的基础上引入了多尺度注意力模块,在 TT100K 上测试 mAP50 达到了 0.965。尽管算法在原数据集上表现出色,但没有考虑复杂天气条件下的数据质量、光照、背景干扰等因素,导致其在复杂天气下的检测效果可能有限。

目前,复杂天气条件下的交通标志检测有2种做法。一是在目标检测前对图像去噪,比如 Xu 等^[14]提出了一种基于暗通道先验图像的单幅交通图像去雾算法,以暗通道为主,结合其他通道特征进行去雾。这种做法对去噪操作要求较高,必须同时兼顾速度和准确度,而且对不同的天气场景泛用性有限。二是利用深度学习技术进行端到端的交通标志检测,比如上文提到过的 Sakaridis 等^[12]基于高质量合成数据集以提升模型去噪能力,通过对其文中提到的实验进行分析可知,高质量数据集的扩展确实能极大提高检测准确率。这种方法可以更好地适应多变的天气条件和场景变化,但也存在计算复杂度高的缺点,而且当图像能见度降低到一定程度后,单靠扩展数据集的方式会达到瓶颈性能,这时就需要结合其他方式以解决该问题。

1.4 小目标特征捕捉

本文检测任务中的另一个关键问题是对交通标志小目标的特征捕捉。由于小目标尺寸较小、形状多变,直接对整个图像进行特征提取和检测往往效果不佳。为了解决该问题, Lu 等^[15]采用在 RPN 阶段添加注意力模块的方式以增强小目标特征。然而,这种方式鲁棒性不足,当图像质量严重降低时检测准确率会直线下降。还有一种方式是 Alhichri 等^[16]使用图像分块技术对高分辨率的遥感图像进行预处理,将大图像分割成若干小图,然后利用预训练的卷积神经网络从每个小图中提取描述性特征。此方法的优点是采用“分而治之”的思路实现了对大尺度高分辨率图像的检测,以小图代替大图输入到检测网络,降低了小目标在特征提取下采样过程中自身特征的损失,使网络能更好地捕捉小目标特征。值得注意的是, TT100K 的图像分辨率为 2 048×2 048,同样适合该方法。但问题是,如何保证在图像分块过程中不会切割到目标自身,这个议题将在下文进行阐释。

2 TT100K-sync 数据集扩展流程

在上述相关工作的基础上,以平衡类别数量和增加复杂天气场景为目的,本文提出了针对 TT100K 数据集的 TT100K-sync 扩展方法,扩展流程如图2所示。

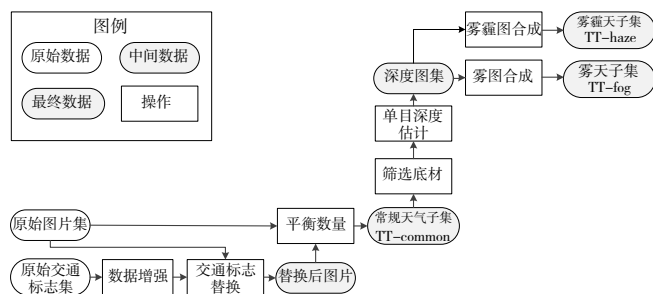


Fig. 2 Process of TT100K-sync data extension

图2 TT100K-sync 数据扩展流程

可以看出,整个流程中有2个原始数据输入、多个操作环节、3个最终数据集输出,其分别为常规天气数据子集 TT-common (TTc)、雾天数据子集 TT-log (TTf)、雾霾天数据子集 TT-haze (TTTh)。

2.1 替换与平衡

TT100K 交通标志样图集如图3所示。在 TT100K 数据集中,交通标志被划分为建议类(i)、禁止类(p)和禁令类(w)三大类。三类标志之间的图像特征,比如颜色、形状等有着较大差别,而大类内部各小类的图像特征区别则不大。

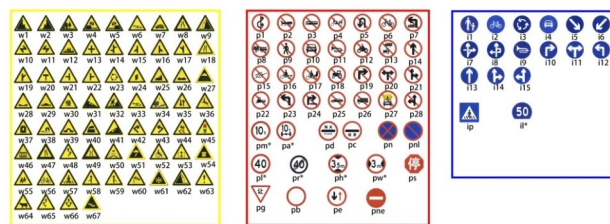


Fig. 3 TT100K traffic sign sample set

图3 TT100K 交通标志样图集

目前替换流程为:①在被替换标志 A 所在图片的标注中找到 A 的边界框;②在需要增加的标志中随机选出需要替换的标志 B (背景透明的 png 图片),算出 B 的尺寸;③将 B 缩放至 A 的大小并按照中心对齐的方式覆盖在 A 上面;④保存为新的图片并生成对应的标注。值得注意的是,在第2步中随机选出替换标志,可能导致替换内容缺乏相关性,如图4(a)所示,用 w59 (上左)替换 i5 (上右),在尺寸不变的情况下无法完全覆盖,从而使得输入模型的信息丢失或混淆。本文方法是将随机替换改为同类替换,即替换与被替换标志须属于同一类别(p类和w类大部分可以相互替换),如图4(b)所示,用 i15 (上左)替换 pn (上右),实现完全覆盖。该方式适用于 TT100K 这种具有明确类别划分的数据集,同类替换能够在替换后不留下痕迹,有效地保持

