吉林大学学报(工学版)
Journal of Jilin University(Engineering and Technology Edition)
ISSN 1671-5497,CN 22-1341/T

《吉林大学学报(工学版)》网络首发论文

题目：机载广域遥感图像的尺度归一化目标检测方法
作者：朱圣杰，王宣，徐芳，彭佳琦，王远超
DOI：10.13229/j.cnki.jdxbgxb.20230034
收稿日期：2023-01-10
网络首发日期：2023-05-23
引用格式：朱圣杰，王宣，徐芳，彭佳琦，王远超．机载广域遥感图像的尺度归一化目标检测方法[J/OL]．吉林大学学报(工学版).
https://doi.org/10.13229/j.cnki.jdxbgxb.20230034

机载广域遥感图像的尺度归一化目标检测方法

朱圣杰[1,2]，王宣[1]，徐芳[1]，彭佳琦[3]，王远超[4]

(1. 中国科学院长春光学精密机械与物理研究所，吉林 长春 130033；2. 中国科学院大学，北京 100049；
3. 驻长春地区第一军事代表室，吉林 长春 130033；4. 上海机电工程研究所，上海 201109)

摘要： 针对机载广域遥感图像的目标尺寸变化大、背景噪声复杂以及局部目标密集给目标检测任务带来的困难，本文通过优化分割方法统一输入图像的目标像素尺寸，并以此简化模型结构提出了一种尺度归一化卷积神经网络模型 MNNet。为增强局部之间的特征关联，本文设计了全局连接块(SGC)，有效提高了检测的精度。针对现有非极大值抑制算法的超参数依赖经验设置的问题，本文提出了一种自适应非极大值抑制方法(DNMS)，降低了模型的部署难度。在 RSF 数据集上的测试结果表明：本文模型的检测平均精度(AP)高于其他模型 5.0%以上，在检测速度上达到了 57.7fps，可以满足遥感图像的检测任务。

关键词： 模式识别与智能系统；计算机视觉；目标检测；遥感图像；卷积神经网络

中图分类号： TP391　　**文献标志码：** A

DOI：10.13229/j.cnki.jdxbgxb.20230034

Multi-scale normalized detection method for airborne wide-area remote sensing images

ZHU Sheng-jie[1,2], WANG Xuan[1], XU Fang[1], PENG Jia-qi[3], WANG Yuan-chao[4]

(1. *Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, Jilin, 130033, China*; 2. *University of Chinese Academy of Sciences, Beijing 100049, China;* 3. *First Military Representative Office in Changchun, Changchun, Jilin 130033, China;* 4. *Shanghai Electro-Mechanical Engineering Institute, Shanghai, 201109, China)*

Abstract: Aiming at the difficulty of object detection caused by the large target size variation, complex background noise and dense targets in airborne wide-area remote sensing images, this paper unifies the target pixel size of the input image by optimizing the segmentation method, and proposes a multi-scale normalized convolutional neural networks model (MNNet). To enhance the feature correlation between localities, this paper designs a space global connection block (SGC), which effectively improves the detection accuracy. For the problem that the parameters of the existing NMS algorithm depend on the empirical setting, this paper proposes a self-adaption non-maxima suppression method (DNMS), which reduces the difficulty of model deployment. The test results on the RSF dataset show that the average precision (AP) of the model in this paper is higher than that of other models by more than 5.0%, and the detection speed reaches 57.7fps, which can meet the detection task of remote sensing images.

Key words: Pattern recognition and intelligent system; Computer vision; Object detection; Remote sensing image; Convolutional neural network

0 引言

随着图像传感技术和航空成像技术的飞速发展，机载遥感图像的获取更加高效。相比卫星遥感，机载航空遥感具有技术成熟、控制灵活、地面分辨率高等优点[1]，是重要的遥感手段。如何快速准确地提取有效信息并加以利用，在交通管理、城市规划、目标监控等任务中有着重要的应用价值[2]。通过人工判读的

收稿日期：2023-01-10.
基金项目：国家自然科学基金项目（NO.61905240）
作者简介：朱圣杰（1997-），男，博士研究生. 研究方向：航空航天成像，计算机视觉. E-mail: shengjie_zhu@foxmail.com
通信作者：王宣（1984-），男，副研究员，博士. 研究方向：机载光电成像测量设备. E-mail: ally637@163.com

方式进行目标识别，数据利用率低且情报时效性差，且易受主观意识的影响。因此，通过计算机视觉技术减少人力的消耗，实现高效准确的遥感图像自动检测尤为重要。

对于目标检测任务，经典算法通常是针对几何特征、空间关系、目标轮廓等特征而设计的，如 Viola-Jones 检测器[3]、梯度方向直方图(Histogram of Oriented Gradient, HOG)检测器[4]、可变形部件模型(Deformable Part Model, DPM)[5]。这些算法只能达到较低的准确率，并且这样的策略信息对于环境的多样性没有很好的鲁棒性，难以很好地适应实际场景的需求。卷积神经网络(CNN)的发展极大地提升了算法对目标的检测能力。以 RFCN[6]和 Faster RCNN[7]为代表的二阶段检测模型展现了优秀的准确率，而以 YOLO[8-11]为代表的端到端的模型具有更高的检测速度。此外，以 DETR[12]为代表的注意力模型揭开了全局感知能力的重要性，在通用数据集中展现了优异的性能。

与一般影像相比，机载广域遥感影像由高空成像设备拍摄，得到的图像一般具有视场大、像素多的特点 (ITCVD[13]: 5616×3744 像素，DOTA[14]: 约 2000×1000 像素)，因此简单的下采样到通用模型要求的输入大小（Faster RCNN: 600×600 像素, YOLO: 608×608 像素）是不合适的，现有算法通常会进行分割的预处理。具体的，YOLT 模型[15]采用滑窗法裁剪图像，并设计了 15%的重叠以确保目标不被截断。然而，使用固定尺寸裁剪后目标像素仍存在较大动态范围，模型仍需要设计较多的锚框，且需要大量的数据支持。本文改进了裁剪策略，使得输入目标像素大小统一，剔除了模型冗余结构，提升了检测效率。

此外，航空遥感图像的感兴趣目标还具有背景复杂、局部密集、角度任意的特点，增加了模型的检测难度。对此，现有算法通常采用特征融合的方式提升模型的检测能力。SSD 模型[16]采用了多尺度特征融合结构（Pyramidal feature hierarchy），Mask R-CNN 模型[17]采用了特征金字塔结构（Feature Pyramid Network，FPN）。这样的设计虽然提升了模型对不同尺度目标的检测能力，但在广域航空遥感图像中，尺度的变换更为复杂，简单的多层特征仍然难以对目标进行有效提取。为此，本文提出了空间全局连接模块以增强局部特征之间的联系，提升模型的检测精度。

