

# 仿真图像作为模板的遥感影像小目标检测方法

曹亚明<sup>1,2</sup>, 肖奇<sup>1,2</sup>, 杨震<sup>1</sup>

1. 中国科学院国家空间科学中心, 北京 100190

2. 中国科学院大学, 北京 100049

**摘要:** 随着传感器技术和航空遥感技术的不断进步, 遥感影像的质量和数量也得到了极大地提高, 而遥感影像中的目标检测是理解和分析遥感影像所面临的一个基本问题。针对神经网络在遥感影像小目标检测任务中难以提取足够多的有效特征、遥感小目标易受云雾遮挡等问题, 提出了一种基于仿真图像模板匹配的方法, 通过特征融合的方式成功地将该方法应用于遥感影像小目标检测任务。成像仿真技术生成的仿真图像包含了遥感小目标的更多特征, 如几何形状、材质等。与深度学习结合的过程中, 更多的特征可以使神经网络对于遥感图像小目标检测的准确率得到提升。结果表明将基于仿真图像的模板匹配方法应用于深度学习之后, 对于遥感影像小目标取得了较好的检测效果, 尤其是针对受云雾等天气干扰的小目标。

**关键词:** 遥感影像; 目标检测; 仿真图像; 模板; 深度学习

**文献标志码:** J    **中图分类号:** TP 183

## Remote sensing image small target detection method using simulation image as template

CAO Yaming<sup>1,2</sup>, XIAO Qi<sup>1,2</sup>, YANG Zhen<sup>1</sup>

1. National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China

2. University of Chinese Academy of Sciences, Beijing 100049, China

**Abstract:** With the continuous progress of sensor technology and aerial remote sensing technology, the quality and quantity of remote sensing images have been greatly improved. Target detection in remote sensing images is a basic problem in understanding and analyzing remote sensing images. Aiming at the problems that neural network is difficult to extract enough effective features in remote sensing image small target detection task, and remote sensing small target is easily blocked by cloud and fog, a method based on simulation image template matching is proposed, which is successfully applied to remote sensing image small target detection task by feature fusion. The simulation image generated by imaging simulation technology contains more features of remote sensing small target, such as geometry, material and so on. In the process of combining with deep learning, more features can improve the accuracy of neural network for small target detection in remote sensing image. The results show that the template matching method based on the simulation image is applied to the deep learning, and it achieves good detection results for small targets in remote sensing images, especially for small targets disturbed by the weather such as clouds and fog.

**基金项目:** 中国科学院基础前沿科学研究计划, E0223301。

**作者简介:** 曹亚明, 1992 年出生, 性别男, 博士研究生在读, 主研方向: 计算机视觉、深度学习。肖奇, 1996 年出生, 性别男, 硕士研究生在读, 主研方向: 对抗学习、风格迁移。杨震, 1972 年出生, 性别男, 博士研究生导师, 主研方向: 复杂系统仿真。

**收稿日期:** 2021-00-00; **修回日期:** 2021-00-00

**Key words:** Remote sensing image; object detection; template; deep learning

**引言:** 随着航天航空技术的突飞猛进, 遥感技术也得到了快速成长, 遥感技术在许多领域中都发挥着巨大作用, 在交通、环境、农业等领域中遥感技术都得到了广泛的应用。遥感技术的发展使得遥感影像的精度越来越高, 其视角较高, 所以这些影像包含的位置、特征等信息也越来越丰富<sup>[1,2]</sup>。因此, 遥感影像中的目标检测受到了越来越多的关注。传统的遥感影像目标检测任务是通过基于人工筛选特征、滑动窗口、分类器的方法来完成的<sup>[3-5]</sup>。深度学习在众多计算机视觉任务中都取得了很好的效果之后, 基于深度学习的方法已经被广泛的应用于目标检测任务中, 并且基于深度学习的目标检测模型的准确率远远高于基于人工筛选特征的方法<sup>[6,7]</sup>。然而, 由于遥感影像中的目标具有视角特殊、目标小、特征少、背景复杂、方向多变以及易受云雾遮挡等问题, 导致将常规目标检测任务中表现优秀的基于深度学习的目标检测模型直接应用于遥感影像目标检测时效果较差, 存在大量漏检、误检<sup>[8]</sup>。与此同时, 成像仿真技术越发成熟<sup>[9]</sup>, 很多 3D 建模软件在工业、建筑等领域中的使用也变得越来越频繁。虚拟图像, 尤其是将仿真图像应用于深度神经网络训练与学习的应用变得越来越多。在成像仿真技术的支持下, 可以得到小目标的更多特征, 在与深度学习结合的过程中, 更多的特征可以使得神经网络的结果得到提升。成像仿真的结果相当于人类的常识加入到神经网络中, 会使得神经网络的训练更快的收敛、提高效率, 并且加入仿真图像以后, 可以增强检测结果的可解释性。因此, 本文提出一种基于仿真图像的深度学习法——TMSI-Net(Template Matching based on Simulation Image), 利用模板匹配方法将仿真图像所包含的特征与神经网络提取的特征进行融合, 来解决遥感影像小目标检测问题。成像仿真技术生成的仿真图像包含了遥感小目标的更多特征, 如几何形状、材质等。与深度学习结合的过程中, 更多的特征可以使得神经网络对于遥感图像小目标检测的准确率得到提升。用大

型遥感数据集 VEDAI<sup>[10]</sup>以及加云雾之后的 VEDAI-Cloud 对模型进行了测试。结果表明将基于仿真图像的模板匹配方法应用于深度学习之后, 对于遥感影像小目标取得了较好的检测效果, 尤其是针对受云雾等天气干扰的小目标。

