文章编号:1001-9014(2021)03-0369-12 DOI:10.11972/j.issn.1001-9014.2021.03.014

复杂背景的夜光遥感建筑区检测

李 海[1], 李 洋[2], 左峥嵘[1*]
(1. 华中科技大学 人工智能与自动化学院 多谱信息处理技术国家级重点实验室,湖北 武汉 430074;
2. 上海交通大学 机器人研究所,上海 200240)

摘要:提出了一种新的解决夜光遥感复杂背景问题的单阶段深度卷积检测网络,首先通过提取高维特征再特征选择的思想设计分类网络提取语义特征,并研究不同的通道数网络对降噪的影响;提出灰度能量的先验框匹配,将低噪声高质量的匹配框输入SSD检测网络,并使用积分图思想简化计算;使用可变形卷积以适应目标的形变,并获取更强的几何特征表达能力;通过加入顺序连接与密集连接改进全局语义模块,引入了网络的跨层信息交互,其注意力图综合考虑了高低感受野以有效区分小型目标和背景噪声。在夜光遥感数据集上通过实验验证了所设计的网络相比于其他单阶段网络具有优势,对于复杂背景下的建筑区具有较好的检测效果。

关 键 词:模式识别与智能系统;夜光遥感;深度卷积网络;建筑区检测;复杂背景
中图分类号:TP753 **文献标识码**:A

Detection of building area with complex background by night light remote sensing

LI Hai[1], LI Yang[2], ZUO Zheng-Rong[1*]
(1. National Key Laboratory of Multi-spectral Information Processing Technology, Huazhong University of Science and Technology, Wuhan 430074, China;
2. Institute of Robotics, Shanghai Jiaotong University, Shanghai 200240, China)

Abstract: A new single-stage deep convolution detection network is proposed to solve the complex background problem of night light remote sensing. Firstly, a classification network is designed by extracting high-dimensional features and then selecting features, and the influence of different channel number networks of noise reduction is studied. A prior box matching of gray-scale energy is proposed, inputting a low-noise and high-quality matching box into SSD detection network, and the idea of integral diagram is used to simplify the calculation. By adding sequential connection and dense connection to improve the global semantic module, the cross layered information interaction of the network is introduced, and its attention map comprehensively considers the high and low receptive fields to effectively distinguish small targets and background noise. Experimental results of the night light remote sensing data set show that the designed network has advantages over the rest single-stage network, which has a better detection effect of the building area under the complex background.

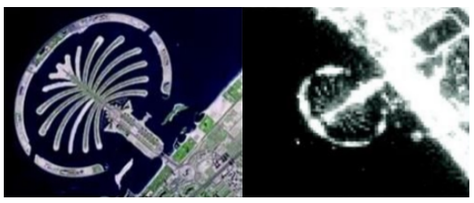
Key words: pattern recognition and intelligent system, night light remote sensing, depth convolution network, building area detection, complex background

引言

夜光遥感是近年来越来越被重视的一种新型遥感技术,不同于常见的遥感,夜光遥感是获取无云条件下地表发射的可见光-近红外电磁波信息,因此夜光遥感所反映的信息与人类活动高度关联[1]。由于夜光遥感数据量丰富、覆盖范围大、时间空间可以连续分布、能够定期快速获取大尺度夜间灯光辐射亮度数据及空间观测数据等的优势,在客观研究经济发展[2]、城市土地优化[3,4]以及电力估算[5]等领域使其成为许多学者专家的研究热点

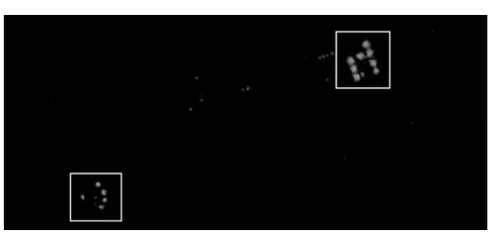
收稿日期:2020-04-05,**修回日期**:2020-10-26 **Received date**:2020-04-05,**Revised date**:2020-10-26
基金项目:国家自然科学基金(61773389)
Foundation items:Supported by the National Natural Science Foundation of China (61773389)
作者简介(Biography):李海(1995-),男,新疆乌鲁木齐人,硕士研究生,主要研究领域为红外目标检测、模式识别与智能系统. E-mail:darksosol26@163.com
*通讯作者(Corresponding author): E-mail:zhrzuo@hust.edu.cn

之一[6]。



(a)





(b)

图1 建筑物在夜光遥感影像中突出于可见光遥感的特点
(a) 局部结构相似性(b) 远距上下文关系

Fig. 1 Buildings are characterized by visible light remote sensing in night light remote sensing images(a) Local structural similarity(b) Remote context

夜光遥感图像的处理[7, 8]已经有部分研究成果，包括基于特征算子[9, 10]的检测、滤波变换[11, 12]和提取势函数[13]等。这些检测方法大部分是基于可见光遥感图像处理的方法改良得到，例如文献[14]在进行形态学滤波等方法处理后，使用K-means方法渲染新的灰度值，提取道路的形状特征；文献[15]使用城市夜间灯光指数提取城市区的形状特征，其基本思想是对于每个城市建成区，选取不同的阈值范围进行城市指数归一化，获取不同区域内像素的概率信息；文献[16]使用多元线性回归模型，将3个谱段的光谱强度作为自变量检测出不同区域的房屋空置率，而文献[17]使用区域内的像元DN的均值等特征算子输入机器学习模型计算房屋的潜在价格。上述方法的不足在于其检测精度受制于建筑区的密度和排列方式，并且当目标信噪比较低时，其检测性能较差，需人为调整的参数较多[18]。

相对于可见光遥感影像，夜光遥感影像的局部特征并不具有较好的分类能力，缺乏颜色特征，主要表现为纹理与几何分布特征；某些地物表现出与自身结构相类似的特征模式(图1(a))；而某些地物不具有自身结构相似性，但与周边地物有较强的上下文关系(图1(b))；夜光遥感背景以自然光的干扰为主，而可见光遥感具有更丰富的背景信息。因此，对夜光图像上的目标进行检测，应注重目标的宏观结构特征及其与环境地物的上下文关系的应用。利用其强大的非结构化数据的表征能力，深度学习相比于传统检测算法有更强的泛化性，无论在检测精度还是速度上都达到了较好的水平。深度学习方法在夜光遥感图像的建筑区检测方面的研究还不多。利用深度学习方法提取夜光遥感影像语义特征，捕捉目标宏观结构特征和上下文高层语义特征，可望有效提升夜光图像中的目标检测能力。

夜光遥感检测中的难点在于对目标检测干扰较大的复杂背景，需要网络在减弱噪声的同时少损失有用信息，提取有效的语义特征；并且在加强纹理与几何特征的同时，关注到建筑区中丰富的全局上下文信息。针对上述问题设计了网络结构，主要包括分类网络、先验框匹配与注意力模块。

1 网络结构

1.1 网络框架

提出了一种可以高性能地对夜光遥感图像中的建筑区进行检测的深度卷积网络，检测方案的出发点在于深度卷积神经网络通过对训练样本的学习，在降低图像背景噪声的情况下保留更多信息并学习到建筑区的纹理形状特征与丰富的上下文语义信息，进而加强对建筑区检测的效果。