最后，本文还关注到模型筛选冗余检测框的非极大值抑制算法(Non Maximum Suppression, NMS)采用的是固定阈值参数。经典 Greedy-NMS[18] 算法通过计算置信度最高的检测框与其他检测框的交并比(Intersection Over Union, IoU)，简单地通过阈值剔除冗余框。Soft-NMS 算法[19]通过 IoU 值设计惩罚函数对置信度进行修正，但仍利用阈值划分筛选。此外，CUDA-NMS 算法[20]优化了算法的运行效率，Softer-NMS 算法[21]引入方差的计算提升了定位效果，但均需预设相关参数。固定的阈值并不适合存在复杂场景的包围框筛选任务，对此，本文设计了基于密度分析的自适应策略以适应场景变换。

1 算法设计

本节将详细介绍所提模型的架构，并讨论如何提升对航空遥感图像的目标检测能力。

本文提出的航空遥感图像目标检测流程如图 1 所示。对于高像素的遥感图像，我们设计了自适应像素尺度的滑窗法对图像进行裁剪。该方法结合成像参数计算合适的分割像素大小，使得目标尺寸更为统一的同时降低了模型的计算复杂度。在提出的尺度归一化卷积神经模型(Multi-scale Normalized Detection Network, MNNet)，首先利用已标注的数据集训练并验证评估。测试时，将检测结果重新拼接为原始图像大小，得到最终输出结果。

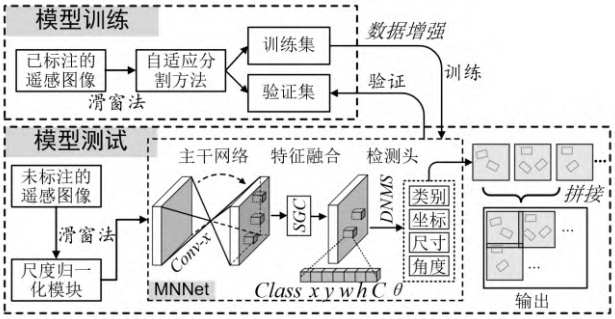

图 1 MNNet 框架的整体架构

Fig. 1 Overall architecture of the MNNet framework

提出的 MNNet 模型由 3 部分组成，分别是主干网络、特征融合模块和检测头。对于已通过尺度归一化模块预处理的高像素广域遥感图像，将分割图像输入到主干网络之中进行特征提取，进而利用特征融合模块对不同网络深度和空间位置的特征进行信息融合，最后通过检测头输出得到目标的位置及大小信息。

1.1 尺度归一化模块

首先，我们对现有航空遥感数据集进行了统计，

如表 1 所示。对比可以发现，大量数据集采用截取谷歌地球影像的方式获取图像，大部分都属于卫星遥感图像，目标的像素尺寸取决于图像的缩放程度。以 ITCVD 和 DLR 3K 为代表的航空遥感影数据集采用飞行器直接拍摄，真实反映了航空遥感时的图像特点，更具有实际应用价值，但由于目标的像素尺寸取决于拍摄时飞行器的飞行高度、成像设备等的相关参数，目标的像素尺寸存在较大范围波动的情况无法避免。

如前所述，广域航空遥感图像高像素的特点使得图像无法直接输入到检测模型中。若将原始图像以固定像素大小进行裁剪，则分割后图像中目标的像素尺度将保持不变。由于航空遥感图像的成像特点，焦距等参数变化会使得相同目标的像素大小也具有较大的连续变化区间，这要求检测模型应具有复杂的多尺度目标检测能力，也需要大量的多尺度训练集的支持，极大的降低了模型的训练与检测效率。统一同类目标像素尺度对简化模型的网络结构，提升检测效率具有重要意义。为解决这一问题，我们提出了一种尺度归一化分割方法。

表 1 光学遥感数据集对比表

Table 1 Optical remote sensing dataset comparison

数据集	图像来源	图像数量	目标数量	图像尺寸	类别数	单张图像类别数	成像参数
ITCVD[13]	机载成像	135	23543	5616×3744	1	1	×
DLR 3K[22]	机载成像	20	14235	5616×3744	7	2~5	×
DIOR[23]	谷歌地球	23463	192472	800×800	20	1~3	×
UCAS-AOD[24]	谷歌地球	910	6029	~1280×680	2	1	×
HRRSD[25]	谷歌地球	21761	55740	~1000×1000	3	1	×
DOTA[14]	谷歌地球	2806	188282	~2000×1000	15	1~3	×
LEVIR[26]	谷歌地球	22000+	10000+	800×600	3	1	×

航空遥感图像通常采用垂直于地面的方式进行拍摄，通过光学成像分析可以得到：

$$\frac{w}{v}=\frac{w_t}{h},(h>>v) \tag{1}$$

其中，w 为在 CMOS/CCD 上成像的物理尺寸，w_t 为目标物体的实际尺寸，v 为像距，h 为物距，即为飞行器的飞行高度。其中焦距 f、像距 v、物距 h 的基本关系可以用高斯成像公式表示：

$$\frac{1}{v}+\frac{1}{h}=\frac{1}{f} \tag{2}$$

由于在航空遥感图像中，物距(h)远大于像距(v)，则可认为存在：

$$v=f \tag{3}$$

根据公式(1)和(3)，若像元尺寸为 p，则目标实际占像素数量 k 可表示为：

$$k=\frac{w}{p}=\frac{w_t\cdot f}{h\cdot p},(h>>v) \tag{4}$$

结果表明，目标像素数与飞行高度和相机焦距有关。据此，我们计算目标像素的数量，并将任意大小的图像划分为可管理的切面。分区是通过重叠的滑动窗口进行的，如图 2 所示。滑窗裁剪后，对于较小的图像（R1<R），我们采用双线性插值放大的方式使其像素尺寸变大；对于较大的图像（R3>R），我们采用等间距采样的方式进行缩小；相等（R2=R）则保持不变。此时，缩放后图像中的同类目标的像素尺寸相对一致，便于检测网络的计算，也为后续模型剪枝提供了依据。

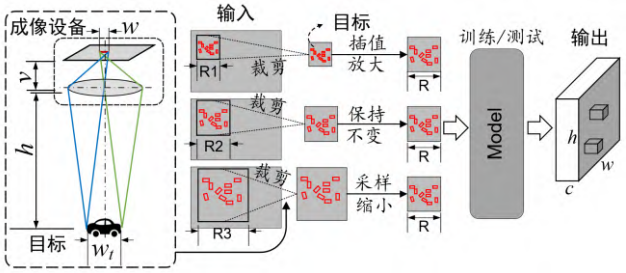

图 2 尺度目标归一化流程示意图

Fig. 2 Schematic diagram of multi-scale normalized process

其中，分区像素尺寸L_P和重叠像素尺寸L_o的大小定义为：

$$\begin{cases}L_p=a\cdot k & (a=N_{gird})\\ L_o=b\cdot k & (1<b<a)\end{cases} \tag{5}$$