## 1 算法概述

### 1.1 TMSI-Net 网络结构

TMSI-Net 结构如图 1 所示。TMSI-Net 的主干网络为去掉最后一个残差模块的 Darknet-53, 共 53 个卷积层, 包含 5 个残差模块, 每个残差模块包含多个残差单元。Conv 表示神经网络中的一个学习单元, 具体为一个卷积层、一个批归一化(batch normalization)层和一个池化层。池化层是通过改变卷积核的步长来实现的。每个残差模块包含不同数量的残差单元, 具体数字在图 1 中进行了标注。Concat 操作为指定两个向量的某一维度进行拼接, 而其他维度固定。

通过将 TMSI-Net 和其他模型分别在 VEDAI、加云雾遮挡作为干扰的 VEDAI 数据集上进行测试, 来验证 TMSI 算法的有效性。在实验过程中, 两个数据集都按照 5:1 的比例将图像随机的分为训练集、验证集。

### 1.2 仿真图像与数据集

TMSI 算法完成了仿真图像与神经网络特征图的匹配, 并且将匹配过程中得到的有效信息加入到原来的特征图中。TMSI 算法的整个过程如图 1 所示。在 TMSI 算法过程中用到的仿真图像如图 2 所示。这些仿真图像是在可见光波段利用成像仿真技术生成的, 仿真图像的颜色和尺寸都进行了适应性的修改。首先, 根据数据集中待检测目标的大小对仿真图像的分辨率进行调整。在 VEDAI 数据集中平均每幅图像包含 5.5 个车辆, 大概占据了一幅图像总像素的 0.7%。在实验过程中所使用的 VEDAI 图像的分辨率为  $512 \times 512$ , 经过简单的计算可以得到在图像中的每个目标所占据的像素大概为  $20 \times 20$ 。因此, 将生成的仿真图像分辨率调整为  $30 \times 30$ , 可以更好地匹配特征图中的目标。考虑到遥感影像中的目标背景通常比较复杂并且多变,

例如,在 VEDAI 数据集中的很多船是在陆地上而不是常见的以水为背景,所以在仿真图像的生成过程中不会添加特定的背景,仿真图像的主体即为待检测目标。为了尽可能的消除匹配时同一类别目标的类内颜色差异性,同时考虑到与仿真图像进行匹配的是特征图,

不是输入图,所以在将所有的仿真图像分辨率进行调整之后,将它们全部转为灰度图。在仿真图像的生成过程中,通过调整待检测目标三维模型的水平角与俯仰角,可以得到如图 2 所示的多角度、多尺度的待检测目标的仿真图像。

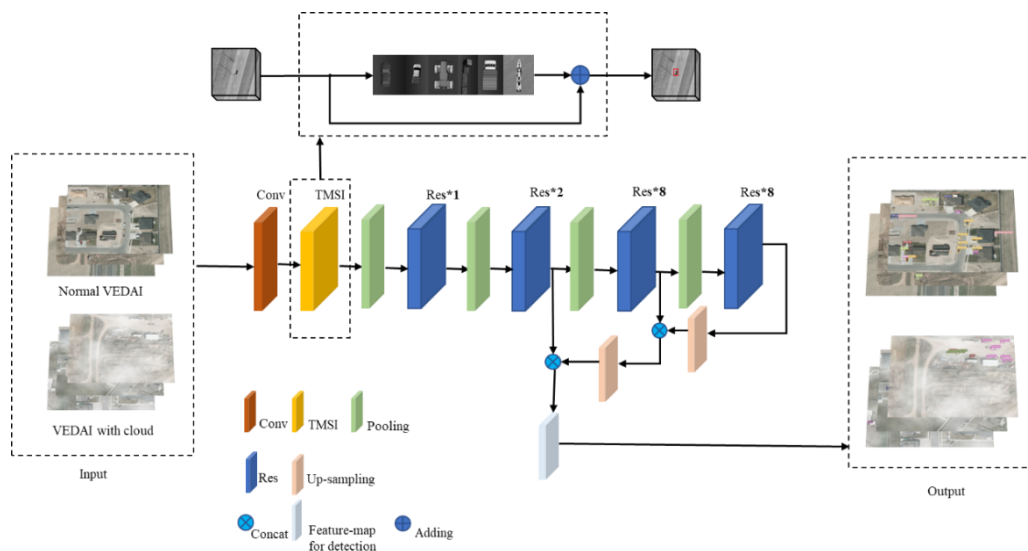


图 1 TMSI-Net

Fig.1 TMSI-Net

VEDAI 数据集中一共有 9 个类别,分别是:飞机、船、小轿车、卡车、拖拉机、露营车、货车、皮卡和其他类别。为了比较加入对应类别的仿真图像对该类别的检测准确率影响,并且减少匹配时的计算消耗,在仿真图像的生成时并不是每个类别都有对应的仿真图像作为模板,仅对其中的 6 类进行了仿真,如**错误!未找到引用源。**所示,从左到右分别是:小轿车、皮卡、拖拉机、卡车、货车和船。结合不同的角度、视角,最后得到了 24 张仿真图像作为模板,与神经网络特征图进行匹配。

