



工学硕士学位论文

**基于视觉提示驱动与边缘感知的冰川遥感变化
检测方法研究**

Research on Remote Sensing Change Detection Method
of Glacier based on Visual Prompt Driven and Edge-
Awareness

作 者 石智深

导 师 崔建明 教授

山东科技大学

二〇二五年六月

中图分类号_____

学校代码_____

UDC_____

密 级_____

山东科技大学

工学硕士学位论文

基于视觉提示驱动与边缘感知的冰川遥感

变化检测方法研究

Research on Remote Sensing Change Detection Method of
Glacier based on Visual Prompt Driven and Edge-Awareness

作 者 石智深 入学时间 2022 年 9 月

导 师 崔建明 职 称 教授

申请学位 工学硕士学位 所在学院 计算机科学与工程学院

学科（类别） 计算机应用技术 方向（领域） 计算机应用技术

答辩日期 答辩委员会主席

答辩委员会成员

论文提交日期

学位论文使用授权声明

本人完全了解山东科技大学有关保留、使用学位论文的规定，同意本人所撰写的学位论文的使用授权按照学校的管理规定处理。

作为申请学位的条件之一，学校有权保留学位论文并向国家有关部门或其指定机构送交论文的电子版和纸质版；有权将学位论文的全部或部分内容编入有关数据库发表，并可以以电子、网络及其他数字媒体形式公开出版；允许学校档案馆和图书馆保留学位论文的纸质版和电子版，可以使用影印、缩印或扫描等复制手段保存和汇编学位论文；为教学和科研目的，学校档案馆和图书馆可以将公开的学位论文作为资料在档案馆、图书馆等场所或在校园网上供校内师生阅读、浏览。

（保密的学位论文在解密后适用本授权）

作者签名：

日 期： 年 月 日

导师签名：

日 期： 年 月 日

学位论文原创性声明

本人呈交给山东科技大学的学位论文,除所列参考文献和世所公认的文献外,全部是本人攻读学位期间在导师指导下的研究成果。除文中已经标明引用的内容外,本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出贡献的个人和集体,均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

若有不实之处,本人愿意承担相关法律责任。

学位论文作者签名:

年 月 日

学位论文审查认定书

研究生石智深在规定的学习年限内，按照培养方案及个人培养计划，完成了课程学习，成绩合格，修满规定学分；在我的指导下完成本学位论文，论文中的观点、数据、表述和结构为我所认同，论文撰写格式符合学校的相关规定，同意将本论文作为申请学位论文。

导师签名：

日 期：

摘要

在全球气候变化日益加剧的背景下，冰川变化检测成为一项紧迫的科学任务。冰川作为气候变化的敏感指示器，其快速退缩不仅推动海平面上升，还威胁依赖冰川融水的生态系统和人类社会，尤其在干旱和高山地区。鉴于冰川多位于偏远且人迹罕至的区域，传统监测方法受限，而遥感技术凭借其大范围覆盖、高时间分辨率及对恶劣环境的适应性，成为冰川监测的理想工具。深度学习技术结合遥感技术成为冰川变化检测的重要手段，然而，成像条件差异导致的光谱不一致性以及冰川边界的复杂模糊性，然而，现阶段冰川变化检测仍面临一些挑战：一方面，由于冰川复杂的地理形态，以及光照角度、成像条件等的限制，冰川地形中分布着大量不规则的阴影，导致变化区域提取难度变大，精确度难以提升；另一方面，冰川边界的复杂性和模糊性，冰川的边缘通常不是清晰的几何线条，而是受到碎屑覆盖、冰碛湖、融水通道或地形起伏的影响，呈现出不规则且模糊的特征。这种边界的不确定性导致难以准确定位冰川范围，增加了在边缘区域进行研究和监测的难度，上述问题严重制约了遥感影像在冰川变化检测中的应用。

(1) 针对由于光照角度、成像条件等的限制，冰川地形中分布着大量不规则的阴影导致的准确性降低的问题，本文提出了一种基于视觉提示的冰川变化检测网络 (Visual Prompt Driven Network for Glacier Change Detection, VPGCD-Net)，该网络基于视觉提示工程和 Transformer 架构，采用双支结构设计。一个分支利用阈值分割和特征差分生成的视觉提示引导检测过程，另一个分支通过 ResNet18 主干和 Transformer 块提取时空特征。特征线性调制 (FiLM) 模块在视觉提示指导下优化特征表达，有效减少阴影等伪影引起的误检，提升变化区域的语义一致性。

(2) 针对冰川的不规则形状以及模糊的边界问题，本文提出了基于边缘感知的冰川变化检测网络 (Edge-aware glacier change detection networks, EACD-Net)，该网络专注于边缘信息增强，基于卷积主干网络主干结合边缘感知模块 (EAM)，利用 Sobel 算子和跨层级注意力机制建模多尺度边缘特征。变化检测解码器通过多尺度特征聚合和变化注意力机制 (CAM) 增强时序对比，精准捕捉复杂边界变化，提高检测精度。结果表明，EACD-Net 能够精确检测冰川的复杂形状和边界信息，从而准确的识别位于边缘处的微小变化冰川区域。

关键词：冰川变化检测；遥感影像；深度学习；视觉提示；边缘感知

Abstract

In the context of increasing global climate change, glacier change detection has become an urgent scientific task. As sensitive indicators of climate change, the rapid retreat of glaciers not only pushes up the sea level, but also threatens the ecosystems and human societies that depend on glacier meltwater, especially in arid and high mountain areas. Given that glaciers are mostly located in remote and inaccessible areas, traditional monitoring methods are limited, while remote sensing technology has become an ideal tool for glacier monitoring due to its wide coverage, high temporal resolution, and adaptability to harsh environments. Deep learning technology combined with remote sensing technology has become an important means of glacier change detection, however, the spectral inconsistency caused by differences in imaging conditions and the complexity and ambiguity of the glacier boundaries, however, at this stage, glacier change detection is still faced with a number of challenges: on the one hand, due to the complex geographic morphology of the glacier, as well as the limitations of the angle of the light, imaging conditions, etc., the distribution of a large number of irregular shadows in the glacier topography, which leads to changes in the region extraction becomes more difficult, and the accuracy is difficult to improve; on the other hand, the complexity and ambiguity of the glacier boundary, the edge of the glacier is usually not a clear geometric line, but is affected by the debris cover, moraine lakes, meltwater channels, or topographic ups and downs, which presents irregular and ambiguous features. This uncertainty of the boundary leads to difficulties in accurately locating the glacier extent, which increases the difficulty of research and monitoring in the edge area, and the above problems seriously restrict the application of remote sensing images in the detection of glacier changes.

(1) Aiming at the problem of reduced accuracy due to the large number of irregular shadows distributed in the glacier terrain due to the limitations of light angle and imaging conditions, this paper proposes a Visual Prompt Driven Network for Glacier Change Detection (VPGCD-Net), which is based on visual cue engineering and Transformer architecture and designed with a two-branch structure. One branch utilizes visual cues generated by threshold segmentation and feature differencing to guide the detection process, while the other branch extracts spatio-temporal features through the ResNet18 backbone and Transformer block. The feature linear modulation (FiLM) module optimizes the feature expression under the guidance of visual cues, effectively reduces false detections caused by shadows and other artifacts, and improves the

semantic consistency of the changing regions.

(2) Aiming at the irregular shape of glaciers and fuzzy boundaries, this paper proposes Edge-aware glacier change detection networks (EACD-Net), which focuses on edge information enhancement, and is based on the convolutional backbone network backbone combined with the edge awareness module (EAM), using the Sobel algorithm, which models multi-scale edge features using Sobel's algorithm and cross-layer attention mechanism. The change detection decoder enhances temporal contrast through multi-scale feature aggregation and change attention mechanism (CAM) to accurately capture complex boundary changes and improve detection accuracy. The results show that EACD-Net is able to accurately detect the complex shape and boundary information of glaciers, thus accurately recognizing the tiny changing glacier regions located at the edges.

Keywords: glacier change detection; remote sensing imagery; deep learning; visual prompt; edge aware

目 录

图清单	I
表清单	III
变量注释表	IV
1 绪论	1
1.1 研究背景与意义.....	1
1.2 国内外研究现状.....	2
1.3 问题的提出.....	5
1.4 主要研究内容.....	6
1.5 本文组织结构.....	7
2 相关技术	8
2.1 Vision Tansformer 网络.....	8
2.2 卷积神经网络.....	12
2.3 注意力机制.....	14
2.4 特征线性调制.....	18
2.5 视觉提示工程.....	20
2.6 本章小结.....	21
3 基于视觉提示的冰川变化检测方法	22
3.1 引言.....	22
3.2 冰川变化检测流程.....	22
3.3 基于视觉提示的冰川变化检测网络.....	24
3.4 实验设计.....	27
3.5 实验结果与分析.....	32
3.6 本章小结.....	39
4 基于边缘感知的冰川变化检测方法	40
4.1 引言.....	40
4.2 基于边缘感知的冰川变化检测网络.....	40
4.3 实验结果与分析.....	46
4.4 本章小结.....	51
5 总结与展望	53
5.1 总结.....	53

5.2 展望.....	53
参考文献	56
作者简介	61
致谢	62
学位论文数据集	63

Contents

List of Figures.....	I
List of tables	III
List of Variables	IV
1 Introduction.....	1
1.1 Research background and significance	1
1.2 Current Domestic and International Research Landscape	2
1.3 Statement of the Problem.....	5
1.4 Main Research Content.....	6
1.5 The Structure of This Paper	7
2 Related techniques	8
2.1 Vision Transformer Network	8
2.2 Multi-Head Self-Attention Mechanism	10
2.2 Convolutional Neural Network.....	12
2.3 Attention mechanisms.....	14
2.3 Feature-wise Linear Modulation.....	18
2.4 Visual Prompt Engineering	20
2.5 Chapter Summary	21
3 Visual Prompt Driven Method for Glacier Change Detection.....	22
3.1 Introduction.....	22
3.2 Glacier Change Detection Workflow	22
3.3 Visual Prompt Driven Network for Glacier Change Detection	24
3.4 Design of experiments	27
3.5 Experimental Results and Analysis.....	32
3.6 Chapter Summary	39
4 Glacier Change Detection Method based on Edge Awareness.....	40
4.1 Introduction.....	40
4.2 Multi-scale Edge Information Enhancement Network	40
4.3 Experimental Results and Analysis.....	46
4.4 Chapter Summary	51
5 Summary and Prospects.....	53

5.1 Summary	53
5.2 Prospects	53
References	56
Author's Resume.....	61
Acknowledgements	62
Dissertation Data Collection	63

图清单

图序号	图名称	页码
图 1.1	双时相冰川遥感影像	1
Fig.1.1	Bitemporal remote sensing imagery of glaciers	1
图 1.2	同一地点不同时间拍摄的冰川影像对比	5
Fig.1.2	Comparison of glacier images taken at different times at the same location	5
图 1.3	冰川形状的复杂性	6
Fig.1.3	Complexity of glacier shapes	6
图 2.1	Vision Transformer 网络模型示意图	9
Fig.2.1	Vision Transformer network model diagram	9
图 2.2	Transformer 块的内部结构示意图	10
Fig.2.2	Specific structure of Transformer Block	10
图 2.3	可变形卷积示意图	13
Fig.2.3	Schematic diagram of Deformable Convolutional.	13
图 2.4	空间注意力模块示意图	16
Fig.2.4	Schematic diagram of the spatial attention module	16
图 2.5	SE 模块结构示意图	17
Fig.2.5	Schematic diagram of the SE module structure.	17
图 2.6	CBAM 模块结构示意图	17
Fig.2.6	CBAM module structure schematic.	17
图 2.7	通道注意力结构示意图	18
Fig.2.7	Schematic diagram of channel attention structure	18
图 2.8	特征线性调制过程示意图	19
Fig.2.8	Schematic diagram of the characteristic linear modulation process	19
图 3.1	基于视觉提示的冰川变化检测网络流程图	23
Fig.3.1	Flowchart of glacier change detection network based on visual prompt	23
图 3.2	VPGCD-Net 网络结构示意图	24
Fig.3.2	Schematic diagram of VPGCD-Net network structure.	24
图 3.3	特征线性调制过程示意图	26
Fig.3.3	Schematic diagram of the characteristic linear modulation process	26
图 3.4	两个地区的冰川局部遥感图像。	29
Fig.3.4	Localized remote sensing images of glaciers from two regions.	29
图 3.5	VPGCD-Net 变化检测结果可视化	32
Fig.3.5	Visualization of VPGCD-Net change detection results	32
图 3.6	VPGCD-Net 与传统方法的实验结果定性分析对比	34
Fig.3.6	Comparison of qualitative analysis of experimental results	34

	between VPGCD-Net and traditional methods	
图 3.7	VPGCD-Net 与深度学习方法实验结果定性分析对比	36
Fig.3.7	Qualitative analysis of experimental results comparing VPGCD-Net with deep learning methods	36
图 3.8	VPGCD-Net 中关键模块消融实验定性分析	38
Fig.3.8	Qualitative analysis of ablation experiments of key modules in VPGCD-Net	38
图 3.9	阈值参数对网络性能的影响分析性能	38
Fig.3.9	Analysis of Threshold Parameter Effects on Network Performance.	38
图 4.1	EACD-Net 网络结构示意图	41
Fig.4.1	Schematic diagram of EACD-Net network structure.	41
图 4.2	边缘注意力门控机制结构示意图	43
Fig.4.2	Schematic diagram of the structure of the edge attention gating mechanism	43
图 4.3	多尺度聚合模块结构示意图	44
Fig.4.3	Schematic diagram of the structure of the multi-scale aggregation module	44
图 4.4	混合模块的结构示意图	45
Fig.4.4	Schematic structure of the Mix Block	45
图 4.5	EACD-Net 与传统方法的实验结果定性分析对比	47
Fig.4.5	Comparison of qualitative analysis of experimental results between EACD-Net and traditional methods	47
图 4.6	EACD-Net 与深度学习方法的实验结果定性分析对比	48
Fig.4.6	Comparison of qualitative analysis of experimental results between EACD-Net and deep learning methods	48
图 4.7	EACD-Net 中关键模块消融实验定性分析	51
Fig.4.7	Qualitative analysis of key module ablation experiments in EACD-Net	51

表清单

表序号	表名称	页码
表 3.1	数据集介绍	29
Table 3.1	Introduction of the dataset	29
表 3.2	VPGCD-Net 与传统变化检测方法的定量对比	34
Table 3.2	Quantitative comparison of VPGCD-Net and traditional change detection methods	34
表 3.3	VPGCD-Net 与深度学习冰川变化检测方法结果的定量对比	36
Table 3.3	Quantitative comparison of results from VPGCD-Net and deep learning glacier change detection methods	36
表 3.4	VPGCD-Net 中关键模块消融实验定量分析	37
Table 3.4	Quantitative analysis of key module ablation experiments in VPGCD-Net	37
表 4.1	EACD-Net 与传统变化检测方法的定量对比	48
Table 4.1	Quantitative comparison of EACD-Net and traditional change detection methods	48
表 4.2	EACD-Net 与深度学习变化检测方法的定量对比	49
Table 4.2	Quantitative comparison of EACD-Net and deep learning change detection	49
表 4.3	EACD-Net 中关键模块消融实验定量分析	51
Table 4.3	Quantitative analysis of key module ablation experiments in EACD-Net	51

变量注释表

变量	注释	初现页
ch_{in}	输入通道数=	48
ch_{out}	输出通道数	48
$CAttn_i$	T1 时相和 T2 时相的特征相似度	49
D	双时相差异图	28
E_{T1}^i	T1 时相影像在边缘感知模块第 <i>i</i> 阶段的边缘响应图	47
T2	第二时相图遥感影像	28
F_{masked}	拼接后的特征图	29
F_{visual_prompt}	视觉提示特征向量	29
$F_{l;d}$	第 <i>d</i> 维的第 <i>l</i> 个特征	30
$feat_{T1}^1$	T1 时相影像在特征编码器第 1 阶段的输出特征	46
$feat_{T1}^2$	T1 时相影像在特征编码器第 2 阶段的输出特征	46
$feat_{T1}^3$	T1 时相影像在特征编码器第 3 阶段的输出特征	46
$feat_{T1}^4$	T1 时相影像在特征编码器第 4 阶段的输出特征	46
$feat_{T1}^i$	T1 时相影像在特征编码器第 <i>i</i> 阶段的输出特征	46
F'	经过 Sigmoid 函数激活后的特征	47
F''	经过混合注意力之后的特征	47
F^i	第 <i>i</i> 级处理后的融合特征	47
$f'eat_{T1}^{i-1}$	第 <i>i</i> 级处理后的增强特征	47
F_{hybrid}^i	第 <i>i</i> 级加权后的融合特征	50
G1	第一时相影像灰度图	29
G2	第二时相影像灰度图	29
G_x	水平方向卷积核	47
G_y	垂直方向卷积核	47
L_{seg}	动态自适应的交叉熵损失函数	31
$L_{bilinear}$	双线性插值算子	31
L_{DACE}	修正后的损失函数	31
\mathcal{L}_{total}	总损失函数	50
\mathcal{L}_{change}	变化区域损失函数	50
\mathcal{L}_{wBCE}	加权二元交叉熵损失函数	50
\mathcal{L}_{Dice}	<i>Dice</i> 系数损失函数	50
\mathcal{L}_{align}	特征对齐损失函数	51
P	预测张量	31
M	阈值分割掩码图	29
Mix_i	第 <i>i</i> 级的混合特征	49
N_c	类别为 <i>c</i> 的有效像素数	31
$P_{i,j}^{(b,c)}$	表示第 <i>b</i> 个样本中位置(<i>i,j</i>)的像素属于类别 <i>c</i> 的预测概率	31

$T1$	第一时相图遥感影像	28
Up_0	上采样特征	50
V_i^1	T1 时相影像的归一化特征向量矩阵	49
V_i^2	T2 时相影像的归一化特征向量矩阵	49
w_p	动态权重	50
x	影像像素x坐标	29
y	影像像素x坐标	29
$Y_{i,j}^{(b)}$	表示第 b 个样本中位置(i, j)的像素的标注值	31
$y_{x,y}$	像素点(x, y)的真实标签值	50
α	比例系数	50
β_d	d 维线性调制因子	30
β_d	d 维线性调制因子	30
γ_d	d 维线性调制因子	30
ε	数值稳定因子	50
ϵ	平滑因子	31
θ	差异图分割阈值	29
$\Omega_{nochange}$	未变化区域的像素坐标集合	51

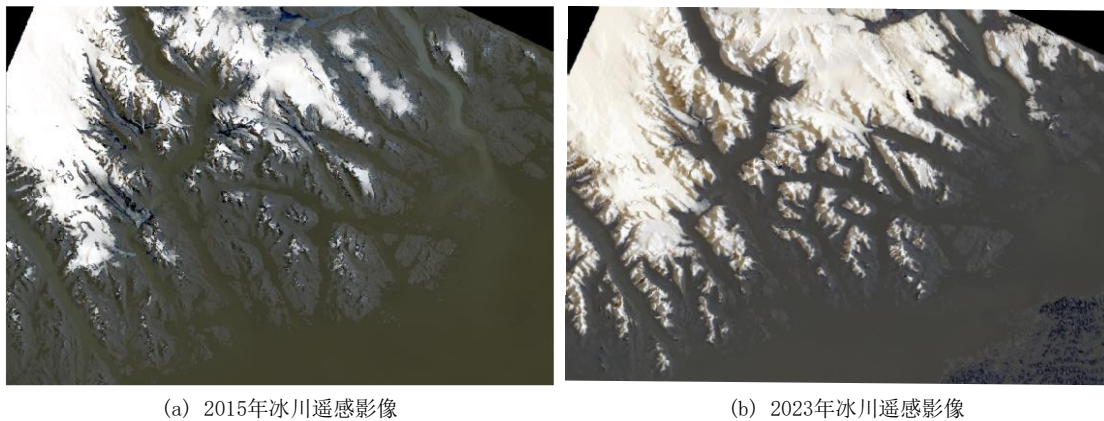
1 绪论

1 Introduction

1.1 研究背景与意义 (Research background and significance)

冰川作为自然界重要的水资源和地质景观,其变化不仅是气候变化的显著指示器,也涉及到多个科学领域^[1]。冰川的动态变化反映了地球系统各个子系统的交互作用,包括气候变化、冰雪物理过程、地质活动以及水文循环等。

冰川变化是气候变化的指示器,近年来,全球气候变暖已成为全球环境研究的焦点。冰川的缩小或消融现象是气候变暖的直接结果。气温升高导致冰川融化的速度加快,许多冰川出现退缩现象,这种变化在全球范围内普遍存在^[2]。通过监测冰川的变化,可以为评估气候变化的程度和趋势提供直接证据;冰川的变化是长期气候变化的直接反映,因此,冰川作为气候监测工具具有重要意义^[3]。冰川的存在和退缩能为研究全球气候变化提供更直观、长期的数据支持,帮助我们更好地理解 and 预测气候变化的影响。如图 1.1 所示为双时相冰川遥感影像。



(a) 2015年冰川遥感影像

(b) 2023年冰川遥感影像

图 1.1 双时相冰川遥感影像

Fig. 1.1 Bitemporal remote sensing imagery of glaciers.