其中 a，b 是超参数。a 对应检测头中输出网格(N_{gird}=72)的大小，以便每个网格只对应一个目标。为了避免目标被分割而导致的遗漏，b 被设置为大于

1 的数字(默认为 1.5)。在拼接过程中，重叠部分需要通过 NMS 方法对切割边缘的检测，具体计算过程将在 1.3 节中详细阐述。

1.2 空间全局连接模块

卷积算子的简单连接将使网络只关注局部邻域，不能敏感地捕捉整个空间之间的全局关系。然而，遥感背景中存在着大量的相似目标，缺乏目标背景信息的融入易导致出现错误识别的情况。基于这一观察，我们设计了空间全局连接模块(Space Global Connection, SGC)。通过同层卷积算子的聚合，SGC 感知块将不同坐标位置的特征相互关联，增强了局部特征之间的关联性。

空间全局连接模块的关键是将某一位置的响应作为所有位置特征的加权和。我们将卷积神经网络中的非局部操作定义为：

$$y_i = \sum_{\forall j} \frac{f(x_i, x_j)}{C(x)} g(x_j) \tag{6}$$

其中，i 是输出位置的编码，j 是所有可能位置的编码，x 是输入特征图，y 是 SGC 模块的输出特征图。f表示特征图中两点之间的相关性函数，$C(x)$为归一化函数，$\mathrm{g}(x)$表示 x 位置的卷积特征图。其中，$f(x_i, x_j)$可以被定义为：

$$f(x_i, x_j) = \theta(x_i)^T \phi(x_j) \tag{7}$$

这里的 θ 和 ϕ 表示要训练的卷积结构，因此，这将x_i和x_j两两联系。为了简化了梯度运算，我们将标准化系数设为$C(x) = N$，其中 N 为输入特征图中的位置数。

具体的，SGC 模块结构如图 3 所示，输入特征图（$M^{C\times H\times W}$）首先采用两个 1×1 的卷积核处理，变维后得到$Q^{HW\times C}$和$K^{C\times HW}$。根据公式(7)，我们可以得到具备特征关联能力的矩阵$T^{HW\times HW}$。另一方面，我们分别采用最大池化和平均池化的方法提取输入特征。为了与矩阵$T^{HW\times HW}$相匹配，我们采用 1×1 的卷积缩小通道，进一步变维后得到$V^{HW\times C}$。1×1 卷积的应用可以提炼输入特征，降低该分支的计算复杂度。

最后，具有空间关联信息的特征图$R^{C/2\times H\times W}$可以表示为：

$$R^{C/2\times H\times W} = softmax[conv(T^{HW\times HW} \otimes V^{HW\times C})] \tag{8}$$

此外，为了防止网络退化，我们定义 SGC 模块的输出$O^{C\times H\times W}$为:

$$O^{C\times H\times W} = R^{C/2\times H\times W} \oplus S^{C/2\times H\times W} \tag{9}$$

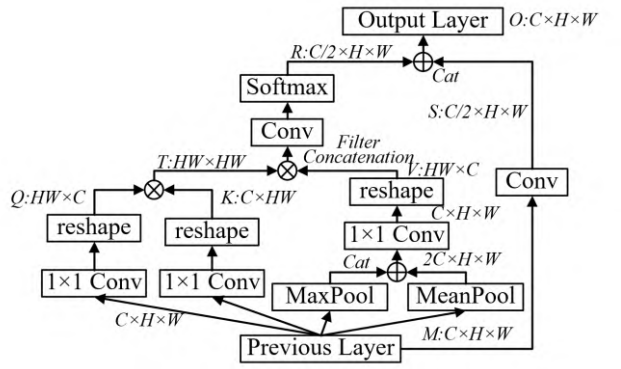

图 3 空间全局连接模块结构示意图

Fig. 3 A space global connection block (SGC Block)

SGC 模块通过设计关系矩阵来增强像素特征的关联，具有完全可微的性质，因此能够通过反向传播进行优化。SGC 模块利用自注意力机制结构，在 V 分支增加了最大池化与平均池化的计算。最大池化有利于突出目标纹理的特征，平均池化有利于背景特征的提取，前景与背景的分离能够更好的使模型学习跨区域的特征关联，最终提升模型的检测能力。此外，我们设计的 SGC 模块具有相同的输入和输出维度，所以它可以很容易地集成到任意检测模型中。

1.3 DNMS 算法

在预测任务中，会出现较多冗余预测框，NMS 算法具有剔除冗余框，提取最优包围框的作用。我们期望当物体分布稀疏时，NMS 可选用小阈值以剔除更多冗余框；而在物体分布密集时，选用大阈值，以获得更高的召回。针对现有 NMS 算法预先设定固定阈值的问题，本文提出了一种基于密度分析的自适应阈值非极大值抑制方法(Density Non Maximum Suppression, DNMS)。

对类别 k 的检测框，选取置信度最高的识别框计算 IoU 值，得到 n 个离散数值：$x_1^k, x_2^k, \ldots, x_n^k$，所处的区间为$[0,1)$，分布函数可表示为：

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} P\{x_i \le x\} \tag{10}$$

其中，P 为对应区间内x_n的数量，以 h 为带宽，通过求差可计算概率密度函数为：

$$\hat{f}_h(x) = \frac{1}{2nh}\sum_{i=1}^{n} P\{x_i - h < x_i \le x_i + h\} \tag{11}$$

由于 P 为离散型函数，因此概率密度函数$\hat{f}_h(x)$呈现冲激函数形式，为了使其具有连续性和平滑性，我们设计核函数$K(u)$：

$$K(u)=\frac{1}{\sqrt{2\pi}}\cdot\exp(-\frac{1}{2}u^2) \tag{12}$$

代入式(11)得：

$$\hat{f}_h(x)=\frac{1}{nh}\sum_{i=1}^{n}K(\frac{|x-x_i|}{h}) \tag{13}$$

带宽 h 值的大小反映了波形的融合程度，根据核密度估计理论，可以最小化平方误差均值的积分(Mean Integrated Squared Error, MISE)来提升$\hat{f}_h(x)$对真实分布的拟合：

$$MISE(\hat{f}_h)=\int E\left[\hat{f}_h-f\right]^2dx \tag{14}$$

根据 Silverman 的最优带宽理论[27]，对式(14)二阶泰勒分解后剔除高次项，得到的 h 可以表示为：

$$h=\left(\frac{4\hat{\sigma}^5}{3n}\right)^{\frac{1}{5}}\approx1.059\hat{\sigma}n^{-\frac{1}{5}} \tag{15}$$

其中$\hat{\sigma}$为样本的标准差：

$$\hat{\sigma}=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2} \tag{16}$$

