

DOI:10.3969/j.issn.1672-2337.2021.05.002

基于GAN的ISAR图像语义分割方法

杜 兰, 吕国欣, 石 钰

(西安电子科技大学雷达信号处理国家重点实验室, 陕西西安 710071)

摘要: 逆合成孔径雷达(ISAR)成像技术能够对空间目标进行远距离成像, 刻画目标的外形、结构和尺寸等信息。ISAR图像语义分割能够获取目标的兴趣区域, 是ISAR图像解译的重要技术支撑, 具有非常重要的研究价值。由于ISAR图像表征性较差, 图像中散射点的不连续和强散射点存在的旁瓣效应使得人工精准标注十分困难, 基于交叉熵损失的传统深度学习语义分割方法在语义标注不精准情况下无法保证分割性能的稳健。针对这一问题, 提出了一种基于生成对抗网络(GAN)的ISAR图像语义分割方法, 采用对抗学习思想学习ISAR图像分布到其语义分割图像分布的映射关系, 同时通过构建分割图像的局部信息和全局信息来保证语义分割的精度。基于仿真卫星目标ISAR图像数据集的实验结果证明, 本文方法能够取得较好的语义分割结果, 且在语义标注不够精准的情况下模型更稳健。

关键词: ISAR图像; 深度学习; 生成对抗网络; 语义分割

中图分类号: TN957.51 文献标志码: A 文章编号: 1672-2337(2021)05-0479-06

ISAR Image Semantic Segmentation Based on GAN

DU Lan, LYU Guoxin, SHI Yu

(National Key Lab of Radar Signal Processing, Xidian University, Xi'an 710071, China)

Abstract: Inverse synthetic aperture radar (ISAR) imaging technology can perform long-range imaging of space targets, depict the information such as the shape, structure and size of the target. Semantic segmentation of ISAR images can obtain the region of interest of the target, which is an important technical support for ISAR image interpretation. Semantic segmentation of ISAR images has very important research value. Due to the poor representation of ISAR images, the discontinuity of the scattering points in the image and the sidelobe effects of strong scattering points make it very difficult to label accurately. Traditional deep semantic segmentation methods based on cross-entropy loss cannot guarantee the robustness of segmentation performance when labels are not accurate. In this paper, an ISAR image semantic segmentation method based on generative adversarial networks (GANs) is proposed. Adversarial learning ideas are used to learn the mapping relationship from the distribution of ISAR images to the distribution of corresponding semantic segmentation images. At the same time, the accuracy of semantic segmentation is ensured by constructing the local information and global information of the segmented image. The experimental results based on the simulated satellite target ISAR image dataset prove that the proposed method can obtain better semantic segmentation results, and the model is more robust when the semantic labels are not enough accurate.

Key words: ISAR images; deep learning; generative adversarial network(GAN); semantic segmentation

0 引言

逆合成孔径雷达(Inverse Synthetic Aperture Radar, ISAR)成像技术是当前对海、陆、空观测的重要方式, 它作为一种能够远距离获取非合作目标图像的技术, 在军事领域起到了重要作用。随

着ISAR成像技术的不断成熟, 对ISAR图像进行快速有效的自动解译, 获得目标丰富的结构、姿态、尺寸等信息, 受到了越来越广泛的关注。ISAR图像语义分割是ISAR图像处理和解译的重要技术支撑, 语义分割指根据某种准则将图像划分为互不重叠的同质区域, 并且给出不同区域中像素的语义类别, 它能够提取图像中感兴趣区域, 为后