冰川在全球水循环扮演这关键角色,冰川是地球上最大的淡水储备体,尤其是在一些依赖冰川融水的区域(如喜马拉雅山脉的流域地区)。这些地区的河流受到冰川融水的影响,融水成为这些流域在干旱季节的重要水源^[4]。随着冰川退缩和融水量的变化,相关地区的水资源供给可能会受到威胁。冰川退缩不仅影响其本身的融水量,还可能改变下游河流的水文过程。在早期阶段,冰川退缩可能增加融水量,但随着冰川进一步消融,融水量将减少,导致下游水资源供应的减少,影响农业灌溉、饮用水供应和生态保护等方面。此外,冰川的融化是全球海平面上升的主要原因之一。尤其是极地地区的冰盖,如格陵兰冰盖和南极冰盖,其融化量直接贡献于海平面上升^[5]。冰川的变化不仅影响近海地区,还对全球气候系统产生深远影响,进而影响全球气候模式和人类社会^[6]。同时,冰川的退缩

改变了海水的分布和温度,影响海洋的环流模式和全球气候系统。通过研究冰川变化与海平面变化之间的关系,可以揭示冰川对海洋生态系统和气候模式的复杂影响^[7]。

冰川变化不仅是气候变化的直接体现,也是全球水循环、生态系统演变及海平面变化的重要因素。随着深度学习等现代技术的发展,冰川变化的监测变得更加精确,能够为气候变化研究、生态环境保护和水资源管理提供强有力的支持。因此,研究冰川变化的科学和生态意义不仅有助于深化我们对自然环境的认识,还有助于预测和应对气候变化带来的挑战^[8]。

在 1930s 以前,冰川末端的变化研究一直依靠实地对冰川末端若干固定点进行定期测量、制图以及计算冰川长度的变化,测量精度一般在几米以内^[9]。由于冰川实测耗资巨大且异常艰辛,截止到 1984 年,全球只有 25 条冰川进行了 50 年左右的实地连续观测^[10]。此外人们还通过构建冰川模型来模拟冰川的流动、融化和积累过程,使用遥感数据结合现场观测,强调了现场观测在验证模型和提高冰盖变化预测准确性方面的作用。模型模拟可以帮助预测冰川未来的变化趋势,但需要准确的气候和冰川参数输入。以上对于冰川变化的相关研究,多是聚焦于流速、厚度以及冰川对于气候变化的影响,使用传统的遥感监测手段,这种手段虽然精度高,但同时也耗费了大量的人力物力资源。

虽然近几年有研究学者结合卷积神经网络进行研究,但都是聚焦于单时相的冰川提取,冰川边界提取等,并没有结合深度学习聚焦于冰川变化监测方面的研究,卫星遥感技术具有拍摄范围广泛,获取信息速度快、周期短、受限条件限制少等优点,因此将遥感技术与深度学习相结合进行冰川变化检测研究具有更高的研究价值。同时,通过数据分析,遥感研究者能够实时获取冰川区的遥感数据,密切监测冰川区的变化信息,为决策提供有力支持。因此,利用遥感影像进行冰川变化的动态监测具有重要的意义。

1.2 国内外研究现状 (Current Domestic and International Research Landscape)

相比于城市农业等传统的变化检测应用领域,极地冰川的变化也同样吸引着研究者的关注,由于极地冰川位置的特殊性,遥感技术更成为冰川监测的有力工具,伴随着遥感技术的不断发展,使用卫星图像和航空照片等遥感数据来监测冰川面积、厚度和流动速度的变化,以评估气候变化对冰川的影响,根据方法的不同可分为两大类,分别为传统图像变化检测方法和基于深度学习的变化检测方法,本节将详细讨论这些工作在冰川变化检测工作中的优势和局限性。

(1) 基于传统方法的冰川变化检测

人工目视解译法是早期人们普遍使用的一种方法,在各种计算机技术不够成

熟的时期是最普遍有效的手段。这种方法是根据遥感影像的直接判读标志和相关领域专家的先验知识,与多种相关资料相结合,使用对比分析的方法来对卫星遥感图像进行对比,将获取的结果与已有的相关材料和方法进行对比分析,最终确定遥感图像中的专题信息。王宁练、刘时银^[11]等学者在 2000 年代初期,基于人工目视解译法对中国西部地区(如新疆、西藏等地)的冰川进行了多时相的对比分析,他们通过对比遥感影像,结合现场观测数据,研究了这些地区冰川的变化情况。虽然人工目视解译法在遥感影像分析的早期应用中发挥了重要作用,但随着遥感数据量的增大和计算机技术的进步,其局限性也日益显现:人为判读会受到解译人员经验、技术水平、环境影响等因素的制约,可能导致分析结果的偏差;对于大范围、长时间序列的遥感图像,人工解译方法效率低下,难以满足实时监测和大规模数据处理的需求;冰川边界的提取和变化量的计算存在较大的误差,尤其是在复杂的地形条件下。

基于代数运算的冰川变化检测方法是一种简单直接的变化检测技术,通常依赖于遥感影像中不同波段的比值,通过设置合适的阈值来区分冰川区域和非冰川区域。这种方法的核心在于选择适当的波段组合和阈值,从而实现冰川与其他地物的有效区分。李忠勤、刘时银^[12]等学者在中国西部地区的冰川变化监测中,使用了基于代数运算的变化检测方法。他们结合 Landsat 卫星影像,通过对比不同时间的遥感图像,利用波段比值(如 NDSI)以及选定的阈值进行冰川变化检测。基于代数运算的冰川变化检测方法在冰川监测中提供了简便有效的工具,尤其适用于遥感数据的初步处理与大范围区域的冰川提取。然而,由于该方法的阈值选择依赖于专家经验,且无法深入描述冰川变化的性质,因此在实际应用中存在一定的局限性。

基于图像分类的变化检测方法是一种常见的遥感影像变化检测技术,尤其适用于多时相遥感影像的对比分析^[13]。该方法通过对不同时相的影像进行单独分类,然后逐像素比较两个时相中的分类结果,从而确定变化信息的位置和类型。图像分类方法相对容易理解和实现,分类过程也相对直接。该方法能够处理多时相的遥感数据,适合用来检测长时间尺度上的地物变化。根据分类规则和特征,可以灵活地进行变化信息的提取,尤其适用于具有明显边界或地物类别的区域。聂勇^[14]等研究人员使用了基于图像分类的变化检测方法,结合专家知识和分类规则,对珠穆朗玛峰国家自然保护区的冰川变化进行了详细研究。然而,这种方法是在地物分类的精度、变化原因的分析以及对数据质量的敏感性方面有着较高的局限性,单纯依靠图像分类无法完全解决冰川变化监测中的所有问题。

(2) 基于深度学习方法的冰川变化检测

遥感技术为冰川变化研究提供了强大的支持,Jeremie Mouginot^[15]等人使用

Landsat-8、Sentinel-1 和 RADARSAT-2 数据在格陵兰和南极进行大规模冰速度测绘,首先根据传感器的不同特性获取位移图,然后使用特定方法对位移图进行滤波、校准和拼接等操作,得到矢量产品,再与原有产品进行对比得出结果。张聪^[16]等人用 Landsat 遥感数据,通过目视解译和半自动分类相结合的方法提取了 2000 年和 2020 年的昆仑山脉冰川轮廓,然后计算了冰川面积的变化率和末端变化率。李海^[17]等研究人员运用波段比值法和 NDWI 指数,提取了 1994-2018 年共 9 年的希夏邦玛峰冰川面积,实现对冰川的动态监测。Winsvold^[18]等研究人员则基于 Sentinel-2 卫星和 Landsat 类型数据的密集时间序列,提出了一种利用光学图像波段比值的稳健方法,以自动算法探索冰川制图应用,包括季节内波段合成、利用雪的光谱季节变化改进冰川制图、研究冰川表面类型的时空变化,以及合成图像用于自动冰川变化检测;随着人工智能技术的不断发展,尤其是卷积神经网络 CNN 在图像处理领域的大放异彩,Vishakha Sood^[19]等人使用基于 U-Net 架构的深度学习分类器,在印度西喜马拉雅山脉的巴拉希格里冰川上成功地识别了冰川边界和雪/冰分布。利用 Landsat 数据进行监测,深度学习在提取冰川边界方面表现出较高的准确性(91.89%的总体精度)。范吉延^[20]等人基于喜马拉雅山脉西端 ALOS2-PALSAR 全极化影像,使用 VGG16 特征提取网络与全卷积神经网络模型 U-net 相结合的 VGG16-unet 对冰川进行识别,并结合地形数据,提高了冰川识别精度。

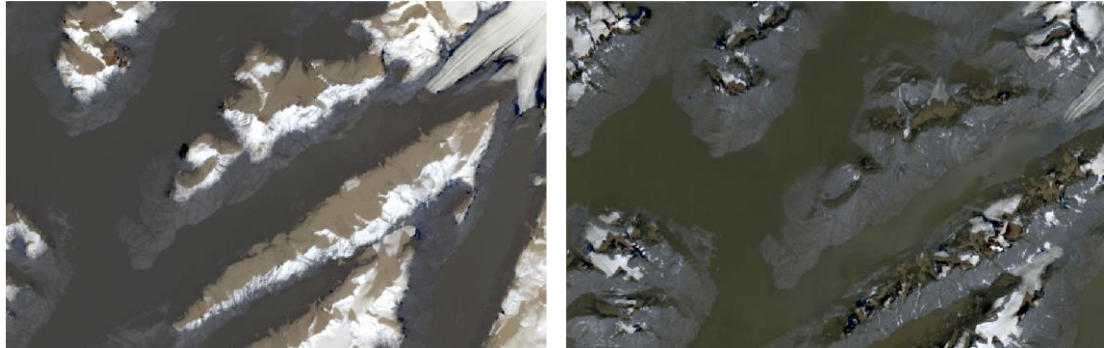
基于传统方法的冰川变化检测通常依赖人工进行特征选择和提取,这种方式容易导致对冰川变化的理解不足或关键信息的遗漏。此外,传统方法对光照变化和气候条件的敏感性较强,这使得在不同时间和气候环境下获取的影像可能出现显著的差异。相比之下,基于深度学习的方法减少了人工干预,能够自动从遥感影像中识别变化特征,展现出更强的鲁棒性。因此,采用深度学习技术进行冰川变化检测不仅具有前瞻性,而且对提高检测精度和深入理解冰川动态变化具有重要的现实意义。

1.3 问题的提出 (Statement of the Problem)

在冰川变化检测中，双时相遥感影像的使用面临多个挑战，特别是在成像时间和成像条件的差异下，冰川在同一位置的光谱特征可能会显著变化。这种光谱差异可能导致深度学习模型在提取双时相图像中的冰川语义信息时出现较大偏差，从而在变化检测结果中引入伪变化。此问题对遥感技术在冰川变化监测中的应用提出了严峻挑战。此外，冰川的动态变化受其底部及周围地形的影响较大，这些复杂的地形特征对冰川形态产生了限制，致使冰川呈现出多样化的形状^[21]。与此同时，冰川自身的动力学机制，尤其是在边缘区域的微小变化，也增加了冰川变化检测的复杂性。

这些问题的产生可归因于以下几个因素：

(1) 冰川区域的特殊性：冰川区域的地理环境和气候条件特殊，且监测周期较长，导致双时相影像中冰川的光谱特征呈现出多样性。如图 1.2 所示，由于极地冰川位置的特殊性（高纬度），太阳高度角直接影响地物在遥感影像中的阴影表现。太阳角度较低时（如清晨和傍晚），影像中的阴影较长，可能会遮掩地物的部分特征，导致部分区域的可见度降低，在一些地区，尤其是山区或地形复杂的区域，光照不均导致的阴影效应可能使得影像中的某些地物被遮挡或难以识别，相同区域的光谱特征可能发生明显差异，导致模型误判，产生伪变化^[22]。



(a) 2015年冰川遥感影像

(b) 2023年冰川遥感影像

图 1.2 同一地点不同时间拍摄的冰川影像对比

Fig. 1.2 Comparison of glacier images taken at different times at the same location.

(2) 冰川流动与地形影响：冰川的流动不仅受到重力的作用，还受到底部和周围地形的强烈影响^[23]。这些地形特征对冰川形态产生了限制，使其呈现出复杂且多变的形状。同时，地形的不均匀性和太阳辐射的分布不均匀性导致冰川边缘的融化速率存在显著差异，进而形成了悬崖、斜坡等复杂地形。这些复杂的地形因素进一步增加了冰川形态的多样性和复杂性。此外，气候变化的影响也加剧了冰川的变化^[24]，尤其是在冰川边缘地区的快速或不均匀融化，复杂的边界地形也增加了冰川变化检测的难度。



(a) 2015年冰川遥感影像

(b) 2023年冰川遥感影像

图 1.3 冰川形状的复杂性

Fig. 1.3 Complexity of glacier shapes.

1.4 主要研究内容（Main Research Content）

针对因成像时间和低光照成像条件下导致的双时相遥感影像中冰川光谱不一致，以及冰川复杂的形状和边缘特征等问题带来的挑战。本文基视觉提示工程和边缘感知模块提出了新颖的冰川变化检测方法。

（1）基视觉提示工程的冰川遥感影像变化检测方法

我们提出了基于视觉提示工程的冰川遥感影像变化检测方法(A Visual Prompt Driven Network for Glacier Change Detection in Remote Sensing Imagery, VPGCD-Net), 这是一种基于 transformer 的网络模型, 利用视觉提示引导实现精确高效的冰川变化检测, 降低由阴影导致的误检。VPGCD-Net 采用双支结构来处理位时图像。一个分支包含视觉提示模块, 该模块结合了阈值分割、连接和特征减法, 以突出重要的变化区域, 这些区域可通过视觉提示为变化检测过程提供高级指导。另一个分支的特点是基于 transformer 的变化检测(TCD)架构, 该架构以 ResNet18 为骨干网络提取位时空间特征。一组 transformer blocks 用于捕捉全局上下文依赖关系, 从而能够对空间和时间关系有更深入的了解。随后, 特征线性调制(FiLM)模块在视觉提示的指导下, 自适应地完善标记以有效代表真正的冰川变化区域。通过利用多头自注意力机制, transformer 解码器可精确识别冰川变化区域, 捕捉时间差异和语义变化, 同时抑制伪变化和背景噪音。

（2）基于边缘信息增强的冰川变化检测方法

本文提出了一种基于边缘信息增强的冰川变化检测网络(Edge-Aware Change Detection Network, EACD-Net), 旨在提高冰川变化区域的检测精度, 特别是针对冰川复杂形状、模糊边界和小尺度变化区域的挑战。EACD-Net 采用 ResNet-50 作为骨干网络, 引入边缘感知模块(Edge-Aware Module, EAM), 通

过 Sobel 算子和跨层级边缘注意力机制建模多尺度边缘信息，从而增强模型对变化区域边界的敏感性，提高检测结果的边界完整性。同时，变化检测解码器（Change Decoder）结合多尺度特征聚合（Multi-scale Aggregation Block, MAB）和混合特征交互模块（Mix Block, MB），并利用变化注意力机制（Change Attention Mechanism, CAM）增强时序特征对比，以提升变化区域的对比度和精准度。实验结果表明，EACD-Net 在边界细节保留、抗噪性、多尺度特征融合及精准变化检测方面均表现优异。消融实验进一步验证了 EAM、CAM 和 MAB 三者的协同作用对整体性能提升的关键作用，使 EACD-Net 在复杂冰川场景下展现出更高的检测精度、边界一致性和抗噪能力，为冰川变化检测任务提供了一种高精度、鲁棒性强的解决方案，具有重要的应用价值。

1.5 本文组织结构（The Structure of This Paper）

本文主要研究内容涵盖六个章节，具体安排如下：

第一章：绪论。本章概述了研究背景，分析了国内外在遥感影像冰川变化检测领域的研究进展，指出现有方法在复杂光照环境、边界模糊及小尺度变化检测方面的不足。此外，本章还阐述了本文的研究目标、主要贡献以及论文的组织结构。

第二章：相关技术。介绍了冰川变化检测涉及的关键技术，包括深度学习在遥感影像分析中的应用、卷积神经网络的原理、Vision Transformer 网络的原理和应用、注意力机制的基本原理等，为后续模型设计提供技术支撑。

第三章：基于视觉提示工程的冰川遥感影像变化检测方法 VPGCD-Net。本章详细介绍了冰川变化检测方法的整体流程，VPGCD-Net 的网络结构及其原理初衷，实验数据、实验配置及评价指标，并通过实验结果和方法对比分析验证 VPGCD-Net 的有效性。

第四章：基于边缘感知的冰川变化检测网络 EACD-Net。本章提出了一种新颖的 EACD-Net 框架，旨在提升冰川变化检测的精度，特别是针对复杂光照条件和模糊边界问题。首先介绍了 EACD-Net 的总体架构，然后详细描述了边缘感知模块（EAM）、变化检测解码器（Change Decoder）及多尺度特征融合模块的设计。实验部分涵盖对比方法、数据集、实验配置及评价指标，最后通过定量和定性分析，验证 EACD-Net 在边界检测、抗噪能力和变化区域识别方面的优越性。

第五章：总结与展望。本章总结了本文研究工作的优势与局限性，并探讨了未来冰川变化检测研究的潜在方向，包括模型的轻量化、对极端环境数据的适应性提升以及多源遥感数据的融合应用。

2 相关技术

2 Related techniques

2.1 Vision Transformer 网络 (Vision Transformer Network)

2.1.1 模型结构

Vision Transformer^[25] (ViT) 是一种基于 Transformer 架构的神经网络模型, 用于处理计算机视觉任务。传统的计算机视觉模型如卷积神经网络 (CNN) 在处理图像任务时取得了很大的成功, 但 CNN 存在一些局限, 例如对于长距离依赖的建模能力较弱。ViT 通过引入 Transformer 的注意力机制来解决这些问题, 并在一些视觉任务上取得了优秀的结果。与传统的 CNN 不同, ViT 将图像分割为一系列的图像块 (或称为图像补丁), 并将每个图像块作为输入序列。然后, ViT 使用 Transformer 的编码器来对这些输入序列进行处理。每个图像块被展平为一个向量, 并与位置编码向量相结合, 形成输入序列。这样, ViT 能够对整个图像进行全局的建模, 而不仅仅是局部区域。ViT 的核心思想是利用 Transformer 的注意力机制来对图像块之间的关系进行建模。注意力机制允许模型根据输入序列中的不同元素之间的关联性来分配不同的权重。通过多层的自注意力机制, ViT 能够对图像块之间的关系进行编码和捕捉, 从而实现对图像的全局理解。其结构如图 2.1 所示, ViT 的处理流程大致可以分为以下几个步骤:

在处理图像输入时, 首先需要对图像进行预处理和嵌入。以 $224 \times 224 \times 3$ 的图像为例, 传统方法是图像拉平成一维数据, 但由于 Transformer 模型的最大输入长度为 512, 这种方法会导致图像的拉平后的维度 ($224 \times 224 = 50176$) 远超输入限制, 从而增加了计算负担。为了应对这一问题, 相关研究提出了一种更为高效的方案——图像分割。

具体来说, 将 $224 \times 224 \times 3$ 的图像划分为 14×14 个 $16 \times 16 \times 3$ 的小图像块 (Patch), 每个图像块的维度为 768, 这样就得到了一个尺寸为 196×768 的输入矩阵, 类似于 Transformer 中的词嵌入。同时, 为了增强全局特征的表示, 该方法引入了一个特殊的 token, 它的维度为 1×768 , 代表了整个图像的全局信息。在分类任务中, 可以通过对该全局特征向量进行进一步的操作 (如 MLP) 来获得最终的分类结果。

最终, 图像的输入矩阵形状为 197×768 , 其中 197 个元素包括 196 个图像块和 1 个全局 token。为了能够捕捉图像块之间的位置信息, 引入了位置编码。

