一种基于深度学习的SAR城市建筑区域叠掩精确检测方法

田 野[①②③] 丁赤飚[①②] 张福博*[①②] 石民安[①②③]

[①](中国科学院空天信息创新研究院微波成像技术国家级重点实验室 北京 100190)

[②](中国科学院空天信息创新研究院 北京 100190)

[③](中国科学院大学电子电气与通信工程学院 北京 100049)

摘要：建筑物叠掩检测在城市三维合成孔径雷达(3D SAR)成像流程中是至关重要的步骤，其不仅影响成像效率，还直接影响最终成像的质量。目前，用于建筑物叠掩检测的算法往往难以提取远距离全局空间特征，也未能充分挖掘多通道SAR数据中关于叠掩的丰富特征信息，导致现有叠掩检测算法的精确度无法满足城市3D SAR成像的要求。为此，该文结合Vision Transformer (ViT)模型和卷积神经网络(CNN)的优点，提出了一种基于深度学习的SAR城市建筑区域叠掩精确检测方法。ViT模型能够通过自注意力机制有效提取全局特征和远距离特征，同时CNN有着很强的局部特征提取能力。此外，该文所提方法还基于专家知识增加了用于挖掘通道间叠掩特征和干涉相位叠掩特征的模块，提高算法的准确率与鲁棒性，同时也能够有效地减轻模型在小样本数据集上的训练压力。最后在该文构建的机载阵列SAR数据集上测试，实验结果表明，该文所提算法检测准确率达到94%以上，显著高于其他叠掩检测算法。

关键词：深度学习；专家知识；3D SAR成像；建筑区域叠掩检测；Vision Transformer模型

中图分类号：TN957.52 **文献标识码**：A **文章编号**：2095-283X(2023)02-0441-15

DOI: 10.12000/JR23033

引用格式：田野, 丁赤飚, 张福博, 等. 一种基于深度学习的SAR城市建筑区域叠掩精确检测方法[J]. 雷达学报, 2023, 12(2): 441–455. doi: 10.12000/JR23033.

Reference format: TIAN Ye, DING Chibiao, ZHANG Fubo, *et al*. SAR building area layover detection based on deep learning[J]. *Journal of Radars*, 2023, 12(2): 441–455. doi: 10.12000/JR23033.

SAR Building Area Layover Detection Based on Deep Learning

TIAN Ye[①②③] DING Chibiao[①②] ZHANG Fubo*[①②] SHI Min'an[①②③]

[①](*National Key Laboratory of Microwave Imaging Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing* 100190, *China*)

[②](*Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing* 100190, *China*)

[③](*School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing* 100049, *China*)

Abstract: Building layover detection is a crucial step in the 3D Synthetic Aperture Radar (SAR) imaging process in urban areas. It affects imaging efficiency and directly influences the final image quality. Currently, algorithms used for layover detection struggle to extract long-range global spatial characteristics and fail to fully exploit the rich features of layover in multi-channel SAR data. To address the issue of insufficient accuracy in existing layover detection algorithms to meet the requirements of urban 3D SAR imaging, this paper proposes a deep learning-powered SAR urban layover detection method that combines the advantages of the Vision Transformer (ViT) model and Convolutional Neural Network (CNN). The ViT model can efficiently extract global and long-range features through a self-attention mechanism, whereas the CNN has strong local

收稿日期：2023-03-11；改回日期：2023-04-02；网络出版：2023-04-24

*通信作者： 张福博 zhangfb@aircas.ac.cn *Corresponding Author: ZHANG Fubo, zhangfb@aircas.ac.cn

基金项目：国家重点研发计划(2021YFA0715404)

Foundation Item: National Key R&D Program of China (2021YFA0715404)

责任主编：张群 Corresponding Editor: ZHANG Qun

feature extraction capabilities. Furthermore, the proposed method in this paper incorporates a module for investigating inter-channel layover features and interferometric phase layover features based on expert knowledge, which improves the accuracy and robustness of the algorithm while effectively decreasing the training pressure on the model in small-sample datasets. Finally, the proposed algorithm is tested on a self-built airborne array SAR dataset, and experimental findings revealed that the proposed algorithm achieves a detection accuracy of >94%, which is significantly higher than other layover detection algorithms, completely revealing the effectiveness of this method.

Key words: Deep learning; Expert knowledge; 3D SAR imaging; Building area layover detection; Vision Transformer (ViT) model

1 引言

建筑叠掩检测在多通道合成孔径雷达(Synthetic Aperture Radar, SAR)城市区域三维成像过程中扮演着关键角色，直接决定了后续流程的选择。如图1所示，城市区域三维成像流程中的建筑叠掩检测环节位于关键位置。在进行多通道数据的图像配准和相位补偿后，需要对叠掩区域进行检测，根据检测结果的不同，针对性地使用不同的处理方法，对叠掩区域使用超分辨算法，对非叠掩区域使用干涉算法，综合两部分区域的结果得到SAR三维成像。

然而，一旦叠掩区域的误识别发生，将会极大地影响SAR三维成像结果的质量。如果将干涉算法错误地应用于叠掩区域，会导致目标信息的丢失，使得应检测到的目标漏检。相反，如果错误地应用超分辨算法于非叠掩区域，则会引入额外的噪声，使成像结果失真，同时还将消耗更多的计算资源，影响SAR三维成像的效率。因此，城市建筑区域的叠掩检测对于城市区域的三维SAR成像具有重要意义。

城市建筑区域的叠掩检测一直是雷达图像识别领域内的重点[1–7]。为此，相关专家设计了一系列基于城市建筑叠掩特征的检测方法。例如，叠掩是由多个信号混叠而成的，叠掩区域的幅度值较高，Soergel等人[8]根据这一特征，设计了基于幅度的叠掩检测方法。Prati等人[9]根据叠掩区域相位梯度为

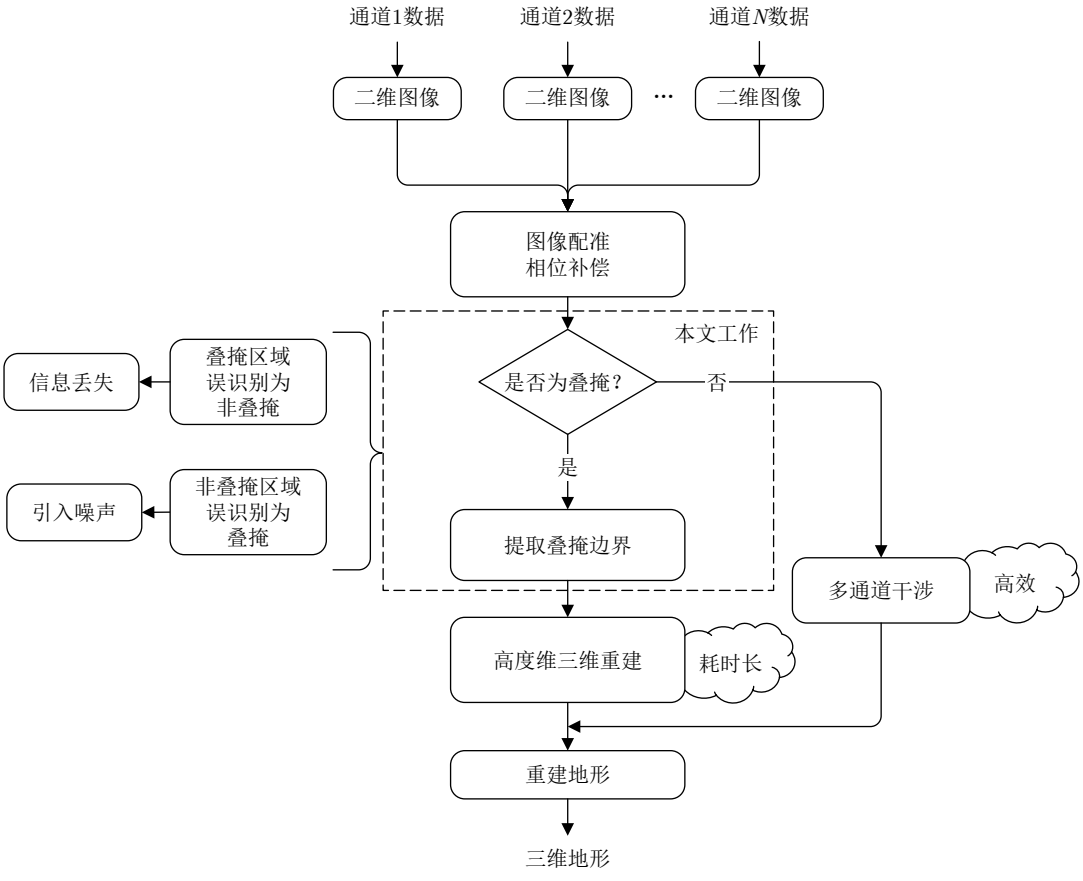


图 1 城市区域SAR三维成像流程图

Fig. 1 The flowchart of 3D SAR reconstruction of the urban area

负的特点对叠掩区域进行识别。Wilkinson[10]通过分析叠掩区域的统计特性，以叠掩区域相干性较差为特征对叠掩进行检测。随着多通道SAR的发展，Chen等人[11]和Wu等人[12]通过特征值分解等方法，根据通道间的信号特征对叠掩区域进行分割检测。传统叠掩检测算法的问题在于其需要大量的专家知识和人工设计的特征。随着深度学习的发展，学者设计了许多基于深度学习的检测方法，该类方法多数都是基于卷积神经网络(Convolutional Neural Networks, CNN)，通过从数据中学习特征表示和分类器，来更好地应对SAR图像中目标的多样性和复杂性，并表现出比传统算法更好的自适应性和鲁棒性。Wu等人[13,14]通过设计多尺度神经网络和基于注意力机制的神经网络来提高模型对于叠掩的检测精度。Chen等人[15]设计了针对InSAR数据的叠掩分割网络。

