

1.1.1 基于机器学习的遥感图像目标检测算法

基于机器学习的传统遥感图像目标检测算法大多将检测问题分成两步实施，候选区域生成和目标分类确认，候选区域生成阶段多采用阈值分割、显著性检测、连通域分析、圆周频率滤波、简单滑窗设计和基于金字塔的滑窗设计等方式，利用简单的图像处理得到存在目标的候选区域，一般产生的候选框数量较多，可以称为“粗检测阶段”。粗检测得到的候选框再次送给目标分类确认阶段，目标分类确认中的目标在这里包括背景，将不含目标的背景当成目标的一种。在目标分类确认阶段主要也有两步，特征提取方法的设计和分类器的选择，目标分类确认阶段可以称为“精检测阶段”，进一步排除虚警目标，相比粗检测阶段检测的精度更高，定位更准确。

1. 特征提取方法

传统的遥感图像目标检测算法更关注于特征工程中的特征提取和特征选择，多年来的实际工程表现也说明了特征的产生与选择对算法效果的提升远超于分类器的选择。

LBP (Local Binary Pattern, 局部二值模式) 特征有诸多的优点，在遥感图像船只目标检测中得到了广泛的应用。作为一种特征描述子，它具备方向不变、灰度不变、计算简单等优良的性能，是一种局部化的特征描述子。其计算思想是在局部化窗口，一般取 3×3 的窗口，将小窗口周围 8 个像素点的值与窗口中心点的像素值分别作差，大于 0 的标记为 1，小于 0 的标记为 0，通常会根据这种 8 位 0,1 组合的序列计算得到一个十进制数，如此便可以得到该窗口的 LBP 描述，该 LBP 值反映了局部窗口的纹理信息，其周围像素点和局部化窗口的大小也可以根据应用环境自由调整。基于该算法的思想，研究者提出了圆形的 LBP 特征描述、LBP 的旋转不变描述、LBP 的等价模式描述等来进一步提升特征描述的能力和减小计算的繁琐性。文章^{错误!未找到引用源。}利用金字塔对 LBP 描述进行池化采样，并且用多种尺度大小的采样窗口和不同的采样点数来丰富 LBP 对局部窗口信息的描述，提升了舰船目标检测的效果。

Gabor 特征能较好的适应船只检测、飞机检测，在这些任务中应用较为广泛。Gabor 特征的提取依赖于加窗短时傅里叶变换 (Windowed Short-time Fourier Transform)，它是把加窗函数确定为高斯函数时短时傅里叶变换的一种特殊情况，

傅里叶变换是将一个信号整体分解成正交频率信号的加权和,一般对具有平稳性、确定性的信号操作,这些信号在时域范围和空域范围的变换小,变化程度不剧烈,所以傅里叶变换的缺点在于缺乏信号在时空域的局部信息,对非平稳信号和时变信号的分析处理效果不好甚至不适用,不能给出频谱成分的时间和分布情况。依赖于傅里叶变换,并在频域上对信号加窗处理,可以得到信号的局部化信息,这样 Gabor 特征就可以描述方向上和尺度上的相关特征。Gabor 的核函数是正弦函数和高斯函数的调制,其滤波器在频率和方向上和人类视觉细胞的滤波方式相似,适合于纹理信息的提取,但另一方面因为 Gabor 的核函数不是正交函数,这将导致其提取的分量之间含有较多的冗余信息。在图像上提取 Gabor 信息时,需要先得到 Gabor 变换后的特征图,因为其特征维数较高,不适合直接处理,这时需要将特征图像分块处理,在每个子特征图像块上进行能量计算,得到某一方向和尺度上的能量信息特征向量。从本质上说, Gabor 特征提取方法也是一种卷积操作,所以卷积的效率一定意义上决定了 Gabor 特征提取的效率。文章^{错误!未找到引用源。}针对遥感图像中船只目标的检测问题,提出应用 Gabor 变换进行特征提取,对提取到的高维 Gabor 特征用核主成分分析(Kernel Principal Component Analysis, KPCA)进行特征的压缩降维,取得了适应度高、检测速度快的检测效果。文章^{错误!未找到引用源。}利用 Gabor 变化提取遥感图像飞机头、翼、尾三个部分的多尺度特征,并依据每部分对全局特征的贡献量来给三个部分分配不同的权重,用加权后得到的融合特征向量来进行飞机目标的检测,达到了高识别率,对遮挡目标有较好鲁棒性的检测效果。