(1)针对夜光遥感图像中的噪声问题，基于特征筛选的思想提取高维特征再进行特征选择以获得低噪的特征，设计了卷积核通道数先加后减的降噪分类网络，并分析通道数对降噪的影响。使用可变形卷积加强特征的几何表达能力。

(2)对于设计的先验框，为了使检测网络获得噪声更小质量更高的匹配先验框，由区域算子选择光斑影响小的先验框，使用灰度能量进行检测网络的先验框匹配，并基于积分图的思想简化计算。

(3)对于建筑区域中丰富的全局语义信息，基于GC搭建了引入连接模型的全局语义模块CGC，

对比了顺序连接与密集连接2种方式,引入上下文信息提高了网络区分小型目标与背景噪声的能力。

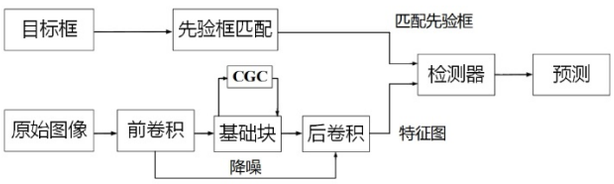


图2　网络流程图

Fig. 2　Flow chart of detection network

网络由分类网络和检测网络构成,其中分类网络由前卷积、基础块和后卷积组成,检测网络由先验框匹配和SSD[19]检测器构成。在夜光遥感数据集上经过实验验证,本文设计的分类网络、灰度能量匹配方法和全局语义模块有效,且网络精度优于经典的单阶段网络Yolov3[20]和最新的单阶段全卷积网络Fcos[21],在保持精度稍优的情况下速度是两阶段网络Faster R-CNN[22]的3倍多,对于小型类别的效果有较大提升。

1.2　分类网络

特征选择的目的是产生具有类内聚类性和类间可分性的特征,对于某一类,需要它与其他类别/噪声具有更好的区分度。特征维度越小,特征选择的特征数越少,降噪效果越好但信息损失越大。为了使分类网络通过通道数变化强化特征选择,分类网络的通道数先增加以提取高维度特征,再减少以进行特征选择获取低噪的特征。所设计的分类网络包含前卷积、基础块和后卷积3部分,如图3所示。其中,基础块先按2倍大小逐级增加通道数,通道数从128逐层提升到512,然后再减小通道数到128。基于上述思想设计分类网络A后,为了研究卷积核通道数对降噪性能的影响,采取递进法修改网络各层通道数,如图3(b,c和d所示)。其中(a)A Net:基线模型,(b)B-Small Net:相比于(a),缩小了前卷积部分,网络的第2阶段修改为32,倒数第2阶段修改为64,(c) C-Small Net:相比于(b),网络的第3阶段修改为16,倒数第1、3阶段通道数修改为64,(d) D-Small Net:相比于(c),基础块各阶段通道数均减半. 这4个网络的深度相同,其前卷积、基础块和后卷积的感受野大小是一样的,最终感受野为99。下文分别简称为A、B、C、D网络分类网络的每个卷积层后接1个Relu(激活函数)层,增强特征表达能力,后卷积的最后一层加入失效比例0.2的drop_out层防止过拟合。

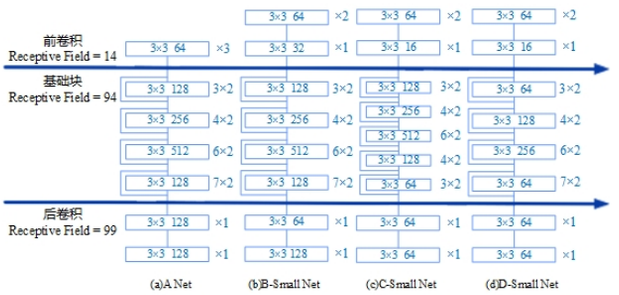


图3　4种分类网络结构(每个方框的右边标注卷积层的数量)

Fig. 3　Four kinds of Classification network structure W

1.3　检测网络

1.3.1　GE(灰度能量)先验框匹配

目标框(ground truth)不能直接输入到检测网络中,检测器需要从预先设计的多个先验框中对目标框进行先验框(anchor)匹配,以划分正负样本。为了设计合理的先验框尺寸,需要考虑目标框占图像的平均比例大小。3.1节中介绍的数据集给出了图像的宽中位数3 056与高中位数2 248,同时表3给出了数据集中5个目标类别的平均边长:72、160、1 010、318和269,故平均边长/宽中位数的范围为2.3%~33.0%,平均边长/高中位数的范围为3.2%~44.9%,这就得到了先验框尺寸需要覆盖的范围,小型目标居多要多考虑50%以下的范围。检测网络中每个网络层中的先验框边长占特征图边长的大小比例随着检测网络的阶段数递增,根据经验取0.05的倍数,设置比例为[0.05、0.1、0.2、0.3、0.5、0.7、0.9](图5中的检测网络共7个阶段,对应于这里的7个比例)。本文宽高比和高宽比均采用这个比例,得到14种先验框,通过这种方法设计的先验框尺寸可以覆盖到多数目标的范围。

网络滑窗得到先验框后,SSD的先验框匹配流

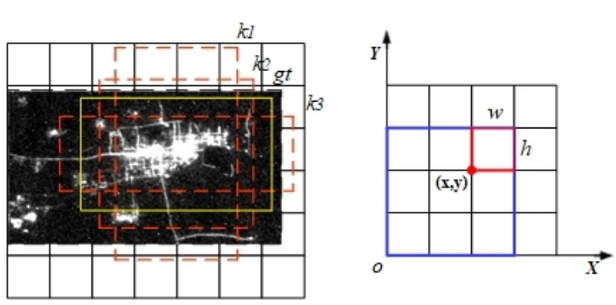


图4　GE先验框匹配 (a) Prior box matching, (b)Dynamic plan diagram

Fig. 4　Prior box matching correction

程是先计算出所有和目标框交并比IOU大于0.5的先验框，然后从中选择IOU最大的先验框作为正样本（称为匹配先验框），其余为负样本。但是对于建筑区来说，这仅仅考虑了面积的重合度而没有考虑建筑区具有的形状特点。从图4(a)可以看到，建筑区周围会存在建筑块，这些小亮块可以作为目标的特征（同时可以作为上下文信息，本文在第2节考虑这个问题），框k3相比于框k2更适合作为匹配的先验框，即使它的IOU可能不是最大的，因此使用灰度能量作为区域描述算子进行匹配修正。灰度能量反映的是灰度分布的均匀程度，当先验框中多出了小亮块时，多数情况会导致分布的趋于不均匀从而提高灰度能量，故得到本文的GE先验框匹配流程：对于每次滑窗的先验框，计算其与重叠目标框的IOU，在大于阈值的所有先验框中，取灰度能量最大的作为匹配先验框，即正样本。如果说灰度直方图是灰度概率密度，灰度能量可以称作它的平方，定义分别如下：

$$H(i)=\frac{n_i}{N},i=0,1,\cdots,L-1 \quad , \quad (1)$$

$$energy=\sum_{i=0}^{L-1}H(i)^2 \quad , \quad (2)$$

式中i表示灰度级，n_i表示图像中具有灰度级i的像素的个数，N表示图像像素总数，L表示灰度级种类数256，$H(i)$是图像灰度级i对应的像素数。对于每个先验框的灰度直方图，每轮训练都重新统计直方图是不可取的，本文基于积分图的思想使用三维数组储存每张图像的灰度直方图，图4(b)是示意图，X轴对应m，Y轴对应n。动态规划数组$dp[i][m][n]$对应示意图中的蓝色矩形，含义是图像灰度等级为i时，从点O(1,1)到右上角(m,n)矩形所包含的像素点个数，动态规划方程如下：

$$H(i)=dp[i][x+w][y+h]-dp[i][x+w][y-1]-dp[i][x-1][y+h]+dp[i][x-1][y-1], \quad (3)$$

