

基于改进残差注意力网络的 SAR 图像目标识别

史宝岱*, 张秦, 李瑶, 李宇环

空军工程大学研究生院, 陕西 西安 710051

摘要 针对合成孔径雷达(SAR)图像噪声较高,导致目标识别率较低的问题,选取 MSTAR 数据作为样本集,首先分析了对网络添加注意力机制的必要性,接着在残差注意力网络中引入残差收缩块,从识别率和参数数量的角度进行实验分析。改进残差注意力网络的第一阶段和输出阶段,得到模型 S,最终模型 S 的识别率达 99.6% 的同时,参数量减少了近 1/2。为测试改进模型的鲁棒性,对图像进行了遮挡和加噪处理,结果显示,在图像被遮挡和有椒盐噪声情况下,模型 S 具有较强的鲁棒性。

关键词 图像处理; SAR 图像; 目标识别; 残差收缩; 鲁棒性

中图分类号 O436

文献标志码 A

doi: 10.3788/LOP202158.0810008

SAR Image Target Recognition Based on Improved Residual Attention Network

Shi Baodai*, Zhang Qin, Li Yao, Li Yuhuan

College of Graduate, Air Force Engineering University, Xi'an, Shaanxi 710051, China

Abstract In this study, MSTAR data were selected as a sample set to solve the problem of the high noise of the synthetic aperture radar (SAR) images leading to a low target recognition rate. First, the necessity of adding an attention mechanism to the network was analyzed. Subsequently, the residual shrinkage piece was introduced in the residual attention network. An experimental analysis was performed from the perspective of the recognition rate and number of parameters. Model S was obtained by improving the first stage and the output stage of the residual attention network. Consequently, the recognition rate of model S was found to be 99.6%, and the number of parameters was reduced by nearly 1/2. Image occlusion and noise processing were conducted to test the robustness of model S. Results show that model S has a strong robustness under the conditions of image occlusion, salt and pepper noise.

Key words image processing; synthetic aperture radar image; target recognition; residual contraction; robustness

OCIS codes 100.2960; 110.4280; 150.1135

1 引言

合成孔径雷达(SAR)采用自身发射脉冲信号的回波来成像,因而突破了光学成像的限制条件,受复杂天气的影响较小,在军事和民用领域应用广泛^[1]。随着时代发展,传统的依靠人工解译 SAR 图像难以满足需求,人工智能的出现,人们逐渐将精力转移到依靠计算机的 SAR 图像目标自动识别上来。

Hinton 等^[2]提出了深度学习的概念,但深度学习一直未在图像识别领域得到成功应用,直到 2012 年, Krizhevsky 等^[3]凭借深度卷积神经网络 15.3% 的错误率刷新了记录,向世人展示了深度学习在图像识别领域的优秀性能。深度学习开辟了机器学习新领域,其核心是通过数据驱动,从众多数据中自动提取多种多角度特征,这种方式使得图像处理更加高效。常见的深度学习方法有深度置信网

收稿日期: 2020-08-24; 修回日期: 2020-08-31; 录用日期: 2020-09-10

基金项目: 国家自然科学基金面上项目(61971438)

* E-mail: 1908509679@qq.com

络^[4]、卷积神经网络 (CNN)^[5]、自动编码器 (DAE)^[6]等。作为一种前馈型神经网络,简单 CNN 已无法满足需要,而对 CNN 进行调整改进时一般从深度和宽度上着手。在深度上,层数的增加会在一定程度上提高网络的性能,基于此,提出了 LeNet-5、AlexNet、VGG^[7]、残差网络 (ResNet)^[8]等。其中 Szegedy 等^[9]提出的 ResNet 在 ILSVRC2015 竞赛中解决了层数加深之后带来的梯度消失和爆炸问题,使得网络层数可以进一步深化。在宽度上,经历了 GoogLeNet^[9]、InceptionV2^[10]再到 Inception ResNetV2^[11]等,网络加宽可以大幅提高特定数据集的精度,但计算量会平方增加,而且模型的泛化能力也会减弱。

基于此,为了增强从输入到输出的映射能力,针对特征图分配时的权重不够准确、冗余参数较多的问题,在神经网络中引入了注意力机制,核心思想是分配给有效特征更多的权重。如 2015 年空间转换网络 (STN) 模型^[12]用空间转换器模块提取出空间域的有效信息;压缩和激励网络 (SENet)^[13]利用通道注意力增加有效特征通道的权重,从而达到更好的训练效果;2017 年残差注意力网络^[14]结合了 Inception 和 ResNeXt^[15],形成了混合域的注意力机制;2018 年非局部神经网络^[16]通过增加 CNN 中远离局部感受野位置特征的权重来提高模型训练效果;2018 年卷积注意力模型 (CBAM)^[17]关注了通道注意力的同时关注了空间注意力;2019 年,双重注意力网络 (DAN)^[18]融合了 CBAM 和非局部神经网络,通过减少手工设计 pooling 等环节提高了训练效率。