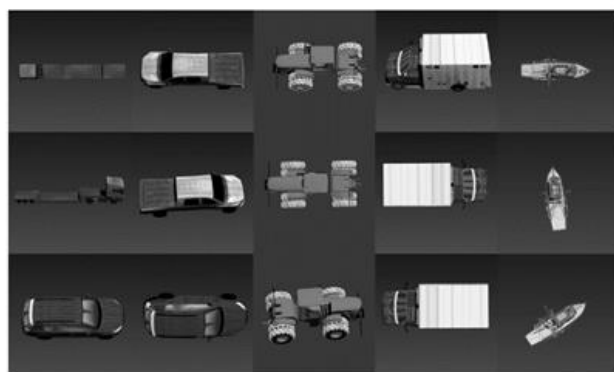


图 2 用作模板来与神经网络特征图进行匹配的仿真图样例

Fig.2 An example of a simulation image used as a template to match the neural network feature map

## 2 TMSI 算法

在 TMSI 算法中使用 OpenCV 中的 `CV_TM_CCOEFF_NORMED` 模板匹配算法进行仿真图像与神经网络特征图的匹配。在匹配前分别对模板  $T(i, j)$  和原图  $I(i, j)$  进行归一化处理, 如式 (1)、(2)

所示。 $(w_T, h_T)$ 和 $(w_I, h_I)$ 分别代表模板与原图的宽、高。原图与模板在经过式(1)、(2)之后,完成了匹配前的标准化。最后由式(3)得到模板与原图对应区域的相关系数,以及原图中与模板最相近的区域。

$$T'(x, y) = T(x, y) - \frac{1}{w_T \times h_T} \sum_{x', y'} T(x', y') \quad (1)$$

$$I'(x, y) = I(x, y) - \frac{1}{w_I \times h_I} \sum_{x', y'} I(x', y') \quad (2)$$

$$R(x, y) = \frac{\sum_{x', y'} (T'(x, y) * I'(x + x', y + y'))}{\sqrt{\sum_{x', y'} T'^2(x, y) \sum_{x', y'} I'^2(x + x', y + y')}} \quad (3)$$

TMSI 算法的本质是通过作为模板的仿真图像来增强特征图中与小目标相关的特征。TMSI 算法的输入为通过 CNN 产生的特征图与作为模板的仿真图像。不同类别与角度的仿真图像用来增强不同的小目标的相关特征。为了避免例如类内颜色多样性等非关键特征对模板的匹配效果产生影响,同时考虑到过多的下采样层与大步长的卷积层会损失很多的小目标特征,采用 Darknet-53 中第一个池化层前的特征图作为 TMSI 算法的输入。在训练的过程中, batch size 设置为 8, 而 Darknet-53 的第一个卷积层有 32 个卷积核,所以 TMSI 算法中输入的特征图  $F$  尺寸为  $8 \times 32 \times 512 \times 512$ 。由仿真图像组成的模板库  $S$  为  $24 \times 1 \times 30 \times 30$ 。在模板匹配的过程中,超参数  $\delta$  是用来衡量相关系数  $r$  的。如果特征图中某个特定区域的相关系数  $r$  大于超参数  $\delta$ , 则认为当前的区域与对应模板类别的小目标最为相近,将该区域利用相关系数  $r$  进行直接扩大。在实验的过程中,设定超参数  $\delta$  的值为 0.6。TMSI 的整个结构与残差结构相似,残差结构最后的输出是通过向量拼接得到,而 TMSI 最后的结果则是通过输入与对应扩大区域的相加得到。当特征图与所有的模板匹配之后,最可能具有小目标的区域得到增强,更多的小目标关键特征与信息被用于目标检测。

Algorithm1 中的伪代码描述了 TMSI 算法的过程。

#### Algorithm 1 TMSI

**Input:** The feature maps  $F$  to be matched, the shape of it is  $bs \times ch \times w \times h$ , where  $bs$  is the batch size,  $ch$  is the number of channels of  $F$ ,  $w \times h$  is the shape

of each channel, a threshold  $\delta$ , where  $0 < \delta < 1$ , the grayscale simulation images as template library  $S$ , the shape of it is  $N \times 1 \times n \times m$ , where  $N$  is the number of simulation images,  $n \times m$  is the shape of each simulation image.

```

1:  get  $N$  (number of  $S$ )
2:  for  $i=1, 2, \dots, bs$  do
3:    for  $j=1, 2, \dots, ch$  do
4:       $I = F(i, j, :, :)$ , use  $I$  as the input image of
        the matching algorithm
5:    for  $t=1, 2, \dots, N$  do
6:       $T = S(t, 1, :, :)$ , use  $T$  as the template of the
        matching algorithm
7:      standardize the template  $T$  using formula
        (1)
8:      standardize the input image  $I$  using for-
        mula (2)
9:      for  $(x, y) = (0, 0), (0, 1), \dots, (w-n, h-m)$  do
10:        use formula (3) to calculate the correla-
          tion coefficient  $r$  between  $I$  and  $T$ ,
          and get the input image  $I$  coordinates
           $(x_r, y_r)$  corresponding to the current  $r$ 
11:        if  $r > \delta$ :
12:           $I(x_r : x_r + n, y_r : y_r + m) +=$ 
             $r \times I(x_r : x_r + n, y_r : y_r + m)$ 
13:        end if
14:      end for
15:    end for
16:  end for
17: end for

```

**Output**  $F$

### 3 实验结果与分析

TMSI 算法通过利用仿真图像作为模板与特征图进行匹配增加神经网络可以学习到的遥感影像小目标的关键特征。本算法同时以知识的形式将仿真图像的光学、物理等模型加入到了神经网络。TMSI-Net 网络的可解释性也得到了提高。通过在 VEDAI 数据集、加云雾遮挡的 VEDAI 数据集上的实验验证了 TMSI 算法的有效性。在实验过程中分别比较了 TMSI-Net、YOLO V3<sup>[11]</sup>、YOLO V5<sup>[12]</sup>和 SEN<sup>[13]</sup>模型的表现,并且通过对内部神经元可视化的方式分析了各个模型在不同数据集上表现差异的原因。