式(3)中$dp[i][x+w][y+h]$对应图4(b)示意图中红色矩形，表示当灰度等级为i，左下角为(x,y)，宽w高h的先验框矩形所对应的像素点个数。这样只需要先用w=1和h=1建立dp数组，然后对于每个先验框计算256次就可以建立灰度直方图，极大简化了计算。修正先验框的匹配可以获得更低噪声更高质量的先验框，改善了检测网络的输入，虽然匹配筛选过程会增加网络训练时间，但网络预测时不需要筛选，不会影响预测时间。

1.3.2 检测器

SSD检测器主要包含2个过程，首先通过设计的先验框匹配得到的正样本，提取分类网络中对应的后卷积特征图，输送到检测器以预测框中目标类别及目标边框，最后将所有预测框通过非极大值抑制NMS（Non-Maximum-Suppression）筛选最佳预测结果。其中预测目标类别的损失函数是常用的交叉熵，在网络中由softmax激活函数实现；预测目标边框用到的损失函数在下文描述。

在所设计的检测网络中，特征提取器的维度与分类网络的先增后减变化趋势一致，对应的4个检测网络的具体结构如下：



图5 4个网络的检测器

Fig. 5 4 kinds of network detectors

提取特征后的检测方法采用SSD的滑窗卷积方法，根据滑窗卷积步长Stride的大小可以划分检测器为7个阶段，检测网络的浅层主要检测小型目标，深层主要检测大型目标。在检测网络的框回归损失函数设计上，使用smooth_l_1损失函数，l_1损失函数加入smooth的主要作用是平滑梯度防止离群点导致的梯度爆炸，但这样导致了其对尺寸变化不敏感，考虑到数据集中小型目标较小的IOU造成的不平衡，本文加入IOU损失进行优化以使回归框更接近目标框，因此损失函数L采用L_{smooth_l1}和L_{IOU}相加的方式，损失函数计算公式如式(5)、(6)、(7)所示：

$$L=L_{smooth_l1}+L_{IOU} \quad , \quad (4)$$

$$L_{smooth_l1}=\frac{1}{batch_{size}}\sum_{i=1}^{batch_{size}}\begin{cases}0.5*z^2 & |z|\le 1\\ |z|-0.5 & z<-1 \text{ or } z>1\end{cases} \quad , \quad (5)$$

$$L_{IOU}=\frac{1}{batch_{size}}\sum_{i=1}^{batch_{size}}-\ln\left(\frac{\bigcap_{i=1}^{4}\left(abs\left(x_{pred}(i)-x_{target}(i)\right),abs\left(y_{pred}(i)-y_{target}(i)\right)\right)}{\bigcup_{i=1}^{4}\left(abs\left(x_{pred}(i)-x_{target}(i)\right)+1,abs\left(y_{pred}(i)-y_{target}(i)\right)+1\right)}\right) \quad (6)$$

$$z=\sum_{i=1}^{4}\left(abs\left(x_{pred}(i)-x_{target}(i)\right)+abs\left(y_{pred}(i)-y_{target}(i)\right)\right) \quad (7)$$

式中***batch_size***代表每轮训练中样本的批大小;式(7)中的$x_{pred}(i)$是和$y_{pred}(i)$分别是X轴与Y轴上样本的网络预测框的4个角点(x_{pred1},y_{pred1})、(x_{pred2},y_{pred2})、(x_{pred3},y_{pred3})和(x_{pred4},y_{pred4}),同理$x_{target}(i)$是网络先验框的4个角点;式(6)中的z对应式(8),z计算出了4个角点的差距的绝对值和,没有考虑它们的相关性。式(7)中L_{IOU}与前文的IOU指标相同,考虑了4个角点的整体性。为了防止负样本中可能出现的除以0,交集面积的宽高均加平滑系数1。

1.4　降噪模块存在的问题

为了采取可视化的方法分析通道数变化与噪声的关系,采用文献[23]的方法可视化第X隐含层权重的输出,其思想是前向传播至第X层,然后使用第X层卷积核迭代若干次已有特征图,以研究图像不同的特征响应,具体方法如下:

1)前向传播:使用训练好的网络前向传播图像,得到第X层卷积核输出。

2)计算损失:计算输出的特征图的输出均值,以均值作为网络的损失,下式中N和M分别是特征图尺寸,$F(i)$是特征图中的像素点。

$$loss=\frac{1}{N*M}\sum_{i=1}^{N*M}F(i) \quad (8)$$

3)后向传播:计算梯度并更新输入图像,***gradient***是梯度,***loss***是损失,I是特征图。α为更新系数,本文取1。

$$gradient=\frac{\propto(loss)}{\propto(I)} \quad (9)$$

$$I=I+\propto *gradient \quad (10)$$

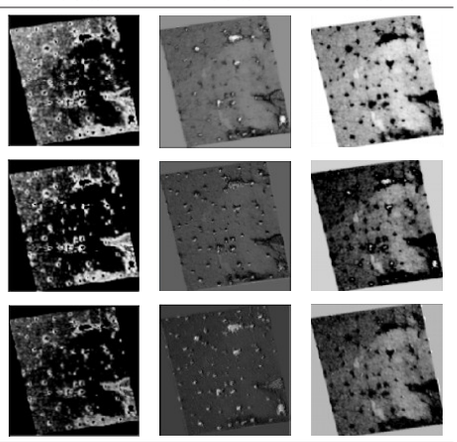
4)重复迭代:修改卷积核的输入通道等于输出通道,重复步骤3和4直到***gradient***小于阈值。

本文选取第3层的卷积核输出作为可视化目标,分别选择其中特征响应最大的4个卷积核,得到最终的可视化图像下表所示:

将图横向比较可以观察到,随着输入通道的减少,图像包含的白点减少,由此可知,输入通道数越低,网络更易学习到包含较少噪声的特征。但同时,还可以观察到图像背景的差异性越来越小,复杂背景信息的区分性变得不那么明显。网络通道数的选择应综合考虑降噪和信息损失。

表1　不同网络的隐含层特征图

Table 1　Hidden layer characteristics of different networks

输入	A网络	B网络	D网络
			

2　局部与全局语义信息

通过改变卷积核的局部感受野,可以加大网络的局部语义信息捕获能力。而通过将感受野扩大到全局,可以使网络学习到全局语义信息。

2.1　可变形卷积DCN&扩张卷积ECN

相比于普通卷积CN,采取可变形卷积[24]可以适应更多目标的形变情况,网络可以具备更强的特征表达能力。本文采用3*3普通卷积与可变形卷积。

卷积核中的每个单元会产生偏移,其直观效果是不同位置的卷积核采样点位置会根据图像内容发生自适应的变化,从而适应不同物体的形状、大小等几何形变。选取目标点中间的3*3卷积核,经过3层后相同位置的采样点变为9^3=729个,选取D分类网络的倒数3层可视化这些采样点,其中绿点表示3*3中心点。

