

基于超分辨率重建的小目标智能检测算法

蔡心悦^{1**}, 周杨^{1,2,3*}, 胡校飞^{1,2}, 吕亮^{1,2,3}, 赵璐颖^{1,4}, 彭杨钊¹¹信息工程大学地理空间信息学院, 河南 郑州 450001;²智慧中原地理信息技术河南省协同创新中心, 河南 郑州 450001;³时空感知与智能处理自然资源部重点实验室, 河南 郑州 450001;⁴河南建筑职业技术学院, 河南 郑州 450001

摘要 针对小目标占有像素少导致检测精确率低的问题, 提出一种基于超分辨率重建的小目标检测算法。首先, 通过图像预处理对高分辨率图像分块并筛选出含有目标的子图像; 其次, 构建超分辨率锐化增强模块, 引入锐化图像和锐化损失, 以获得边缘更清晰的高分辨率子图像; 然后, 采用多尺度锐化目标检测模块检测目标, 该模块添加边缘锐化模型, 在深层特征层中进一步锐化图像边缘, 弥补深层卷积对细节的损失; 最后, 根据子图像编号将小目标检测结果回归到原图像中, 完成小目标图像检测。在 PASCAL VOC 数据集和 COCO 2017 数据集上的实验结果表明, 所提算法的平均精确率 (mAP) 分别为 85.3% 和 54.0%, 对 COCO 数据集的小目标检测精确率为 43.5%, 高于次优值 9.7 个百分点。因此, 所提算法可以有效减少小目标漏检的次数, 提高检测精确率。

关键词 图像处理; 小目标检测; 超分辨率增强; 卷积神经网络; 多尺度特征融合; 边缘锐化

中图分类号 TP391.4

文献标志码 A

DOI: 10.3788/LOP220882

Intelligent Detection Algorithm for Small Targets Based on Super-Resolution Reconstruction

Cai Xinyue^{1**}, Zhou Yang^{1,2,3*}, Hu Xiaofei^{1,2}, Lü Liang^{1,2,3}, Zhao Luying^{1,4}, Peng Yangzhao¹¹Institute of Geospatial Information, Information Engineer University, Zhengzhou 450001, Henan, China;²Collaborative Innovation Center of Geo-Information Technology for Smart Central Plains, Henan Province, Zhengzhou 450001, Henan, China;³Key Laboratory of Spatiotemporal Perception and Intelligent Processing, Ministry of Natural Resources, Zhengzhou 450001, Henan, China;⁴Henan Technical College of Construction, Zhengzhou 450001, Henan, China

Abstract A small target detection algorithm based on super-resolution reconstruction is proposed to solve the problem of low detection accuracy of small targets occupying a few pixels. First, a high-resolution image is segmented via image preprocessing and sub-images containing targets are filtered out. Second, a super-resolution sharpening enhancement module is constructed, and the sharpening image and sharpening loss are introduced to obtain high-resolution sub-images with clearer edges. Subsequently, a multi-scale sharpening target detection module is used to detect the target; it uses an edge-sharpening model to further enhance the image edges of the deep feature layer to compensate for the loss in details due to deep convolution. Finally, the small-target detection results are returned in the original image based on the sub-image number used to complete small target image detection. The proposed detection algorithm is then verified using the PASCAL VOC and COCO 2017 datasets, where the average accuracies (mAP) are 85.3% and 54.0%, respectively. Moreover, the small target detection accuracy of the COCO dataset is 43.5%, which is 9.7 percentage points higher than the suboptimal value. Therefore, the proposed algorithm can effectively reduce the number of times small targets are missed during detection, thus improving the detection accuracy.

Key words image processing; small target detection; super-resolution enhancement; convolutional neural networks; multi-scale feature reuse; edge sharpening

收稿日期: 2022-03-04; 修回日期: 2022-03-29; 录用日期: 2022-05-31; 网络首发日期: 2022-06-10

通信作者: *zhouyang3d@163.com

1 引言

目标检测作为计算机视觉的基础任务之一,在自动驾驶识别^[1]、智能安全防控、医学成像以及遥感影像等方面得到广泛应用。近年来,基于卷积神经网络(CNN)的目标检测算法不断发展,总体可以分为两阶段检测算法和单阶段检测算法^[2]。两阶段检测算法首先通过区域候选网络(RPN)生成候选区域,再进行分类和回归,即经过两阶段依次得到位置和分类结果,如 Faster R-CNN^[3]、R-FCN^[4]等;单阶段检测算法通过神经网络直接对目标进行定位,同时输出目标的类别检测信息,将目标坐标定位问题转化为回归问题,如 SSD^[5]、YOLO 系列^[6-9]等。单阶段检测算法在特征提取过程中可避免生成候选区域,提高检测速度,但检测精确率较低。

目前,基于 CNN 的目标检测算法能较好地检测大、中目标,而小目标由于在图像中所占像素少、分辨率低,且随着网络的加深、图像的抽象化,高层特征层的小目标所占像素更少,导致难以提取到小目标的信息,其平均检测精度仅为大目标的 1/5^[10]。因此,小目标检测成为现阶段计算机视觉领域中的难点。

现有的小目标检测一般是基于主流的目标检测算法进行优化,例如增强小目标样本、优化训练方法、无锚框机制以及特征融合等。Li 等^[11]针对 SSD 中不同尺度特征图难以融合的问题,借鉴 feature pyramid network(FPN)的思想,提出 FSSD 模型,通过在 SSD 算法上添加轻量级的特征融合模块,采用双线性插值法将特征图变为同尺度,并融合生成新的特征金字塔,充分利用各层特征信息,提高小目标检测率。Kisantal 等^[12]通过对数据集中含有小目标的图像进行重复采样和多次复制小目标的策略,提高小目标损失函数的权重,从而提升小目标的检测性能。Tian 等^[13]借鉴语义分割思想,提出一种像素级全卷积单阶(FCOS)目标检测模型,以逐像素预测的方法进行目标检测,同时采用无锚框机制,避免因预定义的锚框过大对小目标造成漏检,且提高了检测速度。Deng 等^[14]针对 FPN 中不同尺度的特征耦合会影响小目标检测性能的问题,提出 extended feature pyramid network(EFPN),通过特征纹理迁移模块融合语义内容和区域纹理特征,并提取可靠的区域细节用于小目标检测,提高准确率。杨其利等^[15]基于语义分割任务,提出一种基于深度学习的弱小目标检测方法,利用全卷积递归网络学习复杂背景下弱小目标的特征,在网络中使用残差学习和递归操作优化网络性能。