HOG 特征主要在遥感图像飞机目标检测应用。HOG (Histogram of Oriented Gradient, 方向梯度直方图) 特征是一种统计图像局部化窗口方向梯度直方图信息来构成特征向量,在图像领域,梯度的概念是指像素值增减变化最为剧烈的方向,目标某点的边缘应该和其梯度方向保持垂直。HOG 与 SVM 的结合在检测行人的方向取得了巨大成功,尽管检测行人的算法不断改进,但是仍多基于该思路的原型。和其他特征描述子相比, HOG 特征有较多优点,对图像的光学变化和几何变化都有一定的稳定性,在归一化的条件下,目标小幅度的非刚性变换对检测结果的影响不剧烈。HOG 特征的提取过程大致包括,图像的预处理,对图像进行 Gamma 校正,计算像素梯度,构建“细胞”的梯度方向直方图,块内归一

化梯度方向直方图,得到 HOG 特征描述向量。文章^{错误!未找到引用源。}利用提取到的 HOG 特征结合级联 Adaboost 分类器作为飞机目标的粗检测方法,提取飞机目标的候选区域,接着利用卷积神经网络和 SVM 分类器实现飞机和背景的精准分类作为精检测方法,该方法在检测准确度和计算效率方面都得到了提升。

SIFT 多用于遥感图像飞机目标检测。SIFT(Scale-invariant feature transform, 尺度不变特征转换)特征描述的是局部空间尺度的极值关键点,这些关键点对图像的旋转缩放,亮度变化,仿射变换等表现出一定的鲁棒性,同时对噪声和视角变化等方面的影响也有很大的容忍度, SIFT 特征在某一方面能够较好的描述遮挡目标的特征,达到对遮挡目标较高的检测率。它包含的信息量丰富,不会因为图像大小或目标数量的多少而使得特征数量级骤缩,而且其计算简单快速,符合工程应用要求。SIFT 特征的提取过程包括,在金字塔尺度空间进行极值检测,去除低对比度和不稳定的边缘点,获取特征点的方向并得到特征区域,创建特征描述向量。文章^{错误!未找到引用源。}提出一种利用卷积神经网络提取到的特征和 SIFT 特征相结合的方式丰富被遮挡目标的特征描述,从而实现对遥感图像遮挡目标检测的高判别力,到达优秀的检测效果。

Hough 多用于遥感图像油罐目标的检测。Hough(霍夫)变换通过把图像空间的检测问题变换到参数空间(也叫 Hough 空间),它也是一种特征提取的方法,可以提取图像中特殊形状的特征信息,所以被广泛应用于图像目标检测领域中具有特定几何形状目标的检测识别。它利用在不同变换空间上点和线的对偶性,通过 Hough 变换可以把在图像空间中一种具有几何形状的线变换到参数空间中一个点,这样在原始图像空间对具有特定形状的目标检测就变成了对参数空间上点的检测。Hough 特征提取一般用于描述直线,圆,椭圆,弧线等特定形状的目标,早期的 Hough 变换主要用来检测图像空间的直线,后来随着研究者的改进扩展到多种形状的目标, Hough 特征描述可以容忍图像噪声的干扰。文章^{错误!未找到引用源。}提出了一种利用改进的 Hough 特征提取方法并结合 Canny 算子对边缘的提取来检测遥感图像中的油桶目标,不仅降低了传统 Hough 特征提取的计算复杂度还使得油罐目标检测的检测率高于80%。

在传统的特征提取方面上还有频谱特征提取、圆周频率滤波、视觉显著性特征、不变矩等特征提取方法,因为同一特征提取方法对不同目标有不同的表达能

力,所以经常需要根据目标具有的特点设计不同的特征提取方法来应对不同的检测任务。传统的基于特征提取的目标检测算法不足以满足海量遥感数据信息提取的需求,难以得到稳定性高、普适性好的统一多类目标检测算法。

2. 分类器的选择和设计

在传统基于机器学习的遥感图像目标检测算法上,分类器的选择相比于特征生成与选择显得不是那么重要,常见的分类器主要有, SVM (Support Vector Machine, 支持向量机)、ELM (Extreme Learning Machine, 极限学习机)、AdaBoost 方法等。