从上图可以看到采样点偏离到了目标周围,学习到了目标的几何特性具有更强的特征表达能力,如果采用CN,采样点会构成一个正方形。接下来本文从感受野的角度研究网络与目标尺度的关系,使用扩张卷积[25]是一种简单改变感受野的方法。特征图上的一个点对应输入图上的区域大小被称为

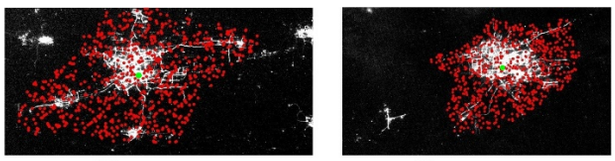


图6 单元数为9的可变形卷积经过3层的采样点图

Fig. 6 Sampling point graph of deformable convolution with unit number 9 passing through 3 layers

感受野，神经元感受野的值越大表示其能接触到的原始图像范围就越大，也意味着它蕴含更大范围的局部上下文语义信息；相反，值越小则表示其所包含的特征越趋向局部和细节。

扩大卷积核尺寸可以增大感受野增加局部语义信息，若步长为1扩张率为2则3*3卷积的感受野可以扩大到近2倍等同于5*5卷积，并且模型的时空复杂度不变。缺点是卷积核的膨胀会损失信息的连续性，模糊边缘细节信息。因此通过扩张卷积可以有效判断感受野存在的大小问题，进而综合考虑局部语义信息和边缘细节信息。

2.2 全局语义模块GC

扩张卷积只能局部扩大感受野，GC模块[26]可以使感受野扩大到全局，其原理如下：

$$y_i = x_i + \delta\left(\sum_{j=1}^{N}\omega\left(i,j,N\right)\right) \quad , \quad (11)$$

$$\delta(\cdot) = W_3 ReLU\left(Ln\left(W_2\cdot\right)\right) \quad , \quad (12)$$

$$\omega\left(i,j,N\right) = \frac{e^{W_1 x_i}}{\sum_{m=1}^{N} e^{W_1 x_m}} x_j \quad , \quad (13)$$

式(12)$\boldsymbol{x}_i$表示特征图上点i的输入值，$\boldsymbol{y}_i$是对应输出值，有$\boldsymbol{N}$个点，$\boldsymbol{N}$是特征图大小h×w×c，高×宽×输入通道；$\boldsymbol{\delta}$对应图9的转换模型；$\boldsymbol{\omega}$对应图9的语义模型，其中的1*1卷积核满足高斯分布函数，这是注意力模型常用的语义函数之一，$\boldsymbol{i}$代表注意力图上的查询点，$\boldsymbol{j}$代表另一个点，是一种查询点对关系；式(13)和式(14)中的$\boldsymbol{W}_1$，$\boldsymbol{W}_2$和$\boldsymbol{W}_3$是特征转换分别对应图9中的1*1卷积。图9中Y代表矩阵乘法，⊕代表同位元素对应相乘的和，也就是式(12)中$\boldsymbol{y}$是$\boldsymbol{x}$和注意力图融合的结果，这样经过全局信息的加强，新的输出值$\boldsymbol{y}$代替了分类网络卷积层的特征图$\boldsymbol{x}$。

GC模块建立了一种模型长范围依赖关系[27]，其中语义模型产生全局上下文权重，是一种位置注意力机制，关注特征图的整体关系；转换模型对权重进行通道依赖关系转换，是一种通道注意力机制，关注通道之间的依赖关系。然而模型的背景经过降噪模块后依然会保留部分噪声，这些信息会影响依赖关系的建立，尤其是背景噪声和小型目标的混淆问题，为了改善这种情况本文对GC模块加入连接(Connect)模型，组成CGC模块，使高感受野的注意力图引入低感受野的边缘信息。

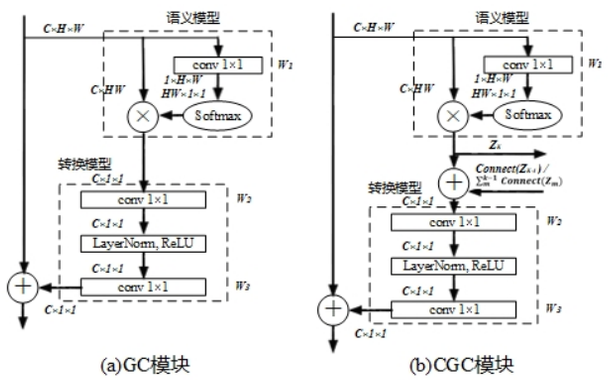


图7 GC模块与CGC模块结构对比

Fig. 7 Structure comparison between GC module and CGC module

图9(b)中z_k代表当前GC模块的注意力图输出，同理z_{k-1}表示上一层注意力图输出，经过式(15)的连接模型转换后变为z_k。连接方式1是按顺序串接，除了第一层每一层的注意力图会包含上一层的注意力图输入，连接对象是每一个特征维度的最后一层的卷积核；连接方式2是密集连接，除了第一层每一层注意力图的输入包括前面所有层注意力图的输出，连接模型由1*1卷积、批标准化BN和激活函数ReLu共3部分组成，其中1*1卷积核进行通道数变换使注意力图的尺寸匹配，BN可以加快梯度收敛，激活函数ReLu可以加强特征提取。

$$Connect(\cdot) = convBNReLU(\cdot) \quad , \quad (14)$$

对于新的连接方式分别更新式(12)：

$$y_i = x_i + \delta\left(\sum_{j=1}^{N}\omega\left(i,j,N\right) + connect\left(z_{k-1}\right)\right) \quad , \quad (15)$$

$$y_i = x_i + \delta\left(\sum_{j=1}^{N}\omega\left(i,j,N\right) + \sum_{m=1}^{k-1} connect\left(z_m\right)\right) \quad . \quad (16)$$

每层卷积后面的GC模块的注意力机制只考虑了本层的注意力图，通过上述的方法引入了注意力图跨层之间的信息交互，使其综合考虑高低感受野。高感受野更关注全局信息，学习图像丰富的关联语义信息，低感受野更关注局部信息，学习目标

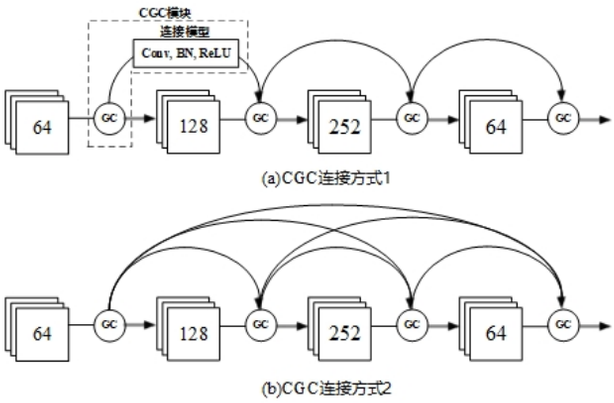


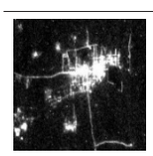
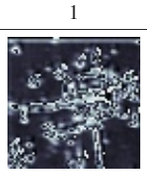
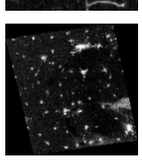
图8　两种CGC连接方式