近些年,超分辨率(SR)算法以其强大的图像重建能力而受到广泛关注,该算法可以有效扩大小目标的分辨率、增加特征信息量,为提高小目标检测精确率提供了一种可行的方案。现有 SR 算法的目的是生成视觉效果更优的高分辨率图像。对于小目标检测问题,

SR 的重建结果不仅要顾及视觉效果,更要考虑小目标边缘清晰度,保证在特征提取时浅层特征层能更好地提取小目标轮廓、形状等信息。

针对小目标像素占比少导致检测率不高的问题,本文基于生成对抗网络,设计了一种 SR 重建的小目标智能检测算法,通过扩大小目标的相对像素数量、增加小目标的特征信息,提高小目标的检测精确率。主要贡献有:1)设计了一种小目标检测思路,通过图像预处理和图像 SR 重建将小目标检测问题转化为非小目标检测问题;2)提出 SR 锐化增强模块,使小目标特征信息更丰富,有利于特征提取;3)提出多尺度锐化目标检测模型,提高小目标检测精确率。

2 算法设计与实现

所提算法主要包括图像预处理、图像 SR 锐化增强、目标检测和子图像回归等 4 个模块,流程如图 1 所示。首先,通过图像预处理,对包含小目标的图像进行分块处理,设置一定的重叠率,保证小目标的完整性,并通过图像分类模型筛选出含有目标的子图像;然后,将子图像输入 SR 锐化增强模块,获得边缘更清晰的图像,使神经网络能够提取到更丰富的目标特征;随后,采用多尺度锐化目标检测模型对生成的高分辨率子图像进行目标检测,该模型将浅层特征与高层语义相结合,同时加强图像锐化,提高小目标检测的精确率;最后,根据子图像编号将小目标检测结果回归到原图像中,完成小目标图像检测。

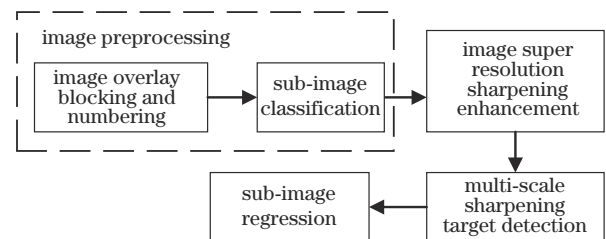


图 1 总体流程图

Fig. 1 Entire flow diagram

2.1 图像预处理

首先对含有小目标的图像进行预处理,将小目标检测问题转化为非小目标检测问题。图像预处理主要包括图像分块和图像分类这 2 部分。

2.1.1 图像分块

图像分块方法主要分为 2 种:图像直接分块和图像重叠分块。图像直接分块方法简单、效率高,但其不考虑切块导致目标被分割的情况。若两相邻子图像之间存在目标,则该目标会因被分割而导致无法被检测到,如图 2(a)所示。因此,该方法一般用于目标不在边缘区域、目标较少等情况。图像重叠分块考虑目标在切块缝隙的情况,通过设置重叠率、增加子图像重叠区域的方法来保证目标不被损坏,避免由于图像分块



图2 图像分块。(a)直接分块;(b)重叠分块

Fig. 2 Image blocking. (a) Direct blocking; (b) overlap blocking
导致目标丢失、降低小目标漏检率的情况,如图2(b)所示。常应用于较复杂的图像处理任务中,可在一定程度上保证小目标的完整性,有利于提高检测精确率。本实验采用图像重叠分块方法进行图像分块。

设原图像大小为 $W \times H$, 共分割出 $m \times n$ 块子图像, 子图像大小为 $s \times s$, 重叠区宽度为 o , 当重叠率为 α 时:

$$m = (H - 5o) / (s - 2o), \quad (1)$$

$$n = (W - 5o) / (s - 2o), \quad (2)$$

$$o = \alpha s. \quad (3)$$

图3中,红色框区域内为不重合子图像,其余为重叠部分。为保证子图像大小相同,边缘图像用黑色补齐,效果如图3(a)所示,中部图像示意图如图3(b)所示。

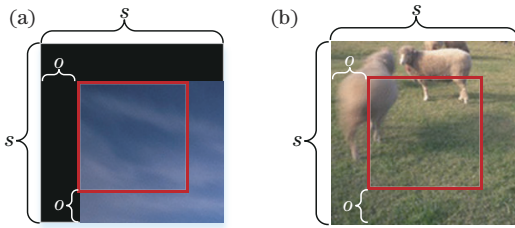


图3 重叠分块示意图。(a)边缘图像示意图;(b)中部图像示意图
Fig. 3 Schematic of overlap block. (a) Schematic of edge image; (b) schematic of middle image

图像分块后,需要对子图像进行编号,以便在目标检测后进行子图像合并操作。子图像命名格式为“XXX_mm_nn”,其中,“XXX”为图像原名,“mm”为子图像所在行数,“nn”为子图像所在列数。

2.1.2 图像分类

图像分类是理解图像的第一步,主要解决图像中有什么的问题。在图像预处理中,图像分类的目的是将含有目标的子图像筛选出来,将只含有背景的子图像剔除,后续只针对含有目标的子图像进行图像增强与目标检测,从而减少工作量。因此,在图像分类时,需将图像背景设为一个类别。