近年来,深度学习算法在计算机视觉领域得到广泛应用,但在 SAR 图像目标识别领域还不够成熟,一方面由于深度学习算法网络层数较高,处理效率较低;另一方面由于 SAR 图像噪声较高。这就要求设计网络时要有针对性,除了精度和参数数量的要求外,还要求有较强的鲁棒性,增加了算法设计的难度。基于此,在 MSTAR 数据集中,首先从网络深度和宽度角度出发,利用 ResNet50、ResNet101 及 Inception ResNetV2 说明简单堆叠的残差网络难以满足对识别率和参数数量的要求,本文在残差注意力网络中引入残差收缩网络,识别率达到了 99.6%,参数数量减少了 1/2。接着进行了鲁棒性分析,对数据集图像进行遮挡和加噪处理,实验证明,所提方法对两种情况都有较强的鲁棒性。

2 MSTAR 数据集

采用美国空军研究室和国防高级研究所合作研究公布的 MSTAR 数据集^[19],由于军事目标的 SAR 图像较难获取,目前国内研究 SAR 自动识别时大部分采用该数据。该数据是 1996 年采用 $0.3\text{ m} \times 0.3\text{ m}$ 高分辨率聚束式 SAR 采集前苏联军事车辆形成系统的数据库,再经过预处理裁减为尺寸为 128×128 的图像,且目标均处于图像中心。其主要包含 BTR-70、BMP2 和 T72 三类目标图像,其中每种类别的目标还有不同型号,以俯仰角为 17° 的数据作训练集, 15° 的数据作测试集,再加上 MSTAR PUBLIC TARGETS CHIPS 中的图像,共含 10 类目标,最终训练集有 2747 张,测试集有 2426 张。图 1 为 10 类目标的光学图像和对应的

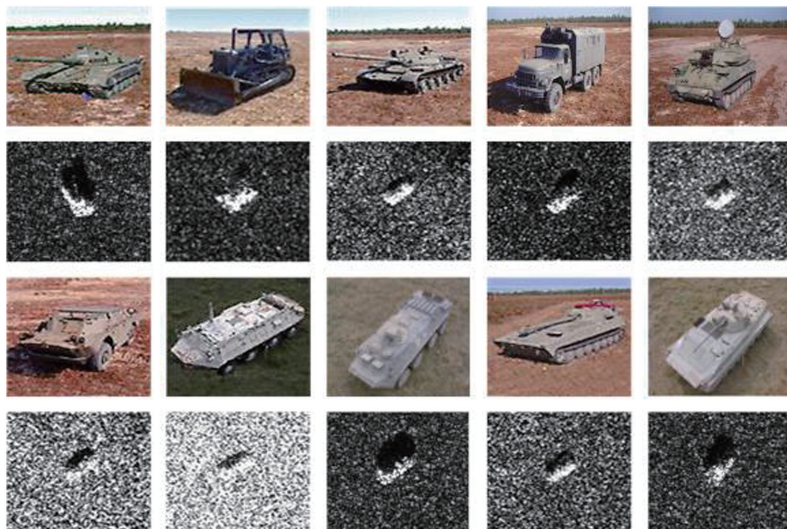


图 1 10 类目标光学图像和对应的 SAR 图像

Fig. 1 10 types of target optical images and corresponding SAR images

SAR 图像,表 1 为实验数据中训练样本和测试样本的型号和对应的数据量。

表 1 数据类别和样本数量

Table 1 Data category and number of samples

| Category | Number of samples in training set | Number of samples in test set |
|----------|--------------------------------------|----------------------------------|
| 2S1 | 299 | 274 |
| BMP2 | 232 | 196 |
| BRDM-2 | 298 | 274 |
| BTR-60 | 256 | 195 |
| BTR-70 | 233 | 196 |
| D7 | 299 | 274 |
| T62 | 299 | 273 |
| T72 | 232 | 196 |
| ZIL131 | 299 | 274 |
| ZSU-23_4 | 299 | 274 |