Fig. 8　Two CGC connection modes

细致的边缘纹理信息。为了观察模块之间的不同,将D分类网络基础块中128维度的最后一层注意力图可视化,亮暗程度可以反应特征点的重要性,从下表2可以看到,使用CGC连接方式1相比于GC可以学习到更多的细节信息,注意力图的边界变得更清晰了;使用连接方式2的密集连接后,低感受野与高感受野的融合可以增加注意力图的多样性使得其对噪声更具鲁棒性,注意力图不再注意背景中的一些噪声白块。

表2　不同全局语义模块的注意力图

Table 2　Attention graphs of different global semantic modules

原图输入	GC模块	CGC连接方式1	CGC连接方式2
			
			

3　数据集与实验结果分析

3.1　夜光遥感数据集

使用中国首颗夜光遥感卫星珞珈一号[28]在广州、河南和湖北的56个城市在19年的不同时段数据,空间分辨率130 m,辐射量化级64 bit,图像经过绝对辐射亮度矫正。目标将建筑区划分为5类:1类小型城市区,2类中型城市区,3类大型城市区,4类街道相连城市区,5类街道区。图像宽度范围:2 825~3 288,中位数3 056,高度范围:2 170~2 326,中位数2 248。文献[29]将每个信息点POI和区域夜光灯光指数结合作为区域特征,类似前者,本文的1、2和3类目标通过百度地图开放平台API划分,具体划分方法是计算每个标注范围内的建筑面积和,然后从小到大排序,设定每个城市的前80%为1类目标,80%~95%为2类目标,后5%为3类目标,地图上没有的建筑物面积信息则由标注人员主观判断类别。这里选择每个城市而不是所有城市的原因在于,数据集是按照每个城市划分的。1、2和3类多属于中心发散的形状,而4类目标是多个较为分散的社区通过街道相连。判断社区是否分散的方法是计算目标框内较大的前5个建筑之间的距离占短边的比例是否大于0.2,标注方法是如果建筑区较为分散,则再由人眼判断,不是中心发散的形状则标注为4类目标。引入4类目标的意义在于区分建筑区中的难样本,避免网络将之识别为多个建筑区。

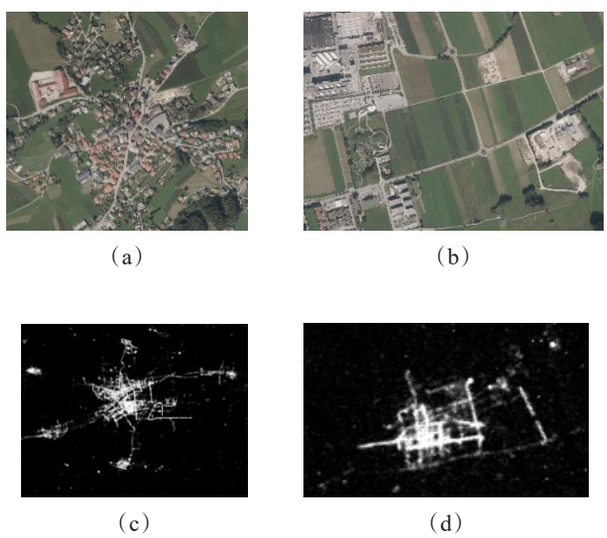


图9　高分遥感图与夜光遥感图(a) 123类目标RGB图, (b)4类目标RGB图(c) 123类目标夜光图 (d)4类目标夜光图

Fig.9　RGB remote sensing picture and night light remote sensing picture(a) RGB of 123 object, (b)RGB of 4 object, (c) Sensing of 123 object, (d) Sensing of 4 object

统计数据集中各个类别的信息,其中平均边长是图像像素总数平方根的均值,平均峰值信噪比(*PSNR*)是比较目标图像与中值滤波后图像峰值信噪比的均值。

从图与表可以观察到,数据集存在2种不平衡,

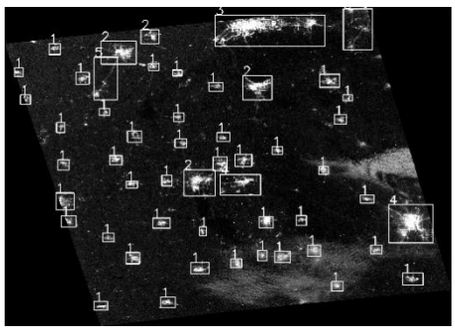


图10 夜光遥感数据集示例图像(图中标注了目标与类别)
Fig. 10 Sample picture of luminous remote sensing data set (object and category are marked in the picture)

表3 数据集中各类别情况

Table 3 All categories in the data set

类别(单位)	1	2	3	4	5
数量/个	5 283	1 138	278	227	166
平均边长/像素	72	160	1 010	318	269
平均*PSNR*/dB	19.26	19.89	21.17	21.71	23.24

其一是尺度不平衡,123类的尺寸差异较大,可以反映网络在不同尺度下的性能;其二是类别数量不平衡,以1类小型目标最多。需要重点考虑的是,遥感中建筑目标的背景语义信息往往会非常丰富,比如4和5类目标周边的街道。同时可以注意到,图像中云层自然光之类的噪声较多,这容易和1类小型目标造成混淆。本文的数据集可以反映多样的真实场景。

数据集共有768张,属于小样本数据集,使用数据增强扩大数据规模以避免神经网络的过拟合,本文使用竖直与水平翻转、旋转、裁剪平移缩放以及光学畸变4种数据增强方式对其进行扩充。夜光遥感数据集的训练集和验证集划分比例为4:1,为了防止验证集增强后的图片泄漏到训练集中,分别对训练集和验证集数据增强,共得到1 536张图像。

网络采用多尺度训练,每个batch输入图像的多尺度随机范围是512~1 024,训练时随机多尺度每隔10次batch迭代随机改变1次。SSD检测器因为全连接的固定单元数而需要固定图像尺度,本文的分类网络的全连接层前有自适应池化层,可以使用不同的尺度进行训练。结合数据增强具体而言,每张图像有0.5的概率保留原图尺寸,0.5的概率进行随机裁剪,裁剪尺度是原图短边的0.5~0.95长度的正方形,得到原图/裁剪图像后,再resize成对应的随机尺度输入网络训练。这样的多尺度训练有2个好处,其一是使网络学习到不同尺度的目标,原图经过resize是较小目标,而裁剪图经过resize相当于扩大了目标;其二是数据集宽高比不一致,不能使用一般的滑窗裁剪方法。本文分别尝试了512和1024尺度预测,指标差别很小,故下文采用512尺度预测。

3.2 消融实验

3.2.1 实验方法

采用NVIDIA GTX 2080 GPU硬件平台,内存64GB显存12GB,使用Pycharm软件操作界面,基于PyTorch的深度学习框架,操作系统为Windows10专业版64位。采用控制变量法设置batch_size=2。如果使用深度越深的网络感受野就越大,可能对位置定位越准确,但缺点是时空复杂度增加与过拟合等,后文的对比实验中网络的预训练模型为34层。