采用基于CNN的方法进行图像分类。相对于多目标复杂分类,区分背景与目标,较为简单。因此,采用网络层数较少的 AlexNet^[16]为骨干网络进行特征提取。在训练过程中,损失函数决定了网络模型参数的更新,对特征表达起到重要作用。采用表现较优的

congenerous cosine(COCO)损失函数^[17],引入特征与类中心的 cosine 距离度量损失,增大类间差距、缩小类内距离,使类内更紧凑。

已知特征 i 与第 j 类中心 c_j 的余弦相似度为 $C(\mathbf{f}^{(i)}, c_j) = \mathbf{f}^{(i)\top} \cdot \mathbf{c}_j / \|\mathbf{f}^{(i)}\| \|\mathbf{c}_j\|$, 则损失函数 L 为

$$L = \sum_{i \in \mathcal{B}} \frac{\exp C(\mathbf{f}^{(i)}, c_j)}{\sum_{m \neq j} \exp C(\mathbf{f}^{(i)}, c_m)}, \quad (4)$$

式中: $\mathbf{f}^{(i)}$ 表示特征 i 的特征向量; \mathcal{B} 表示一个小批量; m 为类别数。

将特征和类中心正则化得到:

$$\begin{cases} \hat{\mathbf{f}}^{(i)} = \frac{\alpha \mathbf{f}^{(i)}}{\|\mathbf{f}^{(i)}\|}, \\ \hat{\mathbf{c}}_k = \frac{\mathbf{c}_k}{\|\mathbf{c}_k\|} \end{cases}, \quad (5)$$

式中: α 为比例因子; 类中心 c_k 为特征平均值: $c_k = \frac{1}{N_k} \sum_{i \in \mathcal{B}} \delta(l_i, k) \mathbf{f}^{(i)} \in \mathbf{R}^D$; N_k 为 k 类的特征数量。

设 $p_k^{(i)} = \frac{\exp(\hat{\mathbf{c}}_k^\top \cdot \hat{\mathbf{f}}^{(i)})}{\sum_m \exp(\hat{\mathbf{c}}_m^\top \cdot \hat{\mathbf{f}}^{(i)})}$, 则损失函数 L^{COCO} 为

$$L^{\text{COCO}}(\mathbf{f}^{(i)}, \mathbf{c}_k) = - \sum_{i \in \mathcal{B}, k} t_k^{(i)} \log p_k^{(i)} = - \sum_{i \in \mathcal{B}} \log p_{l_i}^{(i)}, \quad (6)$$

式中: $t_k^{(i)} \in \{0, 1\}$, 为特征 i 的二进制映射。

2.2 图像SR重建

目前SR优化方法可归纳为以下几类:残差网络学习、递归学习、密集网络连接、注意力机制、生成对抗网络等。其中:残差网络、递归学习和密集网络连接优化方法主要解决由于网络加深带来的模型退化、数据过度拟合以及梯度消失等问题;注意力机制优化利用不同通道之间特征表示的相互依赖性和相互作用,提高模型的特征表达能力;生成对抗网络利用生成对抗学习策略对网络模型进行优化,实现较大放大因子下的图像重建。

在SR领域,现有的生成对抗网络主要研究如何重建更接近真实高分辨率图像,追求视觉效果。针对小目标检测的特点,SR算法不仅要追求视觉效果,更要提升小目标边缘清晰度,以保证在特征提取时浅层特征层能更好地提取小目标轮廓、形状等信息。图像锐化操作能快速聚焦模糊边缘、提高图像清晰度,使图像中的目标形状更明显、色彩更鲜明。因此,在判别器(D)输入中加入锐化图像,使生成器(G)生成的伪高分辨率图像、真高分辨率图像以及锐化图像三者进行博弈,同时在损失函数中引入锐化损失,生成带有锐化风格的图像,以便进行小目标检测。

生成对抗网络主要由生成器(G)和判别器(D)这2个网络模块组成。其中:G负责根据预训练模型将输入的低分辨率图像生成伪高分辨率图像,并试图欺骗判别器;D负责对伪高分辨率图像的真实性进行判别,

试图正确区分真假。 G 和 D 相互对立且交替学习,形成闭环,当 D 输出的伪高分辨率图像与高分辨率图像之比为1:1时,即认为 G 生成的图像接近真高分辨率图像。加入锐化图像后的模型如图4所示。当 D 输出的伪高分辨率图像、高分辨率图像、锐化图像之比为1:1:1时,即认为 G 生成了锐化风格的高分辨率图像。

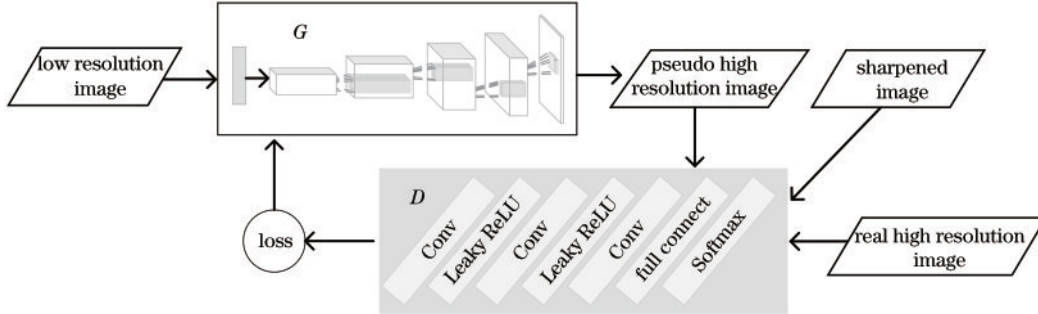


图4 SR锐化模块结构图

Fig. 4 Structure map of SR sharpening module

损失函数对模型输出值和真实值之间的一致性进行评估,是更新生成器参数的重要指标。SRGAN^[18]采用的感知损失函数包含内容损失和对抗损失。不同于直接计算伪高分辨率图像和真高分辨率图像的均方误差(MSE)的损失函数,内容损失在特定的特征图上计算其MSE,从而增强对目标细节的处理。在此基础上,提出感知锐化损失,增强目标的锐化处理,其损失函数如下:

假设图像大小为 $C \times W \times H$, I^{HR} 为高分辨率图像, I^{LR} 为低分辨率图像,则:

$$L^{SR} = L_X^{SR} + 10^{-3}L_{GAN}^{SR} + L_{sharpen}^{SR}, \quad (7)$$

式中: L_X^{SR} 为内容损失; L_{GAN}^{SR} 为对抗损失; $L_{sharpen}^{SR}$ 为锐化损失。

$$L_X^{SR} = \frac{1}{WH} \sum_{x=1}^W \sum_{y=1}^H \left\{ \phi(I^{HR})_{x,y} - \phi[G_{\theta_g}(I^{HR})]_{x,y} \right\}^2, \quad (8)$$

$$L_{GAN}^{SR} = \sum_{n=1}^N -\log D_{\theta_d}[G_{\theta_g}(I^{LR})], \quad (9)$$

式中: G_{θ_g} 表示生成器; $\phi(\cdot)_{x,y}$ 表示特征映射; D_{θ_d} 表示图像是否真实的概率分布; N 表示图像数量。

锐化损失定义为输入与输出图像的特征矩阵的Frobenius范数的平方,用来刻画2个特征图之间的相似性,其计算公式如下:

$$L_{sharpen}^{SR} = \|F_j^\phi(\mathbf{x}_1) - F_j^\phi(\mathbf{x}_2)\|_F^2, \quad (10)$$

$$F_j^\phi(\mathbf{x}) = \frac{1}{CWH} \sum_{w=1}^W \sum_{h=1}^H \phi_j(\mathbf{x})_{c \times wh} \phi_j^T(\mathbf{x})_{wh \times c}, \quad (11)$$

式中: $\phi_j(\mathbf{x})$ 为CNN第 j 层输出的特征图; c 和 c' 表示不同的通道。

在判别器中加入锐化图像,同时采用感知锐化损失函数,加强生成器对目标边缘清晰度以及色彩鲜明的处理,为目标检测奠定基础。

图4中, G 采用转置卷积网络,卷积核大小为 3×3 ,步长为2。设输入图像大小为 $s \times s$,则生成图像大小为 $s \times 2^n$, n 为卷积次数。 D 基于AlexNet进行改进,舍弃池化层,并采用Leaky ReLU激活函数,以最大限度保留前层信息。

2.3 目标检测

经过图像预处理以及图像SR重建后,将小目标检测问题转化为非小目标检测问题。基于CNN的目标检测方法认为:卷积层越深,网络对特征提取的能力越强,但CNN经过多次卷积操作后,输出特征图的分辨率会成倍下降,从而导致网络对图像细节感知能力较差。此外,经SR增强的图像会不可避免地出现部分图像模糊的现象。为保持图像的分辨率,提出多尺度锐化目标检测模型,如图5所示。以融合多尺度特征提取的SSD目标检测模型为骨干网络,采用卷积保持特征图的分辨率,同时,针对图像模糊问题,提出边缘检测锐化模型,通过增强目标边缘像素提高目标检测正确率。

SSD是单阶段目标检测方法,以VGG16前5层卷积层作为基础网络,舍弃第5层(Conv5)后的池化层、全连接层以及Softmax层,同时添加5个特征提取层(Conv6~10)。其中,dilated Conv_2表示采用空洞卷积且空洞率为2,dilated Conv_3表示采用空洞卷积且空洞率为3。多尺度锐化目标检测模型共生成6个特征图,并通过特征金字塔结构融合多尺度目标检测,实现低层网络检测小目标、高层网络检测大目标,从而提高目标检测召回率。

针对图像模糊的问题,提出边缘检测锐化模型,如图6所示。在SSD网络中,较低层网络能保持目标形状,而高层网络中目标较抽象。为提高目标边缘清晰度,在Conv5、Conv6、Conv7、Conv8以及Conv9卷积层后添加边缘检测锐化模型,以提高目标检测正确率。

边缘检测锐化模型通过计算梯度检测出图像的所有边缘信息,并对边缘像素进行锐化,主要目的是补偿图像轮廓、突出图像的边缘信息,以使图像显得更为清

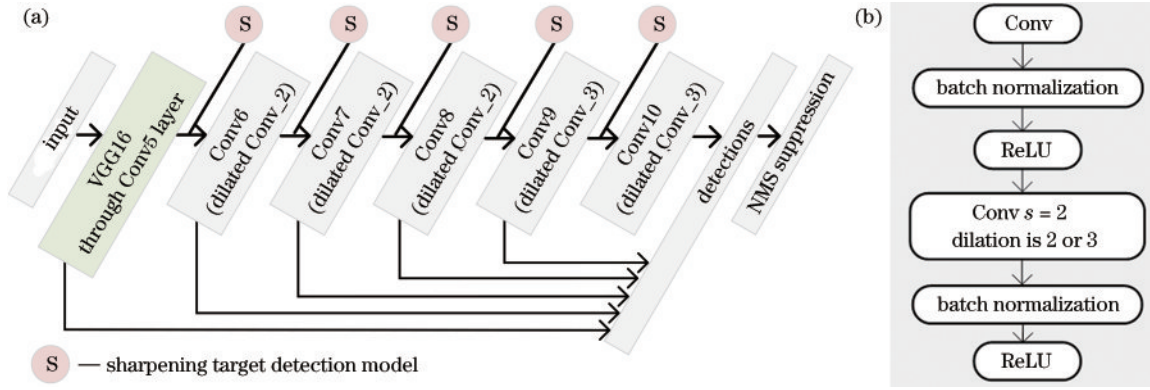


图 5 多尺度锐化目标检测模型。(a)整体模型；(b)添加层结构

Fig. 5 Multi-scale sharpening target detection model. (a) Overall model; (b) structure of added layer