然而，上述方法仍存在不足之处：一方面，随着多通道SAR的发展，多维度数据中蕴含的丰富特征对建筑叠掩区域的特征挖掘提供了更多的选择，但现有方法没有及时将这些特征结合到建筑叠掩检测的网络结构中；另一方面，现有基于CNN的检测方法由于卷积操作的感受野检测范围受限的性质，在提取图像中的远距离依赖特征时，无法充分挖掘大尺度的全局特征，因此无法获得更加精准的分割结果。近年来，Transformer模型的崛起正着力于解决这一问题，越来越多的研究者开始考虑在视觉任务中应用Transformer模型[16–21]。Vision Transformer (ViT)模型在光学图像领域已经取得了巨大的成功[22]，但ViT模型在SAR叠掩检测问题上还未得到应用。在叠掩检测问题上局部特征通常会被SAR图像固有的相干斑噪声干扰甚至破坏，相比之下，建筑叠掩区域所具有的丰富的全局特征则更具有鲁棒性。因此，ViT模型在建筑叠掩识别问题上具有更大的研究潜力。

综上，现有的城市建筑区域叠掩检测方法在多通道SAR数据上未能有效挖掘叠掩的多维度特征，而以CNN为骨架的识别方法虽能有效提取局部特征，但未能充分挖掘叠掩的大尺度空间结构性特征，这导致现有算法的检测识别精度较低。针对这一问题，本文拟通过结合ViT的全局上下文信息感知和CNN的局部特征提取能力，并结合专家知识，提出一种新的基于深度学习的叠掩检测方法。该方法具有以下创新点：首先，它首次将ViT架构应用于SAR图像建筑区域叠掩检测，并与CNN模型框架相结合，利用前者的全局特征提取能力和后者蕴含的局部相似性和平移不变性来挖掘建筑叠掩区域更优的特征表达。这同时保证了模型在小样本情况下对全局和局部特征都具有较强的提取能力。其次，该方法充分利用相关的专家知识，设计了通道间特征模块和干涉相位反偏特征模块，以增强叠掩特征检测的鲁棒性，同时可以在小样本数据集上降低训练难度，提高检测精度。

2　Vision Transformer模型

Transformer模型最早是由Google在2017年提出的一种自然语言处理模型[23]。其提出之初旨在通过自注意力机制来实现对序列信息的全局建模。Transformer模型推出之后，极大地改进了自然语言处理领域中的语言建模问题，取得了很好的效果。Transformer模型的成功引起了计算机视觉领域研究人员的关注。Dosovitskiy等人[20]为了实现对图像全局信息的更好挖掘首次提出了ViT模型，证明了其在图像分类任务中的有效性。此后，有学者以ViT模型为基础提出了Swin-ViT等网络，其分类性能超过了同类的CNN模型[22,24]。为了更好地说明如何将ViT模型融入到叠掩检测模型中，下面对ViT模型的算法流程进行简述，ViT模型流程图如图2。

(1) 嵌入层

对于输入大小为$h \times w \times c$的图像，ViT模型首先将数据分成n个长宽为p的图像块，其中，$n = hw/p^2$，然后将每个图像块展平为$1 \times d$的特征向量。这n个向量组成一个$n \times d$的输入矩阵，记为$\boldsymbol{X}$。经过嵌入层(Embedding Layer)将输入$\boldsymbol{X}$转换为$n \times d$嵌入矩阵$\boldsymbol{Z}$，可表示如下：

$$\boldsymbol{Z} = E(\boldsymbol{X}) = \boldsymbol{X}\boldsymbol{W}_E + \boldsymbol{b}_E \tag{1}$$

其中，$\boldsymbol{W}_E$和$\boldsymbol{b}_E$是可学习参数，嵌入层本质上是一个将输入数据映射到目标特征空间的线性变换。

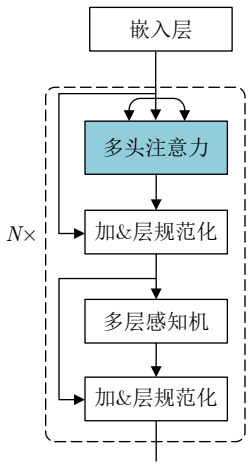


图 2 Transformer模块结构图

Fig. 2　The structure of Transformer module

(2) 多头自注意力机制

多头自注意力机制(Multi-head Self-Attention)旨在通过获取不同子空间的特征编码信息来增强模型的表达能力[23]。具体流程为：将输入的嵌入矩阵$\boldsymbol{Z}$分为h个头，每个头的嵌入向量长度为$d_h=d/h$，对每个头进行独立的自注意力机制计算，最后将各个头的结果拼接起来作为输出。自注意力机制的计算过程包括3个步骤：查询(Query)、键(Key)、值(Value)。3个步骤的公式如下：

$$\begin{aligned}\boldsymbol{Q}&=\boldsymbol{Z}\boldsymbol{W}_Q\\ \boldsymbol{K}&=\boldsymbol{Z}\boldsymbol{W}_K\\ \boldsymbol{V}&=\boldsymbol{Z}\boldsymbol{W}_V\end{aligned}\tag{2}$$

其中，$\boldsymbol{W}_Q$, $\boldsymbol{W}_K$, $\boldsymbol{W}_V$都是可学习的参数矩阵，得到3个特征矩阵后，以缩放点积注意力(Scaled Dot-Product Attention)的方式，得到最终的输出：

$$\text{SelfAttention}\left(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}\right)=\text{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\text{T}}}{\sqrt{d_h}}\right)\boldsymbol{V}\tag{3}$$

其中，$(\cdot)^{\text{T}}$表示矩阵转置，softmax函数的作用是将查询矩阵$\boldsymbol{Q}$与键矩阵$\boldsymbol{K}$的点积转化为注意力矩阵，来表征不同区域的重要性大小，再将其与值矩阵$\boldsymbol{V}$相乘即可得到自注意力模块的输出。$\sqrt{d_h}$是缩放因子，其作用是避免softmax输出的值过大或过小。

(3) 多层线性感知机

经过多头注意力模块后，执行如图2所示的“加&层规范化”模块：对输出的特征张量做层规范化处理，以保证数据的分布易于训练，再将多头注意力模块处理得到的特征张量以元素对应的方式与未被处理的原始特征张量相加。之后的多层感知机模块(Multilayer Perceptron, MLP)由输入层、隐藏层和输出层构成。相邻层所包含的神经元之间使用“全连接”的方式进行连接。该设计可以保证图像中不同区域的特征向量都能以最短路径相互连接。该层也是ViT方法能高效提取全局特征的关键模块，可表示如下：

$$\text{MLP}\left(\boldsymbol{X}\right)=\text{ReLU}\left(\boldsymbol{X}\boldsymbol{W}_1+\boldsymbol{b}_1\right)\boldsymbol{W}_2+\boldsymbol{b}_2\tag{4}$$

其中，$\boldsymbol{W}_1$, $\boldsymbol{b}_1$将输入的特征矩阵映射到高维的隐藏层，经过激活函数后，$\boldsymbol{W}_2$, $\boldsymbol{b}_2$将高维特征重新映射到原始特征空间。再经过一次“加&层规范化”模块处理后，得到ViT网络的输出结果如下：

$$\text{LayerNorm}\left(\boldsymbol{X}+\text{MLP}\left(\text{SelfAttention}\left(\boldsymbol{X}\right)\right)\right)\tag{5}$$

由上述算法流程可知，ViT模型中，每个单元都可以通过自注意力后的MLP层连接到任意其他单元，任意单元间的最大路径距离计算复杂度仅有$O(1)$。在深度学习网络中，模型单元的最大路径距离会影响对远距离依赖关系的特征提取[25]。作为对比，以长度为l，输入输出的通道数都为c的序列为例，卷积核大小为k的卷积网络单元的计算复杂度为$O(klc^2)$，最大路径长度为$O(l/k)$；循环神经网络单元的计算复杂度是$O(lc^2)$，最大路径长度为$O(l)$。对比可知，ViT模型的最大路径长度最小，在提取远距离依赖特征时更有优势。该特性也为计算机视觉领域的图像分割问题提供了新的思路和解决方向。

ViT模型虽然在全局依赖特征提取任务上表现优秀，但它也存在一些不足。比较来说，CNN的结构本身蕴含了图像的局部相似性和平移不变性的先验信息，而ViT模型则缺乏这样先验的偏置归纳(Bias Induction)，这导致缺乏深度训练的ViT模型可能仅仅因为位置不同无法识别相同的局部特征。为此，ViT网络必须在大样本、高质量的训练集上进行深度训练，来构建出目标的局部特征，否则就会导致模型的泛化能力不足，在识别精度上低于蕴含局部先验信息的CNN网络。而缺乏海量标注真值的高质量数据集，正是目前多通道SAR建筑叠掩检测所面临的问题。另外，随着多通道SAR的发展，SAR数据中蕴含的叠掩特征也更加丰富。更多地将基于多通道SAR叠掩特征的专家知识融入到网络模型中，理论上能降低网络的训练难度，帮助网络更好地在小样本数据集上实现收敛。综上，为了在现有的小样本SAR数据集上应用ViT模型取得更好的检测结果，本文结合有效的专家知识，提出了一种新型的建筑叠掩精确检测方法。