续图像识别与分类、场景解析、目标检测等任务作预处理。传统的语义分割方法利用像素的灰度、色彩、纹理等低级语义特征对像素进行分类,图像分割精度有限^[1]。随着深度卷积神经网络的快速发展,基于深度学习的语义分割方法成为语义分割的主流方法^[2-3]。Long 等人^[4]在 2014 年提出的全卷积网络(Fully Convolutional Network, FCN)是首个端到端的深度语义分割模型,目前大多语义分割网络都是基于 FCN 的编码-解码结构实现的。FCN 通过对编码过程中提取的特征图进行双线性插值上采样将其恢复至输入图像尺寸,实现像素级分类。文献[5]基于编码-解码结构,解码器进行上采样时利用编码部分记录的池化索引来恢复图像的细节信息,提升分割精度。文献[6]基于编码-解码结构,引入深度可分离空洞卷积在减少计算量的同时增大感受野,采用金字塔池化模块获取图像多尺度上下文信息,提升图像分割精度。传统的深度语义分割模型通常采用交叉熵损失函数训练模型对图像中的每一个像素进行分类,使模型学习原始图像像素到分割图像像素的映射关系,因此需要非常精准的标注图像来对模型进行训练。ISAR 图像与光学图像相比表征性差,图像中散射点的不连续和强散射点存在的旁瓣效应,使得人工精准标注十分困难,传统深度语义分割方法在训练数据标注不精准的情况下无法保证分割性能稳健。

本文针对上述问题,提出了一种基于 Pix2pixGAN^[7]的 ISAR 图像语义分割方法。本方法将 GAN^[8]的对抗思想引入 ISAR 图像语义分割任务中,通过对抗学习使模型学习 ISAR 图像分布到其语义分割图像分布之间的映射,约束方式相比较于像素到像素的映射更加灵活,从而减弱模型对于精准标注的需求,提升语义标注不够精准的情况下模型的稳健性。

1 基于 Pix2pixGAN 的 ISAR 图像语义分割方法

1.1 基于 Pix2pixGAN 的 ISAR 图像语义分割网络

Pix2pixGAN 是 GAN 的衍生模型之一,在图像生成、图像翻译和图像风格转换等任务上具有十分出色的表现^[7]。所提方法的网络结构如图 1

所示,我们定义 ISAR 图像为 x ,其对应语义分割标签图像集为 y ,将 ISAR 图像作为条件信息约束生成器 G 生成数据的方向,训练生成器 G 以生成判别器 D 难以分辨的分割图像,训练判别器 D 以尽可能地分辨真假分割图像,通过这种对抗训练学习 ISAR 图像到其分割图像的映射关系 $G:(x,z) \rightarrow y$ 。本方法将 ISAR 图像 x_i 和噪声向量 z 作为生成器 G 的输入,输出为生成的分割图像 $G(x_i, z)$;将 ISAR 图像 x_i 和生成的分割图像 $G(x_i, z)$ 或 ISAR 图像 x_i 和真实的分割图像 y_i 作为图像对输入判别器 D ,输出为判别器 D 判断输入图像对为真实图像的概率。采用交替迭代的方式训练该网络,首先固定生成器 G ,训练判别器 D 使其尽可能地区分 $G(x_i, z)$ 与 y_i ;然后固定判别器 D ,训练生成器 G 使其生成的分割图像 $G(x_i, z)$ 与真实分割图像 y_i 尽可能相似;循环上述交替迭代过程,当对抗达到平衡时,判别器 D 无法区分生成的分割图像 $G(x_i, z)$ 和真实的分割图像 y_i ,也就是说生成器 G 能够生成近似于真实分割图像的分割结果。

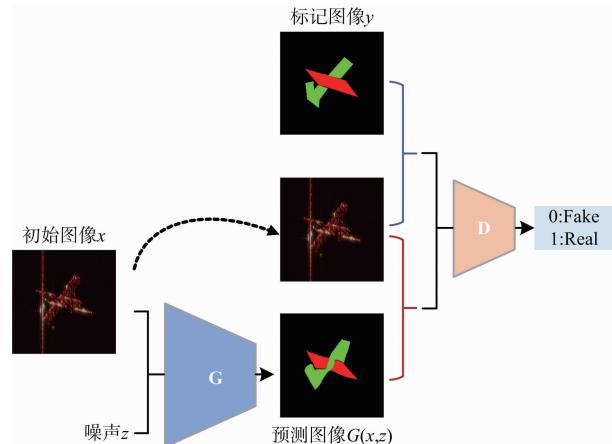


图 1 基于 Pix2pixGAN 的 ISAR 图像语义分割方法框架

1.1.1 生成器网络结构

编码-解码结构是语义分割任务中常用的网络结构,首先使用编码器对输入图像进行特征提取,再使用解码器对特征图进行上采样,将特征图尺寸恢复至输入图像分辨率大小进行像素级分类。本文的生成器如图 2 所示,在编码-解码结构的基础上引入跳线结构,在上采样时引入编码部分的信息,将解码器每层特征与编码器中对称层的特征按通道拼接后再进行上采样,避免直接对特征图进行上采样时丢失大量细节信息,通过融合编