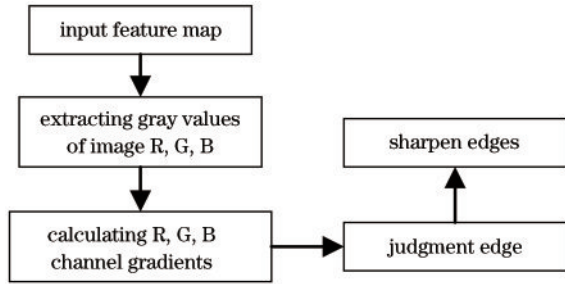


图 6 边缘检测锐化模型

Fig. 6 Model of edge detection sharpening

晰。其计算公式如下：

$$g(x, y) = \begin{cases} f(x, y) + C[f(x, y) - l(x, y)], & e(x, y) > 0 \\ f(x, y), & e(x, y) = 0 \end{cases}, \quad (12)$$

式中： $f(x, y)$ 为原始图像； $l(x, y)$ 为原始图像的模糊图像； C 是常数，用以控制图像锐化程度； $e(x, y)$ 为包含最终边缘信息的图像。对图像上每一像素：若 $e(x, y) > 0$ ，则表示该像素是边缘，需进行锐化；若 $e(x, y) = 0$ ，则表明该像素不是边缘，不需锐化，像素值保持不变。

针对离散的数字图像，采用边缘检测 Sobel 算子作为边缘检测器，对图像 R、G、B 通道的像素点逐一进行卷积计算，即：

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \times f(x, y),$$

$$G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \times f(x, y), \quad (13)$$

式中： G_x 和 G_y 分别为原始图像 $f(x, y)$ 在 x 方向和 y 方向上的梯度。则 $f(x, y)$ 的梯度计算公式为

$$|G| = \sqrt{G_x^2 + G_y^2}. \quad (14)$$

一般认为，目标图像边缘处的梯度值高于其他相对平缓位置。根据像素的梯度值，包含边缘信息的图

像计算如下：

$$e(x, y) = \begin{cases} |G|, & |G| \geq T \\ 0, & |G| < T \end{cases}, \quad (15)$$

式中： T 为常数，用以调整只包含目标图像边缘的梯度值。将 $e(x, y)$ 代入式(12)，即可得到锐化图像。

2.4 图像回归

子图像经过目标检测后，得到目标框在子图像中的坐标。已知子图像的命名格式为“XXX_mm_nn”，现需将已获取的目标框回归到原图像中，回归过程如下：

- 1) 读取子图像分辨率大小以及其所在行数、列数，计算子图像左上角在原图中的坐标；
- 2) 在子图像中，以左上角为原点，读取目标框的坐标，格式为 $(x_{min}, y_{min}, x_{max}, y_{max})$ ；
- 3) 根据子图像在原图像中的坐标和目标框在子图像中的坐标计算出目标框在原图中的坐标；
- 4) 对于重叠的目标，根据最大合并原则，保留检测概率最大的目标框；
- 5) 将所有目标框回归到原图上，完成小目标检测。

3 实验分析

3.1 实验环境

实验在 PyTorch 1.5 的深度学习框架上进行，使用的操作系统为 Ubuntu 18.04 64 位，硬件配置为 Intel Core i9-9900X CPU、NVIDIA GeForce RTX 2080Ti GPU。

3.2 实验设计与结果分析

实验分为两部分，即依次对 SR 锐化增强模块和多尺度锐化目标检测模型的有效性进行验证。

3.2.1 SR 实验分析

3.2.1.1 实验设置

1) 数据集

SR 实验使用 DIV2K 作为训练集。DIV2K 数据集包含 1000 张 2K 分辨率的高清图像，其中，训练集为 800 张，验证集为 100 张，测试集为 100 张。在实验过

程中,对训练集图像进行锐化处理,作为辨别器中锐化图像的输入。此外,为增强样本,对图像进行分割,共生成 40896 张图像,图像分辨率为 480 pixel×480 pixel。其中,32592 张图像用于训练,4152 张图像用于验证,4136 张图像用于测试。另选 Set5、Set14 以及 BSD100 为测试集进行对比,在 PASCAL VOC 2007 和 COCO 2017 数据集上进行验证。

2) 评价指标

SR 实验采用峰值信噪比 (PSNR) 和结构相似性

(SSIM) 这 2 个客观指标评价生成的伪高分辨率图像的质量。

3.2.1.2 实验分析

实验上采样因子采用 ×2、×4、×8 这 3 种,与较先进的 5 种 SR 方法进行对比实验。所提模型初始学习率为 10^{-4} ,衰减率为 0.5,衰减步长为 200,共训练 1000 个 epoch,实验结果如表 1 所示,其中加粗数据为同一尺度因子下的每列最优值,加下划线数据为次优值。

表 1 所提方法与先进 SR 方法对比结果

Table 1 Comparison results obtained by using proposed method and latest SR methods

Scale	Method	DIV2K	Set5	Set14	BSD100
		PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
×2	SRCNN	37.05/0.9458	36.66/0.9299	30.24/0.8688	29.56/0.8431
×2	EDSR	38.55/0.9688	38.20/0.9606	34.02/0.9204	32.57/0.9001
×2	ESRGAN	38.13/0.9664	37.63/0.9588	33.04/0.9118	31.85/0.8942
×2	DRN	37.74/0.9620	37.03/0.9513	<u>33.98/0.9192</u>	<u>32.52/0.8590</u>
×2	LIIF-edsr	34.99/0.9353	<u>38.17/0.9365</u>	33.97/0.8891	32.32/0.8642
×2	Proposed method	<u>38.19/0.9698</u>	37.94/0.9612	33.52/0.9285	32.14/0.9108
×4	SRCNN	32.58/0.9052	30.49/0.8628	27.50/0.7513	26.90/0.7101
×4	EDSR	34.12/0.9264	<u>32.62/0.8984</u>	28.94/0.7901	27.71/0.7006
×4	ESRGAN	34.08/0.9118	32.60/0.9002	28.88/0.7896	<u>27.76/0.7432</u>
×4	DRN	34.16/0.9253	32.68/0.9010	<u>28.93/0.7900</u>	27.78/0.7440
×4	LIIF-edsr	29.27/0.8183	32.50/0.8511	28.80/0.7377	27.74/0.7183
×4	Proposed method	<u>34.14/0.9310</u>	32.52/0.9123	28.90/0.8023	27.70/0.7520
×8	SRCNN	28.85/0.7110	25.33/0.689	23.85/0.5930	22.31/0.5526
×8	EDSR	27.47/0.7913	26.96/0.7750	24.91/0.6400	23.19/0.5680
×8	ESRGAN	25.72/0.7414	26.00/0.7027	23.14/0.6577	25.96/0.6375
×8	DRN	28.96/0.7861	27.41/0.7900	<u>25.25/0.6520</u>	24.98/0.6050
×8	LIIF-edsr	27.09/0.7422	<u>27.14/0.7775</u>	25.15/0.6438	24.91/0.5832
×8	Proposed method	<u>28.93/0.7964</u>	26.98/0.7792	25.42/0.6623	<u>25.66/0.6458</u>