3 结合ViT和CNN的叠掩检测网络

3.1 网络总体框架

本文提出的深度学习模型框架如图3所示。总体架构采用的是图像分割领域经典的“解码-编码”模型。网络结构上，主要的创新是采用了CNN结构和ViT结构交替排布，并引入了基于专家知识的特征模块，这些设计使得该模型既能提取数据中的深层特征，也不会丢失浅层网络中叠掩边界信息。

具体而言，网络训练的正向传播可以分为编码路径和解码路径两个阶段。在编码路径中，多通道复数数据(Multi-channel data)共经过4个编码块(Encoder block)，每个编码块的尺寸逐层减半，维度逐层加倍得到复数特征图(complex-valued map, cv map)。在单独的一个编码块中，复数数据先通过两层复数残差卷积层(complex-valued convolution, cv-conv)得到复数特征图。复数特征图一方面通过降采样作为下一级编码块的输入，另一方面通过3个专门的叠掩特征模块得到3层并联的实数叠掩特征图。这3个特征模块分别是：提取建筑叠掩大

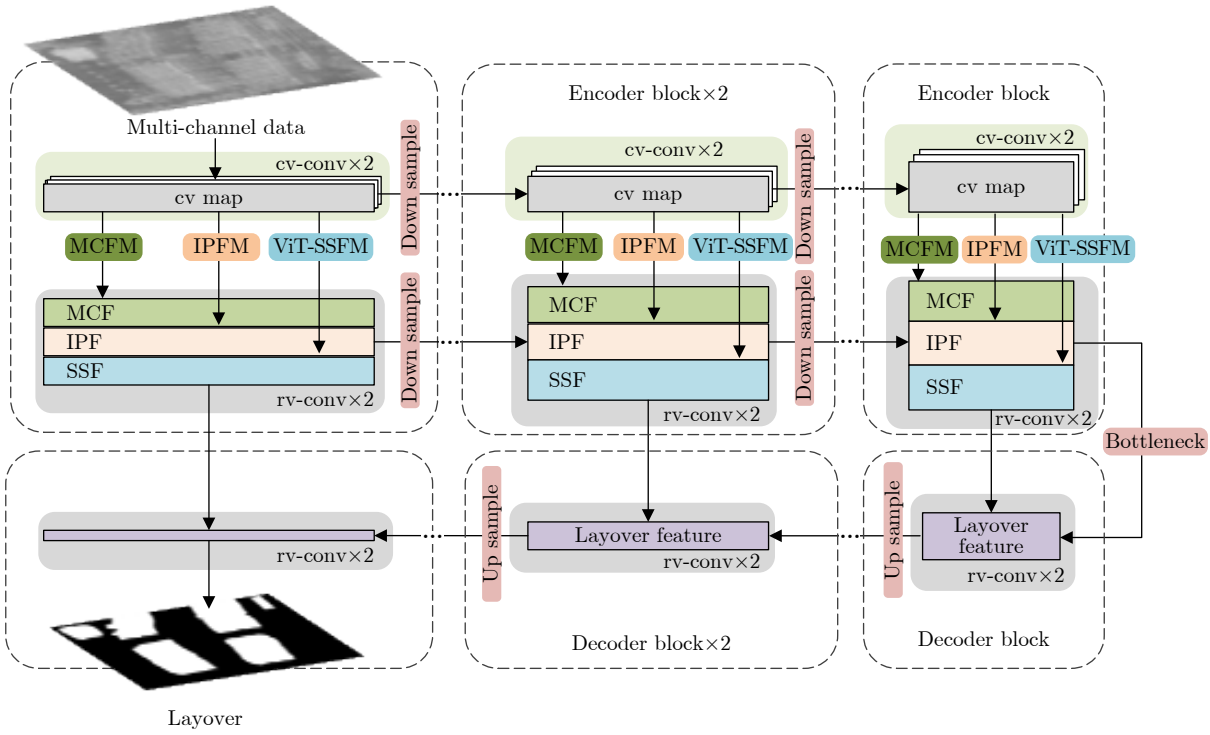


图 3 本文提出的叠掩检测网络的结构示意图

Fig. 3 The architecture diagram of layover detection network proposed in this paper

尺度空间结构特征的ViT空间结构特征模块(ViT Spatial Structure Feature Module, ViT-SSFM)、提取通道间叠掩特征的多通道特征模块(Multi-Channel Feature Module, MCFM)和提取叠掩相位反偏特征的干涉相位特征模块(Interference Phase Feature Module, IPFM)。由3个特征模块得到的实数叠掩特征图通过卷积层(rv-conv)和下采样(down sample)后，作为输入与下一级编码块中获取的实数特征图并联得到新的实数特征图。经过4个这样的编码模块得到编码器的复数特征图和实数特征图，此时通过位于中间的瓶颈层(bottleneck)对复数特征图进行实数化，两个特征图联合作为解码器的输入。在解码器中，通过卷积块(real-valued convolution, rv-conv)与上采样操作(up sample)，将编码得到的特征解码到更大尺寸的特征图上。解码过程中每个解码块(Decoder block)都会通过跳接融合浅层网络保留的叠掩边界特征，最大限度地提取被相干斑噪声严重干扰的边界信息[26]。

在网络的反向传播中，为了缓解正负样本不平衡问题并将学习权重更多地聚焦到难样本检测任务上，模型采用了二元聚焦损失函数(Binary Focal Loss, BFL)计算预测输出与真值间的损失值(Loss)。损失值会延着图3中黑色箭头的反方向进行传播。在每个模块中，自适应矩估计优化器(Adam)根据损失值与学习率进行梯度学习与权重更新。

本文方法旨在实现两个目标：(1)将Vision transformer结构和CNN结构相结合，以此来更好地挖掘叠掩的局部特征和远距离依赖特征；(2)根据专家知识，利用叠掩区域通道间的特征和干涉相位反偏特征在小样本数据集上实现更加高效、精准的识别。本章将具体介绍相关的模块和模型的损失函数。

3.2 ViT空间结构特征模块(ViT-SSFM)

建筑叠掩区域拥有丰富的空间结构特征。这一方面是因为建筑本身具有一定的空间结构，使得建筑叠掩在SAR图像中表现为平行四边形；其次，由于阵列SAR一般为侧视成像，在距离向上，叠掩之后就会出现阴影，叠掩和阴影在SAR图像上呈现相互伴生空间特征；最后，由于建筑物上普遍具有窗户等二面角结构，因此在建筑叠掩中会出现晶格状的亮斑。这些共同组成了建筑叠掩的空间结构特征。

建筑叠掩的空间结构特征与其他局部特征相比，其特征尺度往往更大。CNN中神经单元的最大路径距离过长，使其提取大尺度空间结构特征的能力有限。因此CNN在SAR建筑叠掩特征提取任务上的表现还有待提高。与CNN模型相比，ViT模型在多头自注意力模块的编码下，可以通过MLP在任意两个神经元之间建立依赖关系，进而高效地提取建筑叠掩的大尺度空间结构特征。所以，ViT模型相比于CNN更适合提取建筑叠掩的空间结构特征。

因此，本文采用ViT模型来设计了专门的模块，称之为ViT空间结构特征模块(ViT-SSFM)，其结构如图4所示。输入到ViT-SSFM的特征图，首先通过卷积分块编码得到特征向量序列，然后将特征向量序列输入到多个串联的Transformer block中。Transformer block中的多头自注意力层可以从多个子空间中分别推断像素之间的空间相关性，从全局视角中挖掘空间结构特征。随后的MLP模块会在挖掘出的特征中提取远距离的依赖关系。这二者共同作用，确保有效地挖掘叠掩的大尺度空间结构特征。最后，经过Transformer block的特征向量会在转置卷积(Transpose convolution)解码后重新转化为特征图。

与ViT模型相比，ViT-SSFM在经过特征向量编码和Transformer block挖掘特征后，增加了从特征向量还原到特征图的解码模块，用于连接后续的CNN模型。如图3所示，前一个编码块中的ViT-SSFM模块输出的特征图会在降采样后由下一个编码块的CNN结构即图3中的实数卷积层继续处理，输出的特征图又会传入再下一个编码块的ViT-SSFM模块。ViT-SSFM模块挖掘出了CNN模块难以挖掘的大尺度空间结构特征，而CNN模块包含的局部相似性和平移不变性为ViT-SSFM提供了局部特征的先验信息。本文提出的这种ViT-SSFM模块和CNN模块交替挖掘特征的结构有机地结合了两类模型的优势，在对叠掩的全局和局部特征挖掘上相互补充，可以提高模型对叠掩的检测能力，降低了整体网络的训练难度。

3.3 多通道特征模块(MCFM)