3 残差网络

残差网络的出现给神经网络的发展带来了新的契机。在这之前,神经网络层数除了 VGG 达到的 19 层和 GoogLeNet 达到的 22 层,再无网络可以超越,因为层数加深到一定程度之后出现的过拟合问题还可以通过增加数据量或者采用 Dropout 正则化的方法来减缓,但是无论采用 SGD、AdaGrad^[20],还是 Adam^[21]都无法达到最好的收敛效果,某种意义上也限制了神经网络的发展。根据网络的深浅,残差模块有两种实现形式,50 层以下的网络采用 2 层

表 2 不同算法的识别率

Table 2 Recognition rates of different algorithms

| Algorithm | Number of parameters /10 ⁷ | Recognition rate in training set | Recognition rate in test set | Time /s |
|--------------------|---------------------------------------|----------------------------------|------------------------------|---------|
| ResNet50 | 2.5 | 99.4 | 97.9 | 1760 |
| ResNet101 | 4.5 | 90.7 | 90.9 | 2426 |
| Inception ResNetV2 | 5.5 | 88.1 | 94.5 | 7967 |

有针对性地设计网络时无法预估最优的网络层数,而网络层数过深后会出现冗余层,拟合效果会大打折扣。而如果单纯依靠增加深度和宽度就可以从识别精度和实时性上取得好效果,也就没了改进算法和引进新内容的必要。基于此,针对 MSTAR 数据集,对比了 3 种不同网络模型,ResNet50 和 ResNet101 均由残差块简单堆叠而成,后者在 conv4_x 多出 17 个残差块,采用 ResNet50 进行实

验时,训练集精确度达到了 99.4%,测试集精确度为 97.9%,而样本数量是足够的,可以看出,这种过拟合来自于图像噪声。测试集中 ResNet50 的识别率不算高,2016 年提出的 A-ConvNets 模型^[22]对 MSTAR 数据集的识别率就达到了 99.13%。而层数加深到 101 层之后,测试集中算法的精度为 90.9%,不升反降,精度曲线如图 2 所示,识别率总体呈先上升再下降的趋势,可以看出 101 层已经超出

结构单元,50 层及以上的网络采用 3 层结构单元。3 层结构单元先后用一个 1×1 卷积进行降维和升维,大幅减少参数数量的同时也减弱了层数加深带来的参数量爆炸风险。

针对 SAR 图像目标识别准确率不高的问题,本实验选用 ResNet50、ResNet101 和 Inception ResNetV2 处理 MSTAR 数据集,其中 ResNet50 和 ResNet101 用来纵向研究残差网络的层数对识别率的影响,再加入 Inception ResNetV2 横向比对网络加宽之后对识别率的影响,3 种网络的 Block 均采用 Bottleneck 结构。训练时加载对应的预训练模型,全连接层设计两个 Dense 层,第一个作用是降维,第二个添加 Softmax 函数来分类,两个 Dense 中间设置 Dropout 正则化,参数为 0.5。数据集一共包含 10 类目标,实验结果以 10 类目标的平均识别率为衡量标准。

训练细节有:1)图像预处理,为防止数据较少带来的过拟合问题,把数据尺寸由 128×128 随机裁减为 88×88,样本量扩大了 (128-88+1)×(128-88+1)=1681 倍,每类目标选择 2700 张图片,为方便后续神经网络的处理,对图像进行归一化和错切变化处理;2)采用 Tensorflow 开源框架,收敛方法采用 SGD 梯度下降法,动量设为 0.9,初始学习率设为 0.003,识别率不再上升时,学习率以乘 0.5 的方式减小,min_lr 设为 0.000005,batch_size 设为 16,迭代周期设为 50;3)实验平台为阿里云第四代轻量级服务器,处理器型号为 Intel Xeon(Skylake) Platinum 8163,主频 2.5GHZ,GPU 为 NVIDIA Tesla T4。实验结果如表 2 所示。

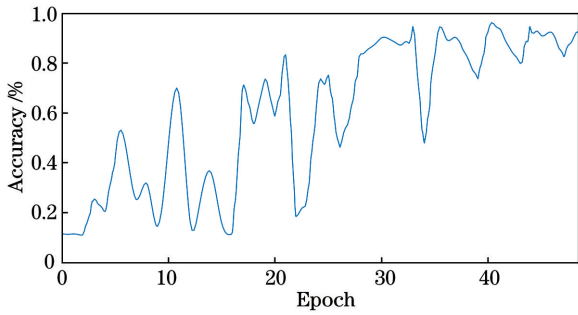


图 2 ResNet101 精度曲线
Fig. 2 ResNet101 precision curve

最优层数。加宽的 Inception ResNetV2 在此数据集上效果不如人意,识别率仅为 94.5%,而且实时性较差。从而得出,对于残差网络而言,简单堆叠层数已经无法满足 SAR 图像对识别率的要求,想在降低参数量的同时提升识别率,就必须添加注意力机制。