#### 3.1 VEDAI 验证 TMSI 算法实验

首先,在 VEDAI 数据集上分别对 YOLO V3、

YOLO V5、SEN 和 TMSI-Net 进行训练, 并进行推理测试。YOLO V3、SEN 和 TMSI-Net 的网络主体都为 Darknet-53, 所以将这三个模型的训练过程进行比较。训练时损失 (loss) 随着训练轮次的变化, 共  $40 \times 125$  轮迭代次数, 从图 3 中可以看出 SEN 和 TMSI-Net 的收敛速度都要优于 YOLO V3, 其中, TMSI-Net 模型收敛最快, 并且在训练的过程中震荡最少。

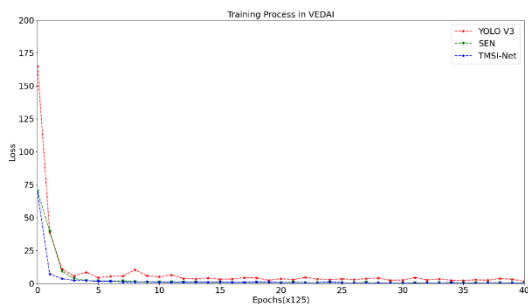


图 3 YOLO V3、SEN 和 TMSI-Net 模型在 VEDAI 数据集训练过程

Fig.3 Training process of YOLO V3, SEN and TMSI-Net in the VEDAI dataset

通过比较表 1 中与其他模型的结果比较, 可以看到提出的 TMSI-Net 在货车 (van) 之外的所有类别的 AP 以及 mAP 上都有很大的提高, 证明了本文提出的 TMSI 算法在遥感影像小目标检测任务上有很好的表现。并且, TMSI-Net 中加入了仿真图像作为模板的类

的准确率提升的更多, 如轿车、卡车、拖拉机和船等。主要原因是加入仿真图像作为模板进行匹配后, 小目标的更多特征被网络所学习到并用于检测。然而, 对于货车, 虽然也有对应的仿真图像作为模板加入, 检测准确率却没有达到所有模型中的最好, 原因可能是该类别的仿真呢图像的风格与待检测的该类小目标相差较大。图像中的该类别目标稀少也可能是其中的原因。还可以从表 1 中观察到有一些类别虽然没有加入对应的仿真图像作为模板, 但检测结果也得到了提升。这个原因可能是网络可以学习到更多的加入仿真图像作为模板的类别的小目标特征, 没有加入仿真图像作为模板的类别的小目标被误检为其他目标的可能性更小了。从表 1 的结果中也可以看出新提出 YOLO V5 的检测准确率要低于 YOLO V3, 这在某种程度上说明对于遥感影像小目标检测来说, 并不意味着网络越复杂、层数越多就对会有更好的效果。针对某种场景设计、具有特定结构的神经网络才可以在遥感影像小目标检测任务中取得更好的效果。总体而言, 在加入仿真图像作为模板之后, 网络的所有类别的小目标的检测准确率都得到了提高, 平均准确率 (mAP) 更是比其他三个模型中最好的结果高了将近 20%。仿真图像作为模板加入到神经网络确实提供了更多的有效的小目标特征。

表 1 所有模型在 VEDAI 数据集上的检测准确率与平均准确率

Table 1 The detection accuracy (AP,%) and average accuracy (map,%) of all models on VEDAI dataset

models	car	truck	tractor	camping car	other	van	boat	plane	pickup	mAP
YOLO V3	69.1	32.8	52.5	54.7	31.4	21.1	4.9	93.2	44.7	44.9
YOLO V5	49.7	31.7	42.3	69.4	35.5	26.9	24.4	39.1	68.1	43.0
SEN	60.0	24.5	37.0	54.8	19.0	<b>56.6</b>	26.5	100.0	52.0	47.8
TMSI-Net	<b>93.9</b>	<b>46.4</b>	<b>84.0</b>	<b>75.4</b>	<b>43.1</b>	51.7	<b>43.4</b>	<b>100.0</b>	<b>69.1</b>	<b>67.4</b>

表 2 所有模型在 VEDAI-Cloud 数据集上的检测准确率与平均准确率

Table 2 The detection accuracy (AP,%) and average accuracy (map,%) of all models on VEDAI dataset

models	car	truck	tractor	camping car	other	van	boat	plane	pickup	mAP
YOLO V3	11.0	26.3	32.4	39.0	27.7	5.5	4.4	87.5	41.8	30.6
YOLO V5	49.9	30.8	37.9	39.8	31.7	18.6	21.7	37.2	44.9	34.7
SEN	53.2	20.4	32.3	40.1	17.8	32.9	23.3	90.0	50.4	40.0
TMSI-Net	<b>89.3</b>	<b>35.2</b>	<b>58.2</b>	<b>78.1</b>	<b>42.1</b>	<b>76.1</b>	<b>43.5</b>	<b>100.0</b>	<b>82.3</b>	<b>67.2</b>

### 3.2 VEDAI-Cloud 验证 TMSI 算法实验

与常规的图像相比较而言, 遥感影像更容易受到云雾遮挡以及光照变化的影响。尤其是考虑到地球表

面的 66%经常被云雾所遮盖<sup>[14, 15]</sup>, 在云雾干扰下准确的检测遥感影像小目标成为必须面对和解决的难题<sup>[16, 17]</sup>。因此, 考虑在 VEDAI 数据集的基础上通过人为的