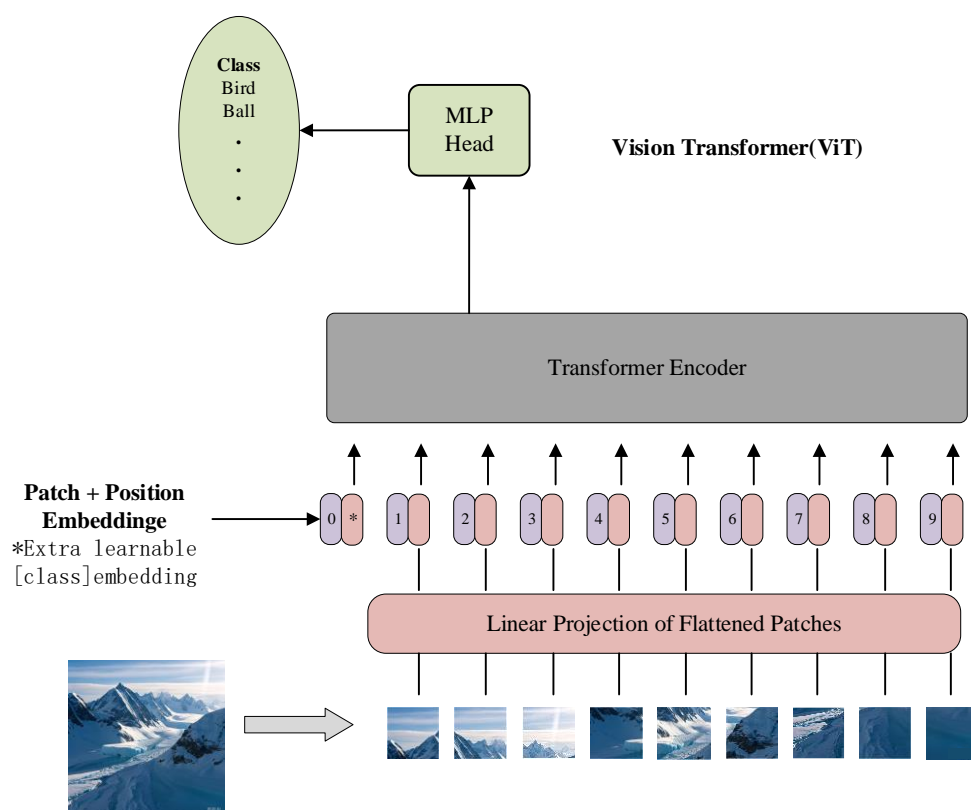


图 2.1 Vision Transformer 网络模型示意图
Figure 2.1 Vision Transformer network model diagram

这些位置编码采用了与 Transformer 相同的 1 维位置编码策略，通过正弦和余弦函数生成，并与图像块的特征向量相加，形成最终的输入矩阵。经过这些预处理步骤，图像被成功转换为适合 Transformer 模型处理的格式，能够有效地捕捉图像的局部和全局特征。

2.1.2 Transformer 模块

视觉 Transformer 块作为 ViT 模型的基本构成单元，其架构和工作原理对图像的特征提取具有重要意义。一个典型的视觉 Transformer 块包括输入图像的分块(Patch Embedding)自注意力机制(Self-Attention)、前馈神经网络(Feed-Forward Network, FFN)、残差连接与层归一化(Residual Connection & Layer Normalization)，如图 2.2 所示。

(1) 输入图像的分块 (Patch Embedding)

在 ViT 模型中，首先将输入图像分割成若干个固定大小的图像块(patches)。每个图像块通过展平操作转换为一维向量，然后通过线性变换(通常使用卷积层)映射到更高维的空间。具体操作如下：

假设输入图像的尺寸为 $H \times W \times C$ (其中 H 为图像的高度， W 为宽度， C 为

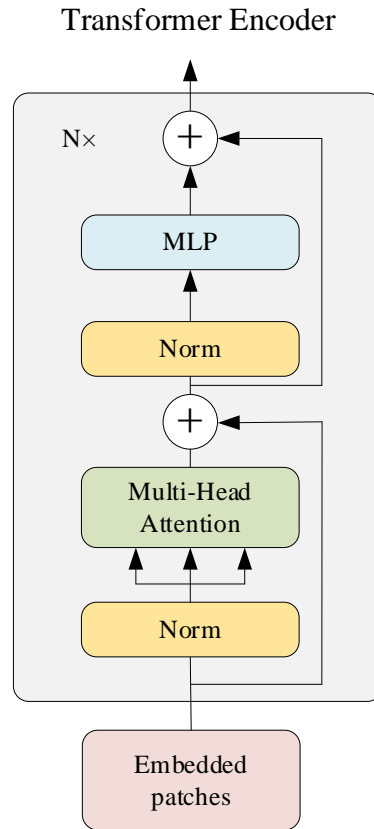


图 2.2 Transformer 块的内部结构示意图

Fig. 2.2 Specific structure of Transformer Block

通道数)。图像被划分为 N 个大小为 $P \times P$ 的图像块（其中 P 为块的大小）。每个图像块被展平成一个一维向量，并通过线性映射将其转换为 d_{model} 维的向量。通过这一过程，输入图像被转换为一个大小为 $N \times d_{model}$ 的矩阵，其中 N 为图像块的数量， d_{model} 为映射后的嵌入维度。这个矩阵作为 Transformer 模型的输入序列，后续会传入 Transformer 进行进一步处理。

（2）多头自注意力机制

多头自注意力机制（Multi-Head Self-Attention）是 Transformer 模型中的核心组件，它的作用是通过并行计算多个注意力头，增强模型对不同上下文信息的建模能力。为了更好地理解这个机制，我们可以分步详细介绍其计算过程和公式。

首先关注自注意力机制，自注意力机制的目的是为每个位置学习一个上下文感知的表示。在序列数据中，每个位置都需要与其他位置进行交互，来捕捉它们之间的依赖关系。自注意力机制通过查询（Query）、键（Key）和值（Value）来实现这一点。具体来说，对于输入的序列（例如，图像分割块或词嵌入），每个元素通过查询、键和值之间的关系来计算其新的表示。

计算注意力分数，通过计算查询向量 Q 与键向量 K 的点积，来衡量各个元素之间的相似度或相关性。

$$\text{Attention}_{\text{Score}} = QK^T \quad (2.1)$$

缩放点积，为了避免在计算点积时数值过大，通常会对点积结果进行缩放。缩放因子是键向量的维度的平方根 $\sqrt{d_k}$ ，其中 d_k 是键的维度。

$$\text{Scaled_Attention_Score} = \frac{QK^T}{\sqrt{d_k}} \quad (2.2)$$

Softmax 操作: 对缩放后的注意力分数应用 Softmax 函数, 得到注意力权重, 表示每个位置对其他位置的关注程度。

$$\text{Attention_Weights} = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) \quad (2.3)$$

加权求和: 最后, 使用注意力权重对值矩阵 V 进行加权求和, 得到最终的输出表示。

$$\text{Attention_Output} = \text{Attention_Weights} \cdot V \quad (2.4)$$

然后关注多头注意力机制通过并行计算多个注意力头来增强模型对不同上下文信息的建模能力。每个注意力头学习一个不同的表示, 模型通过这些不同的表示来捕捉多种注意力模式和上下文信息。其计算过程如下:

初始化多个注意力头: 对于每个注意力头, 都有一组独立的权重矩阵, 这些矩阵用于对输入的查询(Q)、键(K)和值(V)进行线性变换。设有 h 个注意力头, 查询、键和值都经过不同的权重矩阵进行线性变换:

$$\begin{cases} Q_i = XW_i^Q \\ K_i = XW_i^K \\ V_i = XW_i^V \end{cases} \quad i = 1, 2, 3, \dots, h \quad (2.5)$$

其中, W_i^Q, W_i^K, W_i^V 是为第 i 个头设计的权重矩阵。

计算每个注意力头的输出: 对每个头分别进行自注意力计算 (如前所述), 得到每个头的输出表示:

$$\text{Output}_i = \text{Self-Attention}(Q_i, K_i, V_i) = \text{Softmax}\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right) \cdot V_i \quad (2.6)$$

将所有注意力头的输出拼接在一起, 得到一个新的表示:

$$\text{Concat_Output} = [\text{Output}_1, \text{Output}_2, \dots, \text{Output}_h] \quad (2.7)$$

将拼接后的输出通过一个线性变换 (即一个全连接层) 进行变换, 得到最终的多头注意力输出:

$$\text{Final Output} = \text{Concat_Output} \cdot W^O \quad (2.8)$$

其中, W^O 是一个学习的权重矩阵。

多头自注意力机制通过并行使用多个注意力头来增强模型的表示能力, 它不仅能够更好地捕捉序列中各个位置之间的复杂关系, 还能提高计算效率^[26]。在 Transformer 模型中, 多头注意力机制是其成功应用于自然语言处理和计算机视

觉任务的关键所在。通过对输入序列进行不同维度的注意力计算，Transformer 能够有效地建模长程依赖关系，从而使得该模型在各类任务中表现出色。

（3）前馈神经网络（Feed-Forward Network, FFN）

自注意力机制处理完图像块的关系后，通常会接一个前馈神经网络（Feed-Forward Network, FFN）。FFN^[27]由两个全连接层组成，并通过非线性激活函数（如 ReLU）进行变换。其基本操作如下是首先通过第一个全连接层进行线性变换，得到一个较大的维度。随后应用激活函数（如 ReLU），并将结果通过第二个全连接层进行线性变换，最终输出。FFN 的作用是进一步增强模型的表示能力，捕捉图像块的复杂特征。

（4）残差连接与层归一化（Residual Connection & Layer Normalization）

为了提高模型的训练稳定性并避免梯度消失问题，视觉 Transformer 块中的每一层（如自注意力层和前馈神经网络层）都会使用残差连接（Residual Connection）。残差连接^[28]的作用是将输入与子层的输出相加，避免信息在多层堆叠时被逐渐消减。具体公式如下：

$$Output = LayerNorm(x + sub_layer_output) \quad (2.9)$$

其中， x 为子层的输入， sub_layer_output 为该子层的输出，层归一化是为了维持梯度的稳定性并加速训练过程。

（5）位置编码（Positional Encoding）

Transformer 架构没有内建的顺序处理机制，因此需要引入位置编码来为每个图像块提供位置信息。位置编码可以是固定的或可学习的，它的作用是向每个图像块的嵌入向量中加入位置信息。通常，ViT 使用正弦和余弦函数生成的位置编码，这些编码被加到每个图像块的嵌入向量上，使得模型能够感知图像块的相对或绝对位置。

视觉 Transformer 块作为 ViT 模型的核心组件，通过自注意力机制和前馈神经网络的结合，有效捕捉了图像中各块之间的长距离依赖关系。相较于传统的卷积神经网络（CNN），ViT 摒弃了卷积操作，采用了基于 Transformer 的机制来处理图像块。该方法的成功应用表明 Transformer 架构在计算机视觉任务中具有极大的潜力，尤其在大规模数据集上，视觉 Transformer 能够展现出比传统 CNN 更强的建模能力和表现力^[29]。

2.2 卷积神经网络（Convolutional Neural Network）

卷积神经网络（CNN）是一种深度学习算法，广泛应用于计算机视觉任务，如图像分类、目标检测、语义分割等^[30]。CNN 通过模拟生物视觉系统的机制来进行特征提取和学习，能够有效地从数据中自动提取空间层次特征。它的基本单

元是卷积层，这一层通过卷积操作从输入数据中提取特征。CNN 的成功应用主要依赖于其有效的特征提取能力以及减少参数的需求。

2.2.1 可变形卷积

可变形卷积^[31] (Deformable Convolution) 是一种扩展了传统卷积操作的卷积方法。与传统的卷积操作不同，可变形卷积引入了“可学习的偏移量”，使得卷积核可以根据输入图像的内容自适应地改变其形状，从而更好地捕捉图像中具有不规则形态或变形特征的目标。

传统卷积操作通常使用固定大小的卷积核，并对输入图像进行固定模式的扫描。每次卷积操作的感受野是固定的，意味着卷积核的每个位置都是一个固定的局部区域，而不考虑图像中可能存在的形变、旋转、尺度变化等问题^[32]。可变形卷积通过引入动态的卷积核偏移量，允许网络在学习过程中自适应地调整卷积核的扫描位置，从而能够更好地处理图像中复杂的几何变化。

(1) 可变形卷积的工作原理

可变形卷积的核心思想是在传统卷积操作的基础上引入了偏移(offsets)。这些偏移量是通过网络学习得到的，每个卷积核的每个位置会根据输入图像的局部特征自适应调整其采样位置。具体而言，在传统卷积中，卷积核与图像局部区域的元素进行逐点相乘并求和。而在可变形卷积中，卷积核对图像局部区域的采样位置不是固定的，而是通过学习得到的偏移量进行动态调整。其过程如图 2.3 所示。

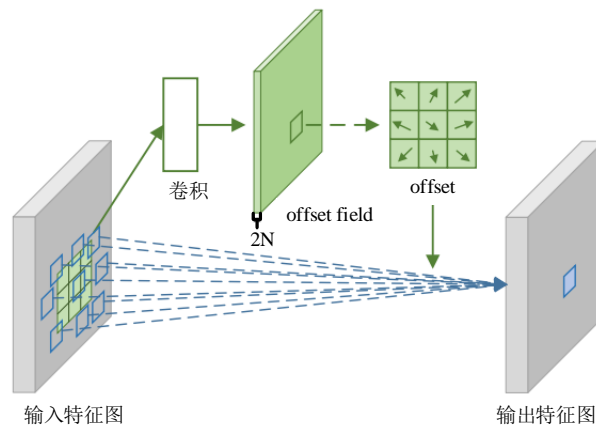


图 2.3 可变形卷积示意图

Fig. 2.3 Schematic diagram of Deformable Convolutional.

(2) 公式化表示

首先介绍对于标准的卷积操作，给定输入特征图 I 和 K 卷积核，其输出特征图 O 的每个元素可以表示为：

$$O(x, y) = \sum_{m=-a}^a \sum_{n=-b}^b I(x + m, y + n) \cdot K(m, n) \quad (2.10)$$

其中, a, b 表示卷积核的大小。接下来我们介绍可变形卷积公式, 可变形卷积在传统卷积的基础上加入了偏移量 Δx 和 Δy , 使得卷积核的采样位置变得灵活。即:

$$O(x, y) = \sum_{m=-a}^a \sum_{n=-b}^b I(x + m + \Delta x(m, n), y + n + \Delta y(m, n)) \cdot K(m, n) \quad (2.11)$$

其中, $\Delta x(m, n)$ 和 $\Delta y(m, n)$ 是由网络学习得到的偏移量, 它们决定了每个卷积核位置的采样偏移。

传统卷积核的感受野是固定的, 而可变形卷积通过学习动态偏移量, 能够自适应地调整感受野。这样, 网络可以根据不同的图像特征选择最合适的卷积核位置。并且, 图像中的物体可能因为旋转、尺度变化或畸变等原因发生形变, 传统的卷积操作很难处理这些问题。而可变形卷积能够通过调整卷积核的位置来适应图像中的形变, 从而有效提高网络对变形物体的处理能力。此外, 可变形卷积有更强的表达能力: 通过自适应地调整卷积核的采样位置, 网络能够捕捉到更多细粒度的特征, 进而提升网络的表现力, 尤其在目标检测、图像分割等任务中表现优异^[33]。

2.3 注意力机制 (Attention mechanisms)

注意力机制在人工智能领域, 特别是深度学习中的应用, 已经成为提升模型性能的重要技术之一。该机制的灵感来源于人类的认知过程, 尤其是视觉注意力系统, 即人在观察复杂场景时, 会优先关注与当前任务最相关的区域, 而忽略无关信息。类似地, 在神经网络中, 注意力机制通过动态调整输入特征的权重, 使模型能够更加精准地捕捉关键信息, 从而提高决策的准确性和计算效率^{[34]-[39]}。在实现上, 注意力机制通过计算不同输入特征的重要性分数, 分配相应的权重, 从而突出对任务有贡献的特征并抑制无关部分。这种机制本质上是模型内部的一种可学习的选择策略, 使其能够在不同上下文中自适应地关注最相关的信息。例如, 在图像分类任务中, 注意力机制可以引导模型聚焦于目标对象的关键区域, 以提取更具判别性的特征。如图 2.2(a)所示, 热力图显示神经网络在识别狗时, 主要关注了其头部区域, 尤其是面部特征, 因为这些区域最能代表狗的类别信息。而在图 2.2(b)中, 热力图则显示模型在识别猫时, 主要集中在猫的头部和尾巴等关键部位。这表明深度学习模型在分类任务中, 会自动调整关注区域, 以更有效地提取目标特征, 实现更精准的识别。在遥感变化检测任务中, 由于场景复杂、光照变化、地物类型多样, 传统的深度学习方法难以精准地捕捉变化区域。因此, 空间注意力 (Spatial Attention, SA) 和通道注意力 (Channel Attention, CA) 被广泛应用于图像处理分析中, 以增强模型对关键区域的关注度。

本节将分别介绍空间注意力和通道注意力的原理、数学建模及其在深度学习中的应用，并探讨如何结合两者以进一步提升模型性能。

2.3.1 空间注意力

空间注意力 (SA) 机制旨在增强模型对空间位置的重要性感知，使其能够关注关键区域并忽略冗余信息，其结构如图 2.4 所示。遥感影像的变化区域往往具有较强的空间特征，例如 建筑物变化、植被生长、冰川消融等，因此，空间注意力能够帮助模型在特征提取过程中动态调整权重，提高对目标区域的检测能力。在计算机视觉任务中，空间注意力广泛应用于目标检测、语义分割、变化检测等任务^[40]。例如，在遥感影像变化检测任务中，空间注意力可以帮助模型聚焦于实际发生变化的区域，而减少对未变化区域的干扰^[41]。

假设输入特征图为 $X \in \mathbb{R}^{H \times W \times C}$ ，其中 C 表示通道数， H 和 W 分别表示图像的高度和宽度。空间注意力的计算过程如下：

首先，使用 全局平均池化 (Average Pooling) 和全局最大池化 (Max Pooling) 在通道维度进行特征提取，以保留关键的空间信息：

$$F_{\text{avg}} = \text{AvgPool}(X) \quad (2.12)$$

$$F_{\text{max}} = \text{AvgPool}(X) \quad (2.13)$$

其中， $F_{\text{avg}}, F_{\text{max}} \in \mathbb{R}^{H \times W \times C}$ 表示分别对所有通道进行平均池化和最大池化操作，得到的两个特征图仅保留空间信息。

接下来，将池化后的特征进行拼接，并通过一个 7×7 的卷积操作 (Convolution) 进一步融合信息：

$$M_s = \sigma(f_{\text{conv}}([F_{\text{avg}}, F_{\text{max}}])) \quad (2.14)$$

其中： $[F_{\text{avg}}, F_{\text{max}}]$ 表示在通道维度上的拼接，得到形状为 $\mathbb{R}^{2 \times H \times W}$ 的特征图， f_{conv} 表示一个卷积核大小为 7×7 的卷积操作， $\sigma(\cdot)$ 为 Sigmoid 激活函数，将权重归一化到 (0,1) 之间，以保证数值稳定性。

最终，空间注意力权重 M_s 作用于输入特征：

$$X' = M_s \odot X \quad (2.15)$$

其中， \odot 表示逐元素乘法，使网络更加关注重要的空间区域，并抑制无关信息

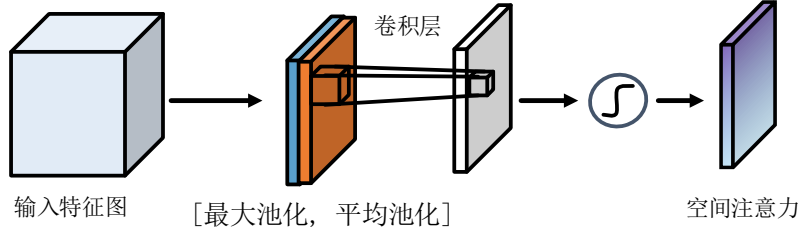


图 2.4 空间注意力模块示意图

Fig. 2.4 Schematic diagram of the spatial attention module

2.3.2 通道注意力

通道注意力（Channel Attention）是一种用于提升神经网络特征表示能力的重要方法，它的核心思想是让网络能够自适应地调整不同通道的重要性，以强化关键特征并抑制冗余信息。传统的卷积神经网络（CNN）对所有通道一视同仁，然而在实际应用中，不同通道所包含的信息可能具有不同的重要性。通道注意力机制能够赋予模型选择性关注能力，使其更加聚焦于对当前任务更具辨识度的特征通道，从而提升性能。这种机制通常通过学习通道权重来自动调整不同通道的贡献，从而有效筛选出对任务最有帮助的特征。通过在网络的各层嵌入通道注意力机制，模型能够更好地理解输入数据的多样性，并增强模型的泛化能力。