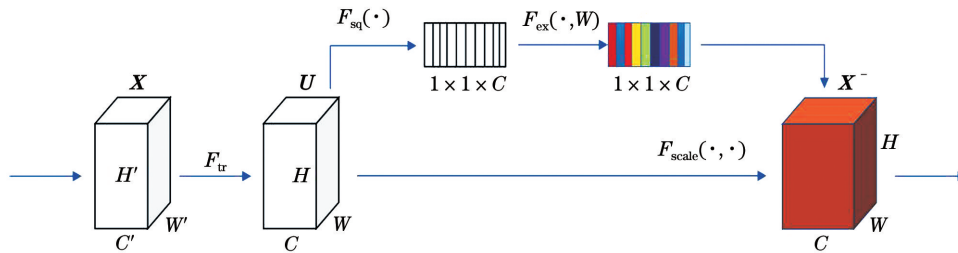


图 3 SENet 结构
Fig. 3 SENet structure

图 3 中,输入 X 的通道数为 C' ,处理后通道数变为 C ,然后并联分支用 Squeeze(特征压缩)和 Excitation(生成权重)操作实现加权权重的自动获取,这意味着可以给每一个特征图赋予相对应的权重,在增加有用通道权重的同时,削弱无关特征对关键信息流动的负面影响。Squeeze 通过全局平均池化操作把特征图压缩为 $1 \times 1 \times C$ 的实数序列,在一定程度上,此实数具有全局感受野,接着 Excitation 操作利用 w 参数(神经网络中的权值矩阵)生成每个特征图对应的权重,最后利用乘法将重新标定的权重加权附加到之前的特征上,这样就实现了在通道领域的注意力。全局平均池化操作的公式为

$$Z_c = F_{sq}(u_c) = \frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H u_c(i, j), \quad (1)$$

式中: Z_c 为压缩操作之后的权值参数; $F_{sq}(\cdot)$ 为特征压缩操作; u_c 为 U 中第 c 个二维矩阵, U 为多个局部特征图的集合; H 为特征矩阵的高度; W 为特征矩阵的宽度。残差收缩网络主要针对的是噪声较高问题,实现方式是将软阈值设置模块嵌入到残差网络中。软阈值函数可以将低于某个门限的特征直

针对 SAR 图像目标识别率不高,且一般网络无法有效剔除噪声的情况,提出把残差收缩网络^[23]引入到残差注意力网络中,增强抗噪能力。

4 残差收缩块

注意力分为两种,包括软性注意力和硬性注意力。硬性注意力侧重于图像上离散的点延伸出的注意力,常被用于注重不同部分差异的细粒度图像分类中,Zhang 等^[24]于 2014 年提出了一种基于对象检测区域卷积网络的部位区域卷积网络(PB R-CNN)。软性注意力更侧重通道和空间等确定性的注意力,其突出特点是可以直接通过反馈机制得到权重,这也是神经网络中常用软性注意力的原因。其中 SENet 通过并联分支实现了对通道的注意,结构如图 3 所示。

接置 0,其他特征也向 0 的方向进行调整,从而实现低于门限的噪声的弱化,所以阈值设置的大小直接决定了去噪效果的优劣。借用 SENet 中通道注意力的实现方式,对阈值进行自适应调整,操作如图 4 右边区域所示,图中 FC 为全连接层,BN 为批量归一化。首先进行全局均值池化操作得到特征 M ,然后将 Sigmoid 函数作为最后一层,将输出调整

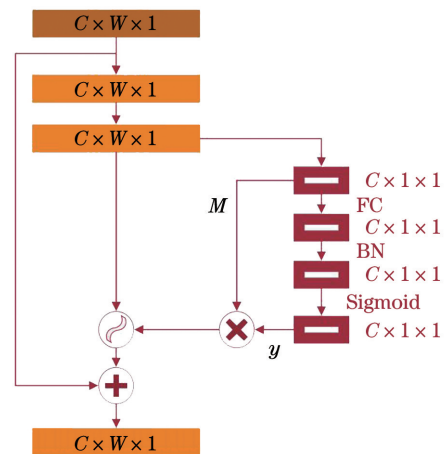


图 4 残差收缩块

Fig. 4 Residual shrinkage block

到 0 和 1 之间,并记为 y ,则阈值为 $y \times M$,阈值设置完毕。为了增强残差注意力网络对噪声较强的 SAR 图像的识别效果,将残差收缩网络引入到残差注意力网络中。