在表 1 中,对于 PSNR 指标,所提方法并不突出,原因是 PSNR 是衡量重建图像与真实图像相似度的指标,PSNR 值越高,在视觉上重建图像和原图越相似,而所提方法考虑图像边缘清晰度,在 GAN 中加入了锐化图像,导致 PSNR 值略低于最优值,但总体结果良好。SSIM 指标衡量的是图像结构、亮度与对比度等信息,所提方法除了尺度因子为 ×8 时,在 Set5 测试集上比 DRN 略低以外,其余 SSIM 值均为最优值,表明所提方法对图像边缘、纹理等高频信息重建度更好,重建图像与真实图像的结构相似程度更高。

图 7 依次展示了尺度因子为 ×2、×4、×8 时各模型的图像重建结果。其中:图 7(a) 为通用数据集中的图像,包含高分辨率图像,可用于结果对比;图 7(b)、图 7(c) 分别为 PASCAL VOC 2007 和 COCO 2017 数据集中的图像,不包含高分辨率图像。通过截取的头饰、人以及网球可以看出:SRCNN 直接使用双三次插值扩大图像分辨率,重建图像较模

糊;基于像素损失的 EDSR 和 DRN 能较好地重建图像图案、形状等信息,但图像边缘细节处理较模糊;加入对抗损失的 ESRGAN 在视觉感官上与原高分辨率图像最为接近;而能连续表示图像的 LIIF-edsr 也具有较高程度的保真度。所提模型的优势在于对图像边缘及图像纹理的重建效果更好,特别是对黑色边缘的重建,使得图像边缘更加清晰,为下一步的目标检测奠定基础。

3.2.2 小目标检测实验分析

3.2.2.1 实验设置

1) 数据集

目标检测实验使用 PASCAL VOC 数据集和 COCO 2017 数据集。PASCAL VOC 数据集包含 PASCAL VOC 2007 和 PASCAL VOC 2012,共有 20 类目标,其中,训练集包含 16551 张图像,测试集包含 4952 张图像。COCO 2017 数据集包含 80 类目标,训练集包含 99310 张图像,验证集包含 4031 张图像,测

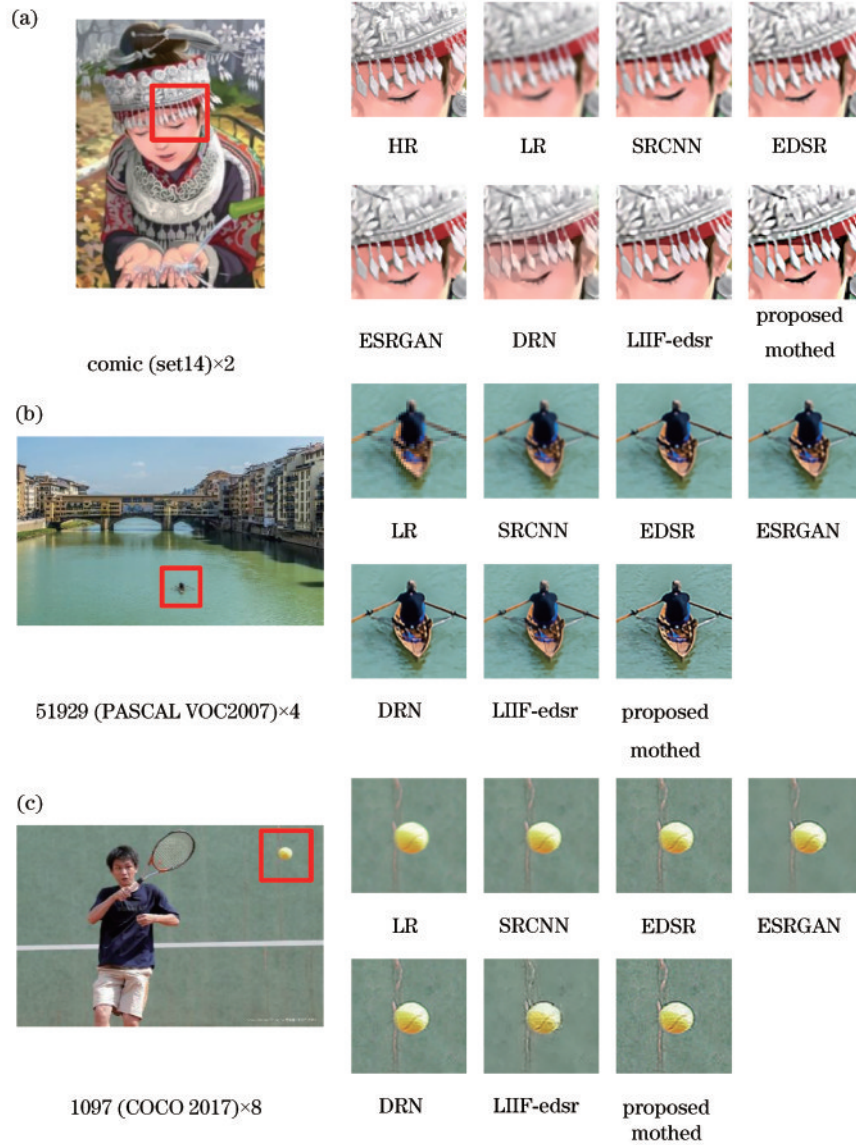


图 7 各模型重建结果。(a)尺度因子为 $\times 2$; (b)尺度因子为 $\times 4$; (c)尺度因子为 $\times 8$

Fig. 7 Reconstruction results of each model. (a) Scaling factor of $\times 2$; (b) scaling factor of $\times 4$; (c) scaling factor of $\times 8$