为更好的表现密度的分布情况，提升数值区分度，对分布曲线进一步采用对数变换：

$$\varphi(x)=\ln\left[\hat{f}_h(x)\right] \tag{17}$$

密度分布如图 4 所示中红色曲线，大量数据趋向于 0，经过对数变换后的蓝色曲线较大的提升了密度的关系。为了得到极小值点，我们通过将数据差分后带入阶跃函数的方法计算：

$$(m_i,n_i)=(x,y)_{\triangle\mathrm{sgn}(\varphi'(x))>0}+1 \tag{18}$$

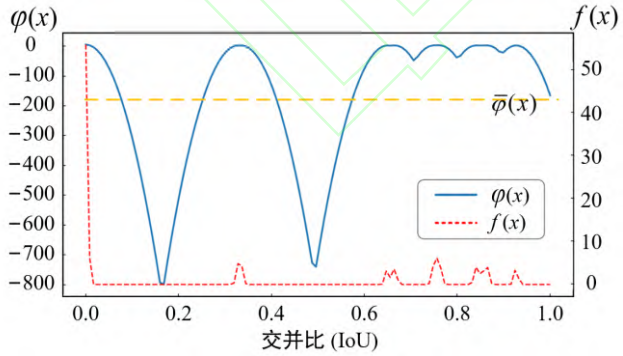

图 4 核密度估计方法处理示意图

Fig. 4 Schematic diagram of kernel density estimation method

最后，比较极小值点与密度均值的点大小筛选出最终的 IOU 阈值：

$$(m_i,n_i)=\begin{cases}(m_i,n_i) & n_i<\bar{\varphi}\\(0,0) & n_i\geq\bar{\varphi}\end{cases},i\in[1,n] \tag{19}$$

通过上述分析，本文提出的 DNMS 算法可以根据模型的检测结果动态调整筛选阈值的设置，能够更好的对包围框进行筛选。提出的 DNMS 算法模型整体程序的伪代码如下图 5 所示：

Algorithm 1 DNMS

Input: $\mathcal{B}_i=\{b_1,b_2,...,b_N\}$; $\mathcal{S}_i=\{s_1,s_2,...,s_N\}$
Output: $\mathcal{B}_o=\{d_1,d_2,...,d_M\}$; $\mathcal{S}_o=\{e_1,e_2,...,e_M\}$

initialize: Set $\mathcal{B}\leftarrow\{\}$;$\mathcal{S}\leftarrow\{\}$;$I\leftarrow\{\}$
while $\mathcal{B}_i\neq empty$ **do**
  $k\leftarrow\arg\max(\mathcal{S}_i)$
  $\mathcal{B}\leftarrow b_k$
  for b_n in $\mathcal{B}_i$ **do**
    $I\leftarrow\boldsymbol{iou}(b_k,b_n)+I$
  end for
  $N_t=\phi_\kappa(I)$
  for i_m in I **do**
    if $i_m>N_t$ **then**
      $\mathcal{B}\leftarrow\mathcal{B}+b_m,\mathcal{S}\leftarrow\mathcal{S}+s_m$
      $\mathcal{B}_i\leftarrow\mathcal{B}_i-b_m,\mathcal{S}_i\leftarrow\mathcal{S}_i-s_m$
    end if
  end for
  $\mathcal{B}_o\leftarrow\psi_\kappa(\mathcal{B})$;$\mathcal{S}_o\leftarrow\psi_\kappa(\mathcal{S})$
end while
return $\mathcal{B}_o,\mathcal{S}_o$

图 5 DNMS 算法伪代码

Fig. 5 Pseudocode of DNMS algorithm

1.4 损失函数

检测头的输出包围框由中心坐标(x,y)、长边(l)、短边(s)、类别置信度(c)以及角度(θ)组成。其中，角度采用分类的方法进行回归。对于输出的各类数据，我们设计了相对应的损失函数，以便对模型进行训练。我们将模型总损失函数分为置信度损失(L_{conf})、角度损失(L_{angle})和回归框损失(L_{box})三个部分。

对于模型$S\times S$的输出网络，每个网格包含 B 个旋转包围框，对于标准包围框$(\hat{x},\hat{y},\hat{l},\hat{s},\hat{c},\hat{\theta})$，我们采用交叉熵的方式描述置信度损失$(L_{conf})$以及角度损失$(L_{angle})$：

$$L_{conf}=\sum_{i=0}^{S^2}\sum_{j=0}^{B}\mathbb{I}_{ij}^{obj}\left[\hat{c}_i\log(c_i)+(1-\hat{c}_i)\log(1-\hat{c}_i)\right]+\lambda_{noobj}\sum_{i=0}^{S^2}\sum_{j=0}^{B}\mathbb{I}_{ij}^{noobj}\left[\hat{c}_i\log(c_i)+(1-\hat{c}_i)\log(1-\hat{c}_i)\right] \tag{20}$$

$$L_{angle}=\sum_{i=0}^{S^2}\sum_{j=0}^{180*B}\left[\hat{\theta}_i\log(\theta_i)+(1-\hat{\theta}_i)\log(1-\theta_i)\right] \tag{21}$$

其中在当前包围框实际存在目标时，$\mathbb{I}^{obj}$=1，$\mathbb{I}^{noobj}$=0；当前包围框实际不存在目标时，$\mathbb{I}^{obj}$=0，$\mathbb{I}^{noobj}$=1。λ_{noobj}

为正负样本损失平衡系数，由于负样本所占比例相对较多，因此λ_{noobj}数值上较小，默认取值为 0.4。

回归框损失需要综合考虑重叠面积、中心点距离以及长宽比，因此我们采用 CIOU[28]的方式进行计算：

$$L_{box}=1-IOU+\frac{(x-\hat{x})^2+(y-\hat{y})^2}{d^2}+\alpha v \quad (22)$$

其中，IOU 衡量了输出框与标准框的重叠度，是重叠面积与联合覆盖面积的比值，d 是覆盖两个盒子的最小封闭盒子的对角线长度。此外，v 代表了横纵比一致性的信息，提高回归精度，α 是权重参数，具体为：

$$v=\frac{4}{\pi^2}\left(\arctan\frac{\hat{s}}{\hat{l}}-\arctan\frac{s}{l}\right)^2 \quad (23)$$

$$\alpha=\frac{v}{(1-IOU)+v} \quad (24)$$

CIOU 考虑到两个框之间的位置信息，提升了对尺度的敏感性，能更好的衡量检测框的坐标定位能力。最终，总损失可表示为：

$$L_{total}=\lambda_{conf}L_{\text{conf}}+\lambda_{angle}L_{angle}+\lambda_{box}L_{box} \quad (25)$$