5 改进混合注意力的残差网络

5.1 结构分析

残差注意力网络包含 3 个注意力模块,每个注意力模块含 2 个分支:上路的主干分支和下路的掩模分支。主干分支包含若干残差块,负责提取主要特征,输出为 $T(x)$ 。掩模分支采用 bottom-up、top-down 结构,负责附加混合域的注意力。首先由降采样获取全局特征,利用线性插值或反卷积将特

征图放大到和 $T(x)$ 一样的尺寸,再用 Sigmoid 函数对权重进行归一化,得到特征对应的权重输出为 $M(x)$,则输出 $H(x)$ 的表达式为

$$H(x) = [1 + M(x)] \times T(x). \quad (2)$$

这样既实现了注意力权重的附加,又不会使特征值随着层数加深而减小,导致网络退化。为了提取不同层次的特征,原网络在 3 个阶段中的 bottom-up、top-down 间添加数量依次为 2,1,0 的 skip connection。本文只介绍残差注意力网络的核心内容,若想进一步了解,可以参考文献[15]。

为研究残差注意力网络在 MSTAR 数据集上的性能,环境设置和训练细节同上,识别结果如表 3 所示。

表 3 残差注意力网络识别率

Table 3 Recognition rate of original residual attention network

| Number of parameters / 10^7 | Recognition rate in training set / % | Recognition rate in test set / % | Training time / s |
|-------------------------------|--------------------------------------|----------------------------------|-------------------|
| 3.2 | 99.5 | 98.9 | 4023 |

实验结果表明,该模型识别率达到了 98.9%,比 ResNet50 和 ResNet101 的精度都要高,尤其在残差网络结构层数低于 101 的情况下,软分支附加的注意力确实起到了增加有效特征权重的效果。但测试集中的精度明显低于训练集,依然产生了过拟合,在噪声环境下该模型性能较差,而且采用残差块,导致整个网络参数量巨大,实时性较差,难以真正得到应用,尤其在 SAR 图像带有大量噪声的情况下。

残差注意力网络实现了混合域的残差注意力,可以增大有效特征的权重,但在强噪声下,从输入到输出的映射能力还不够强,且参数量巨大。基于此,从 3 个方面对模型进行改进。

1) 主干分支采用去噪声能力更强的残差收缩块,这样可以避免只依靠软分支添加注意力带来的误差,尤其在软分支提取特征表征能力不强的情况下。经过主干分支过滤后的特征再附加软分支提供的混合域注意力,提取的特征权重更加准确。

2) 为了进一步增强软分支混合域注意力的效果,依然用残差收缩网络替换掉原有的残差网络,在用软阈值剔除噪声的同时,用混合注意力机制进行特征权重的二次分配。

3) 为了合理地减少参数量,省略阶段一和阶段二之间、阶段二和阶段三之间的残差块,再把最后输出阶段的残差块数量由 4 减为 2,模型结构如图 5 所示。

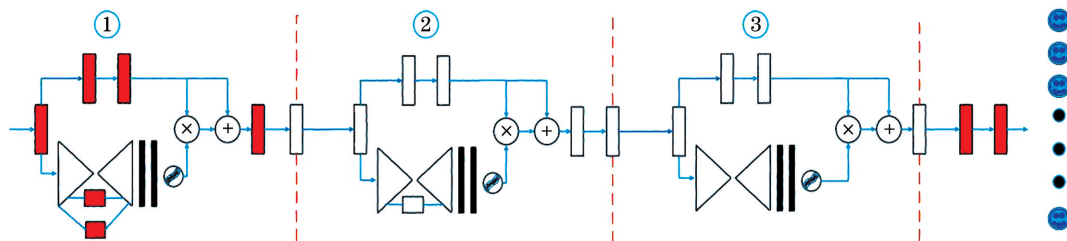


图 5 模型 S 结构图

Fig. 5 Structure diagram of model S

5.2 改进模块主干网络实时性分析

改进之后网络复杂度会有所变化,主要体现在 2 个方面:1) 主干分支和软分支中残差块替换为残差收缩块后,软阈值会把低于门限的特征参数置 0,大幅降低参数量;2) 软阈值设置模块多出的全连接

层会增加参数量。为了找出最佳的改进方式,并提高网络的实时性,主干网络收缩块数量设为 2,对改进阶段的数量进行仿真实验。依然针对 MSTAR 数据集,训练细节同上,阶段数量依次设为 3,2,1,0,改进阶数为 1 代表只改第一阶段,2 代表改进一、