语义连贯性,减少了模型输入信息的丢失和混淆,因而通常不会对检测性能造成显著影响。至于少量形状特殊的标志,比如pg、ip,可以通过手动替换的方式完成。

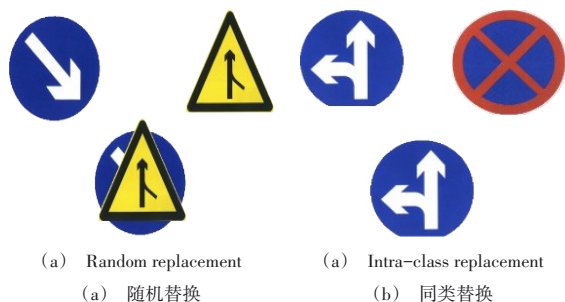


Fig. 4 Replacement

图4 替换

为了更逼真地模拟现实交通场景,本文基于TT100K本身的环境语义特征在标志替换时进行了7种数据增强,其效果如图5所示,按顺序依次为1原图、2噪点、3旋转、4扭曲、5遮挡、6颗粒化、7模糊、8亮度。通过这些增强方式的组合应用能够创建一个更加多样化和具有挑战性的交通标志图像数据集,为训练模型提供更广泛的场景覆盖和更丰富的信息。



Fig. 5 Enhanced effect of traffic signs

图5 交通标志增强效果

即使采取了同类替换策略,仍存在常见交通标志类别样本数量过多的问题,特别是处于头部的类别占比偏高,影响模型训练效率和最终性能。为了进一步平衡数据集,可以考虑在头部标志所在的图片中过滤掉一些重复度较高的图片,减少数据集中的冗余信息,从而避免模型在训练过程中过度适应这些重复样本,也即俗称的“过拟合”问题。

该阶段的过滤主要以手动审查为主,具体操作会考虑以下条件:①尽量避免过滤非晴天图片;②尽量过滤头部类别所在的图片;③尽量避免过滤到尾部类别所在图片;④过滤掉标志大小类似和所处街道场景类似(比如高速路)的图片。

通过上述替换和过滤步骤得到一个数量平衡的常规天气数据集TTc,它是后面合成TTf和TTb的基础。

2.2 生成深度图

参考Sakaridis在Foggy Cityscapes数据集上的工作,相对于使用随机数创建雾气效果的传统合成方案,本文采用更为先进的深度图合成方案^[12]。深度图是一种灰度图数据,其中每个像素的值表示该像素距离观察者即摄像机的距离,根据深度图可以计算出摄像机对雾气的感知特性,包括远处物体的模糊效果、景深变化等,从而推断出不同位置的雾的密度。然而,TT100K数据集没有提供现成的深度图。为了解决该问题,本文使用单目深度估计算法Monodepth2以生成单幅图像的深度图^[17]。

Monodepth2是一个用于单目深度估计的模型,基于自监督学习和卷积神经网络,通过编码器—解码器结构和最小化左右两个视图之间视差的一致性损失函数设计,在训练时最小化光度再投影误差,进而从单目图像中学习并预测每个像素点的深度信息。其主要计算公式如下:

$$L_p = \sum_i pe(I_i, I_{i' \rightarrow i}) \quad (1)$$

$$I_{i' \rightarrow i} = I_{i'} \langle proj(D_i, T_{i' \rightarrow i}, K) \rangle \quad (2)$$

其中, L_p 代表光度再投影误差, $I_{i' \rightarrow i}$ 代表源视图 $I_{i'}$ 相对于目标视图 I_i 的姿态, pe 是一个光度重建误差函数。 $proj()$ 是投影深度 D_i 在 $I_{i'}$ 中的二维坐标,而 $\langle \rangle$ 是采样算子,为了简化符号,假设所有视图预先计算的内参 K 都相同。

从图6可以看出,在视觉成像的深度图中,物体的轮廓和结构特征保留较好。深度图常以灰度图的形式表示(这里为了展示边缘细节进行了着色),其中不同的灰度值对应到场景中不同的深度。一般而言,较亮的区域(高灰度值)表示距离相机较近的物体,而较暗的区域(低灰度值)表示距离相机较远的物体。

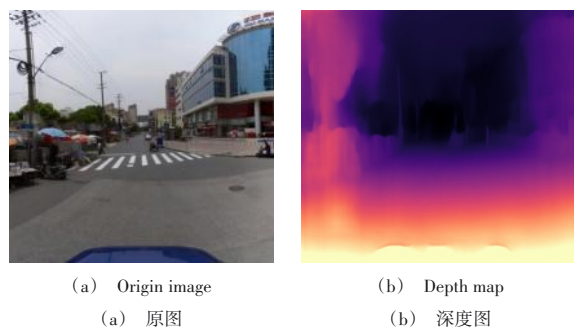


Fig. 6 Comparison of depth map

图6 深度图效果对比

2.3 合成雾图与雾霾图

在得到深度图后,本文参考Zhang等^[18]的工作,即基于大气散射模型,通过深度图和透视投影变换以估计图像中每个像素离摄像机的距离和海拔。合成流程主要包括如下步骤:首先,通过大气散射模型估算每个像素的海拔高度,计算与颗粒层相互作用的距离;其次,利用3D Perlin噪声模拟不均匀的雾效果,以呈现更加自然的结果;再次,根据图像中像素的海拔高度,估算物体到相机通过颗粒层的距离,以获取更准确的距离信息;最后,通过主观评分和