试集包含 40670 张图像,其中,有超过一半的图像中包含分辨率小于 $32 \text{ pixel} \times 32 \text{ pixel}$ 的小目标。

2) 评价指标

使用平均精确率(mAP)作为模型检测精度的评价指标,frame per second(FPS)作为检测速度的评价指标,模型计算量(GFLOPs)和模型大小分别作为时间复杂度和空间复杂度的评价指标。此外,在 COCO 2017 数据集中, AP^{50} 、 AP^{75} 分别表示交并比(IoU)设为 0.5、0.75 时的 mAP, AP^S 、 AP^M 、 AP^L 分别表示小分辨率目标、中分辨率目标和大分辨率目标的 mAP。

3.2.2.2 实验分析

为了验证所提目标检测模型的有效性,将所提模型与较先进的小目标检测方法在 PASCAL VOC 和 COCO 2017 数据集上进行对比,实验结果如表 2、表 3 所示,其中,加粗数据为每列最优值,加下划线数据为每列次优值。

由表 2 可知,在 PASCAL VOC 数据集上,所提模型的 mAP 取得了最优值,比次优值高 2.3 个百分点,但 FPS 明显低于骨干网络 SSD 网络,且 GFLOPs 和模型大小表现稍逊于 SSD 网络,表明所提模型能有效提高精确率,但边缘检测锐化模型以及图像分块方法增加了计算量、降低了检测速度。所提模型大小为 7.8 MB,相对其他模型空间复杂度较小,能够满足安装在终端设备的要求。由表 3 可知,在 COCO 2017 数据集上,所提模型的 AP^{50} 和 AP^L 略低于 TSD 模型,其余指标均为最优值。其中, AP^S 高于次优值 9.7 个百分点,表明所提模型对小目标检测效果较好,而对大目标检测精确率较低。其主要原因是图像分块将大目标分割成多个子目标,导致大目标不具备原来的特征,网络模型难以识别目标,从而降低了检测精确率。相对于骨干模型 SSD300,所提模型在 2 个数据集上的 mAP 分别提升 8.1 个百分点和 28.9 个百分点。其中,在

表 2 所提方法与其他方法在 PASCAL VOC 数据集上的结果比较

Table 2 Comparison results among proposed method and other methods on PASCAL VOC dataset

Method	Backbone	mAP / %	FPS	GFLOPs	Model size / MB
YOLOv3_Tiny	Darknet-Tiny	58.2	25.0	0.48	2.3
FCOS	ResNet	76.4	14	<u>3.9</u>	30
SSD300	VGG16	77.2	46	31	<u>4.8</u>
FSSD	VGG16	80.9	<u>35.7</u>	40	6.5
DSSD	ResNet101	81.5	5.5	79	122
TSD	SENet154+DCN	<u>83.0</u>	2.7	7.3	58.9
Proposed method	SSD300	85.3	28	35	7.8

表 3 所提方法与其他方法在 COCO 2017 数据集上的结果比较

Table 3 Comparison results among our method and other methods on COCO 2017 dataset

unit: %

Method	Backbone	mAP	AP ⁵⁰	AP ⁷⁵	AP ^S	AP ^M	AP ^L
YOLOv3_Tiny	Darknet-Tiny	33.0	57.9	34.4	18.3	35.4	41.9
FCOS	ResNet	44.7	64.1	48.4	27.6	47.5	55.6
SSD300	VGG16	25.1	43.1	25.8	6.6	25.9	41.4
FSSD	VGG16	31.8	52.8	33.5	14.2	35.1	45.0
DSSD	ResNet101	33.2	53.3	35.2	13.0	35.4	51.1
TSD	SENet154+DCN	<u>51.2</u>	74.9	<u>56.0</u>	<u>33.8</u>	<u>54.8</u>	64.2
Proposed method	SSD300	54.0	74.2	58.7	43.5	55.8	<u>60.7</u>