码部分的底层特征来保证输出图像在细节上能够有较好的表现。生成器网络结构如图2所示,编码器由8个卷积层构成,对一张宽度、高度和通道数为 $256 \times 256 \times 3$ 的图像经过特征提取后输出为

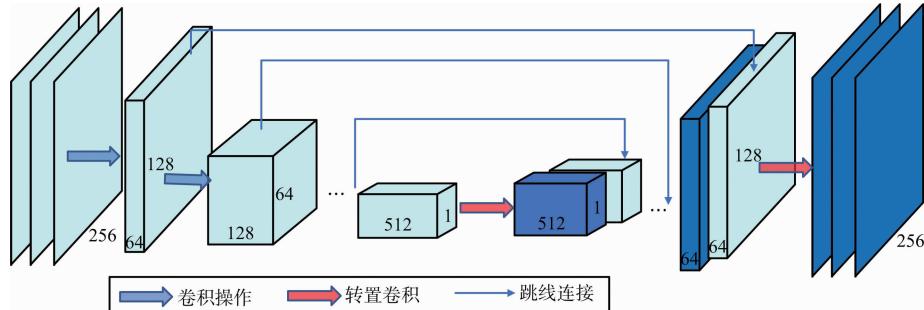


图2 生成器网络结构示意图

1.1.2 判别器网络结构

本文采用马尔科夫判别器来构建生成图像的局部信息,进一步提升生成的语义分割图像在细节上的表现。马尔科夫判别器的结构如图3所示,其为一个全卷积网络,采用四层卷积层对输入图像进行特征提取,得到一张 $n \times n$ 的特征图,网络的输出为一个 $n \times n$ 矩阵,其中矩阵中每一个结果对应输入图像中的一个感受野,即输出矩阵为输入图像中 $n \times n$ 个Patch的判别结果,最终以输出矩阵的均值判断输入图像为生成图像还是真实图像。马尔科夫判别器和生成器均是全卷积网络,因此可以处理任意尺寸的图像,增强了网络的扩展性。

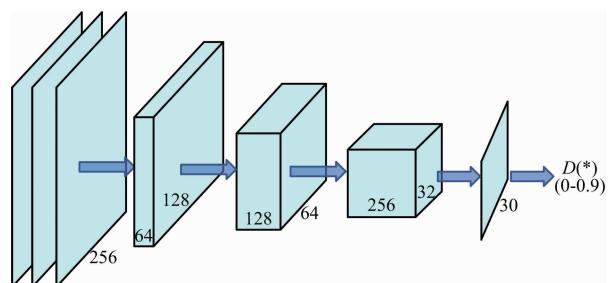


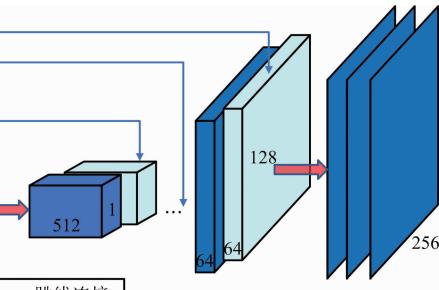
图3 判别器网络结构示意图

1.2 损失函数

本文方法的损失函数由对抗损失和重构损失两部分组成。对抗损失函数 $L_{\text{CGAN}}(G, D)$ 如式(1)所示:

$$\begin{aligned} L_{\text{CGAN}}(G, D) = & E_{x,y} [\log D(x, y)] + \\ & E_{x,z} [\log(1 - D(x, G(x, z)))] \end{aligned} \quad (1)$$