客观评估方法,如平均意见分数(MOS)和皮尔逊线性相关系数(PLCC),以评估合成图像的真实性和效果。整个流程旨在模拟出更加逼真的雾图和雾霾图效果,其主要计算公式如下:

$$\frac{dI_o}{I_o} = (-\beta_a(\omega) - \beta_s(\omega))d_x \quad (3)$$

$$\beta_{ex} = \frac{3.912}{V}(m^{-1}) \quad (4)$$

$$I = I_o \exp(-\beta_{ex}d) \quad (5)$$

$$O = 1 - \exp(-\beta_{ex}d) \quad (6)$$

式(3)描述了通过厚度为 d_x 的介质,在方向 ω 上的辐射 I_o 的分数变化。该变化由吸收和散射引起,用吸收系数 β_a 和散射系数 β_s 描述,这两种效应导致能见度的损失。将这两个系数结合成一个灭光系数 β_{ex} ,即与大气可见范围 V 相关的灭光系数。式(4)给出了灭光系数 β_{ex} 的计算方式,其中3.912是一个经验常数。根据Beer-Lambert定律^[19],式(5)表示光束在穿过粒子层行进距离 d 后的强度,初始辐射为 I_o 。光束强度随着距离的增加而呈指数衰减,衰减率由灭光系数 β_{ex} 决定。式(6)表示透明度 O 与灭光系数 β_{ex} 和穿过距离 d 之间的关系。透明度 O 的值范围为0~1,其中0表示完全不透明,1表示完全透明。

在合成过程中有2个关键参数:受大气水分子影响的能见度和受大气气溶胶影响的能见度。雾和雾霾的合成受这两个参数影响,数值越低的参数将更多地影响合成效果。雾图中大气水分子占主导地位,而雾霾图中大气气溶胶分子占主导地位,可从视觉上比较雾图和雾霾图的区别,如图7所示。



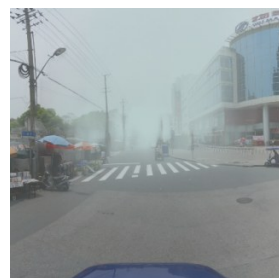
(a) Fog image(visibility about 200 m)
(a) 雾图(能见度约200 m)



(b) Fog image(visibility about 100m)
(b) 雾图(能见度约100 m)



(c) Haze image(visibility about 200 m)
(c) 雾霾图(能见度约200 m)



(d) Haze image(visibility about 100 m)
(d) 雾霾图(能见度约100 m)

Fig. 7 Comparison of fog and haze images

图7 雾图、雾霾图效果比较

具体而言,两者的主要区别是:雾是自然环境中的水汽凝结形成的气态水滴,通常表现为远处景物逐渐模糊、近处景物仍然清晰可见,整体呈现出一种柔和、朦胧的感觉。而雾霾是大气中悬浮颗粒物和有害气体混合形成的一种现象,通常表现为整体景色灰暗、模糊不清,远处景物和近处景物都缺乏清晰度,色彩也可能受到影响,呈现出一种阴郁、沉闷的感觉。因此,对生成的雾霾图又进一步添加上一个半径为1的高斯模糊滤镜和半透明灰色覆盖层,最终模拟出较为逼真的雾霾效果。

3 算法改进

3.1 U-subnet

在完成对数据集的扩展后,本文对基线CenterNet2网络^[20]进行改进。如图8所示,本文参考Huang等^[21]的研究成果采取了一种创新的策略,通过在经典的主干网络ResNet50后附加一个与其结构相对偶的解码器分支,共同构建了一个U形的网络结构,将主干的特征增强训练和分支的去噪训练联合起来。通过输入合成的有雾图像,并用无雾图像监督网络输出,在训练网络去雾的同时,增强主干网络对复杂天气交通场景图像特征提取的能力。

其主要流程为:①针对图像恢复分支(U形网络的下半部分解码器分支)进行训练,以获得预训练的网络权重;②解码器分支将被忽略并得到的编码器权重赋予主干特征提取网络(U形网络的上半部分编码器分支),并将其参数冻结,以此为基础继续训练检测网络;③在训练过程接近尾声时,解冻主干网络,允许网络保持之前训练成果的基础上,再进一步优化调整;④通过U形网络上下部分联合训练,优化能见度增强、目标分类和目标定位3个不同的任务,实现整个网络在能见度降低导致图像质量下降时检测性能的改进。

在图像恢复分支损失函数的选择上,本文采用了均方误差(MSE Loss),它能够有效地量化网络输出图像与目标图像之间的差异,从而指导网络权重的优化方向,进一步提高去雾效果的准确性和图像的视觉质量。其公式如下:

$$L_{ve} = \frac{1}{n} \sum_{j=1}^n (Y_j - \hat{Y}_j)^2 \quad (7)$$

其中, n 是批处理大小, Y_j 是真实图像, $\hat{Y}_j$ 是估计的恢复图像。需要强调的是,尽管图像恢复分支可以直接生成去噪图像,但其主要目的是通过学习能见度增强任务共同优化能见度增强、目标分类和目标定位这3个任务的检测性能。具体而言,由于框架只关注目标检测任务,故“特征恢复模块”仅在训练期间被激活,用以产生恢复的图像供所提出的多任务学习过程使用。这意味着,在训练目标检测模型的同时,特征恢复模块致力于从损坏或质量较低的图像中恢复出清晰、有用的特征,从而提高识别准确性和效率,但该过程不会在实际的目标检测应用阶段(模型部署后的推理阶段)中进行。