多通道SAR数据有着丰富的叠掩特征。对于叠掩区域中的任一点像素$P(m,n)$，其中混叠了多个不同高度地物目标的回波信息，而这些不同的地物目标回波之间的干涉相位是不同的，并且与它们的高程相关[27]，可表示如下：

$$\begin{aligned}
y_0(m,n) &= \sum_{i=0}^{N_s} x_0^i(m,n)\exp(\mathrm{j}\psi_i'(m,n)) + v_0(m,n)\\
&\vdots\\
y_{n_c}(m,n) &= \sum_{i=0}^{N_s} x_{n_c}^i(m,n)\exp(\mathrm{j}(\psi_i'(m,n)+\psi_i^{n_c}(m,n)))\\
&\quad + v_{n_c}(m,n)\\
&\vdots\\
y_{N_c}(m,n) &= \sum_{i=0}^{N_s} x_{N_c}^i(m,n)\exp(\mathrm{j}(\psi_i'(m,n)+\psi_i^{N_c}(m,n)))\\
&\quad + v_{N_c}(m,n)
\end{aligned} \tag{6}$$

其中，$y_1,\cdots,y_{n_c},\cdots y_{N_c}$表示通过阵列SAR获取的多通道复值数据。$n_c$为通道序号，$N_c$为阵列SAR的天线数量，即数据的通道总数。$N_s$为该点像素中叠加的不同高度和不同区域的回波数量。$i$为不同区域的序列标号，$x_{n_c}^i$表示第$n_c$个通道信号混叠的多个区域回波的地面散射特性和相干斑噪声，其服从高斯分布，且不同区域之间独立互不相关。$v_1,\cdots,v_{n_c},\cdots v_{N_c}$表示互不相关的加性复高斯白噪声。$\psi_i'$是第$0$个通道信号从发射到接收的相位历程，$\psi_i^{n_c}$是第$n_c$个通道信号与第0个通道信号的干涉相位差。当雷达的下视角、轨道高度等因素恒定不变时，分析单点信号，可以得到$\psi_i^{n_c}$为

$$\psi_i^{n_c} = \langle \alpha B_{\perp n_c} h_s + \omega_{n_c} \rangle_{2\pi} \tag{7}$$

其中，$\langle\cdot\rangle_{2\pi}$表示对相位的解缠绕操作。$\alpha$的值为$4\pi/(\lambda R_0 \cdot \sin\theta)$，其中$\lambda$为波长，$R_0$是主天线与场景中心之间的距离，$\theta$为下视角。$B_{\perp n_c}$是阵列SAR的主天线与第$n_c$个天线的正交基线，$h_s$为散射体的高

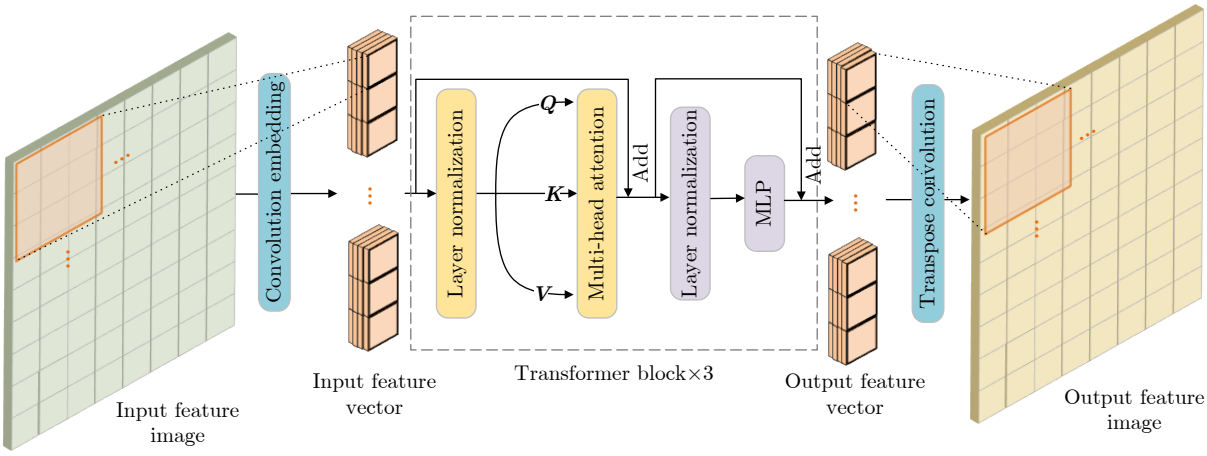


图 4 ViT空间特征模块(ViT-SSFM)网络结构示意图

Fig. 4 The network structure of the ViT-Spatial Structure Feature Module (ViT-SSFM)

度，下标s标识了该像素点在距离方位图中的坐标位置。ω_{n_c}是去相关的相位噪声。

对式(7)进行分析，可以看出ψ_{N_c}和 $B_{\perp N_c}\cdot h_s$之间是线性的。也即多通道SAR数据的频率和高度值相关。根据这一特性，可以利用空间谱估计技术来得到高度维上的目标分布[28–31]。具体的操作步骤是：(1)将多通道SAR数据时域转换到频域，通过比较幅值来确定主信号频率分量；(2)提取出主信号频率分量后反演回时域获得主信号的时域序列，在原信号中去掉得到的主信号；(3)将剩余信号分量的能量与噪声水平比较，剩余信号能量如果超过一般噪声水平即可认为原信号包含了多个有效信号分量，根据叠掩区域混叠多个地物目标回波这一特征，即能判定该区域为叠掩区域，否则，该区域只包含了0个或1个主要信号成分，可以判定为非叠掩区域。

总结以上的专家知识，本文中将该特征提取流程设计为专门的叠掩多通道特征模块，如图5所示，该模块中的输入是每个编码模块中的多通道复数特征图。首先，模块会在复数特征图上做通道间的FFT，得到频域上的目标分布。然后，通过各分量幅值的大小来获取主要的信号分量。接着把除主要信号分量外其余分量置0后，将频域特征层做IFFT反演回时域。最后，把反演的特征层与原特征层做共轭相乘，并再次对结果进行FFT，将直流分量置0后，求取剩余分量的能量总和，作为最后提取的实数特征图进行输出。该模块可以一定程度上减少噪声对于识别的干扰，提高叠掩检测的置信度，增加检测精度。

该特征模块基于专家知识设计，并没有引入额外需要训练的参数。根据第2节所述，ViT模型在小样本数据集上提取特征难度较大，因此本文中引入这一模块旨在通过领域中的专家知识降低模型的训练难度，进一步提高模型在小样本集上的表现。

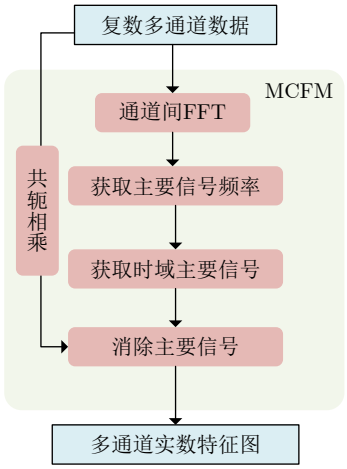


图 5 多通道特征模块流程示意图

Fig. 5 The flowchart of multi-channel feature extraction module

3.4 干涉相位特征模块(IPFM)

除了通道间的特征，多通道数据在干涉相位上也有着丰富的叠掩特征。以双通道的InSAR模型为例，在如图6所示的几何条件下，Wilkinson等人[10]计算出干涉相位的表达式如下：

$$\Delta\phi_{12}=\frac{2\pi B\cos(\theta-\alpha)\times r_p}{\lambda r\times\tan(\theta-\beta)} \tag{8}$$

其中，r_p是地物目标点P_1和P_2的斜距；B是InSAR系统的基线长度；θ是InSAR的下视角；β是点P_1与点P_2之间的坡度；α是InSAR系统的基线倾角。分析式(8)可知，在非叠掩区域，$\beta<\theta$，此时干涉相位梯度为正；而在叠掩区域，$\beta>\theta$，此时干涉相位梯度为负。由此可知，叠掩区域的干涉相位具有相位反偏的特征。

单幅干涉相位图可能由于信噪比低等多种原因而导致某些叠掩区域的相位反偏特征不明显，而使用多通道数据可以得到多个干涉相位图，更有利于特征的提取，防止遗漏。根据叠掩的相位反偏特征，本文设计了干涉相位特征模块(IPFM)，如图7所示。干涉相位特征模块首先使用编码块中的多通道复数数据计算出不同通道间的共轭相乘矩阵，矩阵中的值即为干涉相位，可以适当增加共轭矩阵的个数来提高检测的鲁棒性。之后在距离方向做多个尺度的FFT，得到了多个通道间干涉相位的频率特征图，特征图中的正负表示相位是否反偏。将多个频率特征图进行卷积得到最终的叠掩干涉相位特征图。该模块只有最后一层用到了简单的1×1卷积层，因此引入的参数数量很少，减轻了模型在小样本数据集上的训练压力，并且可以与其他特征模块形成互补，进一步提高叠掩检测的鲁棒性和准确率。