增加云雾干扰构造新的数据集, VEDAI-Cloud, 来进行实验, 验证各个模型在云雾干扰下对于遥感影像小目标检测的效果。数据集中的云雾通过随机位置、随机厚薄的方式来加入, 使得 VEDAI-Cloud 数据集尽可能的接近真实场景中被云雾遮挡之后的效果。通过将 YOLO V3、YOLO V5、SEN 和 TMSI-Net 模型在 VEDAI-Cloud 数据集上进行训练并测试, 来观察所有模型的表现。在训练的过程中, 依然比较了 YOLO V3、SEN 和 TMSI-net 的损失曲线, 共  $40 \times 125$  轮迭代次数。从图 4 中可以看出, TMSI-Net 收敛最快、震荡最少, 依然是所有模型中训练过程中表现最好的模型。

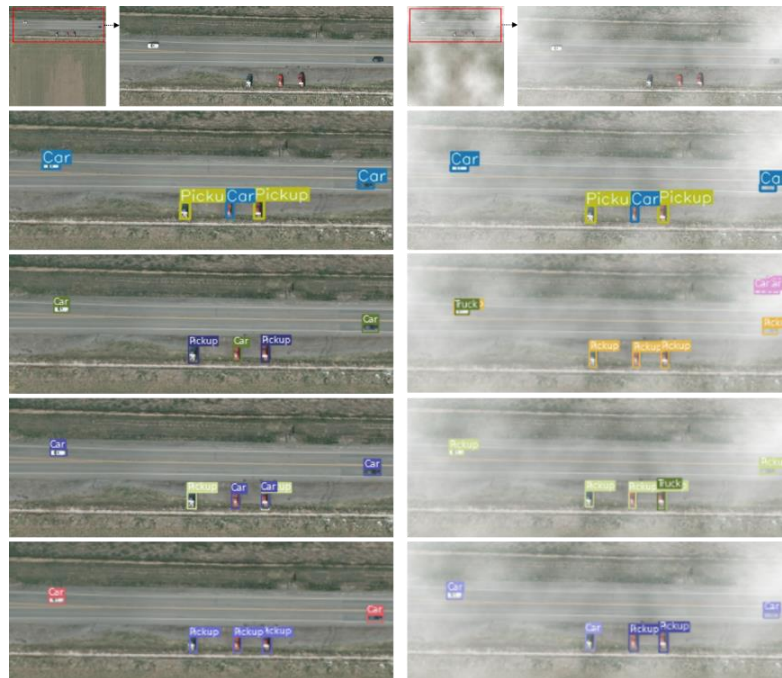
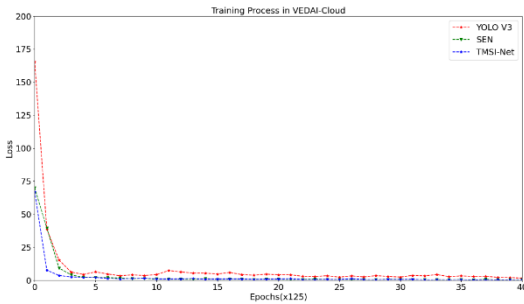


图 5 TMSI-Net、SEN、YOLO V3 和 YOLO V5 方法的检测结果比较 1

Fig.5 Comparison of detection results of TMSI-Net, SEN, YOLO V3 and YOLO V5

图 4 YOLO V3、SEN 和 TMSI-Net 模型在 VEDAI-Cloud 数据集训练过程

Fig.4 Training process of YOLO V3, SEN and TMSI-Net in the VEDAI-Cloud dataset

表 2 列出了所有模型在 VEDAI-Cloud 数据集上所有类别的准确率。与表 1 相比较, 除了 TMSI-Net 模型之外, 所有模型在加入云雾遮挡作为干扰之后的准确率都下降了很多。从表 2 中可以看出加入云雾作为干扰之后, 基于仿真图像模板匹配的 TMSI-Net 表现最好, 平均准确率比排名第二的模型 SEN 高出了 37%。表 2 中结果表明加入了仿真图像作为模板的对应类别的准确率提升最为明显, 如轿车。相反, 其他模型在加云雾遮挡之后所有类别的检测准确率都受到了极大地影响, 下降很多。在有干扰的时候 TMSI 算法依然可以为小目标检测提供很多有效的关键特征。