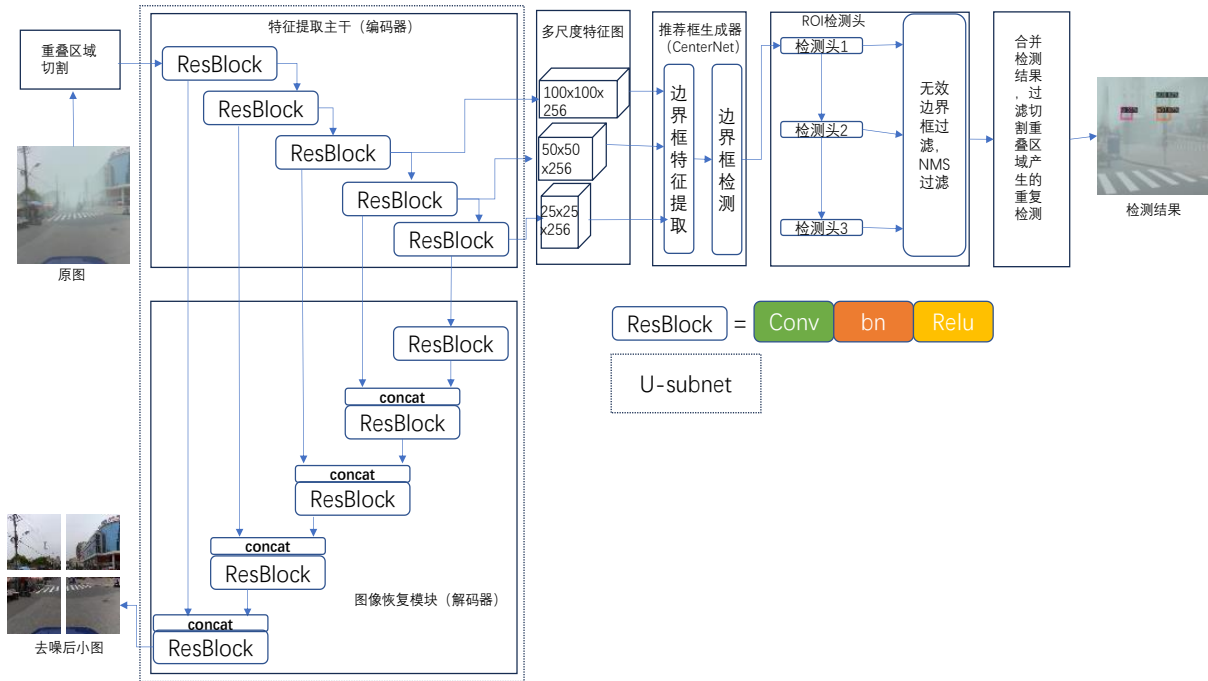


Fig. 8 Improved network structure

图8 改进后的网络结构

完成图像恢复预训练后,将解码器分支的训练成果转移到U形网络的上半部分,即编码器分支。编码器分支负责特征提取,它作为检测网络的主干特征提取网络。此时,将预训练的编码器权重用于初始化,并在一定阶段内冻结这些权重。这样做是为了利用已经训练好的图像恢复能力,为下一步检测任务提供稳定和有效的特征表示。

在接下来的训练过程中,重点转移到目标分类和定位的检测网络上,此时继续使用之前冻结的特征提取主干网络。训练接近尾声时,逐渐解冻主干网络权重,允许整个网络(包括图像恢复分支和特征提取分支)进行联合优化。这一步骤旨在调整网络参数,优化能见度增强、目标分类和目标定位3个不同的任务,以实现在复杂天气条件下图像质量显著下降时目标检测性能的整体提升。

3.2 Overlap-tiling

在Alhichri等^[16]工作的基础上,本文对图像切割方案进行改进。在确定方案前,为了能够更好地理解数据集中交通标志的总体分布特点,对TT100K中目标大小分布进行了统计,结果如图9所示。

分析可知,90%以上目标的尺寸都不会超过80×80。原作者是直接对大图进行分块,在分块过程中可能会切到目标导致目标特征丢失。为了解决该问题,本文提出了一种新的方法,基于数据集目标的尺寸分布特征进行重叠区域分块,具体流程如下:以图10为例,首先在2048×2048的原图像上预先设置 $x=1024$ 和 $y=1024$ 两条十字切割线(图10中2条红线);其次,加上长度为80的重叠区域,将图像切割为1104×1104的4张小图(X_2 和 Y_2 围成的小图1, X_1 和 Y_2 围成的小图2, X_2 和 Y_1 围成的小图3, X_1 和 Y_1 围成的小图4),这样可以保证目标至少能完整地出现在一张小图中;之后,将这4张小图批量输入特征提取网络,可以在兼顾检测效率的同时较好地保留其特征。检测完成后将检测结果映射回源图像上,并用非极大抑制将可能出现的重复检测去掉,得到最终检测结果。

小图4),这样可以保证目标至少能完整地出现在一张小图中;之后,将这4张小图批量输入特征提取网络,可以在兼顾检测效率的同时较好地保留其特征。检测完成后将检测结果映射回源图像上,并用非极大抑制将可能出现的重复检测去掉,得到最终检测结果。

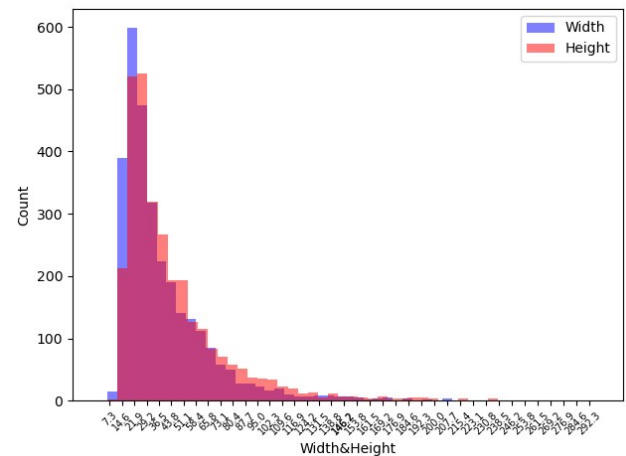


Fig. 9 Statistical distribution of target sizes in TT100K

图9 TT100K目标尺寸分布统计

4 实验与结果分析

4.1 实验环境

处理器: Intel(R) Xeon(R) CPU E5-2678 v3; 显卡: P40-24 G; 内存: 32 G; 操作系统: Ubuntu 20.04.3。