给定输入特征图 $X \in \mathbb{R}^{H \times W \times C}$ ，其中 C 为通道数， H 和 W 分别表示特征图的高度和宽度。通道注意力的基本目标是为每个通道分配一个权重，使得模型能够自动强调重要通道，同时削弱不重要通道。SE^[42]（Squeeze-and-Excitation）模块是一种引入注意力机制的网络组件，旨在增强深度神经网络对输入特征的建模能力，其结构如图 2.5 所示。典型的通道注意力机制如 Squeeze-and-Excitation (SE) 模块采用以下三步计算过程：

首先是全局信息聚合，为了提取全局信息，SE 模块对输入特征图的每个通道进行全局平均池化（Global Average Pooling, GAP），计算得到通道的全局描述：

$$S_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X_c(i, j) \quad (2.16)$$

其中 S_c 代表第 c 个通道的全局统计信息， $X_c(i, j)$ 代表输入特征图在第 c 个通道的像素值。

然后进行通道权重计算，为了计算每个通道的重要性，SE 模块采用一个两层全连接网络（Fully Connected, FC），通过非线性变换学习通道之间的依赖关系：

$$z = \sigma(W_2 \delta(W_1 s)) \quad (2.17)$$

其中 $W_1 \in \mathbb{R}^{\frac{C}{r} \times C}$ 和 $W_2 \in \mathbb{R}^{C \times \frac{C}{r}}$ 是两个全连接层的权重， $\delta(\cdot)$ 表示 ReLU 激活函数， $\sigma(\cdot)$ 表示 Sigmoid 激活函数， r 是降低维度的缩放因子。通过这种方式，模型

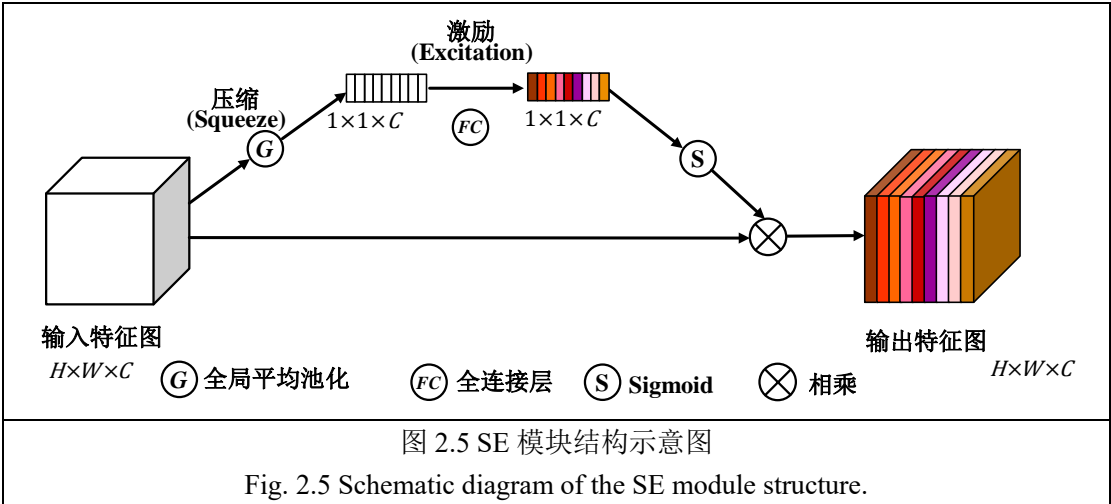
能够学习到每个通道在不同任务中的权重分布。

最终，计算得到的通道注意力权重 z 被用于对原始输入特征进行加权：

$$\hat{X} = z_c \cdot X_c$$

(2.18)

通过这种方式，通道注意力模块能够增强网络对关键通道的关注度，从而提高特征表达能力。



在这里要特别介绍 CBAM^[43]（Convolutional Block Attention Module），该模块是由韩国首尔大学的 Sanghyun Woo 等人在 2018 年 ECCV 会议上提出的，旨在通过融合通道和空间维度的注意力机制来增强特征表达。其结构如图 2.6 所示，受 SENet（Squeeze-and-Excitation Network）启发，CBAM 通过级联通道注意力和空间注意力机制，能够自适应地增强关键特征并抑制冗余信息。

如图 2.6 所示，这里的通道注意力通过全局平均池化和最大池化捕捉全局信息，并通过轻量级 MLP 网络生成通道权重，强化重要通道；空间注意力则通过沿通道维度的池化操作生成空间特征图，并通过卷积层融合生成空间权重，突出关键区域。通道和空间注意力通常按顺序级联优化，以全面提升特征表达。

CBAM 的能够同时优化通道和空间维度，比单一维度的注意力机制更全面；计算成本极低，仅增加约 1% 的参数数量，且可以轻松嵌入主流 CNN 架构，如 ResNet、

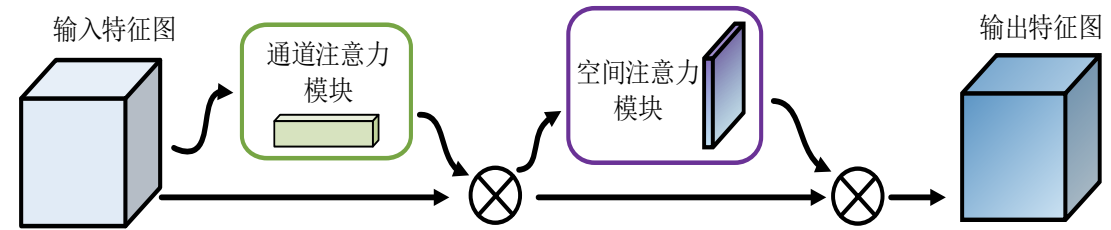


图 2.6 CBAM 模块结构示意图

Fig. 2.6 CBAM module structure schematic.

MobileNet、YOLO 等；能够抑制无关通道和空间区域的噪声，从而提升模型性

能，在分类、检测和分割任务中能平均提升 1-3% 的精度。

在应用上，CBAM 被广泛应用于图像分类（如在 ResNet-50 中加入 CBAM 提升 ImageNet 分类任务准确率）、目标检测（如 YOLOv4 中增强小目标检测能力）、语义分割（如在 U-Net 中提升医学影像的边缘分割精度）、轻量化模型（如 MobileNetV3 提升效率与性能）和视频分析（如动作识别中的关键帧聚焦）。

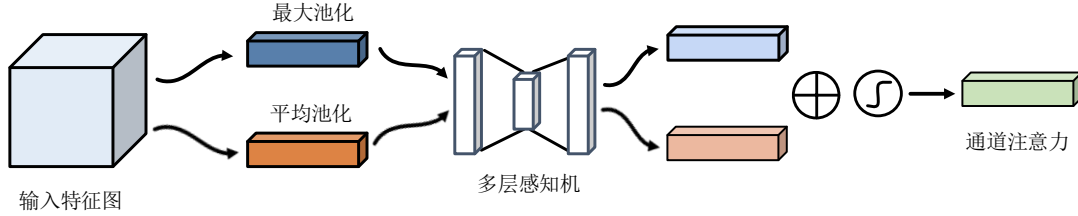


图 2.7 通道注意力结构示意图

Fig. 2.7 Schematic diagram of channel attention structure

2.4 特征线性调制（Feature-wise Linear Modulation）

FiLM^[44]（Feature-wise Linear Modulation）的提出源于对多模态学习和动态特征适应的需求。其设计初衷是为了解决传统神经网络在融合外部条件信息（如文本、时间序列、地理坐标等）时的僵化性问题。在 2017 年之前，条件生成模型（如条件 GAN）主要通过拼接输入（如将标签向量与噪声拼接）或条件批归一化（如 AdaIN）引入外部信息。这些方法对复杂条件的表达能力有限。

例如，在图像生成任务中，若需根据文字描述生成细节（如“冰川上的裂缝”），简单拼接难以精确控制局部特征。2018 年，DeepMind 团队首次提出 FiLM 模块，如图 2.8 所示。其核心思想是通过仿射变换动态调制网络中间特征，而非静态调整输入或归一化参数，从而实现对特征空间的细粒度控制。

2.4.1 FiLM 的基本原理

FiLM 的核心思想是通过条件输入动态生成一组调制参数，对目标神经网络的中间特征进行特征级的线性变换，从而实现对网络行为的条件控制。具体而言，FiLM 通过两个函数 f 和 h （通常由神经网络实现）生成调制参数 $\gamma_{i,c}$ 和 $\beta_{i,c}$ ，这些参数基于条件输入 x_i 计算：

$$\gamma_{i,c} = f_c(x_i) \quad (2.19)$$

$$\beta_{i,c} = h_c(x_i) \quad (2.20)$$

其中， $\gamma_{i,c}$ 和 $\beta_{i,c}$ 分别表示第 i 个输入的第 c 个特征通道的缩放和偏移参数。目标网络的特征图 $F_{i,c}$ 随后通过以下仿射变换进行调制：

$$FiLM(F_{i,c} | \gamma_{i,c}, \beta_{i,c}) = \gamma_{i,c} F_{i,c} + \beta_{i,c} \quad (2.21)$$

这一变换对每个特征通道独立操作，赋予了 FiLM 对特征图的细粒度控制能

力。 f 和 h 可以是任意函数，但在实践中通常被实现为一个共享参数的神经网络，称为 $FiLM$ 生成器。被调制的目标网络则称为 $FiLM-ed$ 网络。 $FiLM$ 的调制能力体现在其可以根据条件输入 x_i 对特征图进行多种操作，例如放大、缩小、取反、关闭（当 $\gamma_{i,c} = 0$ 时），甚至结合后续的 $ReLU$ 激活函数实现选择性阈值操作。这种灵活性使得 $FiLM$ 能够根据任务需求动态调整网络的特征表达。

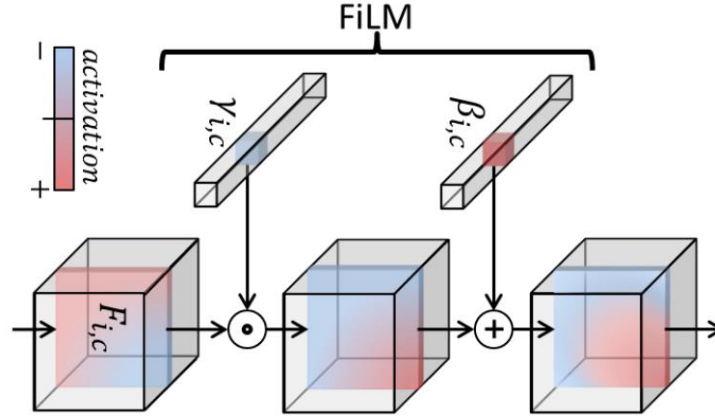


图 2.8 特征线性调制过程示意图

Fig. 2.8 Schematic diagram of the characteristic linear modulation process

2.4.2 FiLM 的模型结构与实现

（1）FiLM 生成器

负责处理条件输入（例如问题文本）。在视觉问答任务中， $FiLM$ 生成器通常采用循环神经网络（如 GRU ）处理问题文本[45]，生成问题嵌入向量。以Perez等人的实现为例， GRU 网络包含4096个隐藏单元，输入为200维的词嵌入向量， GRU 的最终隐藏状态被用作问题嵌入，通过仿射投影生成每个残差块（Residual Block）的 (γ_i^n, β_i^n) 参数对[46]。

（2）FiLM-ed 视觉管道

负责处理图像输入并应用 $FiLM$ 调制。视觉管道首先通过CNN提取图像特征，例如使用预训练的ResNet-101提取conv4层的特征，或者从头训练一个包含4层的CNN（每层具有128个 4×4 卷积核，结合ReLU和批归一化）。提取的特征图（例如128个 14×14 的特征图）随后通过多个 $FiLM-ed$ 残差块（ResBlocks）处理，每个残差块包含 1×1 卷积和 3×3 卷积，并在特定位置应用 $FiLM$ 层。最终，特征图通过分类器（包括 1×1 卷积、全局最大池化和两层MLP）生成答案的softmax分布[47]。

FiLM 层的实现非常高效, 每个特征通道仅需两个参数 (γ 和 β), 因此其计算成本与图像分辨率无关。这种特性使其在高分辨率任务中具有显著优势, 相比之下, 其他方法 (如 Relation Networks) 的计算成本可能随分辨率呈二次方增长。

2.5 视觉提示工程 (Visual Prompt Engineering)

视觉提示工程^[48] (Visual Prompt Engineering, VPE) 起源于自然语言处理 (Natural Language Processing, NLP) 领域的提示学习 (Prompt Learning), 其核心思想是通过构造特定指令 (如文本模板) 引导预训练模型适配下游任务。在 NLP 任务中, 提示学习通过精心设计的上下文信息, 使预训练语言模型在少量或零样本的情况下高效执行目标任务。随着多模态预训练模型 (如 CLIP、ALIGN 等) 的发展, 这一思想被引入计算机视觉 (Computer Vision, CV) 领域, 以显式视觉信号 (如差异图、掩膜、边界框) 注入先验知识, 动态引导模型关注关键区域, 提高模型的泛化能力和鲁棒性。

视觉提示工程已在计算机视觉的多个任务中展现出广泛的适用性^[48]。在目标分类任务中, 通过向输入图像添加边界框提示, 可以有效增强模型对目标区域的关注度, 提高分类准确率; 在目标检测任务中, 可学习的位置编码能够提供更精确的空间信息, 从而提升检测性能; 在医学影像分析领域, 利用病灶提示信息可以提升病变区域的分割精度, 减少误检率。此外, 近年来涌现的多模态模型 (如 Flamingo、GATO 等) 借助图文混合提示, 实现了更复杂的跨模态推理任务, 进一步扩展了视觉提示工程的应用边界。

目前的视觉提示多依赖人工设计或固定规则, 未来可探索基于元学习 (Meta-learning) 或强化学习 (Reinforcement Learning, RL) 的动态优化方法, 以根据不同环境条件自适应调整提示信息。例如, 在遥感变化检测任务中, 可通过自适应调整阈值策略, 使模型能够针对不同地理区域或季节特征生成最优提示, 提高其适应性和检测精度^[49]。同时, 随着多模态学习的进步, 视觉提示可与文本、语音等模态信息相结合, 以增强模型的综合感知能力^[51]。例如, 结合地理文本描述 (如地形特征、气候信息) 可使遥感变化检测模型对复杂场景的理解更加精准。此外, 在变化检测、医学影像分割等任务中, 目标区域通常具有明显的边缘特征, 引入边缘感知优化 (Edge-aware Optimization) 能够进一步提升模型的检测能力^[52]。基于梯度引导的边缘强化 (Edge-aware Attention) 或曲率约束的边界细化 (Curvature-aware Refinement) 等技术, 可增强模型对目标边界的捕捉能力, 有效减少伪变化区域的影响, 从而提高整体检测的准确性和鲁棒性。

综上所述, 视觉提示工程通过引入显式的先验信息, 有效提升了深度学习模型在复杂场景中的泛化能力。在冰川变化检测、医学影像分析、目标检测等多个领域, 视觉提示工程均展现出显著优势。未来, 随着自适应优化、多模态融合及

边缘感知等技术的发展，视觉提示工程有望进一步拓展其应用范围，并在资源受限环境下实现更高效的部署，为计算机视觉任务提供更加精准解决方案。

2.6 本章小结（Chapter Summary）

本章对冰川遥感变化检测领域的相关技术进行了系统性综述，为后续研究奠定了理论和技术基础。本章通过对 Vision Transformer、卷积神经网络、多头自注意力机制、FiLM 技术以及工程应用的全面分析，梳理了冰川遥感变化检测领域的技术现状和发展趋势。这些技术为本研究提出基于多任务深度学习的冰川变化检测框架提供了重要的理论支持和技术储备，同时也指明了模糊边界检测等关键问题可能的技术解决方案，为后续模型设计和实验验证奠定了坚实基础。

3 基于视觉提示的冰川变化检测方法

3 Visual Prompt Driven Method for Glacier Change Detection

3.1 引言 (Introduction)

冰川是地球气候系统的重要组成部分,在自然环境和人类社会的可持续发展中发挥着关键作用^[52]。作为全球水资源的重要来源,冰川对干旱与半干旱地区的农业生产、工业用水及居民生活具有不可替代的支撑作用。冰川变化直接影响海平面上升、气候调节和生态系统稳定性,这使得研究其对气候变化的响应成为重要科学议题。传统冰川变化监测主要依赖实地勘测方法,自 20 世纪 30 年代起通过固定点位测量冰川长度变化。随着遥感技术的快速发展,高分辨率卫星与航空影像已近年来,基于深度学习的遥感变化检测技术取得显著进展。通过融合多源多时相影像与地理空间数据,该技术可精准分析地表特征变化,有效提取有效信息并滤除环境干扰。卷积神经网络(如 UNet++^[54])在端到端检测中表现优异,GET-NET^[55]等专用框架在超光谱影像分析中取得突破。其他方法包括基于生成对抗网络(GANs)^[56]的数据增强技术和利用长短期记忆网络(LSTM)进行时序趋势分析。孪生网络凭借其双通道结构,在双时相影像特征差异比对方面具有独特优势,已成为变化检测的主流架构。Transformer 与 CNN 的融合架构进一步提升了检测精度^[57],而图结构网络通过建模像素级关系增强了空间特征表达能力^[58]。因此,将深度学习方法应用于冰川遥感变化检测将成为有效技术手段。然而,由于极地冰川地理位置的特殊性,目前尚未有公开的深度学习专用冰川遥感变化检测数据集。针对上述问题,本研究构建了首个面向极地冰川的遥感变化检测专用数据集,并提出视觉提示驱动的冰川变化检测网络。通过构建基于阈值的冰川差异图,引导基于 Transformer 的特征提取网络聚焦双时相冰川影像的关键区域,同时抑制阴影干扰。本章创新性地将视觉提示机制引入冰川变化检测领域,提出基于视觉提示工程的检测网络。通过特征线性调制(FILM)技术对主干网络实施仿射变换,实验证明该方法能有效引导网络关注低照度成像条件下的冰川变化区域,降低阴影及复杂环境导致的误检率。成为冰川监测的重要技术手段。

3.2 冰川变化检测流程 (Glacier Change Detection Workflow)

本节详细阐述了基于视觉提示工程的冰川变化检测方法的完整流程,如图 3.1 所示。该流程包含三个核心阶段:数据集预处理、模型训练以及模型测试,每个阶段均经过精心设计,以确保数据质量和模型性能的最优表现。

在数据集预处理阶段,主要任务是对原始遥感影像进行系统化处理,以满足

后续深度学习模型的需求。这一阶段包括数据裁剪、样本标注以及数据集的最终制作等步骤。首先，利用 ArcGIS 和 ENVI 等专业地理信息处理软件，对北极地区获取的双时相冰川遥感影像进行精确裁切，将大范围影像分割为适合模型输入的子图像块。随后，通过 ArcGIS 软件对裁切后的影像进行手工标注，标注过程中将冰川区域标记为 1，非冰川区域标记为 0，生成双时相冰川的真值图（Ground Truth）。为进一步提取变化信息，对双时相真值图执行交集取反操作，即通过逻辑运算识别出冰川发生变化的区域，最终生成变化标签图。这一过程确保了标签的准确性，为后续模型训练提供了可靠的监督信号。最后，将预处理完成的数据集按照 8:2 的比例随机划分为训练集和测试集，其中训练集用于模型参数优化，测试集用于评估模型的泛化能力。

在模型训练阶段，采用 Pytorch 深度学习框架搭建 VPGCD-Net 模型，该模型集成了视觉提示工程和 Transformer 架构，旨在高效检测冰川变化。将预处理后的训练集输入 VPGCD-Net 进行前向传播，模型通过特征提取和视觉提示引导，生成冰川变化的预测结果。为了优化模型性能，利用精心设计的损失函数（如动态交叉熵损失）计算预测结果与真值图之间的误差，衡量模型输出与实际变化的偏差。通过反向传播算法，根据损失梯度更新模型的权重参数，逐步提升检测精度。训练过程中还可引入数据增强技术（如随机旋转、翻转等），以增强模型对复杂场景的鲁棒性。此外，优化器（如 AdamW）和学习率调度策略也被用于加速收敛，确保模型在有限迭代内达到最佳状态。