SVM 的目标在于在加入正则化项后求解经验风险最低的一个最优化超平面,使得每类的样本数据能够尽量远离这个超平面,得到的支撑向量是最靠近这个支撑平面的点组成的向量,当样本找不到一个效果足够好的超平面分开正负样本时,可以利用核函数将样本数据映射到其他特征空间,在其他特征空间数据可能变得线性可分。SVM 中常见的核映射有线性核 (Linear Kernel)、多项式核 (Polynomial Kernel)、径向基核 (RBF Kernel)、拉普拉斯核 (Laplacian Kernel),在此基础上通过研究者的努力,发展了 MKL (Multiple Kernel Learning, 多核学习) 的方式,利用训练阶段自动学习多个核函数加权的权重,来组合不同的核函数来实现分类器的作用,减少分类器的设计时间,提高分类准确度。

ELM 是一种类似于神经网络的分类器,它只具有前向传播过程,没有反向传播,而且其隐藏层节点的权重和偏置是人为给定或者随机初始化的,在分类器的使用过程中这些参数不再调整,在学习过程中只学习输出层的权重和偏置,这种学习是通过解方程的方式得到,所以 ELM 具有较好的泛化能力,计算效率特别高。

AdaBoost 方法是集成学习 (Ensemble Method) 或元学习 (Meta Learning) 方法的一种,其思想在于弱弱联合,以弱胜强,对于同一个训练数据集,训练得到多个性能较低的分类器通过平均投票或者加权投票的方式得到最后的分类结果,它是一种迭代形式的算法,可以得到较好的分类效果。

除此之外,也有其他分类器用于遥感图像目标检测技术中,如非监督聚类 (unsupervised clustering)、决策树 (Decision Trees)、随机森林 (Random Forest) 等。

1.1.2 基于深度学习的遥感图像目标检测算法

基于机器学习的传统遥感图像目标检测算法的检测效果几乎完全依赖于大量且有效的特征提取过程，而在光学遥感图像的背景下，遥感图像的拍摄条件和目标所处环境复杂多变，天气干扰严重，传统算法常常难以到达工程化对海量遥感数据信息快速提取的要求。

近年来随着深度学习技术的迅猛发展，遥感图像目标检测算法的性能得到了阶跃式提升。基于深度学习的遥感图像目标检测技术不需要手工进行特征的提取，可以自动地、自适应地学习到目标所具有的特征，整个模型在学习过程中不需要人工干预，是一种数据驱动的模式，在训练数据充足且分布合理、模型参数足够的情况下，对复杂多变条件下的数据仍有很好的泛化能力，对不同任务之间的整合自适应性高；有着更为强大的特征表达方式与能力，通过构建较大深度和广度的学习结构，深度学习提取的特征可以由浅层到深层不断抽象和强化，不断挖掘数据中的潜在特征和关联。深度网络结构相对浅层网络结构的神经网络和传统的特征提取方法而言可以实现更加复杂的函数映射并具有更好的模型泛化能力，从而可以实现更加精确的分类和回归效应；同时具有对增量数据进行增强学习的特性，基于深度学习方法的学习过程具有持续性和动态性的特点，可以在已有网络模型基础上根据新的样本数据进一步对网络进行强化训练，使得其性能得到进一步提升，对训练数据的分布拟合的更好。

基于深度学习的方法，在遥感图像目标检测等计算机视觉领域的问题上表现出的效果远超传统算法。基于深度学习的遥感图像目标检测算法大多借鉴自然图像目标检测的算法基础，但是根据遥感图像自身存在成像方式不同、宽尺度范围、分辨率多样、复杂环境干扰等问题，在目标检测上也有区别于自然图像目标检测的特殊之处。从大体上来讲基于深度学习的遥感图像目标检测算法主要依托于自然图像的算法基础分为基于候选区域生成的两步检测算法和基于一体化卷积的一步检测算法两大阵营。其模型的原型主要有：基于候选区域生成的两步检测代表算法 R-CNN（2014）、SPP-Net（2014）、Fast R-CNN（2015）、Faster R-CNN（RPN，2015）、R-FCN（2016）、Feature Pyramid Net（FPN，2017）、Mask R-CNN（2017）等；基于一体化卷积的一步检测代表算法 OverFeat（2014）、YOLO v1（2016）、SSD（2016）、YOLO v2（2017）、RetinaNet（2017）、YOLO v3（2018）、