其中，λ 为对应权重参数，用于平衡损失之间的数值差距。

2 实验与结果

为了更好地验证本文所述方法的有效性，在本节中，我们首先对多尺度归一化模块进行验证，然后对 MNNet 模型做出细致的评估，并与当前最先进的目标检测模型进行比较。

2.1 数据集与预处理

由于现有数据集中缺乏遥感成像时飞行高度、相机焦距等相关数据，导致难以通过尺度归一化的方法计算分割像素尺寸。为了能够完整验证算法的逻辑，我们制作了具有相关飞行参数的航空遥感图像 RSF 数据集对其进行验证，数据集利用无人机拍摄完成，包含了 185 张分辨率为 3840×2160 的可见光航空遥感图像，具备雨雪等不同的气象环境。由于实验条件限制，为了避免出现正负样本不平衡问题，我们设计的 RSF 数据集仅标注了小型轿车，拥有 6445 个车辆目标。数据集虽然检测目标种类单一，但由表 1 可以发现，现有部分遥感数据集中也仅包含单类目标，其他数据集虽具有多类目标，但不同类别的目标所处的环境具有较大不同，单张图像往往不具备检测尺度差异过大的目标。例如，标注有普通客机的图像中，车辆目标的图像信息模糊，肉眼难以辨认，无法有效标注。因此，我们设计的 RSF 数据集在实际机载广域遥感图像的检测任务中仍具有重要使用价值。最终，不同分割方式得到的目标尺寸分布结果如图 6 所示。

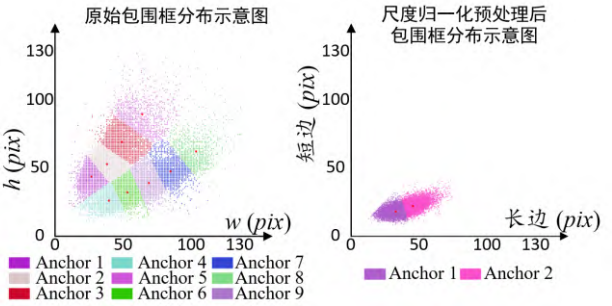

图 6 目标尺寸分布示意图(RSF 数据集)

Fig. 6 Schematic diagram of target size distribution (RSF Dataset)

基于卷积神经网络的目标检测模型中，通常采用设置锚框 Anchors 的方式为检测模型加入先验经验，进行尺寸约束。Anchors 的大小和数量信息取决于目标尺寸聚类后的区域大小。在图 6 中，我们通过 Kmeans 方法进行聚类，图中的每一种颜色代表一个聚类。可以直观的发现，对比 YOLT 模型的固定像素大小的裁剪方法，所提出的自适应像素尺度的分割方法能使切片中目标像素尺寸更为集中，只需要更少的预设锚框 Anchors 即可满足检测的需要，对网络剪枝提供了依据。

为了提升所述尺度归一化方法的说服力，增强数据集的复杂度，本文进一步对具有多类目标的 DOTA 公开数据集进行测试。由于数据集缺乏计算所需的相关飞行参数，我们通过统计数据集中特定类别目标像素尺寸的方法进行仿真。我们首先利用具有固定物理尺寸的检测目标（车辆）作为基准成像尺寸，通过逐张统计图像中标准目标的平均像素大小以确定需要的缩放和裁剪尺寸。经过尺度归一化的预处理方法计算后，目标像素尺寸的分布如图 7 所示。

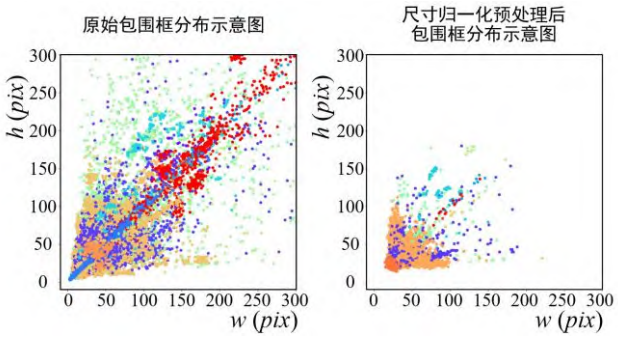

图 7 DOTA 数据集分割后目标的像素尺寸分布图

Fig. 7 Pixel size distribution of objects after segmentation in DOTA dataset

在图 7 中，我们采用不同的颜色标注出了不同类别目标的像素分布范围。通过对比可以发现，预处理前的目标像素尺寸分布呈现整体分散，局部聚集的情况，经过尺度归一化后，各类目标的尺寸均更为集中，证明所提出的方法能够有效的解决目标尺度分散的问题。

得益于目标尺寸的统一，本文提出的 MNNet 模型可以剔除多余的检测结构，使得网络具有更小的模型体积，我们对当前流行的目标检测网络进行了参数统计，结果如表 2 所示。通过对比可以看出，我们提出的 MNNet 网络参数量更少，得到的模型体积也更小。

表 2 检测模型的参数量对比

Table 2 Parameters of neural networks for target detection

模型	锚框数	检测头数	参数量	模型大小(MB)
YOLOv3	9	3	61,949,149	236.32
YOLOv4	9	3	63,943,071	245.53
YOLOv5m	9	3	22,229,358	84.80
YOLOv5l	9	3	48,384,174	184.57
YOLOv5x	9	3	89,671,790	342.07
SSD300	9	3	23,745,908	90.58
Faster-RCNN	9	3	137,078,239	522.91
MNNet	3	1	31,443,246	119.95

2.2 训练环境及参数配置

为了定量地对模型性能进行评估，我们在 Intel Core i7-10700F 2.9GHz CPU、16GB 内存和 NVIDIA GeForce GTX 2060Ti GPU (6GB 内存)的计算机上进行了测试，使用开源 Pytorch 框架实现。在训练期间采用随机弹性梯度下降(SGD)的方法优化参数，基础学习率设定为 0.0001，网络采用 kaiming 方法初始化。

为了更好地验证该方法的有效性，我们对当前流行的目标检测器进行了定量的比较。使用精度（P）、召回率（R）、F1 值和平均准确率（AP）作为度量标准，定义为：

$$Recall = \frac{\text{number of true detections}}{\text{number of existing objects}} \tag{26}$$

$$Precision = \frac{\text{number of true detections}}{\text{number of detected objects}} \tag{27}$$

$$\text{F1} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{28}$$

P-R 曲线下的面积定义为 AP，由于积分计算比较困难，引入插值方法，AP 的计算公式如下：

$$AP = \sum_{k=1}^{N} \max_{\tilde{k} \geq k} P(\tilde{k}) \Delta R(k) \tag{29}$$

2.3 对比实验

首先，为了验证本文提出的 DNMS 算法的有效性，我们在 MNNet 模型上利用不同的非极大值抑制方法进行评估，检测效果如表 3 所示，表中已加粗标注出最优的效果。