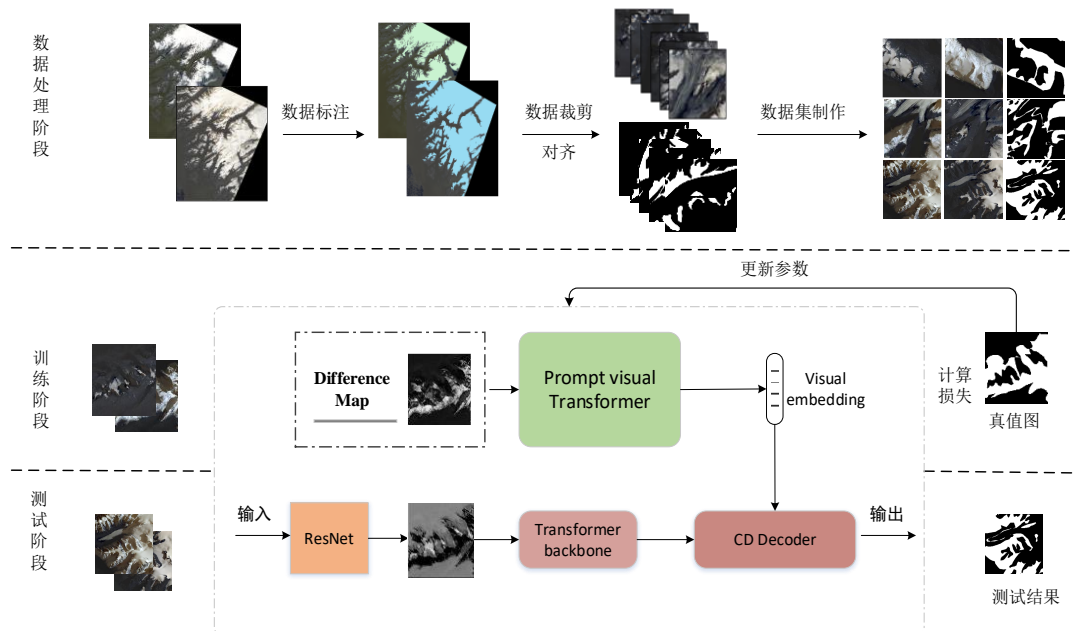


图 3.1 冰川变化检测流程图

Fig. 3.1 Flow chart for glacier change detection

在模型测试阶段,利用训练阶段得到的优化权重参数,对测试集影像进行推理分析。具体而言,将测试集中的双时相影像输入训练好的 VPGCD-Net 模型,通过正向传播生成冰川变化预测图。预测结果以像素级分类形式输出,其中每个像素被标记为变化(1)或未变化(0),直观反映冰川的动态演变。为了评估模型性能,可通过精确率、召回率、F1 分数和平均交并比(mIoU)等指标对预测图与真值图进行定量比较,从而验证 VPGCD-Net 在实际应用中的有效性和可靠性。

通过上述三个阶段的协同工作,该方法实现了从原始遥感数据到冰川变化检测结果的完整流程,为冰川监测提供了高效、精准的技术支持。

3.3 基于视觉提示的冰川变化检测网络 (Visual Prompt Driven Network for Glacier Change Detection)

3.3.1 整体网络架构

VPGCD-Net 的核心架构基于 Transformer 框架,通过视觉提示工程增强模型对冰川变化区域的识别能力。网络构建流程如下:首先对双时相冰川影像(T1 与 T2)进行差值运算生成差异图(D),经阈值处理得到二值化差异图作为视觉提示基础。该视觉提示经视觉提示转换器处理后生成视觉嵌入向量,其中视觉提示转换器采用冻结参数的 VIT-B/16 模型^[25](基于 16×16 图像块划分的基准版视觉 Transformer)。

为在输入阶段最大限度保留多源特征融合信息,本研究选用 ResNet18 作为各时相影像的特征提取器。双时相特征经拼接后输入 Transformer 编码器,该架构通过自注意力机制能有效捕捉冰川变化区域的时空演变规律。视觉嵌入向量通过特征线性调制层(FiLM)与 Transformer 编码器输出进行线性投影融合,动态调整特征表征权重以强化对变化区域的关注。解码器由多级 Transformer 模块构

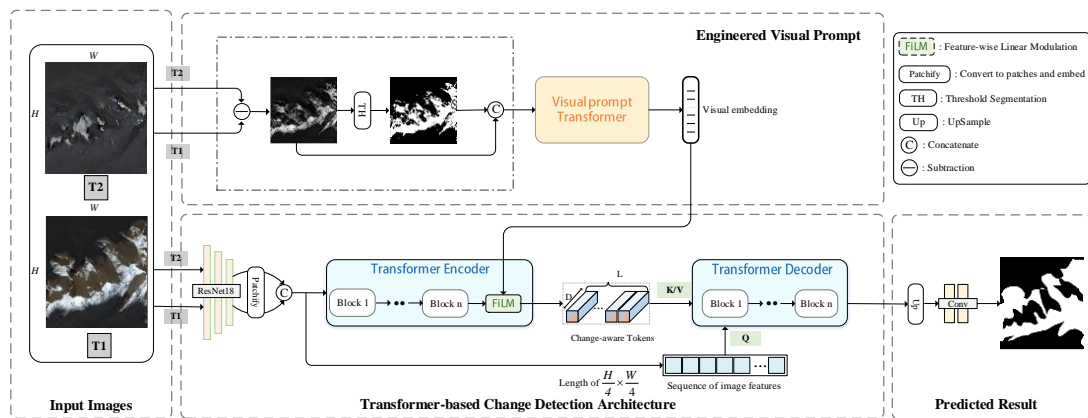


图 3.2 VPGCD-Net 网络结构示意图

Fig. 3.2 Schematic diagram of VPGCD-Net network structure.

成，采用查询-键-值机制对特征空间进行迭代优化，实现对冰川消融边界的精确定位。需要特别说明的是，视觉提示转换器在训练过程中保持参数冻结状态，仅作为固定特征提取器使用。

3.3.2 视觉提示工程

在基于 CNN 的单次语义分割方法中，掩膜池化技术通过下采样支持掩膜（support mask）、与特征图进行逐元素相乘，并在空间维度实施池化操作，以此构建包含目标特征的原型向量。然而，该方法无法直接适用于基于 Transformer 的架构，因为语义信息不仅存在于特征图层面，还会通过层级结构中的 CLS 令牌（class token）进行全局信息聚合。Luddecke 等^[59]提出将掩膜与原始图像拼接形成复合输入图像，通过视觉 Transformer 进行联合处理，该方法与自然语言处理中的提示机制具有相似原理，故称为视觉提示。本研究模型通过视觉提示工程提供的条件向量，接收关于变化区域的先验提示（即重点关注哪些区域进行变化表征）。具体而言，我们首先实施图像差异分析：给定双时相冰川影像 T1 与 T2，其灰度表示分别为 G1 与 G2，定义差异图 D 的计算公式如下：

$$D(x, y) = |G_1(x, y) - G_2(x, y)| \quad (3.1)$$

该公式通过对像素坐标 (x, y) 处的灰度值进行绝对值运算，量化双时相影像间的局部差异强度，为后续视觉提示的构建提供基础数据支撑。

随后，我们通过阈值处理生成二值化分割图：

$$T(x, y) = \begin{cases} 1 & \text{if } D(x, y) > \theta \\ 0 & \text{otherwise} \end{cases} \quad (3.2)$$

其中 θ 为预设阈值， (x, y) 表示影像像素坐标， θ 的具体取值将在后续实验中说明。将掩膜与影像特征进行融合形成复合输入，其核心在于利用阈值分割图 M 作为空间注意力引导，差异图 D 作为特征基底。该方法通过掩膜机制保留分割图中标识的感兴趣区域特征，抑制无关背景信息。具体融合过程可表示为：

$$F_{\text{masked}} = \text{Concat}(D, M) \quad (3.3)$$

式中 F_{masked} 为融合后的特征图，Concat 表示通道维度拼接操作。此操作可有效构建以变化区域为核心的特征表达，提升后续处理精度。为提取视觉提示特征，我们采用冻结参数的 ViT-B/16 模型作为视觉特征编码器。视觉提示特征向量 $F_{\text{visual_prompt}}$ 的计算可表述为：

$$F_{\text{visual_prompt}} = \text{ViT-B/16}(F_{\text{masked}}) \quad (3.4)$$

该过程通过 ViT-B/16 的层级自注意力机制，从融合特征中提取具有全局上下文感知的视觉提示嵌入，其 12 层 Transformer 编码器结构（包含 768 维隐藏层与 12 头注意力机制）可有效建模长程空间依赖关系。

3.3.3 特征线性调制

特征线性调制（Feature-wise Linear Modulation, FiLM）^[44]通过向网络中间特征施加仿射变换实现自适应特征调控，本文进行了改进。具体而言，FiLM 层学习基于视觉提示特征 $F_{\text{visual_prompt}}$ 生成调制因子 γ_d 与 β_d 的函数 f 和 h ：

$$\gamma_d = f(F_{\text{visual_prompt}}) \quad \gamma_d \in R^D \quad (3.5)$$

$$\beta_d = h(F_{\text{visual_prompt}}) \quad \beta_d \in R^D \quad (3.6)$$

其中 γ_d 和 β_d 分别表示第 d 维特征的缩放因子与偏置项。调制操作按如下方式作用于特征 $F_{l,d}$ ：

$$\text{FiLM}(F_{l,d}|\gamma_d, \beta_d) = \gamma_d F_{l,d} + \beta_d \quad F_{l,d} \in R^{L \times D} \quad (3.7)$$

该机制通过通道级仿射变换动态调整编码器输出特征的分布，使网络聚焦于视觉提示标识的冰川变化敏感区域，同时抑制由阴影、云层等干扰因素引起的噪声响应。

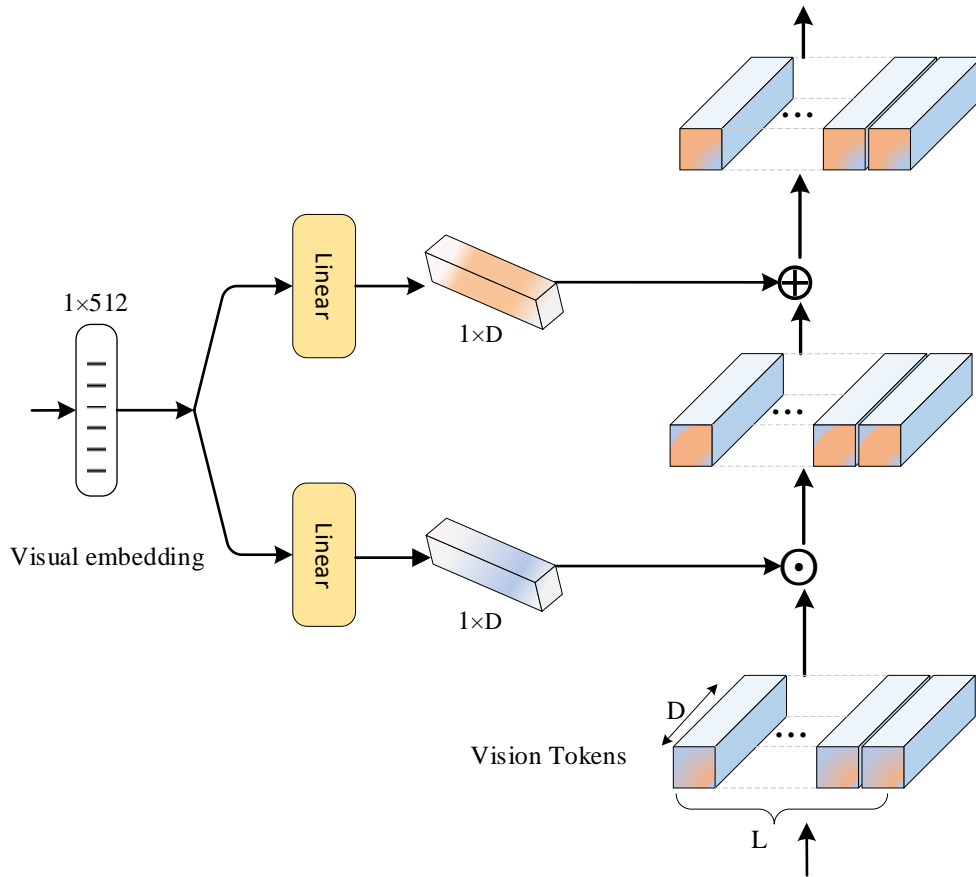


图 3.3 特征线性调制过程示意图

Fig. 3.3 Schematic diagram of the characteristic linear modulation process

3.3.4 损失函数

本研究使用一种动态自适应的交叉熵损失函数，针对遥感影像分割任务中常见的多尺度预测与标注尺寸不匹配问题，定义损失函数为：

$$L_{seg} = L_{DACE}(P, Y) \quad (3.8)$$

其中 $P \in \mathbb{R}^{B \times C \times H \times W}$ 为预测张量， $Y \in \mathbb{R}^{B \times H \times W}$ 为标注标签， L_{DACE} 包含以下两个核心机制：

(1) 动态尺寸适配机制

为解决编解码结构中特征图尺度动态变化问题，建立预测与标签的空间对齐准则：

$$P_{aligned} = \begin{cases} L_{bilinear}(P, S(Y)) & \text{if } S(P) \neq S(Y) \\ P & \text{otherwise} \end{cases} \quad (3.9)$$

式中 $S(\cdot)$ 表示空间维度， $L_{bilinear}$ 为双线性插值算子。

(2) 无效像素屏蔽机制

针对标注边界存在的未标记像素（如 $Y(i, j) = 255$ ），设计掩膜函数：

$$M(Y) = I(Y = 255) \quad (3.10)$$

修正后的损失计算为：

$$L_{DACE} = -\frac{1}{\sum M(Y)} \sum_{b=1}^B \sum_{i,j} M(Y_{i,j}^{(b)}) \cdot Y_{i,j}^{(b)} \log\left(\frac{e^{P_{i,j}^{(b,c)}}}{\sum_{c=1}^C e^{P_{i,j}^{(b,c)}}}\right) \quad (3.10)$$

(3) 类别平衡策略

针对遥感数据的长尾分布特性，引入可学习的类别权重因子 ω

$$L_{DACE} = \omega_C \cdot L_{CE}^{(e)}, \quad \omega_C = \frac{N_{total}}{N_C + \epsilon} \quad (3.11)$$

其中 N_C 为类别 c 的有效像素数， $\epsilon = 1e^{-5}$ 为平滑因子。

3.4 实验设计 (Design of experiments)

3.4.1 研究区域

本研究选取格陵兰岛东南海岸和北极斯瓦尔巴群岛作为实验区域，这两个地区均处于极地环境，具有典型的冰川地貌特征。格陵兰岛东南海岸位于北纬 64° 至 66° 、西经 51° 至 56° 之间，覆盖了格陵兰冰盖东南部的部分区域。该地区的地貌由大面积的冰川和冰原构成，冰川前缘与海岸线相接，部分区域存在裸露山地以及原始的苔原植被。由于受到北大西洋洋流的影响，该区域的冰川消融速率较快，冰川前沿变化显著，使其成为研究极地冰川动态变化的重要区域。此外，该区域的冰川运动受气候和地形因素影响，包括季节性积雪、融水径流以及冰川动力学过程。斯瓦尔巴群岛位于北纬 74° 至 81° 、东经 10° 至 35° 之间，属于北极圈

内的高纬度区域。该群岛约 60% 的面积被冰川覆盖, 30% 为裸露岩石或永久冻土区, 剩余 10% 为低矮的苔原植被。受北冰洋气候影响, 该区域的冰川主要以温冰 (polythermal glacier) 和冷冰 (cold-based glacier) 形式存在, 其消融过程受极端低温、降雪、海冰变化等因素控制。近年来, 受全球变暖影响, 斯瓦尔巴群岛的冰川后退趋势明显, 部分地区的冰川消融速率显著加快, 为研究极地冰川响应气候变化提供了关键数据支持。

3.4.2 实验数据

(1) 实验数据

本研究使用的实验数据包括两对 Landsat-8 卫星获取的光学遥感影像, 一对采集时间分别为 2015 年 8 月 1 日和 2020 年 7 月 27 日, 另一对采集时间分别为 2015 年 9 月 3 日和 2023 年 8 月 31 日, 如图 3.5 所示。

实验影像由 Landsat-8 卫星拍摄, 该卫星是 Landsat 系列的第八颗卫星, 由美国地质调查局 (USGS) 和美国国家航空航天局 (NASA) 联合研制, 并于 2013 年 2 月 11 日成功发射。Landsat-8 继承并增强了前代 Landsat 任务的监测能力, 为全球环境遥感提供了长期、稳定的高质量影像数据。Landsat-8 搭载了两台主要传感器, 其中 Operational Land Imager (OLI) 负责获取可见光、近红外 (NIR) 和短波红外 (SWIR) 波段数据, 具有更优的信噪比 (SNR) 和更高的动态范围, 能够减少大气散射的影响。Thermal Infrared Sensor (TIRS) 提供两个热红外 (TIR) 波段, 用于测量地表温度, 支持地表热异常和能量平衡分析。

OLI 传感器包括 9 个光谱波段 (可见光、近红外和短波红外), TIRS 传感器提供 2 个热红外波段, 光谱覆盖范围广泛, 能够有效支持地表覆盖类型分类、环境监测、气候变化研究以及变化检测任务。Landsat-8 的空间分辨率为 30 米 (多光谱波段) 和 100 米 (热红外波段), 同时具有较高的几何精度, 能够满足大尺度环境变化研究的需求。

(2) 数据预处理

在数据处理过程中, 所有影像经过辐射校正、大气校正和几何配准, 以确保时序影像的光谱一致性和空间对齐精度。其中, 辐射校正用于将原始影像转换为物理测量值, 以消除传感器特性对影像亮度值的影响; 大气校正通过去除大气散射和气溶胶影响, 提高影像的真实反射率; 这些预处理步骤保证了不同时间的影像在后续分析中的可比性, 提高了基于遥感影像的变化检测精度。

此外, 为了提高模型的泛化能力和鲁棒性, 我们对数据进行了数据增强处理, 包括随机旋转、水平翻转、亮度调整等操作, 以模拟不同观测条件下的影像变化, 从而增强深度学习模型在不同场景中的适用性^[65]。同时, 为了减少数据不均衡问题, 我们在数据采样过程中注重不同类型变化区域的平衡性, 确保数据集能够

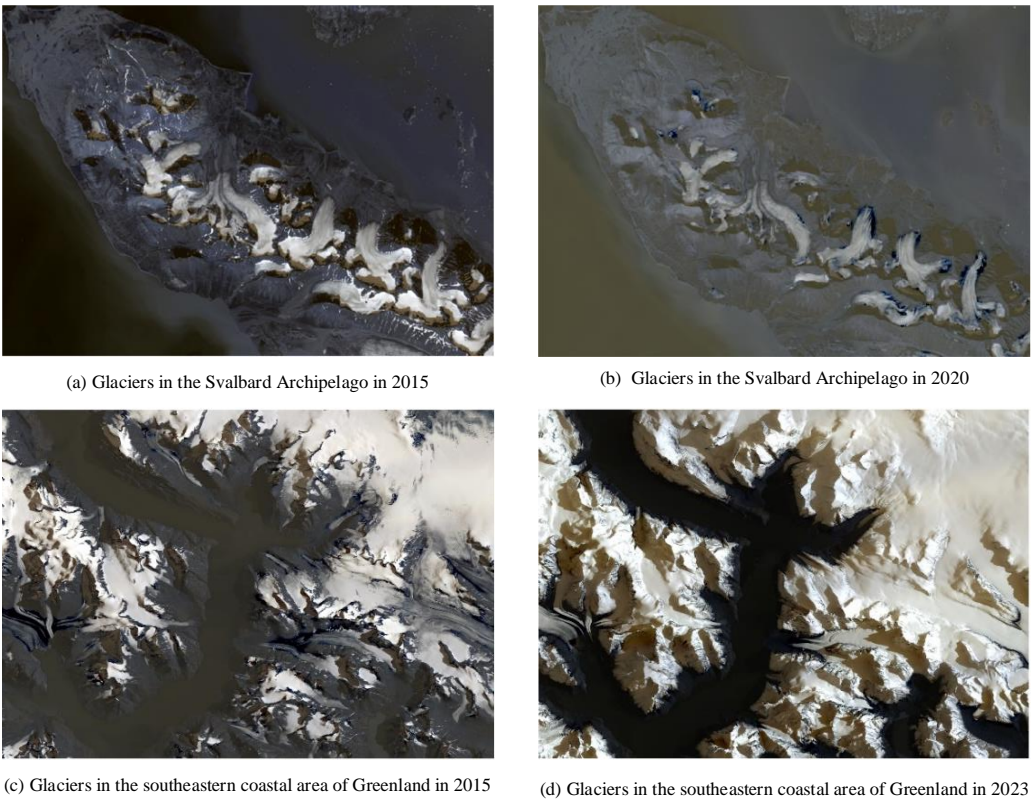


图 3.4 两个地区的冰川局部遥感图像。

Fig. 3.4 Localized remote sensing images of glaciers from two regions.