二阶段,3 同理。当改进阶段数为 0 时,即为残差注意力网络,结果如表 4 所示。

表 4 不同改进阶段数下的实验结果

Table 4 Experimental results of different improvement stages

| Stage | Number of parameters / 10^7 | Recognition rate in training set /% | Recognition rate in test set /% | Training time /s |
|-------|-------------------------------|-------------------------------------|---------------------------------|------------------|
| 3 | 1.1 | 99.5 | 98.9 | |
| 2 | 1.2 | 99.5 | 99.4 | |
| 1 | 1.5 | 99.7 | 99.6 | 2850 |
| 0 | 3.2 | 99.5 | 98.9 | 4023 |

由表 4 可以看出:参数量随着改进阶段数的增加而减少,阶段数为 1 时,参数量减少了 1/2,测试集识别率达到了 99.6%;改进阶段数为 2,3 时,测试集识别率反而降低。可见当置 0 无关特征时,也会使得特征表征能力减弱,从而降低映射能力。因此选择改进阶段数为 1,则模型称为模型 S,结构如图 5 所示。阶段一、二、三结构基本相似,第一阶段的具体结构如图 6 所示,R-S block 代表残差收缩

块,收缩块中主干网络均采用 3×3 卷积,第二、三阶段则依次减少软分支左侧的 R-S block。

针对 MSTAR 数据集,模型 S 和不同算法的识别结果如表 5 所示,可以看出,模型 S 的识别结果是最高的。原因主要有两点,第一点是模型 S 采用软分支,附加了混合域的注意力,在消耗有限计算资源的情况下,增强了输入到输出的映射;第二点是主干网络和软分支的残差块替换成去噪能力较强的残差收缩块,比不添加去噪模块的一般注意力神经网络更具有针对性。

表 5 不同模型的识别结果

Table 5 Recognition results of different models

| Model | Average recognition rate /% |
|------------------------------------|-----------------------------|
| Model S | 99.6 |
| CMNet model ^[25] | 99.3 |
| Faster R-CNN model ^[26] | 99.1 |
| A-ConvNets model | 98.1 |
| SVM model ^[27] | 90.0 |

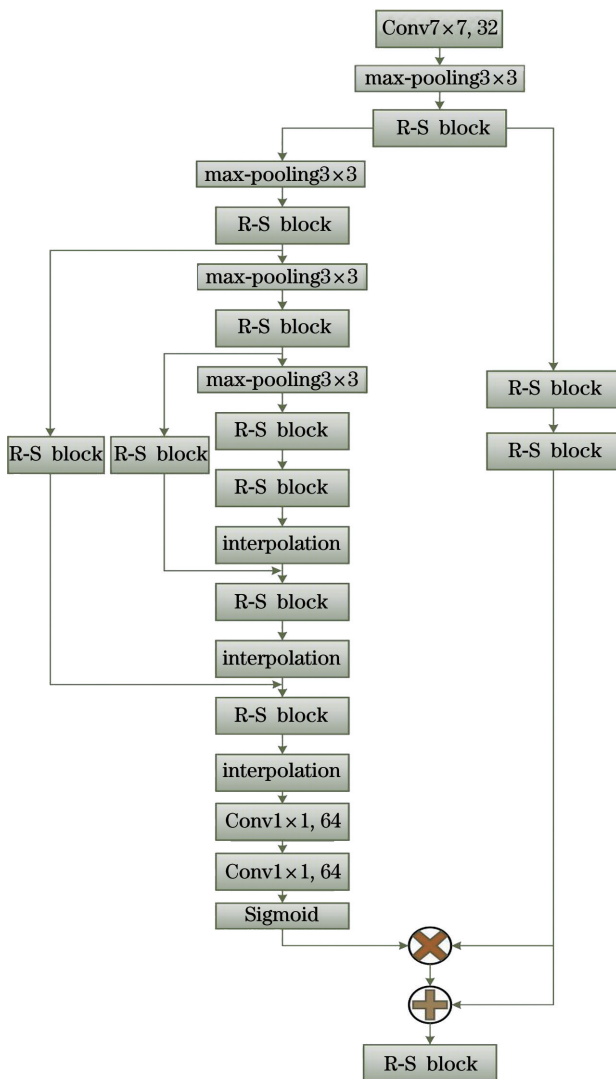


图 6 第一阶段具体结构

Fig. 6 Concrete structure of the first stage

6 鲁棒性分析

基于上述实验,模型 S 在识别精度和实时性上都要优于原网络,而且识别精度在已有算法中最高。但衡量一个算法的好坏,还要看它在恶劣条件下的表现。受外部环境和成像设备的影响,数字图像在形成过程中很容易产生噪声,所以去除这些噪声就有现实意义。本文测试了模型 S 在被遮挡和椒盐噪声环境下的性能,选取 BTR_60、BRDM_2 和 BMP2 共 3 类目标。