开发环境: Python3.8, 深度学习框架 PyTorch 以及基于

PyTorch 的目标检测平台 Detectron。



Fig. 10 Schematic diagram of tiling
图 10 分块示意图

4.2 数据集扩展实验结果分析

4.2.1 数据集扩展前后对比

表 1 展示了 TT100K 扩展前后数据集指标, 通过对比可以看出, 新数据集得到了极大扩展。

Table 1 Statistics of metrics before and after TT100K extension		
表 1 TT100K 扩展前后指标统计		
指标项	原数据集指标	新数据集指标
图片总数	9 176	35 297
平衡后常规天气(TTc)图片数	无	12 637
类别总数	232	240
头部类别数量	3 176	300
尾部类别数量	0	100
雾天(TTf)图片数	无	11 330
雾霾天(TTh)图片数	无	11 330

同时, 以对 TT100K 扩展前后的类别数量分布进行对比(见图 11), 其中有 3 条折线, 蓝色代表原数据集, 红色代表 TTb, 黑色代表包含了 TTb、TTf、TTh 的总数据集(彩图扫描 OSID 码可见)。

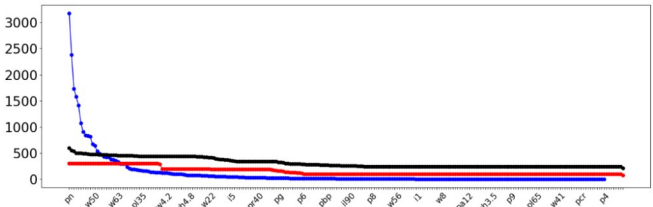


Fig. 11 Distribution of category numbers before and after TT100K extension
图 11 TT100K 扩展前后类别数量分布

可以注意到 2 个细节: ①平衡后数据集类别数量与标志常见程度成正比, 比如常见的 pn 标志数量为 300 而非常见的 w25 标志数量只有 100, 这样的设计可以更好地反映交通标志在实际场景中的分布情况, 使数据集更加贴近真实环境; ②图 11 中蓝色折线尾部比红色折线短, 其原因是在原数据集中有大概 30 种比较少见的交通标志没有标注信息, 比如“w25”连续下坡标志和“w36”注意野生动物标

志, 故采用原数据集实际上训练不到这部分标志, 而平衡后的数据集将这部分标注信息补上, 使得模型能够更全面地学习和识别这些少见交通标志, 从而提高检测覆盖率。

4.2.2 合成效果对比

本文选择 Cityscapes 官方的 Cityscapes-foggy 数据集和 TTf 进行对比。从图 12 可以看出, Cityscapes-foggy 在雾的层次分布上略好于 TTf, 可能原因是 Cityscapes 本身就提供了双目视差图, 可以合成更精准的深度图, 而 TTf 是依靠单目深度估计算法, 难免产生误差。但从应用领域的角度而言, Cityscapes-foggy 旨在更广泛地应用于城市场景的多物体检测和分割, 并且只适用于德国街道, 而 TTf 更专注于中国交通标志检测, 在交通标志类别覆盖率和细节特征呈现上具有优势。

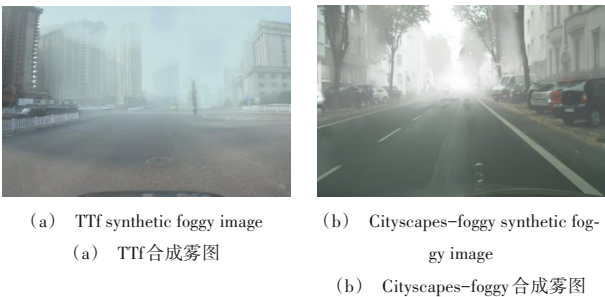


Fig. 12 Synthetic fog image comparison of TTf and Cityscapes-foggy
图 12 TTf 与 Cityscapes-foggy 合成雾图比较

4.3 目标检测实验结果分析

4.3.1 训练参数

本文在训练中采用了混合训练的方式, 参与训练的数据集一共有 3 个, 分别 TTb、TTf、TTh, 都按照 8:1:1 的方式随机划分出训练集、验证集和测试集。其他各种超参数设置如下: 每个训练批次 (Batch) 中包含的图像数目为 38; 初始学习率设置为 0.01, 学习率衰减步数设置为 (6 000, 8 000), 表明在第 6 000 步和第 8 000 步时进行学习率衰减, 衰减为余弦退火法; 优化器使用 ADAM, 训练最大迭代次数设置为 40 000, 学习率预热的迭代次数设置为 2 000, 预热阶段学习率的增长因子设置为 0.000 25, 训练最大迭代步数设置为 100 000。在训练过程接近尾声的最后 10 个 epoch 解冻主干网络, 以允许网络在保持之前训练成果的基础上进一步优化调整, 以达到更好的性能表现。

这里需要指出 2 点: ①在概念上, epoch 指模型循环训练整个数据集的次数, 而迭代步数则指在每个 epoch 中模型更新参数的次数, 本文实验中, 最大迭代步数=100 000, 训练集图片数=35 297, 训练批次=38, 可计算出每个 epoch 包含的迭代步数为 935 (35 297/38≈935), 最大 epoch 数为约 107 (100 000/935≈107), 故最后 10 个 epoch 大概相当于训练最后 935 个迭代 (10/107×100 000≈935); ②batch-size=38 是在训练过程中反复调试得到的最佳值, 这样可以充分利用显卡的显存(训练时显存一直保持在 18 G 左右, 直到最后解冻训练时显卡刚好卡在 24 G 运行, 没有出现因显