全面代表冰川动态变化特征。Landsat-8 遥感卫星图像包括 11 个波段，覆盖可见光、红外线和热红外波段，分辨率为 30 米。本章对红、绿、蓝波段进行融合，生成分辨率为 30 米的多光谱遥感影像。双时相影像使用 ENVI 软件进行配准，ArcGIS 软件完成标注，通过交集取反操作得到真值图，随后使用 Python 编程语言将遥感影像裁切为 256x256 大小的子图像。处理后的数据集以 8:2 的比例随机划分为训练集和测试集。在网络模型训练过程中，对训练集进行了数据增强，包括随机旋转、随机裁减和对比度增强等。数据集详细信息如表 3.1 所示。

表 3.1 数据集介绍

Table 3.1 Introduction of the dataset			
	数据集	图像数量	图像尺寸
训练集	冰川变化检测训练集样本图像	882	256×256×3
	冰川变化检测训练集样本真值图像	882	256×256×1
测试集	冰川变化检测测试集样本图像	221	256×256×3
	冰川变化检测测试集样本真值图像	221	256×256×1

3.4.3 实验配置

本文实验中所用设备为云服务器。其中操作系统为基于 Ubuntu 22.04 LTS 系统的工作站搭载 AMD EPYC 9654 96 核处理器与 512GB DDR5 ECC 内存，显卡为 NVIDIA A100 80GB PCIe，显存 80G，CUDA 为 11.0 版本。编码环境为 Python3.7，网络模型基于 Pytorch2.10 框架实现。利用 AdamW 优化器来优化参数，学习率为动态设置。迭代次数设定为 150 个 epoch，batch-size 设置为 32。

3.4.4 评价指标

本章主要研究基于深度学习的冰川变化检测任务，其中对于双时相遥感影像中每个像素点进行分类，是通过逐像素完成的。本文提出的 VPGCD-Net 是一种端到端的网络模型，用于对双时相冰川遥感影像进行变化检测，并生成包含 0 和 1 像素的预测图像。在这里，0 代表负样本，即双时相图像中未变化的区域，而 1 代表正样本，即双时相图像中变化的区域。为了验证所提出模型的有效性，本文中使用了四个常用的评价指标：精确率（Precision）、召回率（Recall）、F1 分数（F1-score）以及平均交并比（Mean Intersection over Union, mIoU）来对冰川变化检测结果进行定量分析。

混淆矩阵（Confusion Matrix）是一种用于评估分类模型性能的分析工具，广泛应用于机器学习领域，特别是在二分类任务中。它通过矩阵的形式直观地展示模型在不同类别上的分类情况，从而帮助研究者深入了解模型的预测能力和误分类情况。在混淆矩阵中，主要包含以下四个核心指标，真正例（True Positives, TP）：指模型正确地将实际为正类别（即变化的像素）的样本预测为正类别。假正例（False Positives, FP）：指模型错误地将实际为负类别（即未变化的像素）的样本预测为正类别，即模型误判了变化区域。真负例（True Negatives, TN）：指模型正确地将实际为负类别的样本预测为负类别，表明模型准确识别了未变化区域。假负例（False Negatives, FN）：指模型错误地将实际为正类别的样本预测为负类别，即模型未能识别出真正的变化区域。

基于混淆矩阵，可以计算多个性能指标，其中精确度（Precision）是衡量模型在预测正类别时的可靠性的重要指标。它表示模型所预测的正类别样本中，有多少是真正的正类别，数值范围在 0 到 1 之间。精确度越高，说明模型在正类别的预测中误判（假正例）较少，从而具备更好的分类能力。精确度的计算方式如公式 (3.12) 所示。

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3.12)$$

召回率（Recall）是变化检测任务中至关重要的性能指标，也被称为灵敏度（Sensitivity）或真正例率（True Positive Rate, TPR）。它衡量了模型在所有实际

发生变化的区域中，成功检测出的比例，反映了模型对正类别样本（变化像素）的识别能力。

召回率的取值范围在 0 到 1 之间，值越接近 1，说明模型在识别变化区域时的表现越出色。较高的召回率意味着模型能够有效检测大部分实际发生的变化区域，而较低的召回率则表明模型存在较多的漏检情况，即未能识别出所有的变化区域。

在变化检测任务中，召回率的计算方式如公式 (3.13) 所示。

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3.13)$$

F1分数 (*F1 – score*) 是用于评估分类模型性能的重要指标，特别适用于类别分布不均衡的情况。它是精确度 (*Precision*) 和召回率 (*Recall*) 的调和平均值，旨在权衡模型的准确性和灵敏度，以确保模型在检测变化的同时，尽量减少误报。

F1分数的取值范围在 0 到 1 之间，值越接近 1，表示模型在精准识别变化区域的同时，误报较少，从而具备更好的整体性能。在变化检测任务中，**F1 分数**能够有效反映模型在正确检测变化区域 (*True Positive*) 和避免错误分类 (*False Positive*) 之间的平衡性，因此是评估模型稳定性和可靠性的重要指标。

F1分数的计算方式如公式(3.14)所示。

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3.14)$$

平均交并比 (*Mean Intersection over Union, mIoU*)，又称 *Jaccard* 系数，是衡量模型预测结果与真实标签之间空间重叠程度的重要指标，广泛用于变化检测任务。它通过计算预测变化区域与实际变化区域的交集 (*Intersection*) 与并集 (*Union*) 之比，来反映模型的检测精度。

IoU 的取值范围在 0 到 1 之间，值越接近 1，表示模型预测的变化区域与真实变化区域的匹配度越高，从而说明模型在空间定位上的准确性越强。相比于单独的精确度 (*Precision*) 或召回率 (*Recall*)，**IoU** 能够更全面地反映模型的整体检测能力，特别是在评估空间分布一致性方面具有重要意义。

在变化检测任务中，较高的 **IoU** 值表明模型不仅能够正确检测到变化区域，还能确保预测的变化区域范围接近真实情况。因此，**IoU** 被广泛应用于评估深度学习模型在遥感影像变化检测中的性能。

IoU 的计算方式如公式(3.15)所示。

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (3.15)$$

3.5 实验结果与分析 (Experimental Results and Analysis)

3.5.1 变化检测结果分析

VPGCD-Net 在冰川变化检测任务中的表现如图 3.6 所示，其中包括输入的双时相遥感影像（T1 和 T2）、真实变化区域标注（GT），以及 VPGCD-Net 生成的变化检测结果，红色像素为漏检，绿色像素为误检。为了进一步评估 VPGCD-Net 的检测能力，我们在不同测试场景下分析其对变化区域的提取情况，并重点比较其在复杂环境下的稳健性。从图中可以看出，VPGCD-Net 能够精准地提取冰川变化区域，并有效减少误检和漏检现象。在红色和绿色框选区域内，VPGCD-Net 成功捕捉到细小的冰川变化（白色区域），同时避免了由于光照、阴影或地物光谱变化引起的误检（绿色像素）。相比之下，传统方法（如比值法和 PCA）通常在光谱变化较大的区域出现误判，而 VPGCD-Net 通过视觉提示引导增强了模型对真实变化的感知能力，使得结果更加一致。在黄色框选区域中，VPGCD-Net 识别出了复杂地形中的微小冰川变化，而传统方法往往

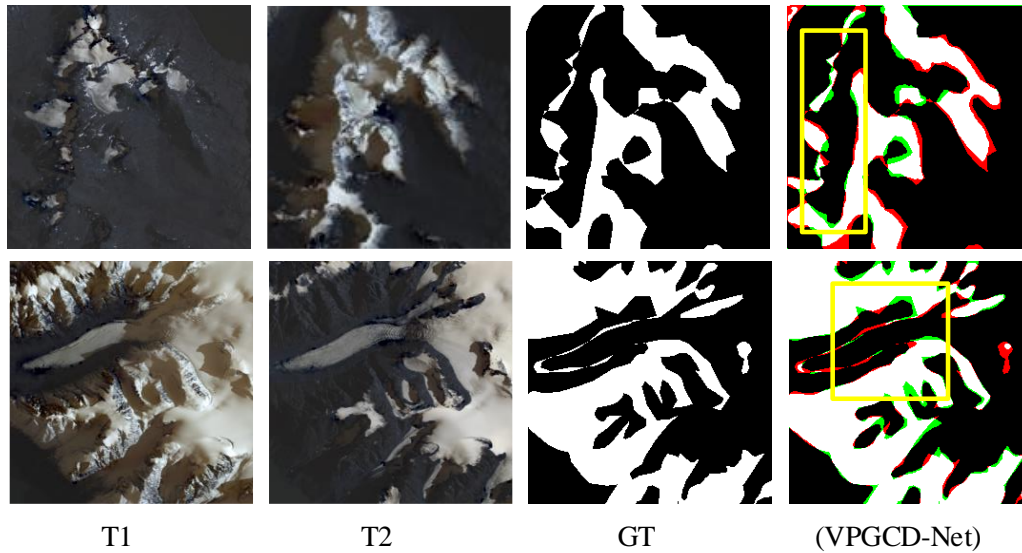


图 3.5 VPGCD-Net 变化检测结果可视化

Fig. 3.5 Visualization of VPGCD-Net change detection results

难以区分这些区域。特别是在低光照条件下，VPGCD-Net 仍然能够保持较高的检测精度，这得益于其 Transformer 结构的全局建模能力以及 FiLM 模块的自适应特征调控能力。

尽管 VPGCD-Net 在绝大多数情况下表现优异，但在某些极端场景（如阴影区域或雪盖反射率变化显著的区域）仍然可能出现一定程度的误检或漏检。然而，相比其他方法，VPGCD-Net 误检率显著降低，表明其在复杂冰川环境下具有更强的适应性。

3.5.2 对比实验分析

1) 与传统变化检测方法对比

为了验证所提出 VPGCD-Net 方法在冰川变化检测任务中的有效性,我们选择了几种传统和经典的变化检测方法进行对比分析,包括比值法(Ratio Method)、插值法(Interpolation Method)、主成分分析(PCA)、变化矢量分析(CVA)等。这些方法的实现原理如下:比值法(Ratio Method)利用两时相影像的光谱比值来检测变化区域,即计算同一像素在 T1 和 T2 影像的光谱值之比,并设定阈值来区分变化区域。该方法对光照变化较为敏感,容易受到阴影和反射率变化的影响。插值法(Interpolation Method)通过对 T1 和 T2 影像进行插值运算来估计未观测时刻的像素值,并计算实际观测值与估计值之间的差异来判断变化区域。插值法能够减少噪声干扰,但在复杂地形环境下精度可能受限。主成分分析(PCA)通过对双时相影像进行特征变换,提取主要变化信息,并利用前几主成分的分布特征来识别变化区域。PCA 具有较强的特征提取能力,但可能存在信息损失,导致检测结果偏差。变化矢量分析(CVA)通过计算像素的光谱变化向量(通常在多维空间中计算欧几里得距离)来量化变化程度,并通过设定阈值进行分类。CVA 具有较高的鲁棒性,但在复杂背景下可能存在误检。

VPGCD-Net 与传统方法的变化检测结果如图 3.7 所示,相比之下,VPGCD-Net 在多种测试场景下均表现出更优的检测能力,尤其是在光谱特征存在显著变化的冰川场景下,能够显著降低漏检率,并提高检测结果的完整性和一致性。

如图 3.7 (c) 和 (d) 部分蓝色框所示,比值法(Ratio Method)与插值法(Interpolation Method)由于对光照变化较为敏感,不仅存在严重的漏检,同时伴有较多椒盐噪声,使得检测结果较为破碎。相比之下,主成分分析(PCA)通过对多维数据进行降维处理,以保留数据的主要变异性,从而减少噪声干扰。如图 3.7 (f) 部分所示,尽管 PCA 在一定程度上降低了椒盐噪声,但仍然存在明显的误检和漏检现象,这些误检区域原影像都存在明显的阴影,表明 PCA 方法无法排除阴影带来的影像干扰。此外,变化矢量分析(CVA)通过计算光谱差异向量的大小和方向来进行变化判断,能够在显著变化区域的检测中表现良好,但对于微小变化的识别能力仍显不足,如图 3.7 (g) 蓝色框所示,部分细微变化区域未能正确识别。从第三行的可视化结果中可以看到,由于成像角度和时间的关系,该场景中双时相影像中冰川沟壑拥有众多阴影板块,这直接导致了传统方法大面积的误判,相比之下,VPGCD-Net 通过融合视觉提示(Visual Prompt)与 Transfmer 架构,能够更深入地建模双时相影像之间的时空特征,能有效降低因为山地阴影导致的误判和漏提,在不同光照和光谱变化条件下保持较强的鲁

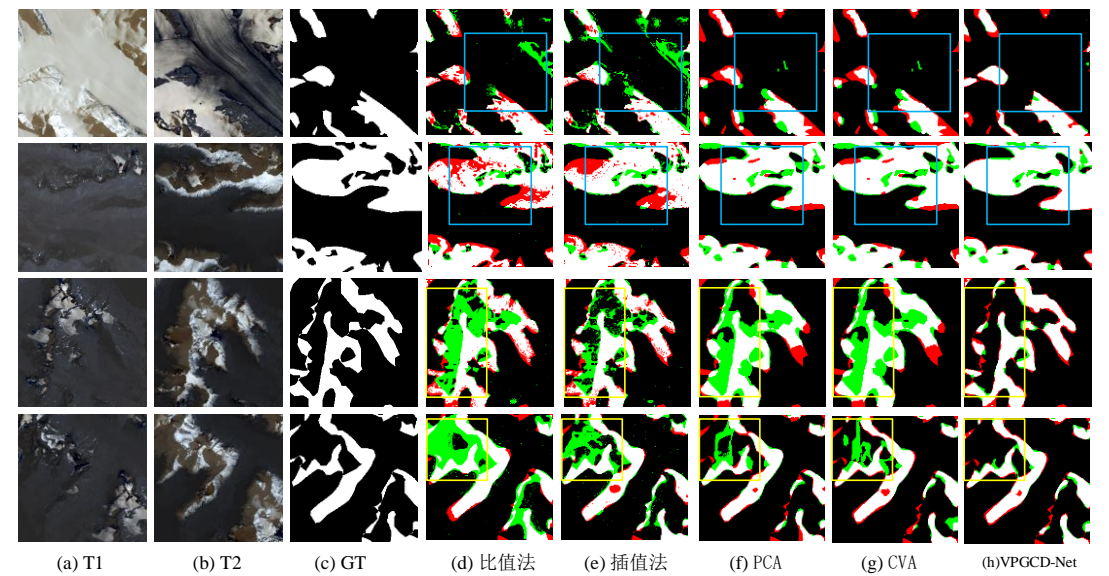


图 3.6 VPGCD-Net 与传统方法的实验结果定性分析对比
Fig. 3.6 Comparison of qualitative analysis of experimental results
between VPGCD-Net and traditional methods

棒性。

综上所述，实验结果表明 VPGCD-Net 在冰川变化检测任务中表现优越，尤其在应对光谱变化显著以及光照角度不同的复杂环境时，其准确性和鲁棒性均优于传统方法。这一结果充分验证了 VPGCD-Net 结合视觉提示机制和 Transformer 结构的有效性，为高精度冰川变化检测提供了新的方法与思路。

如表 3.2 所示展示了 VPGCD-Net 与传统变化检测方法的定量对比结果，从精确度（Precision）、召回率（Recall）、F1-score 和平均交并比（mIoU）四个指标进行评估。从表中数据可以看出，VPGCD-Net 在所有评估指标上均显著优于传统方法，特别是在 F1-score 和 mIoU 指标上的提升尤为突出。这表明该方法不仅能够更精准地检测冰川变化区域，同时在降低误检和漏检率方面也具备明显优势。精确度（Precision）方面，VPGCD-Net 的精确度达到了 95.3%，相比 CVA（90.3%）提升了 5.0%，相比 PCA（89.1%）提升了 6.2%。这说明 VPGCD-Net 在识别真实变化区域的能力上更强，有效减少了误检率，避免了将非变化区域误认为是变化区域的情况。召回率（Recall）方面，在召回率方面，VPGCD-Net 达到了 94.5%，相比 CVA（86.1%）提高了 8.4%，相比 PCA（84.3%）提高了 10.2%。

表 3.2 VPGCD-Net 与传统变化检测方法的定量对比

Table 3.2 Quantitative comparison of VPGCD-Net and traditional change detection methods				
方法	精确度 (%)	召回率 (%)	平均交并比 (%)	F1 (× 100)
比值法	80.2	82.7	72.1	83.9
插值法	87.4	80.9	72.5	84.0
PCA	89.1	84.3	75.4	86.6
CVA	90.3	86.1	77.2	88.1
VPGCD-Net	95.3	94.5	90.9	95.2

较高的召回率表明 VPGCD-Net 能够识别更多真实的变化区域,从而降低漏检率。相较之下,传统方法由于对光谱变化的敏感性较高,往往会导致部分变化区域被忽略,影响检测的完整性。F1-score 方面,VPGCD-Net 的 F1-score 达到了 95.2%,比 CVA (88.1%) 提升了 7.1%,比 PCA (86.6%) 提升了 8.6%。这一提升反映出 VPGCD-Net 在减少误检和漏检的同时,能够更准确地平衡检测结果,适用于复杂冰川场景的变化检测任务。mIoU 方面,VPGCD-Net 的 mIoU 达到了 90.9%,远超 CVA (77.2%) 和 PCA (75.4%)。这一结果表明 VPGCD-Net 所提取的变化区域与真实标注区域的重叠程度更高,检测结果更加精确和完整。相比之下,传统方法受限于光谱变化和噪声影响,难以准确捕捉完整的变化区域,导致 mIoU 较低。

2) 与深度学习变化检测方法对比

我们比较了六种通用的基于深度学习的遥感变化检测方法,用于冰川遥感变化检测对比实验。包括 ChangeFormer^[60],采用了一种基于 Transformer 的孪生网络架构,通过结合分层 Transformer 编码器和 MLP 解码器,有效地提取了多尺度的特征,以提高变化检测的精度。DDPM-CD^[61]创新地将去噪扩散概率模型应用于变化检测任务中。DDPM-CD 通过模拟数据的马尔可夫链扩散过程,学习数据的分布,并在推理时从噪声中重建图像。这种方法能够生成高质量的图像样本,并且在作为特征提取器时,能够为变化检测任务提供强大的特征表示。BIT^[57]利用 Transformer 模型在时空域内有效地建模上下文关系,通过将图像转换为语义标记并在标记空间中进行上下文建模,提高了变化检测的性能。BGSINet-CD^[58],通过图语义交互网络在语义层面上增强了双时相特征的交互,通过软聚类和图交互模块提高了变化检测的准确性和计算效率。EATDer^[62]整合自适应视觉 Transformer 和边缘感知解码器,生成具有精细边缘的变化图。SAM-CD^[63]通过卷积适配器和语义学习分支建模双时相遥感影像的语义表征。

从实验结果对比可以看出如表 3.3,提出的模型在所有评估指标上显著优于现有的基线模型,这表明,所设计的网络在冰川遥感变化检测任务中具备更强的性能。具体而言,该模型在 F1 分数和 mIoU 分别为 0.952 和 90.9%,均明显高于其他模型。此外,较高的召回率(94.5%)和精确度 (95.3%) 表明模型在检测变化区域时不仅能够有效减少漏检,还能够大幅降低误报。实验结果可视化如图 3.8 所示,表格展示了五种方法在多个冰川遥感影像中的变化检测结果。根据误检(绿色)和漏检(红色)区域的分布情况,详细分析如下:BIT 方法在所有实验中表现出较为明显的误检和漏检区域,尤其是在图像边缘部分,误差较为突出。绿色的误检区域较为分散,表明该方法在处理影像中的细节时,容易将非变化区域误判为变化。