神经网络初始参数一般使用变种的高斯分布进行初始化,为了加快网络收敛,对于网络中涉及到的网络层本文使用文献[30]的初始化方法:卷积层使用输入与输出个数综合决定的均匀分布初始化;激活层使用标准正态分布,批标准化层使用常数1,池化层和dropout层没有训练参数。训练轮数epoch取120。使用Adam优化器,初始学习率0.05,使用warm up训练策略,前38轮学习率不断线性上升至0.1,然后在70轮,100轮分别下降至原先的1/3,这样做的好处是先用小的学习率稳定收敛以避免过拟合获得更为稳定的模型表现,评测分数基于验证集的308张图像。

本文使用的多分类指标mAP(mean Average Precision),并参考综合评价指标F-measure,这2个指标均是由召回率和精度组成的。

实验流程的思路是首先比较不同分类网络的性能,选择最佳网络;然后对比不同阶段添加扩张卷积与GC模块的性能,选择最佳阶段;最后进行模块的消融实验,即删除与添加部分网络结构以对比其对网络性能的影响。

3.2.2 通道数的消融试验

采用具有不同通道数的不同分类网络对对夜光遥感数据进行检测试验,以揭示通道数变化对检测性能的影响。

由上表可以看出A网络通过减少卷积层的通道数,得到的B和C网络的性能不断上升,D网络减少了更多的通道数其相比于A网络参数量下降到了1/4以下,检测效果依然高于A网络,性能相比于C网络有小幅的下降,这说明降噪方法发挥了作用,分

表4　4种网络在夜光遥感数据集上的性能(检测器网络参数量和目标类别相关,这里只统计了分类网络的参数量)

Table 4　The performance of 4 kinds of networks on the night light remote sensing data set

分类网络	参数量(10^7)	mAP
A	2. 563 750 4	0. 387 9
B	2. 522 928 0	0. 389 6
C	2. 191 003 2	0. 419 9
D	0. 636 686 4	0. 404 0

类网络的通道数升高再降低是一种有效的低噪特征提取方法。参数量越少速度越快,综合考虑性能与速度,选择D网络作为基线继续模块的消融实验。

3. 2. 3　扩张卷积的消融实验

将网络模型D中基础块对应阶段的普通卷积修改为扩张卷积,基础块共8个阶段,分别在5、6、7、8以及全部8个阶段进行修改。

表5　不同阶段添加扩张卷积

Table 5　Adding expansion convolution in different stages

阶段数	5	6	7	8	1-8
mAP	0. 400 9	0. 401 3	0. 400 8	0. 401 2	0. 400 8

试验表明,扩张卷积虽然将感受野加大了近2倍,但检测性能未有明显提升,需要进一步引入图像中的上下文全局语义信息。

3. 2. 4　GC模块的消融实验

在网络基础块的对应阶段添加GC模块,根据经验不在最后一个即第8阶段添加,分别在4、5、6、7以及1~7阶段进行修改。

从上表可以发现添加全局语义模块是有效的,在所有阶段添加全局语义模块的基础上,按照顺序连接方式加入3个连接,GC模块成CGC1模块,按照密集连接方式加入6个连接,GC模块成CGC2模块,以引入低感受野获得目标细致的边缘信息。

表6　不同阶段添加GC模块

Table 6　Add GC module in different stages

阶段数	4	5	6	7	1-7
mAP	0. 408 9	0. 406 3	0. 409 1	0. 413 7	0. 416 7

3. 2. 5　多个模块的消融实验

下表结合了前文的结果进行了各模块的消融实验对比。实验1和实验2,实验3和实验4均是比较不同的匹配,区别在于前两者和后两者使用不同的卷积。实验2相比于实验1,实验4相比于实验3,验证了GE匹配的有效性;实验3相比实验1,实验4相比实验2,验证了DCN卷积的有效性。实验5在匹配的基础上测试了扩张卷积。在使用最优的匹配与卷积模块的基础上,实验6、7和8分别比较了不同的语义模块,验证了密集连接优于顺序连接。

结果验证了本文做出的3方面改进是有效的,其中DCN与CGC2的效果较好,GE匹配次之,最优网络是加入CGC2模块与GE匹配的D网络(CGC-D)。部分检测结果对比图如下图所示,其中实线框代表真实目标点位置且检测到,虚线框代表未检测到:

第1幅图背景中有云层等干扰噪声,由此产生了虚警情况,第2幅图展示了多个相邻城市区的检测情况,在加入CGC2模块后改进的网络结构还是取得了更好的检测效果。

3. 3　与典型检测网络的性能比较

为了比较本文设计的网络相对于其他检测网络的差别,使用经典的单阶段网络Yolov3、最新的单阶段全卷积网络Fcos和两阶段网络Faster R-CNN进行对比,其中Yolov3的分类网络使用53层的DarkNet,默认epoch=100;Faster R-CNN和Fcos

表7　D网络各模块的消融实验对比

Table 7　Comparison of ablation experiments of D-network modules

实验	CN	ECN	DCN	SSD匹配	GE匹配	GC	CGC1	CGC2	mAP
1	√			√					0. 383 3
2	√				√				0. 387 6
3			√	√					0. 398 2
4			√		√				0. 404 0
5		√			√				0. 400 8
6			√		√	√			0. 416 7
7			√		√		√		0. 424 5
8			√		√			√	0. 429 6

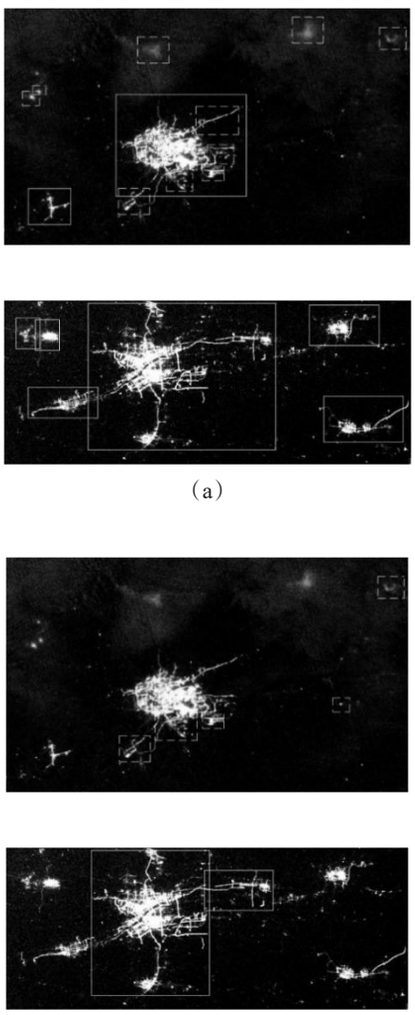


图11 网络检测图(a)未增加CGC2模块(b)增加CGC2模块

Fig. 11 Network detection image (a) No module added (b) Add CGC2 module

的分类网络使用预训练的ResNet34，默认epoch=12。4种网络的P-R曲线，F-measure和mAP指标如下图表所示，其中P-R曲线越靠近右上角，或者F-measure和mAP越高效果越好。F-measure表现了类别之间的不平衡问题，相对越高的类别表现越好。