512个大小为 1×1 的特征图。解码器由8个转置卷积层组成,对512个大小为 1×1 的特征图经过上采样后输出为一张宽度、高度和通道数为 $256 \times 256 \times 3$ 的图像。



其中,生成器 G 的目标是生成与真实分割图像尽可能相似的图像以最小化损失函数,判别器 D 的目标是尽可能区分生成的分割图像与真实的分割图像以最大化损失函数,该过程可表示为

$$G^* = \operatorname{argmin}_G \max_D L_{\text{CGAN}}(G, D) \quad (2)$$

为了进一步提升生成器的性能,减小生成的分割图像与真实的分割图像的差异,本文方法采用 L_1 损失函数减少生成图像的模糊程度,所以本方法使用 L_1 损失函数构建生成的分割图像的低频部分,促使生成图像在全局上更接近于真实图像, L_1 重构损失如式(3)所示:

$$L_{L1}(G) = E_{x,y,z} [\| y - G(x, z) \|_1] \quad (3)$$

本文方法的目标函数可以表示为式(4)所示的形式:

$$\min_G \max_D V(G, D) = L_{\text{CGAN}}(G, D) + \lambda L_{L1}(G) \quad (4)$$

式中, λ 表示两个损失函数的相对重要程度。

2 实验结果与分析

2.1 实验数据

本文使用由电磁仿真软件FEKO对卫星目标建模、再通过BP成像算法得到的卫星目标ISAR图像数据集进行实验。该数据集包含五类卫星目标,分别为阿波罗卫星、北斗卫星、锁眼卫星、美国陆地卫星和天宫卫星,仿真CAD模型和其对应的ISAR图像示例分别如图4和图5所示。为了满足深度神经网络对于训练数据量的要求,我们对

训练数据集进行平移和镜像操作, 将训练图像数量扩充至原始数据量的 36 倍, 五类卫星 ISAR 图像数量如表 1 所示。

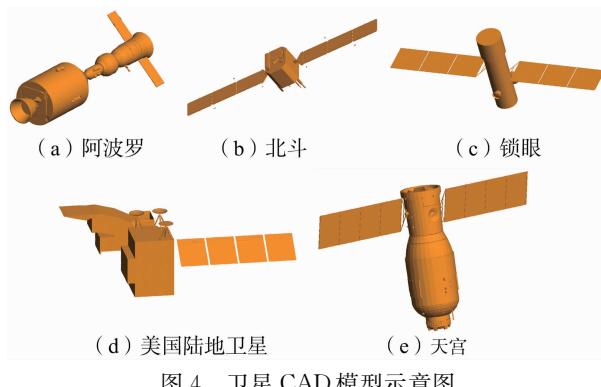


图 4 卫星 CAD 模型示意图

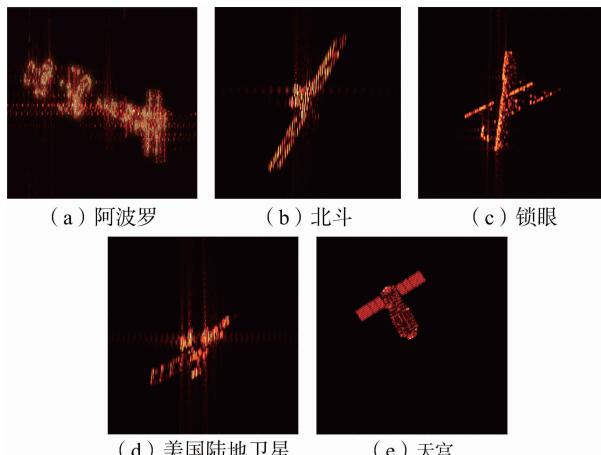


图 5 卫星目标 ISAR 图像示例

表 1 五类卫星目标的样本数信息

卫星类型	原始训练 图像数量/幅	扩充后训练 图像数量/幅	测试图像 数量/幅
阿波罗	90	3 240	36
北斗	93	3 348	33
锁眼	83	2 988	27
美国陆地卫星	85	3 060	26
天宫	88	3 168	30
总和	439	15 804	152