表 3.3 VPGCD-Net 与深度学习变化检测方法的定量对比

Table. 3.3 Quantitative comparison of VPGCD-Net and deep learning change detection

methods				
方法	精确度 (%)	召回率 (%)	平均交并比 (%)	F1 ($\times 100$)
Ice-UNet	89.5	87.6	81.0	88.1
BIT	91.0	90.5	83.1	90.4
BGSINet	89.7	89.4	85.2	89.5
DDPM-CD	90.4	91.5	84.0	91.0
ChangeFormer	91.2	90.1	83.1	90.6
EATDer	93.1	96.7	86.5	93.1
SAM-CD	93.3	93.1	85.4	93.2
VPGCD-Net	95.3	94.5	90.9	95.2

如图 3.8 中(d)部分所示, Ice-UNet 对于细节和变化区域的提取能力较差, 存在大量的误检和漏检, 如图 3.8 中(e)、(f)、(g)、(h)部分黄色框所示, BIT、DDPM-CD、Changeformer、EADter 方法在应对影像中阴影干扰时效果较差, 存在着大量的误检区域。虽然 Changeformer 方法在误检和漏检控制方面略有改善, 但仍然可以看到明显的红色和绿色区域, 该方法在中尺度的冰川区域表现尚可, 但在处理边缘和细节变化时, 存在较多误检。如图 3.8 中(h)部分黄色框所示, BGSINet 方法在误检和漏检方面的控制较为出色, 特别是在细节处理上表现出一定的优势, 有着不错的抗阴影干扰能力, 但仍然存在一定程度的漏检, 仍有进一步优化的空间。值得注意的是, EATDer 在定量分析中表现欠佳, 但在连续狭窄变化区域检测中表现突出, 这得益于其自适应视觉 Transformer 和边缘感知解码器对细粒度边缘特征的提取能力。尽管 VPGCD-Net 也能检测狭窄边界, 但其定位性能仍有提升空间。总体而言, 本文提出的 VPGCD-Net 在抑制阴影干扰方面表现卓越, 显著降低了误检和漏检率, 展现出最优的综合性能。

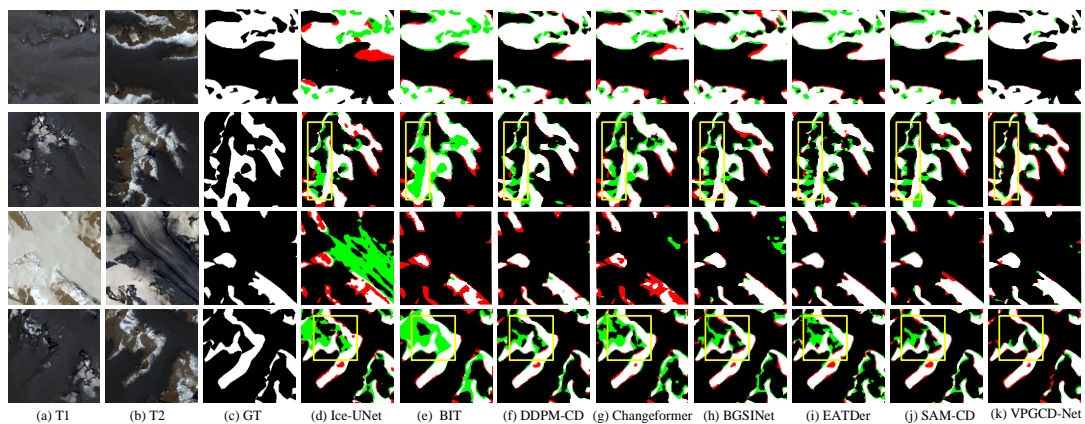


图 3.7 VPGCD-Net 与深度学习方法的实验结果定性分析对比

Fig. 3.7 Qualitative analysis of experimental results comparing VPGCD-Net with deep learning methods

3.5.3 消融实验

为了进一步验证所提出 VPGCD-Net 的有效性，我们设计了一系列消融实验，通过逐步移除关键模块来分析其对模型性能的影响。实验在本文提出的冰川变化检测数据集上进行，并采用 Precision、Recall、F1-score 和 mIoU 作为评估指标。

我们设计了以下消融实验：如表 3.4 所示，w/o VP 表示去除视觉提示（Visual Prompt）模块，仅使用 Transformer-based Change Detection（TCD）进行冰川变化检测。w/o FiLM 表示去除特征线性调制（FiLM）层，不进行基于视觉提示的自适应特征调制。w/o Transformer Decoder 表示移除 Transformer 解码器，仅使用编码器以评估解码器对模型性能的影响。w/o Threshold Mask 表示去除视觉提示中的阈值分割掩码，仅使用原始差分图进行特征引导。

VPGCD-Net w/o Transformer 表示用 CNN 结构（如 UNet）替代 Transformer 进行特征提取和变化检测。

表 3.4 VPGCD-Net 中关键模块消融实验定量分析

Table. 3.4 Quantitative analysis of key module ablation experiments in VPGCD-Net				
模型	精确度(%)	召回率(%)	平均交并比(%)	F1 (×100)
w/o VP	92.1	91.7	86.5	91.9
w/o FiLM	93.0	92.3	87.8	92.6
w/o Transformer Decoder	92.8	92.1	87.4	92.4
w/o Threshold Mask	94.0	92.8	88.5	93.3
VPGCD-Net	95.3	94.5	90.9	95.2

如表 3.4 所示，去除视觉提示（w/o VP）后，模型性能显著下降，平均交并比下降 3.3%，F1-score 下降 3.3%。这表明视觉提示在强调变化区域、降低伪变化干扰方面起到了重要作用。去除 FiLM 组件（w/o FiLM）后，平均交并比下降 2.1%，F1 分数下降 2.6%，表明 FiLM 通过自适应调制增强了特征表达能力。去除阈值分割掩码（w/o Threshold Mask）后，平均交并比下降 2.4%，说明阈值分割掩码有助于增强变化区域的检测精度。用传统 CNN 替代 Transformer Decoder（w/o Transformer Decoder）后，平均交并比下降 5.2%，F1-score 下降 4.1%，表明 Transformer 能够更有效地建模时空变化关系，提高冰川变化检测的准确性。

如图 3.9 所示为消融实验的定性分析结果。可以看到当去除了视觉提示模块或者 FiLM 组件之后，模型对于阴影干扰的抑制能力减弱，如 3.9 中(d)、(e)部分所示，也就是说，当没有了有效的视觉提示后，模型的整体检测效果会下降。如图 3.9 中(f)部分所示，在使用传统 CNN 替代 Transformer Decoder(w/o Transformer

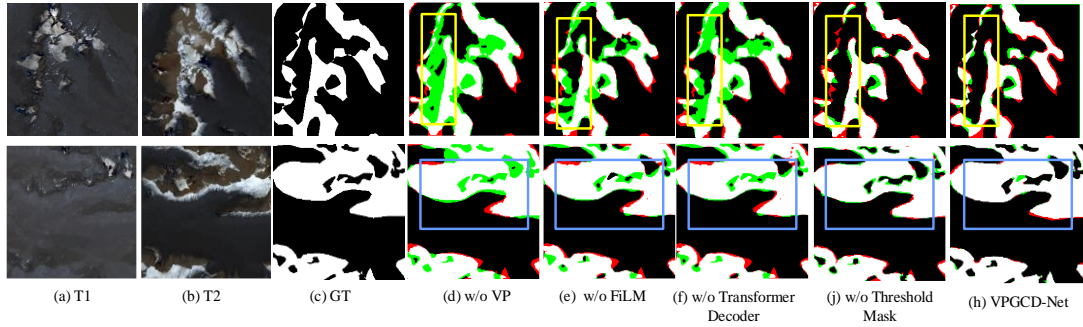


图 3.8 VPGCD-Net 中关键模块消融实验定性分析

Fig. 3.8 Qualitative analysis of ablation experiments of key modules in VPGCD-Net

Decoder) 后, 模型对于整体变化区域的提取能力下降, 出现了一定城的误检和漏检, 在去除阈值分割掩码 (VPGCD-Net w/o Threshold Mask) 后, 如图 3.9 中 j 部分黄色框所示, 模型漏检增多, 说明阈值分割掩码对于模型精度的提升起到了一定的作用。

本小节消融实验表明, VPGCD-Net 的每个组件均对整体性能有显著贡献。其中, 视觉提示模块和 FiLM 组件在应对复杂阴影影响、光照变化等挑战方面起到了关键作用。实验结果验证了所提出方法在冰川变化检测任务中的有效性。

3.5.4 参数实验分析

如图 3.10 所示展示了不同阈值下平均交并比 (mIoU/%) 和 F1 分数的变化趋势。当阈值从 0.05 增至 0.20 时, mIoU 从 84.6%快速提升至 90.1%, 随后在 0.20 至 0.50 阈值区间内稳定维持在 90.4%至 91.0%之间。类似地, F1 分数从 0.895 持续上升, 在阈值为 0.25 时达到峰值 0.953, 随着阈值继续增大, 其数值始终稳定在 0.950 左右。这表明模型对阈值选择具有强鲁棒性, 且性能稳定性显著优于

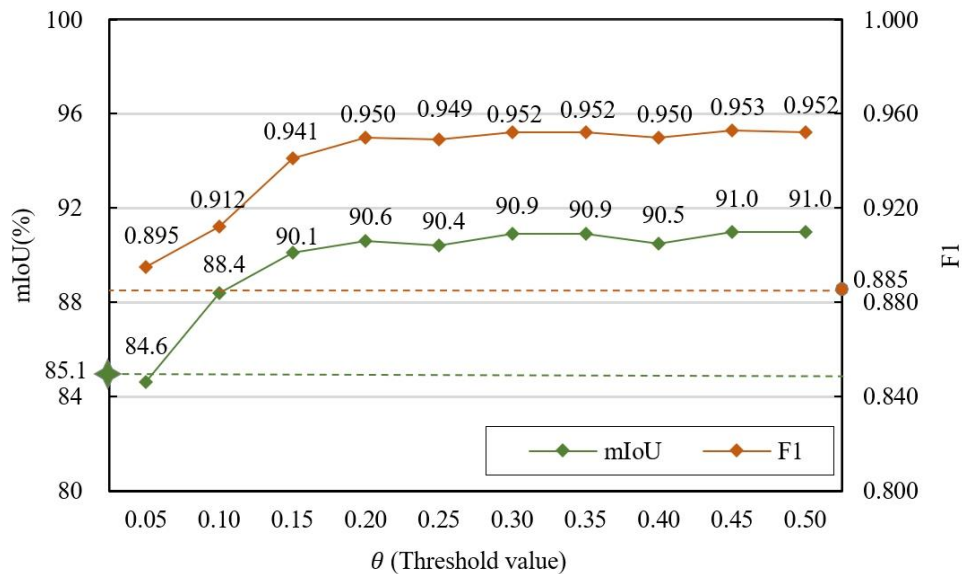


图 3.9 阈值参数对网络性能的影响分析

Fig. 3.19 Analysis of Threshold Parameter Effects on Network Performance.

基线网络。本实验最终设定阈值为 0.30。值得注意的是，引入视觉提示机制后，模型性能始终优于基准网络（如 0.30 阈值下 mIoU 提升 3.8%），这验证了视觉提示工程在特征空间调制中的有效性。具体而言，视觉提示通过引导网络关注冰川消融区域的显著性特征，显著降低了阴影干扰导致的误判率。

3.6 本章小结（Chapter Summary）

本章首先详细介绍了冰川变化检测流程，包括数据预处理、深度神经网络训练和测试。然后，详细阐述了基于视觉提示的冰川变化检测网络 VPGCD-Net，该网络由主干网络、transformer 编码器-解码器，视觉提示工程模块构成。通过引入视觉提示指导网络模型专注于获取变化区域的视觉表征，VPGCD-Net 能够准确捕捉变化分布，抑制阴影区域干扰。接着，本章对包含各种情形的冰川遥感影像进行了实验设计。最后，与其他冰川变化检测方法进行了对比分析，并进行了消融实验和参数分析实验以验证网络性能表现。

4 基于边缘感知的冰川变化检测方法

4 Glacier Change Detection Method based on Edge Awareness

4.1 引言 (Introduction)

冰川变化检测在地理环境监测、气候变化研究及生态保护等领域具有重要意义。然而,由于冰川变化区域的边界模糊、尺度多样性以及复杂光照条件(如阴影和低光照成像质量下降),传统变化检测方法在精确捕捉冰川边界和小尺度变化方面仍存在挑战。因此,如何增强模型对变化区域边缘的敏感性,并在多尺度特征表达中有效融合局部细节与全局语义,成为提升冰川变化检测精度的关键问题。

为此,本文提出了一种基于边缘感知的冰川变化检测网络(Edge-Aware Network for Glacier Change Detection, EACD-Net),旨在通过显式建模多尺度边缘信息,提高模型对变化区域边界的感知能力,并增强对复杂光照环境和模糊边界的鲁棒性。EACD-Net 由特征编码器(Feature Encoder)、边缘感知模块(Edge-Aware Module)和变化解码器(Change Decoder)三个核心部分组成,形成“特征编码→边缘增强→多尺度融合解码”的递进式检测框架。其中,特征编码器采用 ResNet-50 作为骨干网络,并通过结构调整优化特征提取能力,以确保双时相特征对齐;边缘感知模块通过多尺度边缘信息建模,有效增强变化区域的边界敏感性;变化解码器融合多尺度特征与边缘先验,实现像素级精细变化检测。

EACD-Net 采用孪生网络架构独立提取双时相遥感影像特征,并在深层引入跨时相注意力交互与对比损失约束,以提升变化区域的对比度和特征聚焦能力。同时,在特征提取过程中,采用空洞卷积扩展感受野,避免特征分辨率过度损失,从而在保持边界细节的同时增强全局语义信息。实验结果表明, EACD-Net 在多个冰川变化检测数据集上均取得了优异性能,尤其在边界模糊和小尺度变化区域的检测精度方面,相较于现有方法具有明显优势。

4.2 基于边缘感知的冰川变化检测网络 (Edge-aware glacier change detection network)

4.2.1 网络整体架构

本文提出了一种边缘感知的冰川变化检测网络(Edge-aware glacier change detection network, EACD-Net),旨在准确捕捉冰川变化区域,尤其是模糊边界和小尺度变化,通过结合局部与全局特征表达,增强模型对复杂光照、阴影和模糊

冰川边界的鲁棒性。如图所示，EACD-Net 由特征编码器(Feature Encoder)、边缘感知模块 (Edge-Aware Module)和变化解码器(Change Decoder)三个核心组件组成，整体结构如图 4.1 所示，特征编码器 (Backbone)：负责提取双时相遥感图像的多层次语义特征；边缘感知模块 (Edge-Aware Module)：显式建模多尺度图像边缘信息，增强变化区域的边界敏感性；变化解码模块 (Change Decoder)：融合多尺度特征与边缘先验，生成像素级变化检测图，模型通过 “特征编码→边缘增强→多尺度融合解码” 的递进式流程，实现从底层细节到高层语义的协同优化。

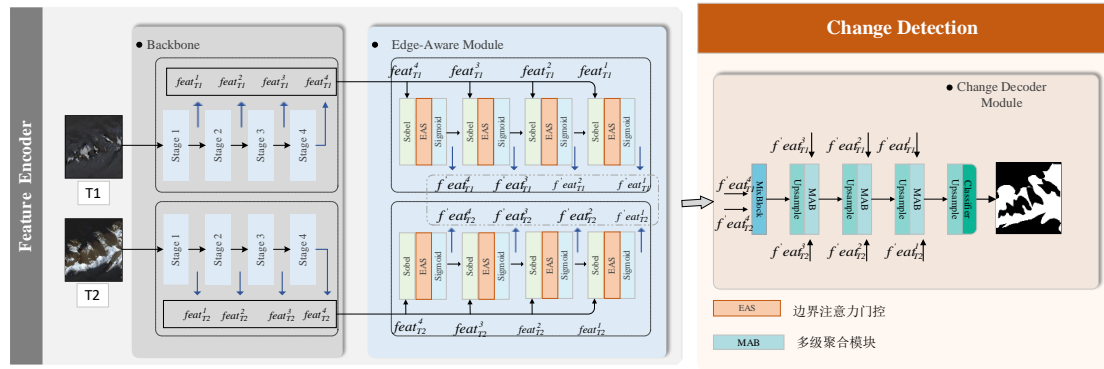


图 4.1 EACD-Net 网络结构示意图

Fig. 4.1 Schematic diagram of EACD-Net network structure.

4.2.2 特征编码器设计

特征编码器作为模型的核心特征提取单元，其设计需兼顾多层次语义信息的捕获能力与双时相数据的特征对齐需求。本节从骨干网络架构、特征层级特性及双时相处理机制三方面展开详细说明。采用 ResNet-50 作为默认骨干网络，其残差连接设计可缓解梯度消失问题，支持深层网络的稳定训练。输入层调整：将原始 ResNet 的 7×7 卷积核替换为 3×3 小卷积核 (步长=2)，减少下采样初特征编码器，以 ResNet-50 为骨干网络进行适应性改造，输入为 $256 \times 256 \times 3$ 的遥感图像，通过分层下采样与特征变换生成多尺度特征图。以 T1 时相图像为例，具体而言，首先在 Stage1 阶段，通过 3×3 卷积 (stride=2) 和 3×3 最大池化 (stride=2) 进行初始下采样，输出 $64 \times 64 \times 64$ 的特征图 ($feat_{T1}^1$)，保留高频边缘与纹理细节；随后 Stage2 堆叠残差块并进一步下采样至 32×32 分辨率，通道数增至 256 ($feat_{T1}^2$)，用于捕捉中级语义结构 (如地物局部组合模式)；Stage3 通过更深的残差块堆叠和下采样操作，输出 $16 \times 16 \times 512$ 的特征图 ($feat_{T1}^3$)，编码全局上下文信息 (如大尺度地物分布)；最后在 Stage4 中，引入空洞率=2 的 3×3 空洞卷积替代常规下采样，扩大感受野的同时保持分辨率，再通过步长=2 的卷积得到 $8 \times 8 \times 1024$ 的特征图 ($feat_{T1}^4$)，建模场景级抽象语义 (如空间依赖关系)。

各阶段通道数按 $64 \rightarrow 256 \rightarrow 512 \rightarrow 1024$ 逐层倍增，兼顾浅层细节保留与深层语义抽象，其中 Stage4 的空洞卷积设计避免了分辨率过度损失，平衡了细节与全局信息的提取需求。对于双时相输入，编码器采用参数共享的孪生网络独立提取特征，并在深层引入跨时相注意力交互与对比损失约束，确保特征对齐与变化区域的精准聚焦。期细节丢失风险。

4.2.3 边缘感知模块

本方法提出一种双路径边缘建模机制，其内部结构如图 4.2 所示，结合显式边缘检测与可学习语义边缘推理，通过跨层级注意力门控融合策略，增强变化区域边界的定位精度。以 T_1 时相图像为例，边缘感知模块首先对特征 $feat_{T_1}^i$ 施加 Sobel 算子，生成像素级边缘响应图 $E_{T_1}^i$ 。具体而言，Sobel 算子通过水平方向卷积核 G_x 与垂直方向卷积核 G_y 计算梯度幅值：

$$E_{T_1}^i = \sqrt{(G_x * feat_{T_1}^i)^2 + (G_y * feat_{T_1}^i)^2} \quad (i = 1, 2, 3, 4) \quad (4.1)$$

其中 $*$ 表示卷积操作。 $E_{T_1}^i$ 表示第 i 个特征经过处理后得到的像素级边缘响应图。 G_x 和 G_y 分别为水平与垂直方向的 Sobel 核矩阵。此步骤显式捕获输入特征的空间边缘细节。

为融合多层级边缘信息，采用跨层级边缘注意力门控机制，们通过反卷积操作 $D(\cdot)$ 对 Sobel 算子计算得到的边缘图 $E_{T_1}^i$ 进行空间对齐，并与原始特征 $feat_{T_1}^i$ 相乘后进行函数激活：

$$F' = \sigma(D(E_{T_1}^i) \cdot feat_{T_1}^i) \quad (4.2)$$

其中， $feat_{T_1}^i$ 代表层级特征， $E_{T_1}^i$ 代表边缘特征， σ 代表 Sigmoid 激活函数。该操作可以保留丰富的高层语义信息，同时增强模型对边界区域的关注度。

此外，为了进一步优化融合后的特征表示，保留原始特征细节，获得更全面的特征表示，我们对特征编码器阶段输出的原始特征使用了混合的空间注意力与通道注意力，计算如下：

$$F'' = SA(CA(feat_{T_1}^i) \odot feat_{T_1}^i) \otimes CA(feat_{T_1}^i) \odot feat_{T_1}^i \quad (4.3)$$