存不够而提前中止训练的问题),最大限度地缩减训练时间,并在训练中得到更好的泛化性能。

4.3.2 检测结果分析

通过在 CenterNet2 上引入联合训练和图像重叠分块方法,在 TT100K-sync 扩展的数据集上进行训练,然后根据训练得出的权重文件,对数据集的测试图片进行检测,检测结果如图 13 所示。

可以看出,不管是原始标志还是替换标志,不管是常规天气、雾天还是雾霾天,改进后的网络都能准确定位和识别场景中的交通标志,达到预期效果。但有 2 个值得注意的细节:①在图像质量随着能见度下降的情况下,越远的标志检测难度越大,比如在 3 和 5 中 ip 因为距离摄像头更远,所以随着能见度的下降置信度下降幅度更大(雾图 99%→93%,雾霾图 99%→81%);②雾霾天气下空气中的颗粒物和污染物会进一步影响图像质量,使得图像变得更加模糊、色彩不鲜艳,可以看到,图 6 中所有置信度出现了较大幅度的下降。

本文用常规天气数据集 TTb、雾天数据集 TTf、雾霾天数据集 TTh 的测试数据对目前主流模型进行对比测试,实验结果如表 2—表 4 所示。

分析实验结果可知,CenterNet2* 的测试指标模型在不同数据集上的平均精度均有所不同。TTb 的平均精度为 0.831 4,这表明在常规天气数据集图像质量较高、细节特征保留完好的前提下模型能表现良好。TTf 的平均精度为 0.761 0,说明雾天给模型带来了一定影响;而 TTh 的平均精度为 0.706 1,说明雾霾天气带来的视觉浑浊,会导致在

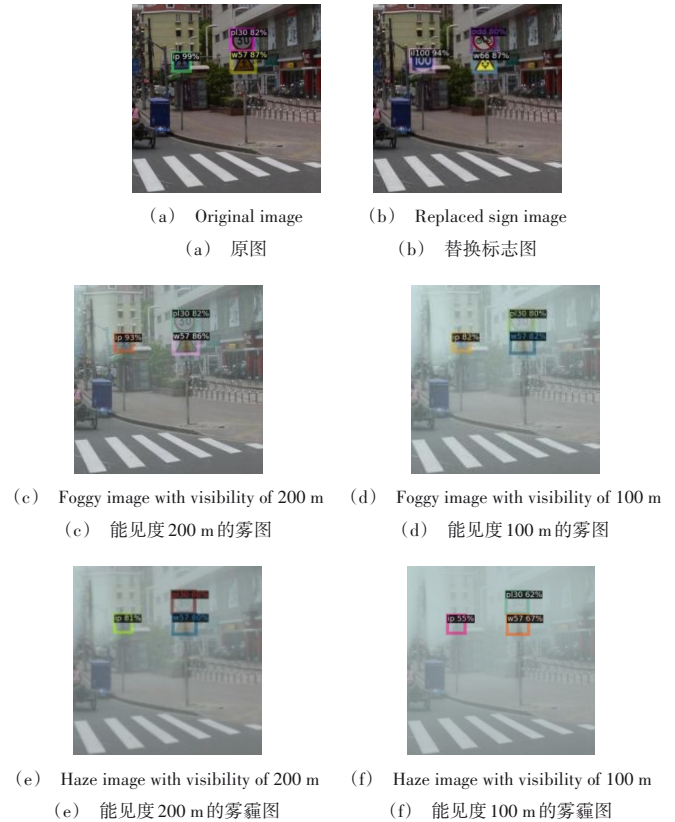


Fig. 13 Traffic sign detection results

图 13 交通标志检测结果

同等能见度条件下其图像质量比雾图更差。

本文对目前主流的目标检测模型进行比较。两种单阶段目标检测算法 YOLOX-S^[22]和 EfficientDet-D0^[23]都支

Table 2 Comparison test results on TTb(* presents our model)

表 2 TTb 数据集上的对比测试结果(*代表本文模型)

网络	Backbone	AP	AP50	AP75	APs	APm	APl	FPS
YOLOX-S	Darknet-53	0.751 2	0.833 6	0.763 4	0.642 3	0.826 7	0.870 1	48
EfficientDet-D0	EfficientNet-B0	0.761 8	0.841 0	0.772 1	0.655 1	0.831 2	0.875 4	45
Cascade R-CNN	Swin-S	0.828 9	0.890 4	0.843 7	0.720 4	0.882 7	0.908 3	29
CenterNet2	ResNet-50	0.825 4	0.881 4	0.839 0	0.705 8	0.888 3	0.914 7	25
CenterNet2*	ResNet-50	0.831 4	0.892 7	0.846 0	0.723 1	0.885 4	0.910 6	24

Table 3 Comparison test results on TTf(* presents our model)

表 3 TTf 数据集上的对比测试结果(*代表本文模型)

网络	Backbone	AP	AP50	AP75	APs	APm	APl	FPS
YOLOX-S	Darknet-53	0.676 0	0.759 4	0.689 2	0.561 8	0.752 4	0.793 7	48
EfficientDet-D0	EfficientNet-B0	0.657 0	0.740 9	0.670 1	0.543 2	0.733 7	0.775 0	45
Cascade R-CNN	Swin-S	0.725 9	0.801 0	0.740 3	0.614 5	0.798 7	0.843 2	29
CenterNet2	ResNet-50	0.735 0	0.805 5	0.741 3	0.611 4	0.774 4	0.856 0	25
CenterNet2*	ResNet-50	0.761 0	0.838 1	0.774 2	0.650 9	0.820 5	0.863 8	24