从图表可以发现，在单阶段网络中CGC-D相比于Yolov3、Fcos与D网络的P-R曲线更靠右上，并且F-measure分数普遍更高，类别之间更为均衡尤其是1类小型目标的改善较大。但某些类别不理想，Faster R-CNN对于某些类的拟合是最优的比如街道区，这可能是两阶段网络需要候选区域的生成导致其与单阶段有一定差别。

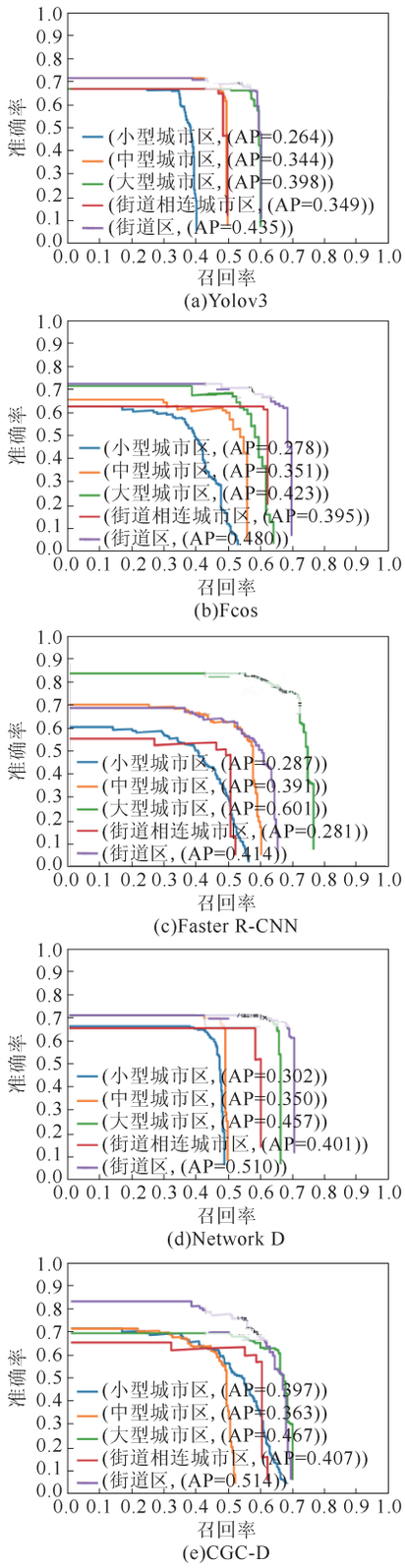


图12 不同网络的P-R曲线

Fig.12 Different networks' P-R curve

统计结果表明，加入了CGC模块的D网络(CGC-D网络)相比于原始D网络在各个类别均有

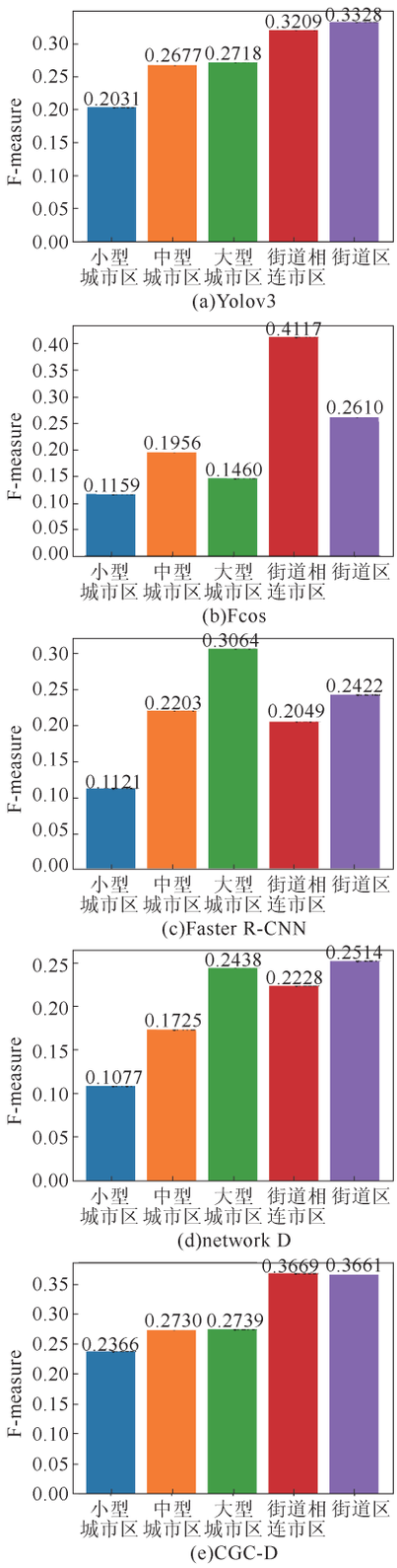


图 13　不同网络的 F-measure

Fig.13　Different networks' F-measure

提升,提升最大的是小型目标;在帧数和 mAP 值上也优于经典单阶段网络 Yolov3 和最新单阶段网络 Fcos,其对复杂背景具有更好的鲁棒性,改善了背景噪声和小型目标的混淆情况;CGC-D 的速度是二阶段网络 Faster R-CNN 的 3 倍多,但是在部分类别包括大型城市区和中型城市区上精度低于后者,这 2 类目标尺寸较大,受背景影响较小,这说明从复杂背景角度设计的网络在其他目标上没有小型目标的提升明显。

表 8　不同网络的检测效果

Table 8　Different network detection effects

网络	1	2	3	4	5	mAP	帧数
Yolov3	0.264	0.344	0.398	0.349	0.435	0.3580	6.7
Fcos	0.278	0.351	0.423	0.395	0.480	0.3854	5.6
Faster R-CNN	0.287	0.391	0.601	0.281	0.414	0.3949	4.3
D	0.302	0.350	0.457	0.401	0.510	0.4040	16.7
CGC-D	0.397	0.363	0.467	0.407	0.514	0.4296	14.3

4　总结

针对夜光遥感的复杂背景问题,提出了一种对建筑区具有较好检测能力的单阶段深度卷积检测网络结构。采用通道数先加后减的特征选择方法设计分类网络,以提取低噪特征,并对比了使用递进法得到的 4 种不同通道数的网络对降噪性能的影响;基于设计的先验框使用区域描述算子-灰度能量进行检测网络的先验框匹配,获得了噪声更小质量更高的先验框,并使用积分图思想简化计算过程;使用可变形卷积提高网络的几何特征表达能力;针对遥感图像中丰富的局部与全局语义信息从感受野的角度分析,基于 GC 提出了 2 种连接方式的 CGC 全局语义模块以有效区分小型目标和背景噪声,其综合考虑了位置注意力,通道注意力与跨层信息交互,密集连接比顺序连接具有更高精度,改善了小型目标与背景噪声的混淆情况。最终设计的网络结构在测试结果上超越了对比的单阶段网络,可以作为一种利用深度卷积网络,在夜光遥感集上对建筑区提取低噪特征并学习上下文语义信息,以有效提升复杂背景检测效果的参考方法。

在夜光遥感领域一个难题是数据集的获取,更多的数据更能证明方法的有效性,使用数据增强后本文的数据集规模依然不够大,在未来的研究中将考虑使用小样本学习方法。

References

[1] Wu Bin, Yu Bailang, Yao Shenjun, *et al*. A surface network based method for studying urban hierarchies by night time light remote

sensing data[J]. *International Journal of Geographical Information Science*, 2019, **33**(7).