2.2 性能评价指标

为了定量评估不同方法的分割性能, 本文采用语义分割标准衡量准则——平均交并比(Mean Intersection over Union, MIoU)作为定量评判的准则, MIoU 为预测结果与真实结果之间的交集与并集之比, 计算示意图如图 6 所示。

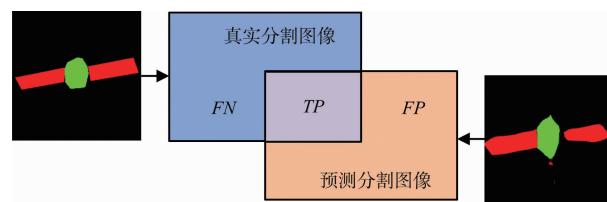


图 6 计算 MIoU 示意图

图 6 中, TP 表示预测分割图像与真实分割图像分类一致的像素数, FP 表示真实分割图像中除去 TP 以外的像素数, FN 表示预测分割图像中除去 TP 以外的像素数。

计算公式如下:

$$\text{MIoU} = \frac{1}{M} \sum_{m=1}^M \frac{TP}{TP + FP + FN} \quad (5)$$

式中, M 表示图像总数。

2.3 实验结果分析

1) 语义分割性能比较分析

图 7 给出了本文方法对不同卫星目标的分割结果, 图中第一、二、三列分别为 ISAR 图像、分割图像和标签, 第四列和第五列分别为从分割图像提取到的卫星太阳能帆板和主体, 并且经过边缘细化、空洞填充等形态学操作后的结果。结果表明, 对于不同类型卫星目标的 ISAR 图像, 本文方法均能够取得较好的分割结果, 经过掩膜分割后得到完整的太阳能帆板和主体。由图可知, 本文的方法在太阳能帆板与主体粘连处能够实现精细的分割, 在细节上有较好的表现, 这表明本文方法中生成器和判别器的结构能够提升语义分割的细节表现, 精细的语义分割结果能够为后续解译工作奠定良好的基础。

为了评估本文方法与现有的深度语义分割方法性能, 表 2 给出了不同语义分割方法对于每类卫星的 MIoU 结果以及五类卫星 MIoU 的平均结果。由表 2 可以看出, 在 ISAR 图像标签不够精准的情况下, 本文方法依然取得了较好的分割精度, 且对于不同卫星目标的分割性能稳定, 说明模型具有较好的鲁棒性。由于 ISAR 图像人工标注无法保证语义标签的精确性, 这种情况下 MIoU 定量评估结果并不能完全说明本文方法在分割精度上一定最优, 但是足以说明其具有较好的分割性能。