1) 任意置 0 像素点,以此模拟目标遮挡情况。置 0 像素点比例设置为 5%,15%,20%,30%,35%,为使结果具有稳定性,多次测量取平均值。加噪声之后的 3 类目标图像效果如图 7 所示。

目标和噪声的灰度值较高,随机添加灰度值为 0 的像素点之后,目标和噪声都可能被置 0,降低噪

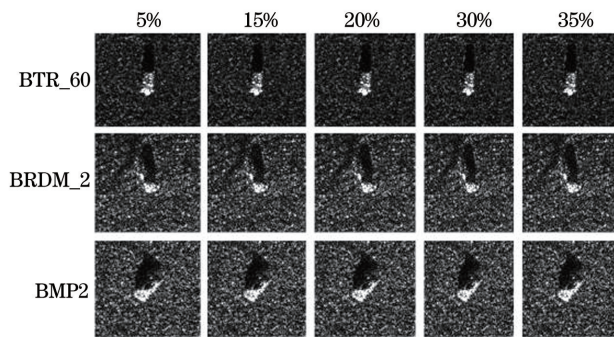


图 7 遮挡效果图

Fig. 7 Occlusion effect pictures

声的同时也降低了目标特性,理论上无法预测识别率的变化,识别结果如表 6 所示。最终 5 类不同噪声比例下的模型平均识别率均为 99%,噪声比例对识别率几乎没有影响,表明在目标被遮挡情况下,模型 S 有较强的鲁棒性。

表 6 遮挡识别结果

Table 6 Occlusion recognition results

| Noise ratio / % | 5 | 15 | 20 | 30 | 35 |
|----------------------|----|----|----|----|----|
| Recognition rate / % | 99 | 99 | 99 | 99 | 99 |

2) 椒盐噪声

一旦获取和信道传输过程受到干扰,图像上就会出现黑或白色的随机像素点,将直接影响特征提取和图像识别的效果,称之为椒盐噪声。因其随机离散的特点,又称为脉冲噪声,常见于数字图像处理中。但这种噪声的产生机制意味着其比例不会太高,选择产生比率分别为 5%,10%,15%,噪声图像如图 8 所示。

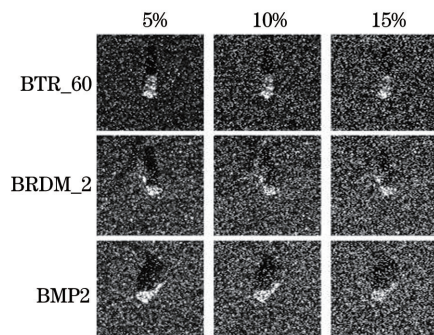


图 8 椒盐噪声效果图

Fig. 8 Pictures of salt and pepper noise

实验设置同上,多次测量求平均值,实验结果如表 7 所示。可以看出:添加椒盐噪声后,模型精度慢慢下降;在噪声产生比率为 15% 时,精度依然达 82%,表明在强椒盐噪声下模型依然有较高的识别率;而在椒盐噪声产生比率为 10% 时,模型精度就

降到了 88%,可见添加椒盐噪声对模型精确度的影响比遮挡情况下更为恶劣。

表 7 椒盐噪声下模型的识别结果

Table 7 Recognition results of model under salt and pepper noise

| Noise ratio / % | 5 | 10 | 15 |
|----------------------|----|----|----|
| Recognition rate / % | 96 | 88 | 82 |

因为本文以随机添加黑色像素点来模拟遮挡情况,情况较简单,而以随机添加白色和黑色像素点模拟椒盐噪声,对目标的影响更加复杂。随机添加白色像素点会对目标识别造成干扰,而添加黑色像素点则会直接减弱目标特性。SAR 图像在军地两方面的应用日益广泛,但其中的高噪声使图像识别等图像处理变得复杂,因此研究去噪性能具有实际意义。实验表明,模型 S 在遮挡和具有椒盐噪声两种情况下依然有较高的识别率,具有较强的鲁棒性。

7 结 论

尽管深度学习方法给计算机视觉领域带来了深刻的变革,但在 SAR 成像上的应用还不够成熟,在不同的应用背景下,还需要不断改进。将残差收缩块和残差注意力网络相结合,构造出了模型 S,该模型在识别率和实时性上都有较大的提升。通过实验证明,模型 S 有较强的鲁棒性,具有广泛的应用前景,但采用深度学习算法进行去噪的效率比传统滤波是否更高,还需进一步分析。