M2Det(2019)等; 两步法和一步法结合的算法如 RON(2017), RefineDet(2018)等, 下面对其进行简要回顾。

1. 基于候选区域生成的两步检测代表算法

R-CNN(2014, Regions with CNN)是在深度学习目标检测方面可谓是打开了新世界的大梦, 掀起了目标检测领域急速发展的浪潮。R-CNN 由加州大学伯克利分校的 Ross B. Girshick 等人提出, R-CNN 的图像目标检测方法利用选择性搜索(Selective Search)的模式来提取目标候选区域, 相比于传统方法, 提取的候选区域遗漏较少, 原文中提取了近 2000 个, 然后把来候选区域和目标的 Ground Truth 缩放到网络所需的图像大小, 并送入到 CNN 特征提取器中进行特征提取, 得到了固定长度的特征描述向量, 把得到的特征向量送入到支持向量机中进行分类预测, 最后利用全连接层对目标的位置进行回归操作。R-CNN 在目标检测领域有着里程碑意义, 其检测效果相较于传统目标检测算法提升50%左右, 但是其缺点十分明显, 特征重复计算、训练过程十分繁琐, 步骤过多、检测效率太低、图片经缩放影响检测效果等。文章[错误!未找到引用源。](#)提出一种 AdaBoost 方法与 R-CNN 方法结合的算法, 用一种由粗检测到精检测的模式来进行遥感图像飞机目标检测, 得到了较好的效果。

SPP-Net(2014, spatial pyramid pooling layer)解决了 R-CNN 网络重复计算图像特征的问题, 提高了算法的检测效率。它是由微软研究院的 Kaiming He 等人提出的, 利用空间金字塔池化的操作避免了对原始输入图像的裁剪、缩放等操作, 从而进一步不需要像 R-CNN 在每一张子图像上进行 CNN 的特征提取, 大大提升了检测算法的速度, 从实质上来讲 SPP-Net 通过引入空间金字塔池化层解决了最后一卷积层输出的特征图大小不统一的问题, 而最后的全连接层依托于固定长度的特征输入, 所以 SPP-Net 的提出将 R-CNN 的多次重复卷积操作变成一次, 相比 R-CNN 的检测效率有 30 多倍的提升。

Fast R-CNN(2015)借鉴了 SPP-Net 的内容, 提出了 ROI pooling layer, 把候选区域通过 CNN 卷积操作得到的不同大小特征图统一成固定尺度的大小。该算法由微软研究院的 Ross B. Girshick 等人提出, 特别的该网络还应用了多任务损失的想法, 在网络特征提取器后面增加了分类的全连接层和回归的全连接层, 直接输出分类结果和回归结果, 不在需要输出 CNN 提取到的特征用 SVM 进行

训练分类，大大提升了检测速度，较 R-CNN 提升了 140 倍左右。

Faster R-CNN (2015) 真正意义上实现了全网络的两步法检测结构，在网络结构中设计了一个 RPN (Region Proposal Netw) 来进行目标候选区域的生成。该算法由微软研究院的 Shaoqing Ren 等人提出，将整个网络主要分成了两个部分，首先由 RPN 网络生成许多锚点 (Anchor Box)，通过 RPN 进行前景背景的二分类，到此阶段 RPN 便生成了目标的候选区域，再通过后续的分类和回归的多任务损失函数进行目标分类和位置修正。这使得目标检测在一个网络结构中完成，实现了端到端的检测效果。文章**错误!未找到引用源**。提出了一种改进的 Faster-RCNN 的遥感图像飞机目标检测算法，通过数据增广的模式扩充原始数据集，进行难样本均衡化处理，并将特征提取的普通卷积层换成深度残差层，得到了很好的性能提升。文章**错误!未找到引用源**。提出一种在 Faster R-CNN 算法的基础上把图像采样过程与图像特征金字塔网络相联合的方法，改进了原始 Faster-RCNN 在舰船小目标检测上表现较差的现象。

R-FCN (2016) 在 Faster-RCNN 算法的基础上提出了改进策略，利用一种位置敏感得分图 (Position-positive score maps) 来解决目标检测平移变化和图像分类平移不变之间的矛盾。该算法由微软研究院的 Jifeng Dai 等人提出，位置敏感得分图把提取到的目标候选区域分成小格子做采样，然后通过投票的方法来对整个的候选区域的类别进行预测。文章**错误!未找到引用源**。提出了以 R-FCN 为原型，通过利用混合尺度卷积核处理 ResNet 的特征提取网络，达到抑制舰船的干扰噪声，取得了良好的检测效果。