[2] Zhu Hui, Zhang Qingling, Zhang Shan. Temporal and spatial characteristics of social and economic development in Central Asia based on night light remote sensing from 1992 to 2017 [J]. *Journal of Earth Information Science*(朱惠,张清凌,张珊. 1992—2017年基于夜光遥感的中亚社会经济发展时空特征分析.**地球信息科学学报**), 2020, **22**(07): 1449-1462.

[3] Cheng X, Shao H, Li Y, *et al*. Urban land intensive use evaluation study based on nighttime light—a case study of the yangtze river economic belt[J]. *Sustainability*, 2019, **11**(3): 675.

[4] Sun Lishuang, Han Yaohui, Xie Zhiwei, *et al*. Neighborhood extremum method for extracting urban built-up area using noctilucent remote sensing data [J]. *Journal of Wuhan University (Information Science Edition*(孙立双,韩耀辉,谢志伟,等. 采用夜光遥感数据提取城市建成区的邻域极值法.**武汉大学学报(信息科学版)**): 1-7.

[5] Li Xi, Xue Xiangyu. Estimation method of remote sensing power consumption based on Boston matrix [J]. *Journal of Wuhan University Information Science Edition*(李熙,薛翔宇. 基于波士顿矩阵的夜光遥感电力消费估算方法.**武汉大学学报(信息科学版)**), 2018, **43**(12): 1994-2002.

[6] Li Deren, Li Xi. On luminous remote sensing data mining [J]. *Acta Sinica Sinica*(李德仁,李熙. 论夜光遥感数据挖掘. **测绘学报**), 2015, **44**(06): 591-601.

[7] Chang Y, Wang S, Zhou Y, *et al*. A novel method of evaluating highway traffic prosperity based on nighttime light remote sensing [J]. *Remote Sensing*. 2019, **12**(1): 102.

[8] Jiang W, He G, Leng W, *et al*. Characterizing light pollution trends across protected areas in china using nighttime light remote sensing data[J]. *ISPRS International Journal of Geo-Information*. 2018, **7**(7): 243.

[9] Wen Kai. Infrared dim small target detection and tracking algorithm based on complex fusion features and gray texture histogram descriptors [J]. *Science, technology and engineering*(闻凯. 基于复杂融合特征与灰度-纹理直方图描述子的红外弱小目标检测追踪算法. **科学技术与工程**), 2016, **16**(34): 83-91.

[10] Yu B , Lian T , Huang Y , *et al*. Integration of nighttime light remote sensing images and taxi GPS tracking data for population surface enhancement [J]. *International Journal of Geographical Information Science*, 2019, **33**(3-4):687-706.

[11] Faouzi B, Washaya P. Tracking Dynamic changes and monitoring socioeconomic parameters in algeria between 1993 and 2012, using nighttime light remote sensing [J]. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci*. 2017, XLII-2/W7: 1127-1135.

[12] Devkota B, Miyazaki H, Witayangkurn A, *et al*. Using volunteered geographic information and nighttime light remote sensing data to identify tourism areas of interest[J]. *Sustainability*. 2019, **11**(17): 4718.

[13] Swathi R, Srinivas A. An improved image registration method using e-sift feature descriptor with hybrid optimization algorithm [J]. *Journal of the Indian Society of Remote Sensing*. 2020, **48**(2): 215-226.

[14] Wang Xing, Zhou Dong, Wang Wenmao. Estimation of the building density of the main urban area of hefei using luojia no.1 luminous image [J]. *Remote sensing information*(王兴,周侗,王文懋. 合肥主城区建筑密度的珞珈一号夜光影像估算. **遥感信息**), 2020, **35**(03): 71-77.

[15] Zhang Q, Schaaf C, Seto K C. The vegetation adjusted ntl urban index: a new approach to reduce saturation and increase variation in nighttime luminosity [J]. *Remote Sensing of Environment*. 2013, 129.

[16] Zhang Guoliang, Zhu Ruifei, Du Yibo, *et al*. Application of jilin-1 high resolution night light remote sensing image in urban monitoring [J]. *Satellite application*(张国亮,朱瑞飞,杜一博,等. 吉林一号高分辨率夜光遥感影像在城市监测中的应用. **卫星应用**), 2020(03): 27-33.

[17] Yongling Yao Y L S O. House vacancy at urban areas in china with nocturnal light data of dmsp-ols[C]. o ZhIEEE International Conference on Spatial Data Mining and Geographical Knowledge Services: 2011.

[18] Wu Shuangchen, Zuo Zhengrong. Infrared small target detection based on deep convolution neural network [J]. *J.Infrared Millim. Waves*(吴双忱,左峥嵘. 基于深度卷积神经网络的红外小目标检测. **红外与毫米波学报**), 2019, **38**(03): 371-380.

[19] Redmon J, Farhadi A. Yolov3: an incremental improvement[J]. *arXiv e-prints*. 2018: 1804-2767.

[20] Liu W, Anguelov D, Erhan D, *et al*. Ssd: single shot multibox detector[C]. Cham: Springer International Publishing, 2016.

[21] Tian Z, Shen C, Chen H, *et al*. Fcos: fully convolutional one-Stage object detection[J]. *arXiv e-prints*. 2019: 1355-1904.

[22] Shaoqing R, Kaiming H, Ross G, *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).

[23] Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: visualising image classification models and saliency maps[J]. *arXiv e-prints*. 2013: 1312-6034.

[24] Dai J, Qi H, Xiong Y, *et al*. Deformable convolutional networks [J]. *arXiv e-prints*. 2017: 1703-6211.

[25] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions[J]. *arXiv e-prints*. 2015: 1511-7122.

[26] Cao Y, Xu J, Lin S, *et al*. Gcnet: non-local networks meet squeeze-excitation networks and beyond [J]. *arXiv e-prints*. 2019: 1904-11492.

[27] Changyong R, Duho P, Kyongchol R, *et al*. Semantic image description and classification based on generalized set [J]. *ICTACT Journal on Image and Video Processing*. 2018, **8**(4).

[28] Li Deren, Zhang Guo, Shen Xin, *et al*. Design and processing of noctilucent remote sensing of luojia-1 satellite [J]. *Acta remote sensing Sinica*(李德仁,张过,沈欣,等. 珞珈一号01星夜光遥感设计与处理. **遥感学报**), 2019, **23**(06): 1011-1022.

[29] Liang Ze, Huang Jiao, Wei Feili, *et al*. Spatial range identification of urban agglomerations in china based on noctilucent remote sensing images and baidu poi data [J]. *Geographic research*(梁泽,黄姣,韦飞黎,等. 基于夜光遥感影像与百度POI数据的中国城市群空间范围识别方法. **地理研究**), 2020, **39**(01): 92-102.

[30] Yang F, Li W, Li W, *et al*. S3od: single stage small object detector from scratch for remote sensing images[C]. Cham: Springer International Publishing, 2019.