Table 4 Comparison test results on TTh(* presents our model)

表 4 TTh 数据集上的对比测试结果(*代表本文模型)

网络	Backbone	AP	AP50	AP75	APs	APm	APl	FPS
YOLOX-S	Darknet-53	0.612 0	0.695 7	0.625 2	0.500 3	0.688 9	0.730 0	48
EfficientDet-D0	EfficientNet-B0	0.607 7	0.691 6	0.620 8	0.495 7	0.684 4	0.725 4	45
Cascade R-CNN	Swin-S	0.645 5	0.735 9	0.660 3	0.533 1	0.728 2	0.770 1	29
CenterNet2	ResNet-50	0.630 1	0.724 3	0.657 9	0.538 9	0.719 7	0.765 9	25
CenterNet2*	ResNet-50	0.706 1	0.788 1	0.720 5	0.590 6	0.782 2	0.706 1	24

持多尺度的训练和检测,以速度和精度的平衡见长,但在处理复杂天气场景时表现不如两阶段方法。而 Cascade R-CNN^[24]由于其二阶段检测的优势,再结合不同 IoU 检测头级联以及 Swin Transformer^[25]层次化特征表示和自注意力机制的优点,因而在图像质量下降时能获得较好的整体检测精度。最后,CenterNet2*在 CenterNet2 的基础上构建了专门针对复杂天气的 U-subnet 联合训练模块,并采用了重叠分块的预处理方法,可以看到,在图像质量大幅下降时,仍能够取得不错的检测精度。此外,从 FPS 方面可以看到,在准确度显著提高的同时,CenterNet2*和 CenterNet2 相差无几,比起 Cascade R-CNN 落后不多,属于可以接受的范围,检测速度上达到了预期。

4.3.3 消融实验

本文通过消融实验,评估 U-subnet 和 overlap-tiling 这两个改进方法对 CenterNet2*检测性能的具体影响。U-subnet 旨在增强模型对复杂天气条件下交通场景图像特征提取的能力,而 overlap-tiling 旨在改善小尺寸目标的检测精度。通过分别以及共同移除这两种策略,可以更深入地理解它们对模型性能的贡献。将实验分为 4 组对比实验:第一组:采用原始 CenterNet2,不包含 U-subnet 和 overlap-tiling 策略;第二组:模型包含 overlap-tiling 策略,但移除 U-subnet;第三组:模型包含 U-subnet 策略,但移除 overlap-tiling;第四组:模型同时包含 U-subnet 和 overlap-tiling 策略作为对照组。

所有模型均在 TT100K-sync 数据集上进行训练和测试,并确保训练参数相同。基于实验设计可以得出消融实验结果,如表 5 所示。

Table 5 Ablation study results
表 5 消融实验结果

网络	TTc	TTf	TTh
第一组(不含 U-subnet 和 overlap-tiling)	0.881 4	0.805 5	0.724 3
第二组(只含 overlap-tiling)	0.885 2	0.817 2	0.736 8
第三组(只含 U-subnet)	0.883 1	0.831 0	0.771 8
第四组(两者都包含)	0.892 7	0.838 1	0.788 1

分析可知:①移除 U-subnet 导致模型在雾图上的 mAP50 从 0.838 1 下降至 0.817 2,在 TTh 上的 mAP50 从 0.788 1 下降至 0.736 8,表明 U-subnet 在增强特征提取能力上起到了提升作用,图像质量越低,提升效果越大;②移除 overlap-tiling 导致模型在所有数据集上的 mAP50 都有所下降,但最明显的是在 TTh 上从 0.788 1 下降至 0.736 8,表明 overlap-tiling 对于提升小目标检测作出了贡献,图像质量越低时在下采样中保留小目标特征的效果越好;③同时包含这两种策略的模型在各项指标上均表现最好,验证了这两种策略的综合效应对于模型性能的提升是正面的。

由此可以得出结论:U-subnet 和 overlap-tiling 两种策略对于提升 CenterNet2 模型在复杂天气条件下的小目标检测能力起到了显著作用。U-subnet 主要通过增强模型的特征提取能力,改善了在降低能见度的复杂天气条件下的

图像处理性能;而 overlap-tiling 通过改进图像分块策略,有效提升了小目标检测精度。这两种策略的结合,为提升复杂条件下的目标检测性能提供了有价值的思路。

5 结语

本文提出了一种基于 TT100K 的新数据集扩展方法 TT100K-sync,有效地扩展了交通标志类别和天气场景,同时提出构建基于联合训练的 U-subnet 网络模块和引入图像重叠分块方法,对复杂天气场景下的目标检测进行了优化。实验结果表明,该方法能够显著提高在复杂天气条件下的交通标志检测识别精度。

尽管本文研究取得了一定成果,但仍存在一定的改进空间。从网络结构的角度,可考虑将特征提取主干从 ResNet50 换成 Swin Transformer,以更好地捕捉全局和局部之间的长距离依赖关系,或者引入自注意力机制或空间注意力机制,以动态地调整网络在不同位置和尺度上的注意力分布,使 U-subnet 在目标特征提取和恢复方面更加有效和灵活。由于时间原因,本文研究只引入了雾图和雾霾图,后续将引入合成雨图和雪图。雨图的合成相较于雾图更为复杂,其合成原理包括随机生成雨滴图、随机生成雨纹图、根据深度图合成雾图^[26]。而目前,关于雪图的主要研究都是基于生成对抗网络(GAN)的模型^[27],使用编码器提取图像中的相关特征信息,再通过解码器将这些特征重构为雪图。这两种图像的加入将进一步增强本文模型在各种天气条件和场景下的适应能力。

参考文献:

[1] ZHU Z, LIANG D, ZHANG S, et al. Traffic-sign detection and classification in the wild[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2110–2118.