3.5 损失函数

交叉熵(Binary Cross Entropy, BCE)损失函数

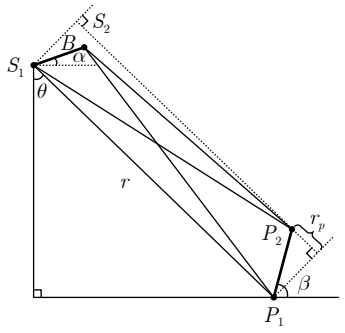


图 6 InSAR几何地理模型

Fig. 6 The InSAR geometry model of layover

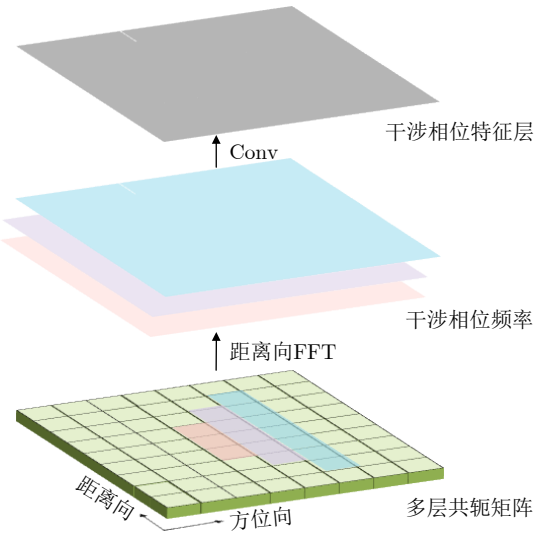


图 7 干涉相位特征模块

Fig. 7 Interference phase feature module

是解决二分类图像分割问题时常用的损失函数，其表示如下：

$$L_{\mathrm{BCE}}=-\frac{1}{M}\sum_{m}\left[y\log p_m+(1-y)\log(1-p_m)\right]$$

其中，log表示以2为底取对数；p_m是一个训练批次中预测第m个像素为正样本的概率；M是该批次中总共包含的像素总和；y是该像素的标签真值。

在建筑区叠掩检测问题中使用交叉熵损失函数存在一些问题。因为叠掩检测的训练集存在较严重的正负样本不平衡问题。这种不平衡表现在两个方面：(1)在整个场景中，非叠掩区域可能远大于叠掩区域，因为建筑区域一般在场景中所占比例较小。(2)叠掩或非叠掩区域在图像中一般是连续的，在局部的训练切片中，单一种类的区域会占据切片的大部分。这两方面的不平衡会使训练过程中的梯度剧烈变化，从而增加训练的难度，甚至使训练中的模型性能退化。

聚焦(Focal)损失函数是多分类任务中常采用的损失函数，一般可以有效地减轻样本不平衡对训练的负面影响[32]。为了提高模型的训练效果，本文对聚焦损失函数做了二元化处理得到二元聚焦(Binary Focal, BF)损失函数。二元聚焦损失函数由平衡因子α_t和调制因子$(1-p_{\mathrm{t}})^{\gamma}$相乘得到，表示如下：

$$L_{\mathrm{BF}}(p_{\mathrm{t}})=-\alpha_{\mathrm{t}}(1-p_{\mathrm{t}})^{\gamma}\log(p_{\mathrm{t}}) \tag{9}$$

其中，p_t是根据模型的估计概率p计算得到的，p_t定义如下：

$$p_{\mathrm{t}}=\begin{cases}p, & \text{叠掩}\\ 1-p, & \text{非叠掩}\end{cases} \tag{10}$$

在损失函数中，α_{t}因子平衡了叠掩与非叠掩样本的权重。在非叠掩样本上，α_{t}设定的值要小于叠掩样本，由此减轻负样本对模型学习影响。$(1-p_{\mathrm{t}})^{\gamma}$因子使模型训练更专注于困难样本而非简单样本。例如，当一个困难样本被错误分类时，p_{t}相对较小，调制因子接近于1，此时损失值几乎不受影响。反之，简单样本下被错误分类时，调制因子接近于0，以此来地降低简单样本对损失值和梯度更新的影响。通过采用该损失函数，可以使模型在正负不平衡样本集下的训练更专注于相对困难的叠掩区域而不是非叠掩区域，从而提高梯度反向传播的更新效率和检测的准确性。

在本文所提方法中，以64个像素为步长，在原始数据中滑动截取出256像素×256像素大小的训练切片。切片输入到模型后得到输出，则模型输出相对于真值的二元聚焦损失值可表示为

$$\begin{aligned}L_{\mathrm{BF}}=-\frac{1}{m}\sum_{j=1}^{m}\big(&\alpha y_j(1-\hat{y}_j)^{\gamma}\log\hat{y}_j\\&+(1-\alpha)(1-y_j)\hat{y}_j^{\gamma}\log(1-\hat{y}_j)\big)\end{aligned} \tag{11}$$

其中，y_j为切片中第j个像素的真值，y_j为1时表示该像素为叠掩目标；y_j为0时表示为非叠掩目标，$\hat{y}_j$为模型对切片中第j个像素的预测输出；m为切片包含的像素总数，即缓解正负样本不平衡问题的平衡因子；α在本模型中取值为0.75，即将叠掩区域与非叠掩区域对损失值的贡献权重调整为0.75 : 0.25，以保证更稀疏的叠掩区域的损失值不会被非叠掩区域的稀释。经多次比较实验，在本模型中取γ值为2，即调制因子为

$$|y_j-\hat{y}_j|^{\gamma}=\begin{cases}(1-\hat{y}_j)^2, & \text{叠掩}\\ \hat{y}_j^2, & \text{非叠掩}\end{cases} \tag{12}$$

无论是叠掩目标还是非叠掩目标，调制因子会使输出与真值差值较大的目标，即难样本，获得相对简单样本来说更大的损失值权重，促进对难样本的训练优化。

4 实验与分析

为了验证本文提出的结合ViT和CNN的叠掩检测网络的有效性，本节选取多个现有流行网络，包括UNet, Unet++, DeepLabV3, DeepLabV3+和ViT，与本文模型进行对比实验。实验的数据集为真实场景中采集的多通道SAR数据，通过人工标注的方式来确定真值。以上所有实验进行多次，取实验结果的平均值作为最终结果。

4.1 实验设置

实验的硬件配置采用了Intel Core i7处理器，

48 GB内部存储器，GPU处理器为NVIDIA GTX 2070Ti。实验平台为Windows 10，软件环境为Python 3.8, CUDA 11.1, CuDNN 8.7。实验以pytorch 1.11为主要的深度学习框架。训练过程中，最大训练epoch设置为200，使用Adam优化器进行参数更新，实验设置的初始学习率为0.004，50个epoch后降为0.001，100个epoch后降为3×10^{-4}。训练的批处理大小选为8。

4.2 数据集介绍

本次实验使用了真实场景数据集，能够测试模型在真实环境下的抗干扰能力与检测能力。测试数据为机载阵列InSAR系统于2022年8月在四川省峨眉山市采集的10通道阵列干涉SAR数据。实测数据的详细参数如表1所示， 叠掩的真值图由人工标注得到。图8所示为一个场景的完整SAR图像。从图8可以直观地感受到建筑叠掩在幅度、干涉相位、空间结构方面的一些特征。为了便于模型训练，对SAR图像以64像素步长的滑窗截取方式裁减为256像素×256像素大小的10通道的复数数据，以256像素×256像素大小的真值图作为标签，得到了200张标注数据集。最后以7:3的比例进行分割，得到训练集和验证集。如图9所示，列出了数据集中的切片示意图。

4.3 评价标准

实验中采用了准确率(Accuracy)、精准度(Precision)、召回率(Recall)、虚警率(False Alarming)和漏警率(Missing Alarming) 5个指标来评价模型性能。

准确率是检测正确的叠掩与非叠掩区域占总体的比率，其表达式为

$$\text{Accuracy}=\frac{\text{TP}+\text{TN}}{\text{TP}+\text{FP}+\text{TN}+\text{FN}} \tag{13}$$

精准度表示的是检测为真的叠掩中实际也为真的叠掩区域的比率，其表达式为

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}} \tag{14}$$

召回率表示的是检测为真的叠掩区域占实际为真的叠掩区域的比率，其表达式为

$$\text{Recall}=\frac{\text{TP}}{\text{TP}+\text{FN}} \tag{15}$$

虚警率表示的是实际为非叠掩却被误识别为叠掩的区域占所有检测为真的叠掩区域的比率

$$\text{False Alarm}=\frac{\text{FP}}{\text{TP}+\text{FP}} \tag{16}$$

漏警率表示的是实际为叠掩却没有被检测出来的区域占所有叠掩区域的比率，其表达式为

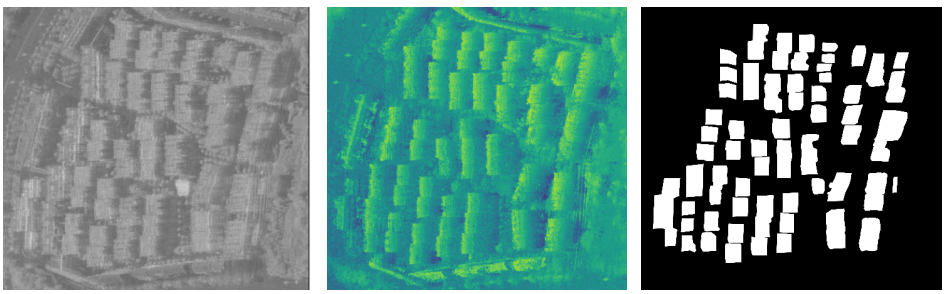
$$\text{Missing Alarm}=\frac{\text{FN}}{\text{TP}+\text{FN}} \tag{17}$$