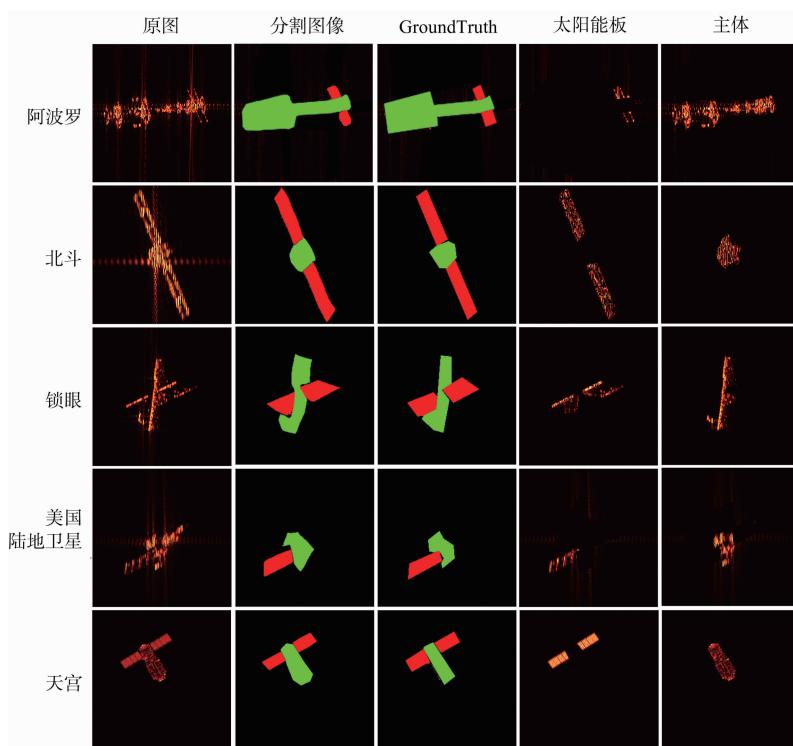


图 7 本文方法的卫星 ISAR 图像分割结果示例(红色:太阳能板;绿色:主体)

表 2 不同语义分割方法的 MIoU %

方法	美国 陆地 卫星					平均 结果
	阿波罗	北斗	锁眼	陆地	天宫	
FCN	68.48	80.64	76.01	66.71	73.62	73.09
SegNet	71.09	82.27	79.55	70.38	76.17	75.89
DeepLabV3+	71.52	81.53	79.12	69.96	76.05	75.63
Pix2pixGAN	74.16	82.52	79.98	74.28	78.81	77.95

图 8 给出了本文方法与三种传统深度语义分割方法在语义标签不够精准的情况下训练数据的分割结果示例。对照图 8 中的原图和 Ground Truth 可以看出,第一行至第三行的 GroundTruth 中均存在太阳能板和主体标注不完整的问题(主

体标注明显不完整),第一、二行的 GroundTruth 中太阳能板和主体之间存在不同程度的断裂。其他三种对比方法的分割结果存在与 GroundTruth 相同的问题,分割结果不够准确;相比于对比方法的分割结果,本文方法的结果对 GroundTruth 的拟合度较低,但通过对照原图和本文方法的结果可以看出,本文方法能够比较完整地分割出主体和太阳能板,且太阳能板和主体间的连续性较好,没有明显断裂的情况。图 8 的结果说明在语义标注不够精准的情况下,其他对比的语义分割方法学到的结果非常接近标注、也不够精准,而本文方法仍然能够准确地、完整地将目标的太

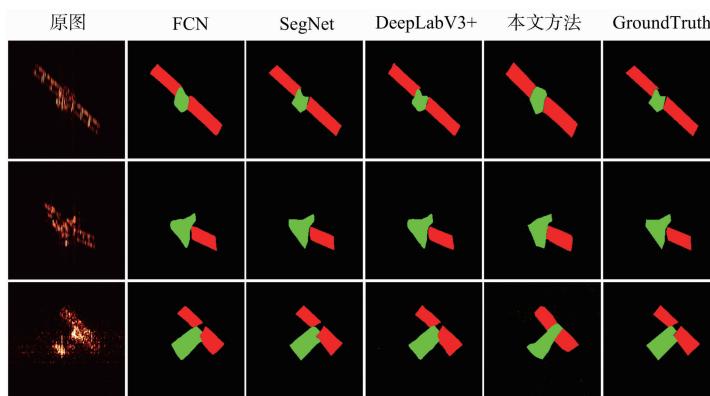


图 8 语义标注不精准情况下不同分割方法的训练结果示例(红色:太阳能板;绿色:主体)

阳能板和主体分割出来。这是因为所对比的语义分割方法采用交叉熵损失函数学习像素到像素之间语义类别的映射,此类方法对于语义标注的精度要求较高,在标注图像不精准的情况下会使得模型学到一些错误信息,导致分割结果不够精准;而 GAN 的对抗学习思想能够训练模型学习数据分布到分布之间的映射,它能够通过生成器与判别器的对抗博弈过程探究数据的分布情况,即使标注本身存在误差,由于约束方式相比较于像素到像素之间的语义类别映射更加灵活,不会对每个像素的预测误差进行惩罚,因而可以减弱模型对于精准标注的需求,使模型具有较好的泛化性。因此,在语义标签不够精准的情况下本文方法的语义分割性能更加稳健。