[2] CHEN F, LIU Y P, LI S Y. Survey of traffic sign detection and recognition methods in complex environment[J]. Journal of Computer Engineering & Applications, 2021, 57(16): 65–73.

陈飞, 刘云鹏, 李思远. 复杂环境下的交通标志检测与识别方法综述[J]. 计算机工程与应用, 2021, 57(16): 65–73.

[3] CUI Y, JIA M, LIN T Y, et al. Class-balanced loss based on effective number of samples[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 9268–9277.

[4] YANG T T, TONG C. Real-time detection network for tiny traffic sign using multi-scale attention module[J]. Science China Technological Sciences, 2022, 65(2): 396–406.

[5] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211–252.

[6] PEREZ L, WANG J. The effectiveness of data augmentation in Image classification using deep learning [DB/OL]. <https://arxiv.org/abs/1712.04621>, 2017.

[7] MENG F X, YU L. Traffic sign recognition based on YOLOv5-EA[J]. Journal of Liaoning University of Technology (Natural Science Edition),

- 2022,42(5):303-310.
- 孟繁星,于臻. 基于YOLOv5-EA的交通标志识别[J]. 辽宁工业大学学报(自然科学版),2022,42(5):303-310.
- [8] WU Y P, FU Y X, SHEN L J, et al. Long-tail distribution traffic sign recognition model based on YOLOX-Tiny[J]. Journal of System Simulation, 2024,36(11):2503-2516.
- 伍云鹏,付应雄,沈丽君,等. 基于YOLOX-Tiny的长尾分布交通标志识别模型[J]. 系统仿真学报,2024,36(11):2503-2516.
- [9] NANNI L, PACI M, BRAHNM S, et al. Comparison of different image data augmentation approaches[J]. Journal of Imaging, 2021,7(12):254.
- [10] FENG Y, HE M, LIU W. A new method for foggy image enhancement [C]//Proceedings of 2009 4th IEEE Conference on Industrial Electronics and Applications, 2009:2416-2419.
- [11] LIN Y, LI Y, CUI H, et al. WeaGAN: generative adversarial network for weather translation of image among multi-domain [C]//Proceedings of 2019 6th International Conference on Behavioral, Economic and Socio-Cultural Computing (BESC), 2019: 1-5.
- [12] SAKARIDIS C, DAI D, VAN G L. Semantic foggy scene understanding with synthetic data [J]. International Journal of Computer Vision, 2018, 126: 973-992.
- [13] CHEN D H, SUN S R, WANG Y C, et al. An improved traffic sign recognition algorithm based on YOLOv3 [J]. Journal of Henan University of Science & Technology(Natural Science), 2022,43(6):31-36.
- 陈德海,孙仕儒,王昱朝,等. 一种改进YOLOv3的交通标志识别算法[J]. 河南科技大学学报(自然科学版),2022,43(6):31-36.
- [14] XU Z, CHEN M. Modified image dehazing algorithm of traffic sign image in fog and haze weather[J]. Journal of Computer Applications, 2017, 37(8): 2329.
- [15] LU Y, LU J, ZHANG S, et al. Traffic signal detection and classification in street views using an attention model[J]. Computational Visual Media, 2018,4: 253-266.
- [16] ALHICHRI H, OTHMAN E, ZUAIR M, et al. Tile-based semi-supervised classification of large-scale VHR remote sensing images [J]. Journal of Sensors, 2018,4:21702261.
- [17] GODARD C, MAC A O, FIRMAN M, et al. Digging into self-supervised monocular depth estimation [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 3828-3838.
- [18] ZHANG N, ZHANG L, CHENG Z. Towards simulating foggy and hazy images and evaluating their authenticity [C]//Guangzhou: 24th International Conference on Neural Information Processing, 2017.
- [19] SWINEHART D F. The beer-lambert law [J]. Journal of Chemical Education, 1962,39(7): 333.
- [20] ZHOU X, KOLTUN V, KRÄHENBÜHL P. Probabilistic two-stage detection [DB/OL]. <https://arxiv.org/abs/2103.07461>, 2021.
- [21] HUANG S C, LE T H, JAW D W. DSNet: joint semantic learning for object detection in inclement weather conditions [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020,43(8): 2623-2633.
- [22] GE Z, LIU S, WANG F, et al. YOLOX: exceeding YOLO Series in 2021 [DB/OL]. <https://arxiv.org/pdf/2107.08430>, 2021.
- [23] TAN M, PANG R, LE Q V. Efficientdet: scalable and efficient object detection [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781-10790.
- [24] CAI Z, VASCONCELOS N. Cascade R-CNN: delving into high quality object detection [C]//Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition, 2018: 6154-6162.
- [25] LIU Z, LIN Y, CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted windows [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 10012-10022.
- [26] HU X, FU C W, ZHU L, et al. Depth-attentional features for single-image rain removal [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 8022-8031.
- [27] YANG H, DING M, CARBALLO A, et al. Synthesizing realistic snow effects in driving images using GANs and real data with semantic guidance [C]//Proceedings of 2023 IEEE Intelligent Vehicles Symposium, 2023: 1-6.

(责任编辑:孙娟)