式(13)—式(17)中，TP表示的是检测与实际都为叠掩的区域中像素点的个数；TN表示的是检测为非叠掩而实际为叠掩的区域中像素点的个数；FP表示的是检测为叠掩但实际为非叠掩的区域中像素点的个数；FN表示的是检测与实际都为非叠掩的区域中像素点的个数。

表 1 机载SAR参数

Tab. 1 The parameters of airborne SAR

参数	数值
飞行高度	5 km
飞行速度	80 m/s
波段	Ku
入射角	40°
分辨率	0.3 m



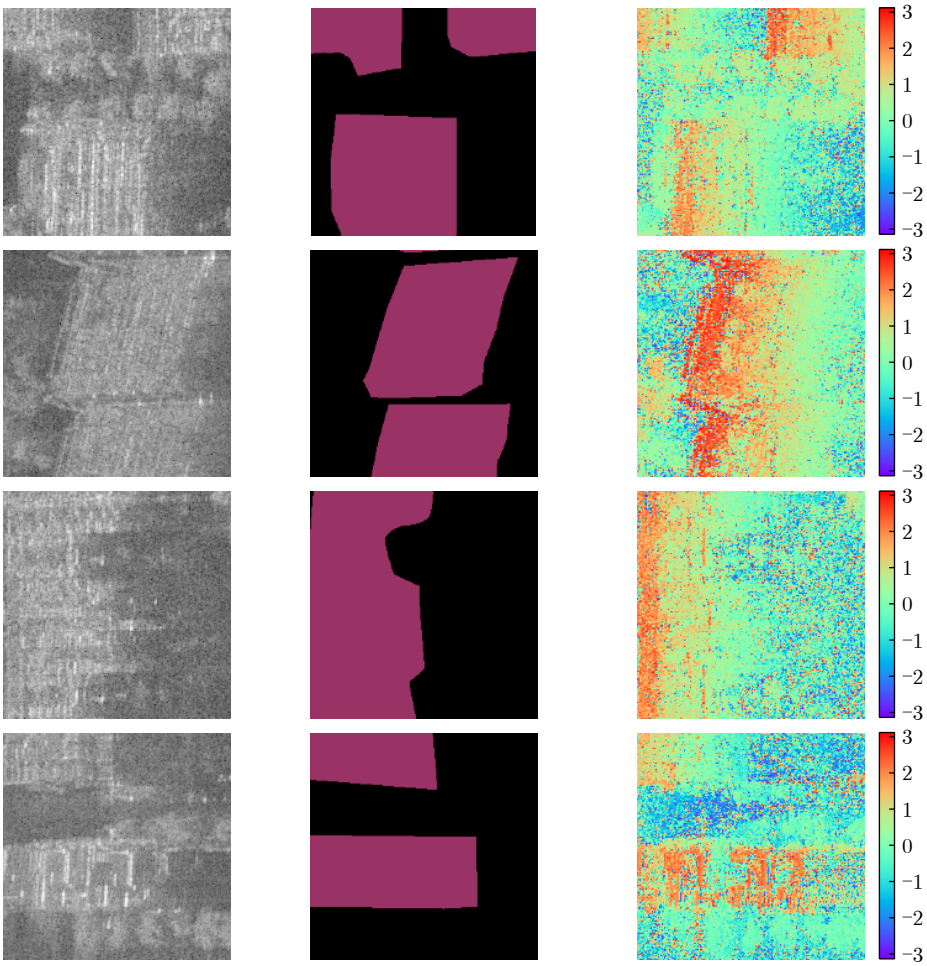
(a) 幅度图
(a) Amplitude of a single channel

(b) 干涉相位图
(b) Rainbow figure of the interference phase

(c) 人工标注的叠掩真值图白色代表叠掩区域，黑色代表非叠掩区域
(c) Denotes the layover as the white and non-layover in black

图 8 数据集场景示意图

Fig. 8 The illustration of a scene in the dataset



(a) 切片的单通道幅度图 (a) Amplitude of a single channel

(b) 叠掩真值图，红色为叠掩区域，黑色为非叠掩区域 (b) Denote the layover in red and non-layover in black

(c) 相邻两通道的干涉相位图 (c) Interference phase of two adjacent channels

图 9 数据集切片示意图

Fig. 9 Image slices of dataset

4.4 对比实验分析

(1) 与传统方法的对比实验

在上述数据集和评价标准下，本节对比了本文方法与其他传统方法的异同。实验主要选取了3种应用广泛具有代表性的传统方法：幅度法、通道间FFT、干涉相位法，结果如表2所示。幅度法是最经典的传统叠掩检测方法，其利用了叠掩区域混叠多个信号而使得幅度较高的特征，所需要的信息量最少，单幅SAR图像即可进行检测。但该方法易受干扰使得其检测精度不高。从如图10所示的检测结果来看，幅度法的检测结果受到了较强的干扰，充满了大量杂点，并且大量误检了城市区域中树木等非建筑地物。通道间FFT方法的原理是通过判断是否混叠了多个目标回波，进而对叠掩进行检测，其具体流程在3.3节进行了介绍。虽然该方法所需要的信息最多，但通道间的叠掩特征具有更强的抗干扰能力，在检测指标上均超过了幅度法。从图10来看，其检测结果有着相对幅度法更少的杂点，并能检出幅度相对较弱的叠掩目标如图10中红圈所示，大幅提高了叠掩检测的召回率。干涉相位法是利用叠掩区域干涉相位反偏的特征进行检测，具体流程如3.4节描述。从表2结果指标来看，该方法性能较差，但从图10可以观察得到，该方法检测的建筑边缘相比其他方法而言更加清晰，适合与其他方法联合起来对叠掩进行多方面特征的提取与识别。总体而言，传统的叠掩检测方法虽然检测性能相对不足，但是其包含了关于叠掩特征的专家知识，无需数据集支撑，可以将其检测的原理融入到深度学习网络中，增加模型的先验信息，降低模型的训练难度，提高检测性能。

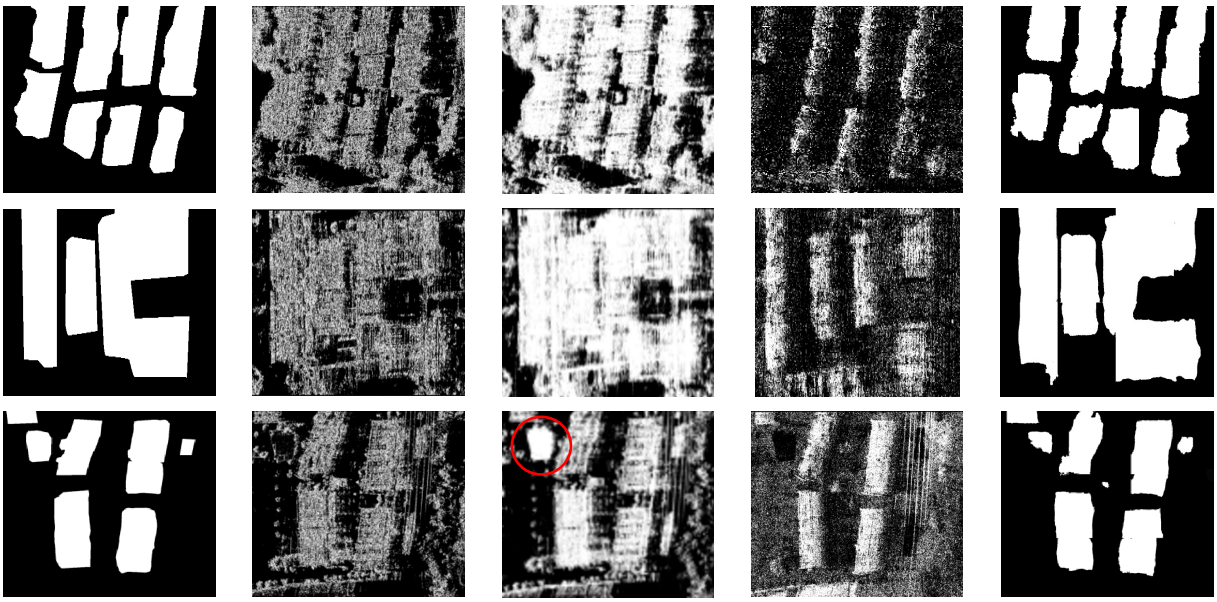
(2) 与其他深度学习方法的对比实验

本节将本文模型与多个流行的图像分割网络进行性能对比。实验结果如表3所示。通过观察可以

表 2 本文方法与传统方法对比实验结果

Tab. 2 Comparison experiment results between the proposed method and traditional methods

实验方法	准确率	精准度	召回率	虚警率	漏警率
幅度法	0.7285	0.6041	0.5912	0.3959	0.4088
通道间FFT	0.7820	0.6295	0.8231	0.3705	0.1769
干涉相位法	0.6502	0.4506	0.4311	0.5494	0.5689
本文方法	**0.9443**	**0.7619**	**0.8699**	**0.2380**	**0.1300**



(a) 真值 (a) The ground truth　(b) 幅度法 (b) Amplitude method　(c) 通道间FFT (c) Inter channel FFT　(d) 干涉相位法 (d) Interferometric phase method　(e) 本文方法 (e) Method in this article

图 10 本文方法与传统方法的叠掩检测图

Fig. 10 Layover detection of the proposed method and traditional methods

表 3 本文方法与其他深度学习算法对比实验结果

Tab. 3 Comparison experiment results between the proposed method and other deep learning methods