参 考 文 献

- [1] Yang G, Chen K X, Zhou M Y, et al. Study evolution of detection and recognition on target in SAR image [J]. Progress in Geophysics, 2007, 22 (2): 617-621.
杨枕, 陈克雄, 周脉鱼, 等. SAR 图像中目标的检测和识别研究进展 [J]. 地球物理学进展, 2007, 22 (2): 617-621.
- [2] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313(5786): 504-507.
- [3] Krizhevsky A, Sutskever I, Hinton G E, et al. ImageNet classification with deep convolutional neural networks [C] // Proceedings of the 25th Interference Conference on Neural Information Processing Systems, December 5-8, 2013, Lake Tahoe, Nevada, USA. New York: Curran Associates, 2012, 1: 1097-1105.
- [4] Bengio Y, Lamblin P, Popovici D, et al. Greedy

- layer-wise training of deep networks[C]//Proceedings of the 19th International Conference on Neural Information Processing Systems, December 4-7, 2006, Vancouver, British Columbia, Canada. New York: MIT Press, 2006: 153-160.
- [5] Le Cun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [6] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 2672-2680.
- [7] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-04-10) [2020-08-23]. <https://arxiv.org/abs/1409.1556>.
- [8] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [9] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions [C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 1-9.
- [10] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift[EB/OL]. (2015-03-02) [2020-08-23]. <https://arxiv.org/abs/1502.03167>.
- [11] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, Inception-ResNet and the impact of residual connections on learning [EB/OL]. (2016-08-23) [2020-08-23]. <https://arxiv.org/abs/1602.07261>.
- [12] Jaderberg M, Simonyan K, Zisserman A, et al. Spatial transformer networks[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems, December 7-12, 2015, Montreal, Quebec, Canada. New York: Curran Associates, 2015, 2: 2017-2025.
- [13] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.
- [14] Wang F, Jiang M Q, Qian C, et al. Residual attention network for image classification[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6450-6458.
- [15] Xie S N, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 5987-5995.
- [16] Wang X L, Girshick R, Gupta A, et al. Non-local neural networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7794-7803.
- [17] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 3-19.
- [18] Fu J, Liu J, Tian H J, et al. Dual attention network for scene segmentation [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 3141-3149.
- [19] Ross T D, Worrell S W, Velten V J, et al. Standard SAR ATR evaluation experiments using the MSTAR public release data set [J]. Proceedings of SPIE, 1998, 3370: 566-573.
- [20] Duchi J, Hazan E, Singer Y. Adaptive subgradient methods for online learning and stochastic optimization [J]. Journal of Machine Learning Research, 2011, 12(7): 257-269.
- [21] Kingma D P, Ba J. Adam: a method for stochastic optimization[EB/OL]. (2017-01-30) [2020-08-23]. <https://arxiv.org/abs/1412.6980>.
- [22] Chen S Z, Wang H P, Xu F, et al. Target classification using the deep convolutional networks for SAR images [J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(8): 4806-4817.
- [23] Zhao M H, Zhong S S, Fu X Y, et al. Deep residual shrinkage networks for fault diagnosis [J]. IEEE Transactions on Industrial Informatics, 2020, 16(7): 4681-4690.
- [24] Zhang N, Donahue J, Girshick R, et al. Part-based R-CNNs for fine-grained category detection [M] // Fleet D, Pajdla T, Schiele B, et al. Computer vision-ECCV 2014. Lecture notes in computer science. Cham: Springer, 2014, 8689: 834-849.
- [25] Hu X, Yao Q L, Hou B Q, et al. Target recognition using convolution neural network for SAR images [J]. Science Technology and Engineering, 2019, 19(21): 228-232.
- 胡显, 姚群力, 侯冰倩, 等. 基于卷积神经网络的合成孔径雷达图像目标识别 [J]. 科学技术与工程, 2019, 19(21): 228-232.

- [26] Wang L, Qin W M, Luo W. SAR target image recognition and classification based on multitask learning and faster region-based convolution neural network[J]. Science Technology and Engineering, 2017, 17(35): 82-86.
王烈, 秦伟萌, 罗文. 基于多任务学习及 Faster R-CNN 的 SAR 目标图像识别分类[J]. 科学技术与工程, 2017, 17(35): 82-86.
- [27] Zhao Q, Principe J C. Support vector machines for SAR automatic target recognition [J]. IEEE Transactions on Aerospace and Electronic Systems, 2001, 37(2): 643-654.