对比表格中不同 NMS 算法的检测结果可以发现，本文提出的 DNMS 算法具有较好的冗余框筛选能力。对比传统的 Greedy NMS 方法 AP@0.75 提升了 2.07%，也优于 Softer NMS 算法。相较于其他算法仍依赖于经验的设置方法，DNMS 由于能够通过鲁棒性计算得到最优的分割阈值，具有更强的适应性，从而得到最优的效果。

表 3 NMS 算法 AP@0.75 性能对比图(RSF 数据集)

Table 3 Comparison of different NMS methods on RSF dataset

阈值	AP@0.75 of NMSs			
	Greedy NMS	Soft NMS	Softer NMS	DNMS
0.25	60.65%	61.90%	61.30%	
0.35	62.70%	63.50%	63.60%	
0.45	64.10%	65.50%	65.60%	
0.55	**64.30%**	65.40%	66.00%	**66.37%**
0.65	64.11%	**65.55%**	**66.25%**	
0.75	62.95%	64.70%	64.00%	
0.85	61.39%	62.00%	62.70%	

进一步，为了评估模型的整体性能，我们分别在 RSF 数据集和 DOTA 数据集上对 MNNet 模型做了评估和对比，最终的检测结果分别如表 4、表 5 所示，其中最优指标已经加粗表示。

通过将本文提出的 MNNet 模型与现有常见模型对比，MNNet 展现了较高检测速度，并具有最优的检测能力。其中，以 HOG+SVM 为代表的传统目标检测方法在具有复杂背景的遥感目标检测任务中精度较差，速度较慢，无法达到应用要求。通过对比具有不同网络深度的 YOLOv5 系列模型可以发现，仅提升网络深度无法有效地解决机载广域遥感图像的目标检测问题，过深的网络结构反而会削弱模型对小目标的检测能力。本文提出的 MNNet 模型具有较高的检测效率，达到了 60 帧左右。与 YOLOv5s 等轻量化模型对比，MNNet 在较小检测速度的损失代价

下检测性能有了较大提升；对比其他模型，MNNet 在检测精度与 AP 值上均展现了较大的优势，不同数据集上测试效果均显示相较于次优模型 AP@0.5 提升了约 5.0%。

表 4 多种模型检测效果对比表(RSF 数据集)

Table 4 Comparison of different network methods on RSF dataset

模型	主干网络	精度	召回率	F1 值	AP@0.50	AP@0.50:0.95	帧率
HOG+SVM	/	6.52%	21.19%	0.0997	/	/	1.3 fps
SSD300	VGG-16	25.55%	47.34%	0.3318	0.2946	0.1245	45.5 fps
Faster-RCNN	ResNet50	39.18%	57.36%	0.4656	0.3921	0.1634	7.2 fps
R-CenterNet	Hourglass	/	/	/	0.4640	0.2021	50.2 fps
RRPN[29]	VGG-16	48.93%	54.72%	0.5166	0.5813	0.2472	39.4 fps
SCRDet[30]	VGG-16	50.37%	58.42%	0.5410	0.6527	0.2537	27.6 fps
DODet[31]	ResNet50 + FPN	54.32%	66.73%	0.5988	0.6677	0.3223	25.1 fps
YOLT	Darknet19	22.94%	61.23%	0.3338	0.5022	0.1681	52.3 fps
R-YOLOv3	Darknet53	21.18%	68.65%	0.3237	0.5334	0.2034	51.3 fps
R-YOLOv4	CSPDarknet53	39.72%	79.25%	0.5291	0.6531	0.2538	56.4 fps
R-YOLOv5s	CSPDarknet53	34.28%	**80.96%**	0.4817	0.6599	0.2720	**71.4 fps**
R-YOLOv5m	CSPDarknet53	36.82%	79.68%	0.5036	0.6356	0.2836	62.1 fps
R-YOLOv5l	CSPDarknet53	40.11%	73.21%	0.5182	0.6033	0.2354	48.5 fps
R-YOLOv5x	CSPDarknet53	33.29%	77.59%	0.4659	0.4825	0.2009	30.7 fps
MNNet (without SGC)	CSPDarknet53	51.34%	74.12%	0.6066	0.6536	0.2934	68.9 fps
MNNet	CSPDarknet53	**59.79%**	71.17%	**0.6498**	**0.7179**	**0.3412**	57.7 fps

表 5 模型检测效果对比表(DOTA 数据集)

Table 5 Comparison of different network methods on DOTA dataset

模型	主干网络	精度	召回率	F1 值	AP@0.50	AP@0.50:0.95	帧率
HOG+SVM	/	9.12%	25.34%	0.1341	/	/	1.6 fps
SSD300	VGG-16	28.46%	50.23%	0.3633	0.3282	0.1531	42.5 fps
Faster-RCNN	ResNet	40.64%	61.20%	0.4884	0.4561	0.1983	9.7 fps
R-CenterNet	Hourglass	/	/	/	0.5024	0.2042	52.2 fps
RRPN	VGG-16	56.93%	63.32%	0.5996	0.6302	0.2624	34.4 fps
SCRDet	VGG-16	57.29%	65.82%	0.6126	0.7221	0.2882	25.9 fps
DODet	ResNet50 + FPN	63.32%	73.73%	0.6812	0.7489	0.3564	27.4 fps
YOLT	Darknet19	22.23%	64.21%	0.3302	0.5542	0.1902	52.3 fps
R-YOLOv5s	CSPDarknet53	33.35%	**83.24%**	0.4962	0.7329	0.3020	**69.2 fps**
MNNet (without SGC)	CSPDarknet53	57.51%	81.45%	0.6741	0.7025	0.3234	67.9 fps
MNNet	CSPDarknet53	**65.29%**	79.78%	**0.7181**	**0.7875**	**0.3912**	60.7 fps

为了验证本文提出 SGC 模块的有效性，我们在上表中补充了不包含该模块的对照组进行消融实验。可以发现，SGC 模块虽使模型检测速度降低 7~10 帧，但检测性能有了较大的提升。在 RSF 数据集中检测精度提升了 8.45%，F1 提升了 4.32%，AP@0.50 提升了 6.43%，AP@0.5:0.95 提升了 4.78%；在 DOTA 数据集中检测精度提升了 7.78%，F1 提升了 4.40%，AP@0.50 提升了 8.50%，AP@0.5:0.95 提升了 6.78%。SGC 模块会使模型召回率略有下降，但能够有效提升模型的检测精度，使得模型综合性能提升。此外，SGC 模块由于增加了额外训练参数，对模型收敛具有一定影响。为了更好的阐述这一过程，我们在训练过程中记录了 RSF 验证集损失值(L_{total})并绘制曲线，如图 8 所示。