实验方法	准确率	精准度	召回率	虚警率	漏警率	参数量(M)
UNet	0.8976	0.7463	0.8391	0.2537	0.1609	7.8
UNet++	0.8963	0.7481	0.8382	0.2519	0.1618	9.8
DeepLabV3	0.8614	0.7112	0.7933	0.2688	0.1767	15.3
DeepLabV3+	0.8831	0.7434	0.8291	0.2566	0.1709	15.6
ViT	0.8091	0.6331	0.6783	0.3668	0.3216	8.6
本文方法	**0.9443**	**0.7619**	**0.8699**	**0.2380**	**0.1300**	**10.0**

发现，本文模型在准确率、精准度和召回率等指标上均超过了其他深度学习算法，证明了本文模型有效地通过ViT和CNN结构挖掘了叠掩的全局和局部特征，同时基于专家知识设计的专有特征模块也成功地降低了模型在小样本集上的训练难度，提高了模型的性能。注意到UNet和UNet++网络获得了次优的识别效果，超过了DeepLabV3和DeepLabV3+网络的表现，这说明在建筑叠掩检测问题上，有助于提取浅层网络中叠掩边界特征的跳接操作更有利于叠掩的检测。相比之下，在DeepLabV3和DeepLabV3+中常用的插值上采样方法则会严重丢失边界特征。这也证明了本文采用跳接连接的正确性。另一方面原因可能在于本数据集中的建筑叠掩区域之间的尺度差异较小，DeepLabV3中提取多尺度特征的空洞空间卷积池化金字塔结构(Atrous Spatial Pyramid Pooling, ASPP)并没有发挥较大作

用。比较UNet网络与UNet++网络在性能上并没有太大差距，说明在叠掩的小样本集上，单纯增加网络的稠密连接并不会对网络的性能有较大提升。ViT网络由于小样本集的缘故，无法对其进行深入有效的训练，所以其检测效果与CNN有较大差距，说明在没有海量数据支撑的前提下单纯使用ViT效果并不理想。

深度学习模型使用256像素×256像素大小的数据切片进行预测，为了更直观地感受不同方法之间检测结果的异同，将测试集中的切片重新拼接成原场景大小，如图11所示。从图11的识别结果来看，DeepLabV3由于缺乏解码模块，在有相干斑噪声影响叠掩检测中表现不佳，叠掩边界的检出率较低，极端情况下可能会造成叠掩区域的完全漏检。DeepLabV3+网络添加了解码模块，这使得其边界检测结果得到了很大改善，叠掩区域中的漏检现象也有所减少。UNet和UNet++网络对于叠掩边界的识别相对较好，但存在部分区域漏检的现象。在非叠掩区域也存在较明显的误识别问题。可以看出上述这两类CNN网络凭借卷积单元的先验信息对叠掩区域进行了有效的识别，但其对于叠掩特征的挖掘还不够充分，尤其遇到易混淆的叠掩区域时，检测效果往往不理想。ViT网络在小样本数据集下明显未得到充分训练，对于局部特征的识别效果不佳，检测结果中空洞较多，较为离散，但识别结果的轮廓信息比较明显，体现了ViT网络能有效提取建筑叠掩的大尺度结构特征。通过比较，本文方法较其他方法取得了更好的检测效果，同时较好地控制了模型的复杂度，与其他模型相比，待训练的参数量没有明显增加，甚至少于DeepLabV3和DeepLabV3+算法的参数量。如图11所示，本文方法对大部分叠掩区域都进行了有效识别，但在叠掩边界处还存在着一定的误差。在非叠掩区域中，由于综合了多方面的特征，有效避免了其他方法中出现的较严重的检测虚警。但也由于专用的特征提取模块，本文方法仅针对多通道SAR叠掩检测问题。对比其他深度学习方法，本文方法以降低模型的通用性为代价，提高了模型在多通道叠掩检测问题上的性能表现。

4.5 消融实验

为了进一步说明各模块对于检测的贡献，本节对模型中的不同模块进行消融实验，结果如表4所示。没有特征提取模块的网络本质上是一个复数UNet网络，但复数网络一般在样本不足的情况下训练难度较大，所以其检测性能略低于UNet网络。添加ViT-SSFM模块后，模型检测的精准度得到了较大的提升，召回率同样得到了改善。由4.4节的实验结果可知，单独的ViT模型检测结果并不理想，说明相对于单独的ViT或者CNN结构，ViT与CNN交替组合后可以挖掘出新的叠掩特征，很好地提升了模型的性能。为了使比较更加清晰，将同样基于专家知识的MCFM模块和IPFM模块作为一组同时添加到模型中。从结果可以看出，仅添加MCFM和IPFM模块也能较好地提升检测性能，说明对模型融入专家知识可以有效地降低模型训练的难度，提升模型的检测精度。本文方法在融合3个特征提取模块后，可以最大限度地挖掘建筑叠掩不同方面的特征，获得了更好的检测效果。

理论上，为模型引入先验的专家知识可以帮助

(a) 真值 (b) DeepLabV3 (c) DeepLabV3+ (d) UNet (e) UNet++ (f) ViT (g)本文方法

(a) The ground truth (b) DeepLabV3 (c) DeepLabV3+ (d) UNet (e) UNet++ (f) ViT (g) The proposed method

图 11 不同深度学习方法的叠掩检测图

Fig. 11 Layover detection of different deep learning methods

模型在小样本数据集上取得更好的训练效果。不同于自然光学领域的识别与分割，SAR多通道叠掩检测中一个比较突出的问题就是缺少高质量标注的样本数据。为了对少样本情况下模型的表现做进一步分析，将训练数据减少到不同百分比后观察模型性能的衰减情况。实验结果如图12所示，总体上复杂度越高的模型受到少样本的影响越大。而由于本文方法融入了基于专家知识的特征模块，模型性能受到样本减少的影响相对较小，在极限情况下，模型将退化到接近传统方法的水平，依然有可观的检测性能。所以，相较于其他方法，本文方法随着样本规模变小而衰退的程度会越来越小，证明在小样本条件下拥有更好的检测性能。

5　结语

面对城市建筑区域叠掩检测问题，本文综合了ViT和CNN两种网络的优点，提出了一种基于深度学习的叠掩检测方法。该方法设计了多个专门的叠掩特征模块，对叠掩的局部纹理特征、全局大尺度空间结构特征、通道间特征以及相位反偏特征进行了综合提取，以实现对于建筑叠掩的高精度检测。通过真实小样本数据集上的对比实验，说明该方法能有效地挖掘多通道SAR数据中叠掩的多方面特征。本文方法实现了优于现有的传统算法和其他深度学习分割网络的表现，将建筑叠掩的检测精度由80%～89%提高到了94%，有助于提高城市区域的3D SAR成像效率与质量。

表 4 消融实验结果

Tab. 4　Results of ablation experiments

ViT-SSFM	MCFM	IPFM	准确率	精准度	召回率
×	×	×	0.8891	0.7263	0.8173
√	×	×	0.9346	0.7512	0.8294
×	√	√	0.9162	0.7387	0.8516
√	√	√	**0.9443**	**0.7619**	**0.8699**

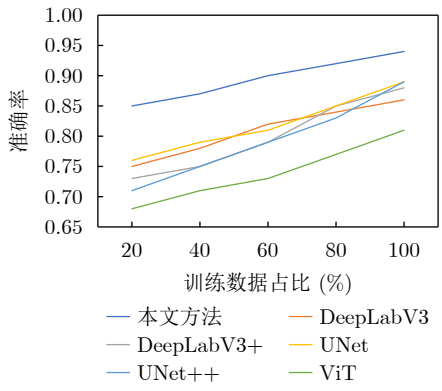


图 12 不同训练数据量下的准确率

Fig. 12　Accuracy with different proportion of training data

参 考 文 献

[1] FU Kun, ZHANG Yue, SUN Xian, *et al*. A coarse-to-fine method for building reconstruction from HR SAR layover map using restricted parametric geometrical models[J]. *IEEE Geoscience and Remote Sensing Letters*, 2016, 13(12): 2004–2008. doi: 10.1109/LGRS.2016.2621054.

[2] CHENG Kou, YANG Jie, SHI Lei, *et al*. The detection and information compensation of SAR layover based on R-D model[C]. IET International Radar Conference 2009, Guilin, China, 2009: 1–3. doi: 10.1049/cp.2009.0346.

[3] 彭学明, 王彦平, 谭维贤, 等. 基于跨航向稀疏阵列的机载下视MIMO 3D-SAR三维成像算法[J]. 电子与信息学报, 2012, 34(4): 943–949. doi: 10.3724/SP.J.1146.2011.00720.

PENG Xueming, WANG Yanping, TAN Weixian, *et al*. Airborne downward-looking MIMO 3D-SAR imaging algorithm based on cross-track thinned array[J]. *Journal of Electronics & Information Technology*, 2012, 34(4): 943–949. doi: 10.3724/SP.J.1146.2011.00720.

[4] 郭睿, 臧博, 彭树铭, 等. 高分辨InSAR中的城市高层建筑特征提取[J]. 西安电子科技大学学报, 2019, 46(4): 137–143. doi: 10.19665/j.issn1001-2400.2019.04.019.