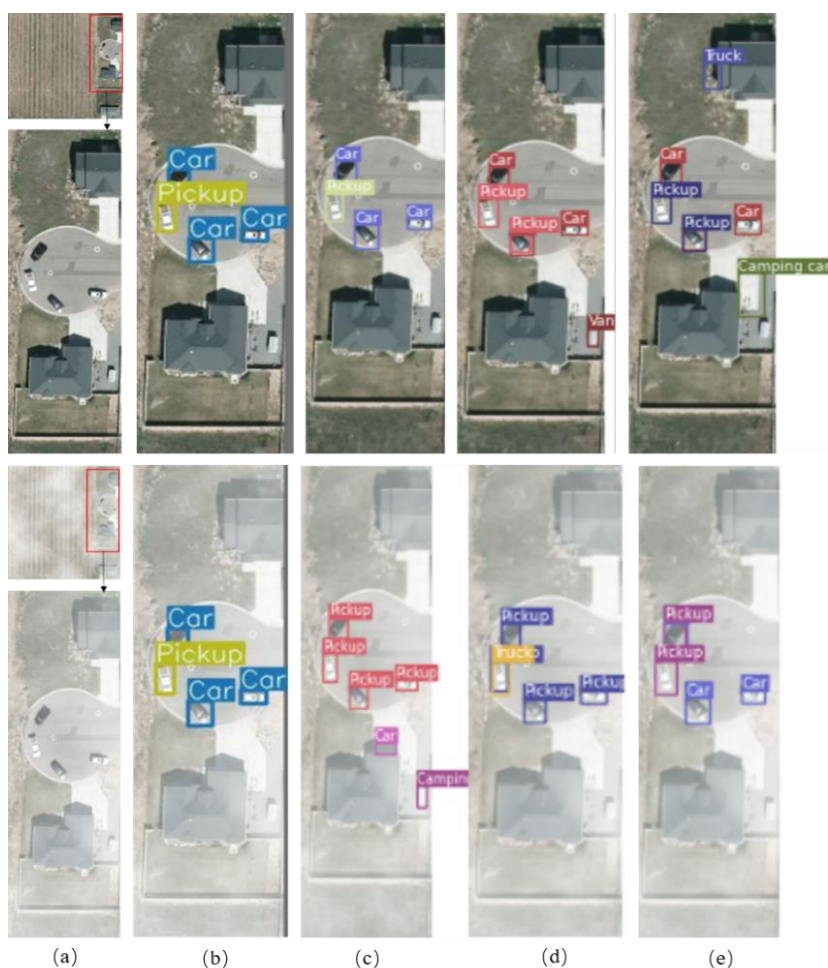


图 6 TMSI-Net、SEN、YOLO V3 和 YOLO V5 方法的检测结果比较 2

Fig.6 Comparison of detection results of TMSI-Net, SEN, YOLO V3 and YOLO V5

图 5 和图 6 是 TMSI-Net、SEN、YOLO V3 和 YOLO V5 在 VEDAI 和 VEDAI-Cloud 数据集上的部分检测结果比较。图 5 中的小目标背景相对简单，而图 6 的背景相对复杂、干扰较多。其中图 5 从上到下分别为：输入 (VEDAI、VEDAI-Cloud)；TMSI-Net 方法检测结果；SEN 方法检测结果；YOLO V3 方法检测结果；YOLO V5 方法检测结果。图 6 从左到右分别为：(a)输入 (VEDAI、VEDAI-Cloud)；(b)TMSI-Net 方法检测结果；(c)SEN 方法检测结果；(d)YOLO V3 方法检测结果；(e)YOLO V5 方法检测结果。正如图 5、图 6 结果所示，TMSI-Net 方法在 VEDAI 和 VEDAI-Cloud 数据集上都取得了最好的检测结果，所有的目标都被检测出来，并做出了正确的分类。SEN 方法在

VEDAI 数据集上的结果与 TMSI-Net 相似，然而，在加入云雾遮挡作为干扰之后，该方法有大量的漏检、误检。尤其是在图 6 背景复杂且有云雾干扰的中，SEN 方法将大部分的轿车检测为皮卡，并且将背景中的屋顶以及地面物体识别为轿车和露营车。YOLO V3 和 YOLO V5 方法则有更多的错误分类与错误检测，尤其是在加入云雾作为这遮挡之后，表现更差。TMSI-Net 方法的效果显著，优于其他方法，观察每个方法在图 5、图 6 中的检测结果，可以定性的得出结论，在遥感小目标检测的过程中，缺乏小目标的关键特征，受云雾遮挡干扰以及复杂背景是很多方法误检、漏检的主要原因。