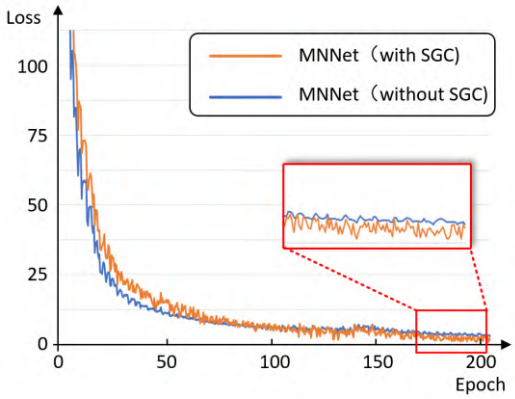

图 8 训练时验证集损失值(L_{total})曲线(RSF 验证集)

Fig. 8 Curve of loss value (L_{total}) during training (RSF validation dataset)

通过曲线可以直观的展现出模型的训练过程，SGC 模块的引入在训练开始阶段减缓了模型的收敛速度，损失值相对较高。但在模型训练的结束阶段，SGC 使模型的输出能够达到更低的误差，最终收敛结果更优。

2.4 效果展示

最后，为了直观地展现 MNNet 模型的检测能力，我们展示了在 ITCVD、RSF 以及 DOTA 数据集的部分检测结果，如图 9 所示。其中，ITCVD 和 RSF 数据集为单类（车辆）机载遥感图像的车辆目标数据集，DOTA 数据集为多类别卫星遥感图像数据集。图中左侧展示的是高像素原始遥感图像，红色框为裁剪的部分图像，裁剪框的像素尺寸大小由 1.1 节描述的尺度归一化模块计算得到。图中右侧展示的是对应的局部裁剪图在缩放后的检测结果，其中红色“×”符号标注出了错检与漏检目标。ITCVD 和 DOTA 数据集通过统计的方法近似得到裁剪像素尺度，而 RSF 数据集为采集的具有成像相关参数的数据集，裁剪像素大小通过计算得到。通过在多个数据集进行测试，本文提出的 MNNet 模型均展现出了较好的检测能力。

特别的，在 RSF 数据集中，由于包含有大量复杂环境的遥感图像，在雨雪天气中白色车辆存在被雪覆盖的问题，而部分路面与黑色车辆颜色基本一致，导致可提取的特征信息下降明显，体现了 RSF 数据集的价值，也为优化遥感检测模型提出了新的方向。在 DOTA 数据集上，由于目标密集难以标注类别信息，我们利用不同颜色的包围框进行了区分。该数据集目标像素数量差异大，图像通过尺度归一化使得各类别像素尺度统一，MNNet 模型仍具有较高的检测能力，能够满足实际任务的需要。

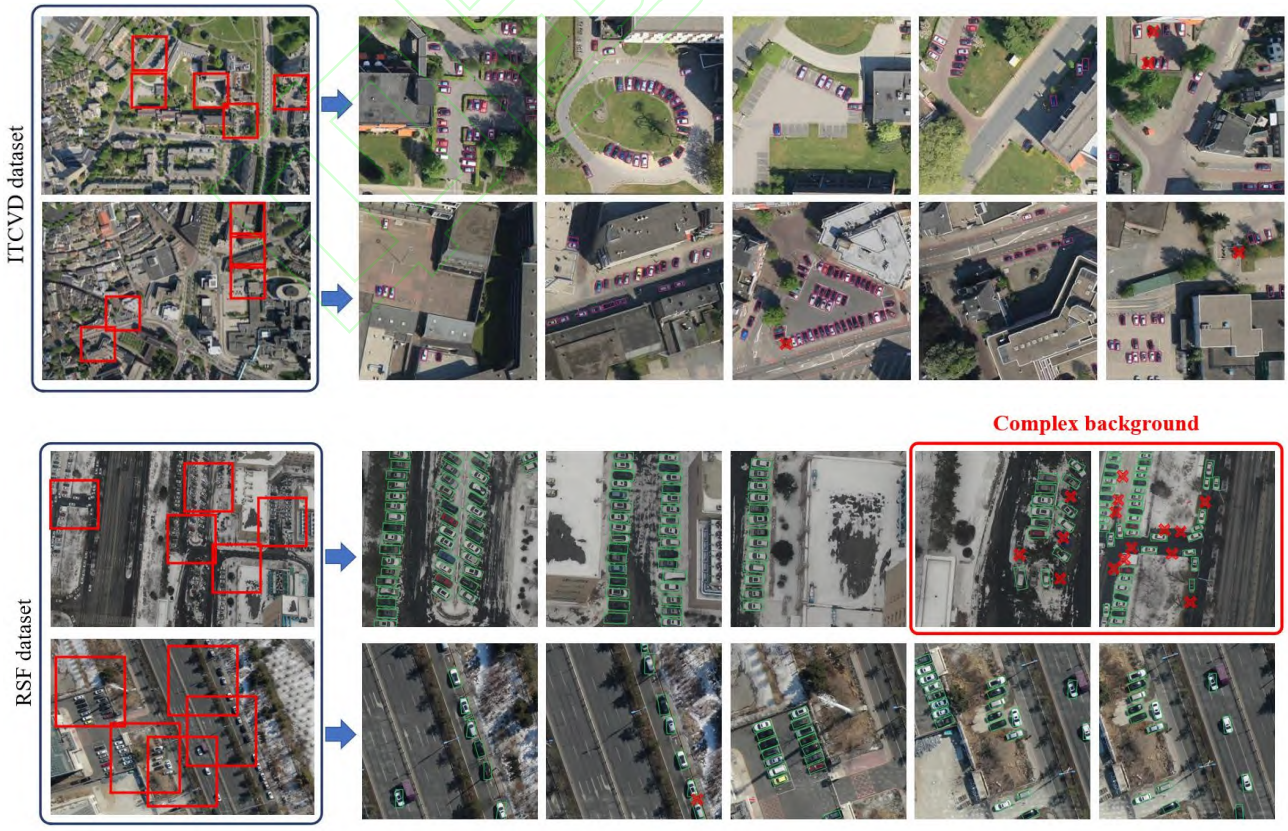

图 9 MNNet 模型的检测结果（ITCVD 数据集、RSF 数据集、DOTA 数据集）

Fig. 9 Detection results of the MNNet model (ITCVD dataset, RSF dataset, DOTA dataset)

4 结论

本文针对机载广域遥感图像中由于背景复杂、目标尺度分布大而导致检测效率低的问题，提出了一种目标尺度归一化的目标检测模型 MNNet。MNNet 模型首先利用飞行数据以及成像参数计算遥感图像的裁剪像素尺寸，通过图像裁剪并缩放等预处理后，待检测目标像素尺度能够保持较好的统一，从而为后续简化模型提供理论基础。我们提出的模型剔除了多余结构，仅保留一个预测输出结构，降低了模型的参数量，提升了模型的检测效率；我们还设计了全局连接块 SGC 模块用于增强特征图不同感受野之间的关联信息，降低模型的误检、漏检现象；此外，我们还设计了自适应阈值的非极大值抑制方法 DNMS，避免了超参数设置对经验的依赖，更好地提升了目标检测的效果。