GUO Rui, ZANG Bo, PENG Shuming, *et al*. Extraction of features of the urban high-rise building from high resolution InSAR data[J]. *Journal of Xidian University*, 2019, 46(4): 137–143. doi: 10.19665/j.issn1001-2400.2019.04.019.

[5] 田方, 扶彦, 刘辉, 等. 多输入多输出下视阵列SAR姿态角误差分析[J]. 测绘科学, 2020, 45(9): 65–71, 110. doi: 10.16251/j.cnki.1009-2307.2020.09.011.

TIAN Fang, FU Yan, LIU Hui, *et al*. Attitude angle error analysis of MIMO downward-looking array SAR[J]. *Science of Surveying and Mapping*, 2020, 45(9): 65–71, 110. doi: 10.16251/j.cnki.1009-2307.2020.09.011.

[6] 冯荻. 高分辨率SAR建筑目标三维重建技术研究[D]. [博士论文], 中国科学技术大学, 2016: 75–99.

FENG Di. Research on three-dimensional reconstruction of buildings from high-resolution SAR data[D]. [Ph. D. dissertation], University of Science and Technology of China, 2016: 75–99.

[7] 韩晓玲, 毛永飞, 王静, 等. 基于多基线InSAR的叠掩区域高程重建方法[J]. 电子测量技术, 2012, 35(4): 66–70, 85. doi: 10.3969/j.issn.1002-7300.2012.04.019.

HAN Xiaoling, MAO Yongfei, WANG Jing, *et al*. DEM reconstruction method in layover areas based on multi-baseline InSAR[J]. *Electronic Measurement Technology*, 2012, 35(4): 66–70, 85. doi: 10.3969/j.issn.1002-7300.2012.04.019.

[8] SOERGEL U, THOENNESSEN U, BRENNER A, *et al*. High-resolution SAR data: New opportunities and challenges for the analysis of urban areas[J]. *IEE*

Proceedings – Radar, Sonar and Navigation, 2006, 153(3): 294–300. doi: 10.1049/ip-rsn:20045088.

[9] PRATI C, ROCCA F, GUARNIERI A M, *et al*. Report on ERS-1 SAR interferometric techniques and applications[J]. *ESA Study Contract Report*, 1994: 3–7439.

[10] WILKINSON A J. Synthetic aperture radar interferometry: A model for the joint statistics in layover areas[C]. The 1998 South African Symposium on Communications and Signal Processing-COMSIG'98 (Cat. No. 98EX214), Rondebosch, South Africa, 1998: 333–338. doi: 10.1109/COMSIG.1998.736976.

[11] CHEN Wei, XU Huaping, and LI Shuang. A novel layover and shadow detection method for InSAR[C]. 2013 IEEE International Conference on Imaging Systems and Techniques (IST), Beijing, China, 2013: 441–445. doi: 10.1109/IST.2013.6729738.

[12] WU H T, YANG J F, and CHEN F K. Source number estimator using Gerschgorin disks[C]. IEEE International Conference on Acoustics, Speech and Signal Processing, Adelaide, Australia, 1994: IV/261–IV/264. doi: 10.1109/ICASSP.1994.389826.

[13] WU Yunfei, ZHANG Rong, and ZHAN Yibing. Attention-based convolutional neural network for the detection of built-up areas in high-resolution SAR images[C]. IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 2018: 4495–4498. doi: 10.1109/IGARSS.2018.8518463.

[14] WU Yunfei, ZHANG Rong, and LI Yue. The detection of built-up areas in high-resolution SAR images based on deep neural networks[C]. The 9th International Conference on Image and Graphics, Shanghai, China, 2017: 646–655. doi: 10.1007/978-3-319-71598-8_57.

[15] CHEN Jiankun, QIU Xiaolan, DING Chibiao, *et al*. CVCMFF Net: Complex-valued convolutional and multifeature fusion network for building semantic segmentation of InSAR images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 60: 5205714. doi: 10.1109/TGRS.2021.3068124.

[16] 崔紫维. 基于Transformer框架的地基SAR边坡监测相位分类方法研究[D]. [硕士论文], 北方工业大学, 2022: 1–63. doi: 10.26926/d.cnki.gbfgu.2022.000002.

CUI Ziwei. Phase classification method of ground-based SAR slope monitoring based on transformer framework[D]. [Master dissertation], North China University of Technology, 2022: 1–63. doi: 10.26926/d.cnki.gbfgu.2022.000002.

[17] 李文娜, 张顺生, 王文钦. 基于Transformer网络的机载雷达多目标跟踪方法[J]. 雷达学报, 2022, 11(3): 469–478. doi: 10.12000/JR22009.

LI Wenna, ZHANG Shunsheng, and WANG Wenqin. Multitarget-tracking method for airborne radar based on a transformer network[J]. *Journal of Radars*, 2022, 11(3): 469–478. doi: 10.12000/JR22009.

[18] AZAD R, AL-ANTARY M T, HEIDARI M, *et al*. TransNorm: Transformer provides a strong spatial normalization mechanism for a deep segmentation model[J]. *IEEE Access*, 2022, 10: 108205–108215. doi: 10.1109/ACCESS.2022.3211501.

[19] DONG Hongwei, ZHANG Lamei, and ZOU Bin. Exploring vision transformers for polarimetric SAR image classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 5219715. doi: 10.1109/TGRS.2021.3137383.

[20] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, *et al*. An image is worth 16x16 words: Transformers for image recognition at scale[C]. The 9th International Conference on Learning Representations, Vienna, Austria, 2021: 1–20.

[21] JADERBERG M, SIMONYAN K, ZISSERMAN A, *et al*. Spatial transformer networks[C]. The 28th International Conference on Neural Information Processing Systems, Montreal, Canada, 2015: 2017–2025.

[22] 张激钟. SAR图像舰船目标快速检测识别技术[D]. [硕士论文], 电子科技大学, 2022. doi: 10.27005/d.cnki.gdzku.2022.002606.

ZHANG Lianzhong. Fast detection and recognition of ship targets in SAR images[D]. [Master dissertation], University of Electronic Science and Technology of China, 2022. doi: 10.27005/d.cnki.gdzku.2022.002606.

[23] VASWANI A, SHAZEER N, PARMAR N, *et al*. Attention is all you need[C]. The 31st Conference on Neural Information Processing Systems, Long Beach, USA, 2017: 6000–6010.

[24] LIU Ze, LIN Yutong, CAO Yue, *et al*. Swin transformer: Hierarchical vision transformer using shifted windows[C]. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, Canada, 2021: 9992–10002. doi: 10.1109/ICCV48922.2021.00986.

[25] HOCHREITER S, BENGIO Y, FRASCONI P, *et al*. Gradient Flow in Recurrent Nets: The Difficulty of Learning Long-term Dependencies[M]. KOLEN J F, KREMER S C. A Field Guide to Dynamical Recurrent Neural Networks. New York: Wiley-IEEE Press, 2001: 401–403.

[26] 王万良, 王铁军, 陈嘉诚, 等. 融合多尺度和多头注意力的医疗图像分割方法[J]. 浙江大学学报:工学版, 2022, 56(9): 1796–1805. doi: 10.3785/j.issn.1008-973X.2022.09.013.

WANG Wanliang, WANG Tiejun, CHEN Jiacheng, *et al*. Medical image segmentation method combining multi-scale and multi-head attention[J]. *Journal of Zhejiang*

University:Engineering Science, 2022, 56(9): 1796–1805. doi: 10.3785/j.issn.1008-973X.2022.09.013.

[27] BASELICE F, FERRAIOLI G, and PASCAZIO V. DEM reconstruction in layover areas from SAR and auxiliary input data[J]. *IEEE Geoscience and Remote Sensing Letters*, 2009, 6(2): 253–257. doi: 10.1109/LGRS.2008.2011287.

[28] WANG Bin, WANG Yanping, HONG Wen, *et al*. Application of spatial spectrum estimation technique in multibaseline SAR for layover solution[C]. 2008 IEEE International Geoscience and Remote Sensing Symposium, Boston, USA, 2008: III-1139–III-1142. doi: 10.1109/IGARSS.2008.4779556.

[29] REIGBER A and MOREIRA A. First demonstration of airborne SAR tomography using multibaseline l-band data[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2000, 38(5): 2142–2152. doi: 10.1109/36.868873.

[30] FORNARO G, SERAFINO F, and SOLDOVIERI F. Three-dimensional focusing with multipass SAR data[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2003, 41(3): 507–517. doi: 10.1109/TGRS.2003.809934.

[31] GUILLASO S and REIGBER A. Scatterer characterisation using polarimetric SAR tomography[C]. 2005 IEEE International Geoscience and Remote Sensing Symposium, Seoul, Korea (South), 2005: 2685–2688. doi: 10.1109/IGARSS.2005.1525619.

[32] LIN T Y, GOYAL P, GIRSHICK R, *et al*. Focal loss for dense object detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(2): 318–327. doi: 10.1109/TPAMI.2018.2858826.

作 者 简 介

田　野，博士生，主要研究方向为多通道SAR叠掩检测与深度学习。

丁赤飚，博士，研究员，中国科学院院士，主要研究方向为合成孔径雷达、遥感信息处理和应用系统等。

张福博，博士，副研究员，主要研究方向为SAR三维成像技术和高分辨率宽测绘带成像技术等。

石民安，硕士生，主要研究方向为微波成像与人工智能。

(责任编辑：高山流水)