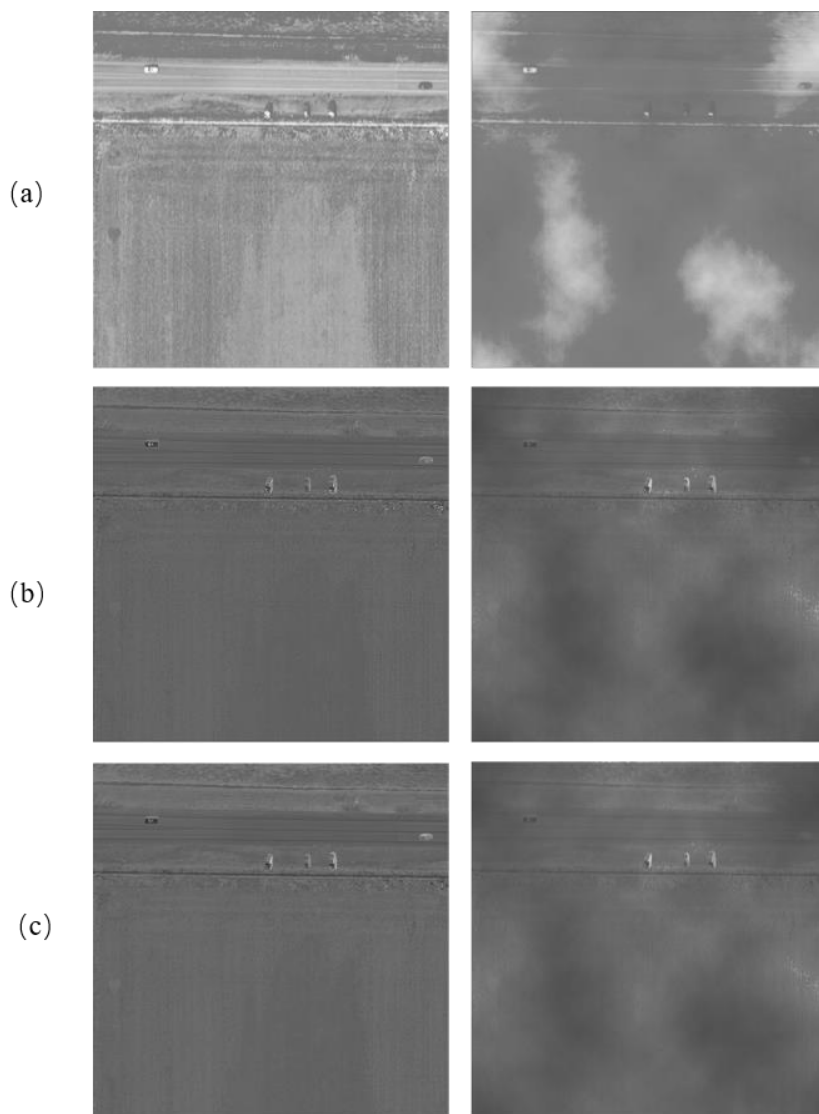


图 7 TMSI-Net、SEN 和 YOLO V3 方法的可视化结果比较

Fig.7 Comparison of visualization results among TMSI-Net, SEN and Yolo V3 methods

在图 7 中通过可视化的方式比较了 TMSI-Net、YOLO V3 以及 SEN 方法的第二个卷积层通道 0 的特征图，第一列是在 VEDAI 数据集下的，第二列是在 VEDAI-Cloud 数据集下的，从上到下分别是：(a)TMSI-Net；(b)SEN；(c)YOLO V3。这三个方法的网络主体都是 Darknet-53，所以它们的特征图具有可比性。如图 7 所示，可以观察到 TMSI-Net 方法的特征图是三个模型在 VEDAI、VEDAI-Cloud 数据集上的所有特征图中包含小目标特征与信息最多的特征图。将仿真图像作为模板加入到神经网络之后，目标与背景之间的差异得

到放大，目标可以被更好的检测到。YOLO V3 与 SEN 方法提取到的小目标的关键特征在加入云雾作为干扰之后急剧减少，不利于小目标检测。与之相反，TMSI-Net 在云雾的干扰之下依然可以提取到足够多的有效小目标特征用于检测。从图 7 中的定性比较可以得出仿真图像作为模板的有效性以及 TMSI-Net 方法在遥感影像小目标检测任务中的优越表现。

#### 4 讨论与分析

遥感影像小目标检测任务是分析和理解遥感影像所面临的基本问题。该问题在很多的民用、军用场



景中频繁出现。由于遥感影像中的目标尺寸小,受云雾遮挡的影响大,角度特殊,所以该问题一直都很难以被很好地解决。本文的实验结果表明利用待检测目标的仿真图像作为模板以模板匹配的方式加入到神经网络可以很好的提高网络在遥感影像中的目标检测任务准确率。

深度神经网络是目前为止用来解决遥感影像目标检测任务最常用的方法。其中,最关键的是如何设计网络,以便使得网络能够提取或学习到足够多的小目标特征用于检测。针对遥感目标与常规目标的视角差异,许多研究者通过修改网络结构来适应遥感影像小目标检测任务。SEN方法中的作者通过修改YOLO V3网络深、采用更浅层的特征图用于检测,在VEDAI数据集上取得了更好的检测结果。本文提出的TMSI-Net的网络结构的选取与设计也采取了同样的思路。通过改变网络深度,可以加快训练过程,并且可以去除掉对于小目标检测没有意义的冗余信息。浅层特征图的用于遥感影像目标检测可以提供更多的小目标特征和信息,如位置、几何等<sup>[18, 19]</sup>。在最后检测过程中的锚点尺寸是通过KNN聚类的方法得到的。有的研究者通过改变RPN网络输入的卷积层、调整RPN网络中的锚点尺寸来使得Faster R-CNN网络更好的适应遥感目标检测<sup>[6]</sup>。由此可以得出,通过改变网络结构来使得模型在遥感数据集上获得更好的检测效果已经成为一种有效的将深度学习应用于遥感影像数据集的基本方法。

同时,也有很多对遥感数据集进行预处理的方法被提出来去解决遥感影像小目标检测困难的问题。如果直接将高分辨率的遥感影像数据用于目标检测,确实会包含更多的目标特征,但是高分辨率的遥感影像数据通常尺寸都很大,对于网络的训练异常耗时,并且也会使得训练时每批次的数据有限,影像网络的训练效果<sup>[20]</sup>。有很多人考虑通过将大尺度的遥感数据集进行剪切和分割,然后再将剪切之后的图片块用于网

络的训练与推理,但是这种方法会导致另一个问题的出现,如何能够使得目标在切割的过程中能够保证不被破坏<sup>[21]</sup>。有的人提出了一种将高分辨率遥感数据进行下采样到合适网络训练的尺寸,然后再用其进行网络的训练与推理<sup>[10]</sup>。这种方法的不足之处在于下采样之后的图像没有足够多的小目标特征可以用于检测。数据增强的方法也经常用来对遥感数据进行预处理。在本文中提出的通过将仿真图像以模板的方式加入到神经网络中的方法可以解决遥感影像数据集中缺乏小目标特征用于检测的问题。实验结果表现优异也可以证明仿真图像的加入确实可以弥补小目标的特征与信息,提高网络检测结果。从本文的实验中也可以看出云雾的遮挡对于遥感影像小目标检测有很大的干扰,而加入仿真图像之后,可以起到很好的抗干扰作用。本文提出的加入仿真图像的方法也为解决神经网络在某些特征学习局限的场景中提供了一种新的思路。将知识与常识以模板的方式加入到神经网络中可以提高网络的表现,并且增加网络的可解释性。