2) 不同模型空间、时间复杂度分析

为了分析本文方法与现有语义分割方法的模型复杂度与运算效率,表 3 计算了用于衡量不同模型时间复杂度和空间复杂度的浮点运算次数(Floating-point Operations, FLOPs)和总参数量,由于 CNN 的 FLOPs 和参数量主要集中在卷积层、反卷积层和全连接层,其余层的 FLOPs 和参数量相比于卷积层、反卷积层和全连接层非常少,可以忽略,又由于所提方法和对比方法均未含有全连接层,因此我们分析时间和空间复杂度时仅考虑了卷积层和反卷积层。

时间复杂度与模型的 FLOPs 成正相关,计算公式如下式所示:

$$\text{Time} \sim O\left(\sum_{l=1}^D K_l^2 \cdot M_{\text{out},l}^2 C_{l-1} \cdot C_l\right) + \sum_{n=1}^N K_n^2 \cdot M_{\text{in},n}^2 C_{n-1} \cdot C_n \quad (6)$$

式中, K 为卷积核与反卷积核的尺寸, C 为卷积与反卷积通道数, D 为卷积层的层数, N 为反卷积层的层数, $M_{\text{out},l}$ 为第 l 卷积层输出的特征图尺寸, $M_{\text{in},n}$ 为第 n 反卷积层输入的特征图尺寸。

空间复杂度与模型的总参数量成正相关,计算公式如下式所示:

$$\text{Space} \sim O\left(\sum_{l=1}^D K_l^2 \cdot C_{l-1} \cdot C_l\right) \quad (7)$$

式中, K 为卷积核与反卷积核的尺寸, C 为卷积与反卷积通道数, D 为卷积层与反卷积层的总层数。

表 3 不同模型总参数量、FLOPs 和 MIoU

方法	FLOPs	总参数量	MIoU/%
FCN	4.526×10^{10}	3.412×10^7	73.09
SegNet	4.497×10^{10}	3.494×10^7	75.89
DeepLabV3+	2.902×10^{10}	4.649×10^7	75.63
Pix2pixGAN	9.970×10^9	5.440×10^7	77.95

由表 3 可以看出,FCN^[3] 和 SegNet^[4] 具有较高的 FLOPs, 这是因为 FCN 和 SegNet 均用普通的卷积组堆叠构成深层编码-解码结构, 并且使用步长为 1 的卷积核使得输出特征图尺寸较大; DeepLabV3+^[5] 模型中使用深度可分离卷积减少了计算量; 本文方法采用步长为 2 的卷积核进行下采样使得输出特征图尺寸大大减小, 因此具有最低的 FLOPs; 然而, 本文方法使用跳线连接保持编码器细节信息的同时使得解码器部分的卷积通道数倍增, 因此具有较高的参数量, 如何进一步优化模型的空间复杂度, 减少模型参数量是下一步需要探索的方向。

3 结束语

本文针对 ISAR 图像语义分割问题, 提出了一种基于 Pix2pixGAN 的 ISAR 图像语义分割方法, 该方法采用对抗学习思想学习 ISAR 图像分布到其语义分割图像分布的映射关系, 同时通过构建分割图像的局部信息和全局信息来保证语义分割的精度。实验结果证明, 本文方法能够对 ISAR 图像取得较好的语义分割结果, 且在语义标注不够精准的情况下模型更稳健。然而, 本文方法的网络存在较高时间复杂度的问题, 如何进一步优化模型的时间复杂度是下一步需要探索的方向。

参考文献:

- [1] JAIN S, LAXMI V. Color Image Segmentation Techniques: A Survey[C]//Proceedings of the International Conference on Microelectronics, Computing and Communication Systems, Springer, Singapore: IEEE, 2018: 189-197.
- [2] TAKOS G. A Survey on Deep Learning Methods for Semantic Image Segmentation in Real-Time[EB/OL]. [2020-09-27]. <https://arxiv.org/abs/2009.12942>.

(下转第 490 页)