在 RSF 数据集上进行检测性能实验验证，精度达到了 59.79%，召回率达到了 71.17%，F1 值达到了 0.6498，AP@0.50:0.95 为 34.12%，帧率达到了 57.7 fps。实验结果表明，本文提出的 MNNet 模型对比现有通用检测器在检测精度和检测速度均有明显优势，对机载广域遥感图像的目标检测任务具有积极意义。对于下一步工作，我们将会在真实机载设备中对本模型进行测试与应用，着重于提升模型对困难样本的检测能力。此外，由于机载计算设备的性能有限，在保持检测精度的同时，进一步降低计算复杂度提升模型检测效率也尤为重要。

参考文献

[1] 董秀军,邓博,袁飞云,等. 航空遥感在地质灾害领域的应用：现状与展望[J]. 武汉大学学报(信息科学版), 2023: 1-19.
Dong Xiu-jun, Deng Bo, Yuan Fei-yun, et al. Application of Aerial Remote Sensing in Geological Hazards: Current Situation and Prospects[J]. Geomatics and Information Science of Wuhan University, 2023: 1-19.

[2] 杜培军.高分辨率遥感影像处理进展与城市应用若干实例[J].现代测绘,2020,43(01):1-9.
Du Pei-jun. Progress of High Resolution Remotely Sensed Image Progressing and Urban Application Examples[J]. Modern Surveying and Mapping, 2020,43(01):1-9.

[3] Viola P, Jonus M J. Robust Real-Time Face Detection [J]. International Journal of Computer Vision. 2004, 57:137–154

[4] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 20-25, 2005, San Diego, USA. IEEE, 2005: 886-893.

[5] Girshick R B, Felzenszwalb P F, Mcallester D. Object Detection with Grammar Models[C]// Neural Information Processing Systems, December 12-17, 2011, Granada, Spain, 2011: 442-450.

[6] Dai J F, Li Y, He K M, et al. R-FCN: Object Detection via Region-based Fully Convolutional Networks[C]// Neural Information Processing Systems (NIPS), December, 5-10, 2016, Barcelona, Spain, 2016.

[7] 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.

[8] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]// IEEE Conference on Computer Vision and Pattern Recognition

(CVPR), June, 27-30, 2016, Seattle, USA, 2016: 779-788.
[9] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger[C]// IEEE/CVF Conference on Computer Vision 77and Pattern Recognition (CVPR), July, 21-26, 2017, Honolulu, USA, 2017: 6517-6525.
[10] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement[J]. arXiv 2018, arXiv:1804.02767.
[11] Bochkovskiy A, Wang C Y, Liao H. YOLOv4: Optimal Speed and Accuracy of Object Detection[J]. arXiv 2020, arXiv:2004.10934.
[12] Carion N, Massa F, Synnaeve G, et al. End-to-End Object Detection with Transformers[J]. arXiv 2020, arXiv:2005.12872.
[13] Yang M Y, Liao W T, Li X B, et al. Vehicle Detection in Aerial Images[J]. Photogramm. Eng. Remote Sens. 2019, 85: 297-304.
[14] Xia G S, Bai X, Ding J, et al. DOTA: A Large-scale Dataset for Object Detection in Aerial Images[C]// IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June, 18-23, 2018, Salt Lake City, USA, 2018: 3974-3983.
[15] Van E A. You Only Look Twice: Rapid Multi-Scale Object Detection In Satellite Imagery. arXiv 2018, arXiv:1805.09512.
[16] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector[C]// the 14th European Conference on Computer Vision (ECCV), October, 8-16, 2016, Amsterdam, Netherlands, 2016: 21-37
[17] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN[C]// IEEE International Conference on Computer Vision (ICCV), October, 22-29, 2017, Venice, Italy, 2017: 2980-2988.
[18] Neubeck A, Van G L. Efficient Non-Maximum Suppression[C]// 18th International Conference on Pattern Recognition (ICPR), 2006(3): 850-855.
[19] Bodla N, Singh B, Chellappa R, et al. Improving Object Detection with One Line of Code[C]// International Conference on Computer Vision (IEEE/CVF), 2017:5562-5570.
[20] 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 & Machine Intelligence, 2017,39(6):1137-1149.
[21] He Y H, Zhu C C, Wang J R, et al. Bounding Box Regression With Uncertainty for Accurate Object Detection[C]// Computer Vision and Pattern Recognition, 2019:2888-2897.
[22] Liu K, Mattyus G. Fast Multiclass Vehicle Detection on Aerial Images[J]. IEEE Geosci. Remote. Sens. Lett. 2015, 12, 1938-1942.
[23] Li K, Wan G, Cheng G, et al. Object detection in optical remote sensing images: A survey and a new benchmark[J]. ISPRS J. Photogramm. Remote Sens. 2020, 159, 296-307.
[24] Zhu H G, Chen X G, Dai W Q, et al. Orientation Robust Object Detection in Aerial Images Using Deep Convolutional Neural Network[C]// IEEE International Conference on Image Processing (ICIP), September, 27-30, 2015, Quebec City, Canada, 2015: 3735-3739.
[25] Lu X Q, Zhang Y L, Yuan Y, et al. Gated and Axis-Concentrated Localization Network for Remote Sensing Object Detection[J]. IEEE Trans. Geosci. Remote Sens. 2020, 58, 179-192.
[26] Zou Z X, Shi Z W. Random Access Memories: A New Paradigm for Target Detection in High Resolution Aerial Remote Sensing Images[J]. IEEE Transactions on Image Processing. 2018,27(3): 1100-1111.
[27] Silverman B W. Density Estimation for Statistics and Data Analysis[M]. Chapman and Hall/CRC, 2018.
[28] Zheng Z H, Wang P, Ren D W, et al. Enhancing Geometric Factors in Model Learning and Inference for Object Detection and Instance Segmentation[J]. IEEE Transactions on Cybernetics, 2022, 52(8): 8574-8586
[29] Ma J Q, Shao W Y, Ye H, et al. Arbitrary-Oriented Scene Text Detection via Rotation Proposals[J]. IEEE Transactions on Multimedia, 2018, 20(11): 3111-3122.
[30] Yang X, Yang J R, Yan J C, et al. SCRDet: Towards More Robust Detection for Small, Cluttered and Rotated Objects[C]//IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea(south), 2019: 8231-8240.
[31] Cheng G, Yao Y Q, Li S Y, et al. Dual-Aligned Oriented Detector[J]. IEEE Trans. Geosci. Remote Sensing, 2022, 60:1-11.