## 5 结论

针对遥感影像小目标检测时存在提取到的特征不足、检测效果差等问题,本文提出了一种将仿真图像作为模板加入到神经网络的方法,可以使网络学习到更多的小目标特征,提高遥感影像小目标检测的准确率。并且,将仿真图像加入到神经网络之后,也提高了模型在面对具有云雾干扰的遥感影像小目标时的检测准确率。在VEDAI、VEDAI-Cloud数据集上的实验结果都证明了TMSI算法的有效性与稳定性。通过对神经网络内部神经元可视化的过程,可以清晰的看到TMSI-Net学习到了更多的小目标特征,并且减弱了背景的干扰。仿真图像是通过明确的物理光学模型生成的,在物理性质上是可以解释的。当仿真图像作为模板加入到神经网络时,像是一种先验知识或是常识,指导网络的学习与推理,不仅提高了网络的准确率,也提高了网络的可解释性。

## 参考文献:

- [1] Aksoy S, Akcay G, Cinbis G, et al. Automatic mapping of linearwoody vegetation features in agricultural landscapes [C]// IEEE International Geoscience & Remote Sensing Symposium. IEEE, 2009.
  - [2] Li D, Wang M, Dong Z, et al. Earth observation brain (EOB): An intelligent earth observation system [J]. Geo-spatial Information Science, 2017, 20(002): 134-40.
  - [3] McKeown Jr D M, Denlinger J L. Cooperative methods for road tracking in aerial imagery [C]//Proceedings of the 1988 DARPA IUS Workshop. 1988: 327-341.
  - [4] Barzohar M, Cooper D B. Automatic finding of main roads in aerial images by using geometric-stochastic models and estimation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996, 18(7): 707-721.
  - [5] d'Oleire-Oltmanns S, Marzolf I, Tiede D, et al. Detection of gully-affected areas by applying object-based image analysis (OBIA) in the region of Taroudannt, Morocco [J]. Remote Sensing, 2014, 6(9): 8287-8309.
  - [6] Sakla W, Konjevod G, Mundhenk T N. Deep multi-modal vehicle detection in aerial ISR imagery [C]// IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2017: 916-923.
  - [7] Nie K, Sommer L, Schumann A, et al. Semantic labeling based vehicle detection in aerial imagery [C]//IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2018: 626-634.
  - [8] Zhu X X, Tuia D, Mou L, et al. Deep learning in remote sensing: A comprehensive review and list of resources [J]. IEEE Geoscience and Remote Sensing Magazine, 2017, 5(4): 8-36.
  - [9] 王志斌. 仿真图像应用于深度学习的可行性方法研究 [D]. 西安电子科技大学, 2020.  
WANG Z B. Research on the feasibility of applying simulated image to deep learning [D] Xi'an University of Electronic Science and technology. 2020.
  - [10] Razakarivony S, Jurie F. Vehicle detection in aerial imagery: A small target detection benchmark [J]. Journal of Visual Communication and Image Representation, 2016, 34: 187-203.
  - [11] Redmon J, Farhadi A. Yolov3: An incremental improvement [J]. arXiv preprint arXiv:1804.02767, 2018.
  - [12] GLENN JOCHER A S, JIRKA BOROVEC. yolov5: v3.1 - Bug Fixes and Performance Improvements (Version v3.1). Zenodo. [J]. 2020.
  - [13] Ju M, Luo J, Zhang P, et al. A simple and efficient network for small target detection [J]. IEEE Access, 2019, 7: 85771-85781.
  - [14] Zhang Y, Rossow W B, Lacis A A, et al. Calculation of radiative fluxes from the surface to top of atmosphere based on ISCCP and other global data sets: Refinements of the radiative transfer model and the input data [J]. Journal of Geophysical Research: Atmospheres, 2004, 109(D19).
  - [15] Luo H, Chen C, Fang L, et al. Ms-rfsegnet: Multiscale regional relation feature segmentation network for semantic segmentation of urban scene point clouds [J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(12): 8301-8315.
  - [16] Lv H, Wang Y, Shen Y. An empirical and radiative transfer model based algorithm to remove thin clouds in visible bands [J]. Remote Sensing of Environment, 2016, 179: 183-195.
  - [17] Li Q, Lu W, Yang J. A hybrid thresholding algorithm for cloud detection on ground-based color images [J]. Journal of atmospheric and oceanic technology, 2011, 28(10): 1286-1296.
  - [18] Penatti O A B, Nogueira K, Dos Santos J A. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? [C]// IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2015: 44-51.
  - [19] Nanjun H, Paoletti M E, Mario H J, et al. Feature Extraction With Multiscale Covariance Maps for Hyperspectral Image Classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, PP:1-15.
  - [20] Courtrai L, Pham M T, Lefèvre S. Small Object Detection in Remote Sensing Images Based on Super-Resolution with Auxiliary Generative Adversarial Networks [J]. Remote Sensing, 2020, 12(19): 3152.
  - [21] Mundhenk T N, Konjevod G, Sakla W A, et al. A large contextual dataset for classification, detection and counting of cars with deep learning [C]//European Conference on Computer Vision. Springer,(ECCV), Cham, 2016: 785-800.
1. 曹亚明 ; 2. 北京市怀柔区杨雁路京密北二街国家空间科学中心 (101400) ;  
3. yafei4554@foxmail.com、13051158526。