COCO 2017 数据集中,所提模型的 AP^S 较 SSD300 模型提高了 32.9 个百分点,表明所提模型针对小目标检测

测精确率有一定的提高。

图 8 为所提模型与 SSD 算法在包含小目标图像上

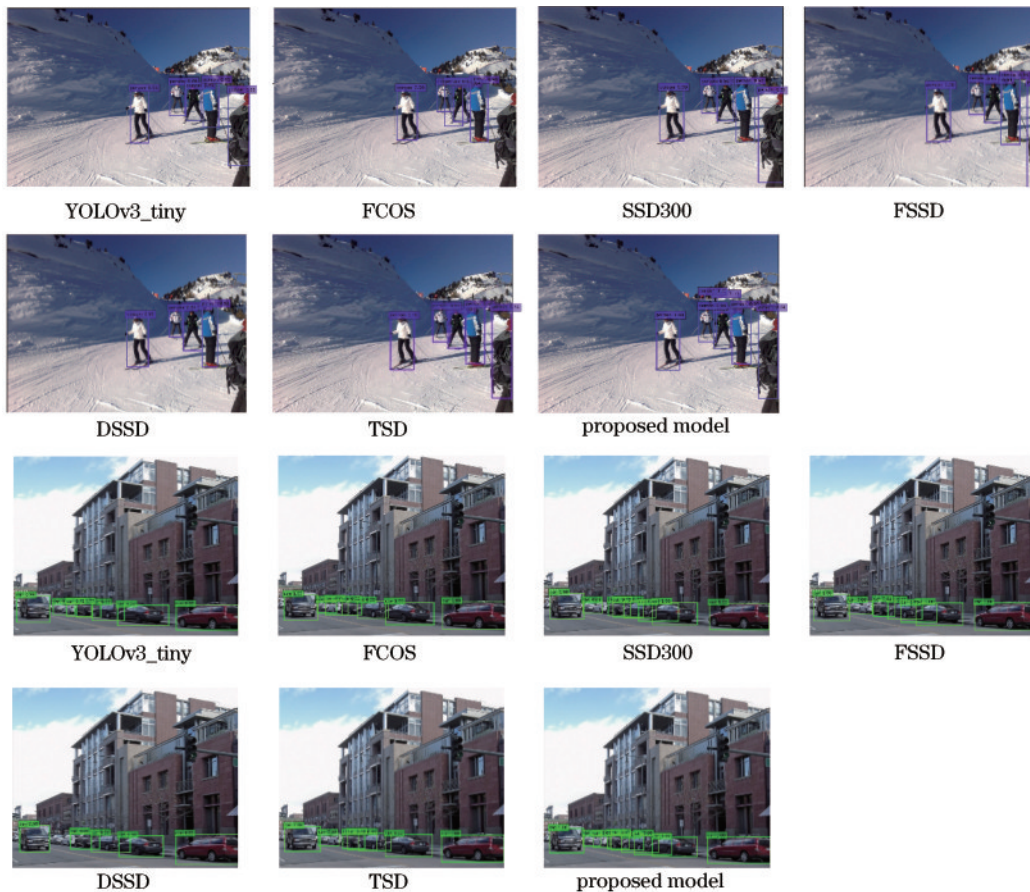


图 8 目标检测结果可视化对比

Fig. 8 Visual comparison of target detection effect

的检测结果对比。从结果来看,所提模型可以有效减少小目标漏检的情况,取得了更好的小目标检测效果。

4 结 论

针对目标检测中小目标检测精确率低的问题,提出一种 SR 锐化增强的小目标检测算法。该算法通过图像预处理,将小目标检测问题转化为非小目标检测问题,并通过 SR 锐化增强模块扩大子图像的分辨率,使小目标覆盖较多像素;然后通过多尺度锐化目标检测模型进行子图像的目标检测;最后将检测结果回归原始图像,完成小目标检测。所提算法通过两次图像锐化操作,增强图像边缘清晰度,丰富小目标特征信息。经过实验对比,SR 锐化增强模块在增大图像分辨率的同时,有效增强了图像边缘清晰度,此外,目标检测模型中的边缘锐化模型能较好地保留图像细节,在一定程度上抵消深层卷积对细节的损失。实验结果表明,所提算法相对于其他先进算法具有较高的精确率。

但是,所提算法也存在不足,主要表现在:1)对大目标检测精确率略低,主要是图像分块弱化了大目标的图像特征;2)由于添加了图像分块和边缘锐化模型,所提算法的检测速度较低。因此,下一步的工作重点是优化图像分块方法和目标检测模型,以提高大目标检测精确率和检测速度。

参 考 文 献

- [1] 张达峰,刘宇红,张荣芬. 基于深度学习的智能辅助驾驶系统[J]. 电子科技, 2018, 31(10): 60-63.
Zhang D F, Liu Y H, Zhang R F. Intelligent assistant driving system based on deep leaning[J]. Electronic Science and Technology, 2018, 31(10): 60-63.
- [2] 段仲静,李少波,胡建军,等. 深度学习目标检测方法及其主流框架综述[J]. 激光与光电子学进展, 2020, 57(12): 120005.
Duan Z J, Li S B, Hu J J, et al. Review of deep learning based object detection methods and their mainstream frameworks[J]. Laser & Optoelectronics Progress, 2020, 57(12): 120005.
- [3] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [4] Dai J F, Li Y, He K M, et al. R-FCN: object detection via region-based fully convolutional networks[C]// NIPS'16: Proceedings of the 30th International Conference on Neural Information Processing Systems, December 5-10, 2016, Barcelona, Spain. New York: ACM Press, 2016: 379-387.
- [5] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 21-37.
- [6] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.
- [7] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6517-6525.
- [8] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2020-07-04]. <https://arxiv.org/abs/1804.02767v1>.
- [9] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2020-07-02]. <https://arxiv.org/abs/2004.10934>.
- [10] 刘洋,战荫伟. 基于深度学习的小目标检测算法综述[J]. 计算机工程与应用, 2021, 57(2): 37-48.
Liu Y, Zhan Y W. Survey of small object detection algorithms based on deep learning[J]. Computer Engineering and Applications, 2021, 57(2): 37-48.
- [11] Li Z X, Zhou F Q. FSSD: feature fusion single shot multibox detector[EB/OL]. (2017-12-04) [2020-07-02]. <https://arxiv.org/abs/1712.00960>.
- [12] Kisantal M, Wojna Z, Murawski J, et al. Augmentation for small object detection[EB/OL]. (2019-02-19)[2020-05-04]. <https://arxiv.org/abs/1902.07296>.
- [13] Tian Z, Shen C H, Chen H, et al. FCOS: fully convolutional one-stage object detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 20-26, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 9626-9635.
- [14] Deng C F, Wang M M, Liu L, et al. Extended feature pyramid network for small object detection[J]. IEEE Transactions on Multimedia, 2022, 24: 1968-1979.
- [15] 杨其利,周炳红,郑伟,等. 基于全卷积递归网络的弱小目标检测方法[J]. 光学学报, 2020, 40(13): 1310002.
Yang Q L, Zhou B H, Zheng W, et al. Dim and small target detection based on fully convolutional recursive network[J]. Acta Optica Sinica, 2020, 40(13): 1310002.
- [16] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [17] Liu Y, Li H Y, Wang X G. Rethinking feature discrimination and polymerization for large-scale recognition[EB/OL]. (2017-10-02)[2020-02-05]. <https://arxiv.org/abs/1710.00870>.
- [18] Ledig C, Theis L, Huszar F, et al. Photo-realistic single image super-resolution using a generative adversarial network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 105-114.