FPN (2017) 的贡献在于提出利用不同特征层进行目标的检测。该算法由脸谱网的 Tsung-Yi Lin 等人提出网络在进行卷积操作时，特征图的大小也会随之改变，整体上来讲特征图的尺寸随着网络层数的加深，高层特征信息更为丰富，特征图的尺寸大小更小，FPN 算法的思想在于利用不同大小的特征图构成特征金字塔。文章**错误!未找到引用源**。通过在遥感飞机数据集上测试 FPN 的效果并于 Faster R-CNN 算法进行比较，得出了结合不同特征图的特征对于遥感图像目标检测来讲，有着不同的检测效果。

Mask R-CNN (2017) 相比于 Faster-RCNN，改进了通过非线性插值的方法替代直接在特征图采样过程中的取整操作，并且将网络进行扩展，增加了实例分

割的输出。该算法由脸谱网的 Kaiming He 等人提出，在原有 Faster-RCNN 的基础上添加了一个 Mask 生成的分支，这个分支同类别预测和边框回归并行预测，互不干扰，使得算法不仅可以做目标检测还可以做实例分割任务。文章**错误!未找到引用源**。提出了一种基于 Mask R-CNN 的舰船目标检测框架，通过网络输出的目标定位框和其实例分割信息来进一步提取舰船目标位置，可以解决由于目标排列过密引发的检测问题。

2. 基于一体化卷积的一步检测代表算法

OverFeat (2014) 利用卷积神经网络特征提取的优势，把得到的目标候选区域框送进神经网络和回归的模型中进行目标的分类和定位。该算法由纽约大学的 Yann Lecun 等人提出，用一个网络完成了多项计算机视觉任务，充分发挥了卷积神经网络的优势。

YOLO 系列 (v1, 2016、v2, 2017、v3, 2018)，YOLO v1 去除了目标候选区域的提取过程，在一条主线网络中实现特征提取、目标分类、边框回归这三个工作，因此网络整体更加简洁，是一种端到端的网络。由华盛顿大学的 Joseph Redmon 等人提出，检测效率比 Faster R-CNN 提高了 10 倍左右，但是由于采用的是固定大小网格预测的方式，存在定位不准，对小目标检测效果较差。YOLO v2 是 v1 的升级版本，引入了有针对性的锚点 (Anchor box)，残差网络来进行特征提取，卷积层使用批量归一化 (Batch Normalization) 等提升检测效果。YOLO v3 引入了结合 FPN 网络结构思想的 Darknet53 的差网络，让 YOLO v3 在保持检测速度极快时，仍能够达到前沿算法的检测精度。文章**错误!未找到引用源**。提出了一种借鉴 YOLO v3 和 Densenet 网络结构的方法，该方法用两个密集网络块去替换 YOLO v3 中的两个残差网络块从而产生一种新的网络，实验结果显示，该算法相比于原始算法有较大提升。

SSD (2016) 对 YOLO v1 定位精度的问题进行了改进，不仅取得的较高的检测精度，能够与两步法相匹敌，而且在速度上方面也 and YOLO 旗鼓相当。该算法由北卡大学的 Wei Liu 等人提出，通过引入预设框 (Prior Box)，并在不同尺寸的特征图上做预测，使用多尺度的特征层信息进行分类。基于 SSD 产生了很多改进算法，如 R-SSD、DSSD、DSOD、FSSD、ESSD 等。文章**错误!未找到引用源**。提出一种基于 SSD 网络架构，将 VGG-16 用密集连接网络代替，并利用

密集连接块之间的特征构建特征图金字塔的方法，在遥感图像飞机目标检测上相比于 Faster-RCNN 和 R-FCN 取得了较大提升。

RetinaNet(2017)通过引入 Focal loss 解决一步法中正负样本不均衡的问题，一步法中样本不均衡的现象会导致大量负样本数据产生的损失函数作用远超过正样本。该算法由脸谱网的 Tsung-Yi Lin 等人提出。

M2Det（2019）通过提出 MLFPN（Multi-Level Feature Pyramid Net）来解决目标检测的难点之一尺度变化问题。在主线网络提取得到特征后，通过 FFM（特征融合块）和 TUM（细化 U 型网络块）来得到更具代表性的多级多尺度特征，然后利用 SFAM（尺度特征融合块）对 TMUs 提取得到的特征进行整合，然后进行类别预测和包围框回归。该算法由来自北京大学和阿里达摩院的 Qijie Zhao 等人提出。