其中 CA 代表通道注意力操作， SA 代表空间注意力操作。 \odot 代表逐通道相乘， \otimes 代表逐元素相乘，然后通过逐元素相乘的形式将前面处理好的两个特征进行融合：

$$F^i = F' \otimes F'' \quad (4.4)$$

我们在特征融合后继续采用多个卷积层进行特征增强。经过两次 3×3 卷积和批归一化处理后，我们最终使用 1×1 卷积进行通道映射，以降低计算复杂度并减少冗余信息，最终输出边界增强特征 $f'eat_{T_1}^{i-1}$ ，代表 i 级处理后得到的增强特

征, 该操作能够有效抑制噪声, 同时确保边缘信息的完整性, 使得网络能够更精确地检测冰川变化区域的边界。

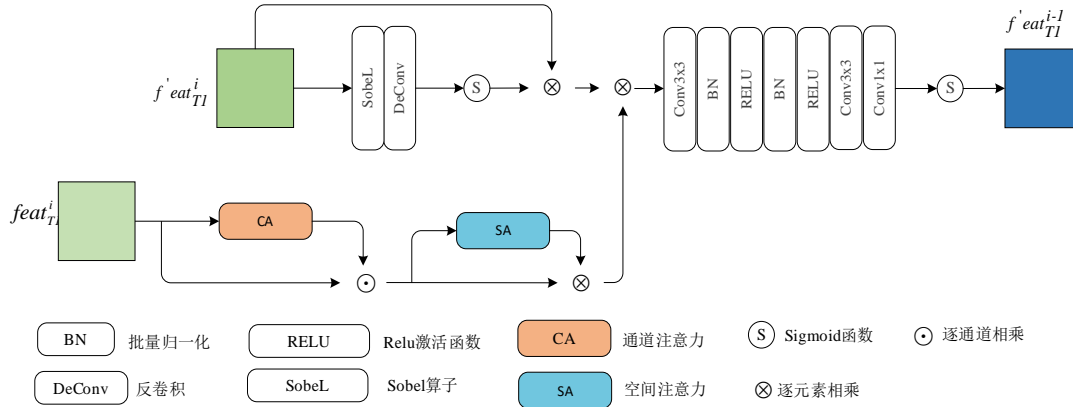


图 4.2 边缘注意力门控机制结构示意图

Fig. 4.2 Schematic diagram of the structure of the edge attention gating mechanism

4.2.4 变化检测解码器

变化解码模块通过自底向上特征融合与边缘引导的多尺度聚合, 该模块由多注意力块与混合块构成, 将编码器提取的语义特征与边缘感知信息融合, 输出高精度变化概率图。如图 4.3 所示展示了多尺度特征聚合块 (Multi-Scale Aggregation Block, MAB) 的详细结构, MAB 的设计结合了变化注意力机制、特征混合以及空间分辨率恢复, 特别适用于冰川遥感影像中模糊边界和变化区域的精确定位。MAB 内部结构如图 4.3 所示, 变化检测解码器详细设计如下:

两幅时间图像之间的特征交互可以从时域的角度对变化信息进行建模, 这一点非常重要^[64], 因此在变化检测模块中采用了特征混合的方法, 即图 4.4 中的混合块。

编码器提取的双时相特征 $f'eat_{T1}^4$ 与 $f'eat_{T2}^4$ 在通道维彼此交叉后获得新的融合特征图, 这样就使得信息可以在时间维度上进行融合。然后, 在通道维度上成对地将特征分成 C 组, 每组通过卷积核聚合信息, 生成混合块输出 $Output \in \mathbb{R}^{H \times W \times C}$ 。混合块的整个操作可以通过可变形卷积核来实现, 计算如下

$$Output = PReLU(IN2d(DefConv(concat(f'eat_{T1}^4, f'eat_{T2}^4) \quad (4.5) \\ ch_{in} = 2C, ch_{out} = C)))$$

其中 PReLU 是激活函数层, IN2d 是实例归一化, concat 是拼接操作。值得注意的是, mix block 的工作原理是与两个特征图的每个通道进行交互。因此, 实例归一化更适合 mix block, 其归一化统计量, 即均值和方差是根据特征图中的通道数计算的。

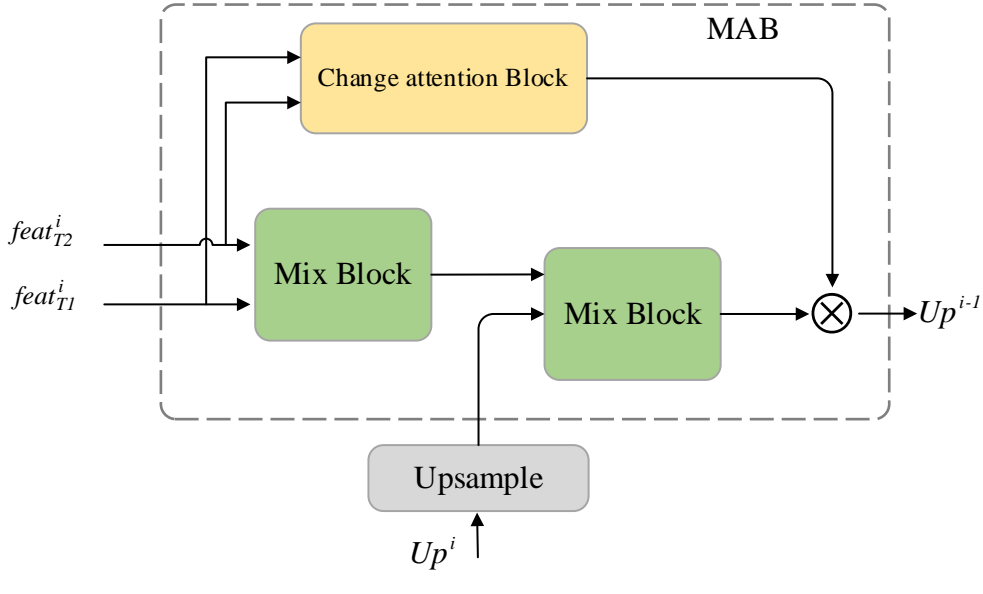


图 4.3 多尺度聚合模块结构示意图

Fig. 4.3 Schematic diagram of the structure of the multi-scale aggregation module

在上采样过程中，使用了用于跳过连接的注意模块 MAB。来自主干的双时态多尺度特征图由混合块混合，并由变化注意力块生成变化注意力图。混合块的输出然后与来自前一上采样阶段的输出混合，最后与变化注意力图相乘以获得特征表示，其计算流程如下：

对于输入的双时相特征 $feat_{T1}^i$ 和 $feat_{T2}^i$ ，首先通过变化注意力块（Change Attention Block）生成变化注意力图 $CAttn_i \in [0,1]^{H_i \times W_i}$ ，首先对输入特征进行展平（Flatten）与 L2 归一化：

$$V_i^{1,2} = \text{L2Norm}(\text{flatten}(feat_i^{T1,T2})) \quad (4.6)$$

其中 $V_i^1 \in \mathbb{R}^{H_i \times H_i \times C}$ 和 $V_i^2 \in \mathbb{R}^{H_i \times H_i \times C}$ 分别为时相 T1 和 T2 的归一化特征向量矩阵。

利用余弦相似度度量特征向量方向差异，公式为：

$$CAttn_i = \frac{V_i^1 \cdot V_i^2}{\|V_i^1\| \|V_i^2\|} \quad (4.7)$$

其中 \cdot 表示矩阵内积， $\|\cdot\|$ 为 L2 范数。该设计通过方向差异（而非幅度差异）捕捉变化特征，增强对真实变化区域的敏感性。

双时相特征经混合块（Mix Block）进行多尺度融合，对 $feat_{T1}^i$ 与 $feat_{T2}^i$ 执行通道拼接（Concat）与 3×3 卷积，生成混合特征 $Mix_i \in \mathbb{R}^{H_i \times H_i \times C}$ 其中 C 为输出通道数；将混合特征与变化注意力图逐元素相乘，进行注意力加权，强化变化区域响应，公式如下：

$$F_{\text{hybrid}}^i = Mix_i \odot CAttn_i \quad (4.8)$$

其中 \odot 表示逐元素乘法， $F_{\text{hybrid}}^i \in \mathbb{R}^{H_i \times H_i \times C}$ 为加权后的融合特征。

上采样过程采用递进式特征聚合策略，公式定义为：

$$Up_{i-1} = \begin{cases} Mix(F_{hybrid}^3) & i = 3 \\ Mix(Up_{up_i}, Mix(F_{hybrid}^i)) & i = 3, 2, 1 \end{cases} \quad (4.9)$$

最终上采样特征 $Up_0 \in \mathbb{R}^{H_i \times H_i \times C}$ (恢复至输入图像分辨率 $H \times W$) 经过 1×1 卷积与 Sigmoid 激活, 生成变化概率图。

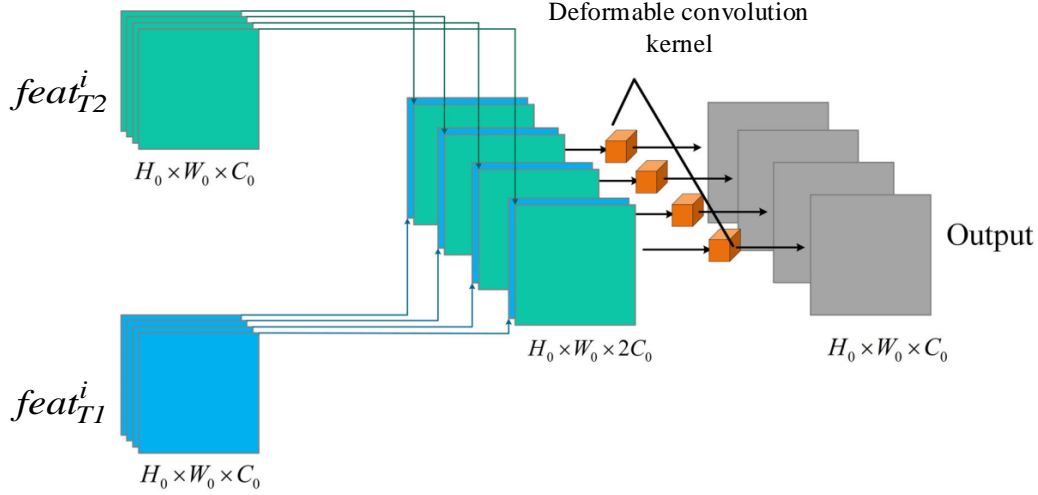


图 4.4 混合模块的结构示意图

Fig. 4.4 Schematic structure of the Mix Block

4.2.5 损失函数设计

本文设计复合损失函数以平衡变化检测精度与特征对齐约束, 总损失函数由变化区域损失与特征对齐损失联合优化, 其数学表达式为:

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{change} + \lambda_2 \mathcal{L}_{align} \quad (4.10)$$

其中 $\lambda_1 = 0.1$ 、 $\lambda_2 = 0.2$ 为平衡权重。

1) 变化区域损失 (\mathcal{L}_{change})

为缓解类别不平衡问题 (变化区域占比通常较小), 采用加权二元交叉熵与 Dice 损失的混合形式:

$$\mathcal{L}_{change} = \alpha \mathcal{L}_{wBCE} + (1 - \alpha) \mathcal{L}_{Dice} \quad (4.11)$$

其中 $\alpha=0.7$, 具体计算为:

$$\mathcal{L}_{wBCE} = -\frac{1}{N} \sum_{x,y} [w_p \cdot y_{x,y} \log p_{x,y} + (1 - w_p) \cdot (1 - y_{x,y}) \log (1 - p_{x,y})] \quad (4.12)$$

$$\mathcal{L}_{Dice} = 1 - \frac{\sum_{x,y} y_{x,y} p_{x,y} + \varepsilon}{\sum_{x,y} y_{x,y} + \sum_{x,y} p_{x,y} + \varepsilon} \quad (4.13)$$

式(4.12)中 $y_{x,y} \in [0,1]$ 为像素点 (x,y) 的真实标签, $p_{x,y} \in [0,1]$ 为预测概率, w_p 为动态权重, 其中 $\varepsilon = 1e - 5$ 用于数值稳定性。

2) 特征对齐损失(\mathcal{L}_{align})

为抑制未变化区域的特征差异,引入对比损失约束双时相特征在深层语义空间的一致性:

$$\mathcal{L}_{align} = \frac{1}{\Omega_{nochange}} \sum_{(x,y) \in \Omega_{nochange}} \|feat_{T1}^4 - feat_{T2}^4\|_2^2 \quad (4.14)$$

其中, $\Omega_{nochange}$ 为未变化区域的像素坐标集合,通过真实变化图反选获取; $\|\cdot\|_2$ 为 L2 范数,用于衡量向量间的欧氏距离; $feat_{T1}^4, feat_{T2}^4$ 为特征编码器 Stage4 输出的深层特征。

4.3 实验结果与分析 (Experimental Results and Analysis)

4.3.1 实验数据与配置

第四章实验设计重点关注于第三章模型的不足部分,及对于冰川边界的高效提取,专注于北极冰川裁切图像上的深度学习模型性能评估,实验采用与第三章同源冰川遥感变化检测测试集,通过时空同步校准与辐射校正保障数据物理一致性。实验硬件配置升级如下:基于 Ubuntu 22.04 LTS 系统的工作站搭载 AMD EPYC 9654 96 核处理器与 512GB DDR5 ECC 内存,实现 PB 级冰川时序数据的高速预处理。图形计算单元采用一张 NVIDIA A100 80GB PCIe 显卡,模型开发框架升级至 PyTorch 2.1,集成 Torch Dynamo 编译器与 A100 专属的 TF32 张量加速模式。优化策略方面,采用 Lion 优化器并引入分阶段学习率机制(0.002-0.0005),权重衰减动态调整为 0.0004-0.0008 区间,同步实施全局梯度归一化(阈值 2.0)与自适应梯度裁剪。实验设置 150 个训练周期,采用混合精度(BF16/FP32)与弹性批处理(32-128 动态调整),集成 A100 独有的 MIG (Multi-Instance GPU) 技术将单卡划分为 7 个计算实例实现细粒度任务调度,并通过 NVIDIA DALI 加速数据流水线,最终推理阶段启用 TensorRT 9.0 的稀疏张量优化功能。

4.3.2 方法对比分析

(1) 与传统变化检测方法对比

图 4.5 展示了 EACD-Net 与传统变化检测方法在不同场景下的冰川变化检测结果对比。差值法和比值法通过简单的代数运算检测变化区域,其方法直观且计算成本低。然而,由于其对背景噪声和光照条件变化的敏感性,检测结果中出现了较多的误检和椒盐噪声。例如,在图 4.5 (d) 和 (e) 中,红色框区域显示出明显的误检现象。此外,这两种方法对边缘部分的细微变化缺乏敏感性,如图 4.5 (d) 和 (e) 蓝色框所示,冰川边缘的微小退缩未能被准确捕捉,导致检测结果不完整。PCA 通过识别最大化方差的主要变化,能够在一定程度上减少噪

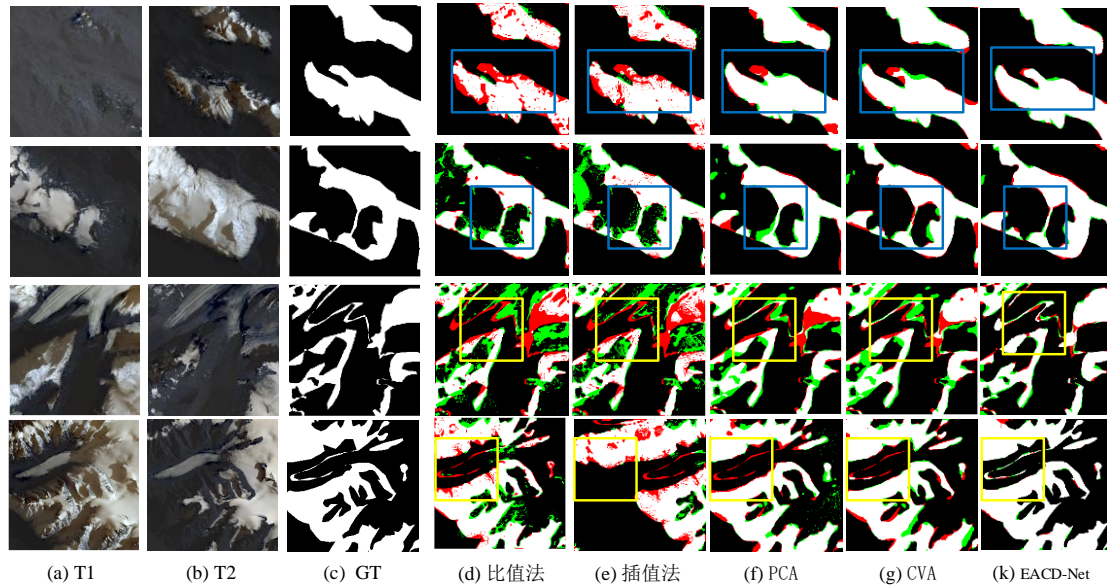


图 4.5 EACD-Net 与传统方法的实验结果定性分析对比

Fig. 4.5 Comparison of qualitative analysis of experimental results between EACD-Net and traditional methods

声的影响，适用于检测大范围的显著变化区域。如图 4.5 (f) 红色框所示，PCA 在冰川主体区域的变化检测上表现较为完整，成功识别了冰川退缩的主要部分。然而，PCA 对小尺度或非主要变化的检测能力有限，尤其是在冰川边缘区域的细微变化上表现不足，如图 4.5 (f) 蓝色框所示，边界处的微小变化被忽略，导致检测结果的边界一致性较差。CVA 通过计算不同时间点影像相同像素点的光谱值差异向量来识别变化，能够量化变化的强度和方向。然而，CVA 对噪声较为敏感，容易受到光照不均或云影等干扰的影响。如图 4.5 (g) 红色框所示，CVA 在冰川主体区域的检测中出现了较多的误检和漏检，同时，CVA 对边缘部分的细微变化检测效果较差，如图 4.5 (g) 蓝色框所示，冰川边缘的微小变化未能被有效识别。相比之下，EACD-Net 通过融合边缘感知模块和多尺度特征提取机制，能够准确捕捉冰川的复杂形状，并显著提升对边缘区域微小变化的检测能力。如图 4.5 (h) 所示，EACD-Net 在所有测试图像中均表现出优异的性能。蓝色框区域显示，EACD-Net 在冰川主体变化的检测上完整且准确，成功抑制了背景噪声和光照变化的干扰；如图 4.5 中的(h)部分，EACD-Net 对冰川边缘的细微变化具有更高的敏感性，边界定位更加精确，复杂地形下的边缘变化被清晰地捕捉。总体而言，EACD-Net 在检测结果的完整性、边界一致性和抗噪能力上均优于传统方法。

(2) 与深度学习变化检测方法对比

如图 4.6 所示在不同场景的变化检测方法中，EACD-Net 在边缘信息提取和变化区域完整性方面展现出了显著优势。对比实验结果如图所示，展示了多

表 4.1 EACD-Net 与传统变化检测方法的定量对比

Table 4.1 Quantitative comparison of EACD-Net and traditional change detection methods

方法	精确度 (%)	召回率 (%)	平均交并比 (%)	F1 ($\times 100$)
比值法	80.2	82.7	72.1	83.9
插值法	87.4	80.9	72.5	84.0
PCA	89.1	84.3	75.4	86.6
CVA	90.3	86.1	77.2	88.1
EACD-Net	94.1	92.5	86.7	93.2

种方法在冰川变化检测任务中的表现。如图 4.6 中(d)、(e)、(f)、(g)中蓝色框所示,对比方法, Ice-Unet, BIT、DDPM-CD 和 Changeformer 方法在一定程度上能够捕捉变化区域的轮廓。但并不能够捕捉连续的精细化边界,尤其在变化区域较为复杂的情况下,仍然存在较多的误检和漏检现象,导致最终检测结果的完整性较差。

相比之下,如图 4.6 的(h)、(i)、(j)中蓝色框所示, EATDer、BGSINet 和 SAMCD 的检测结果更好,误检率和漏检率都有了不同程度的降低,想过更加平滑,在一定程度上也更为连续,并有效减少了孤立的噪声点。然而,由于其在边缘信息建模方面的能力仍然有限,导致部分变化区域的边界较为模糊,特别是在细粒度变化检测任务上表现不够理想。此外,三种方法在部分变化区域的检测结果仍然存在不连续性,影响了整体检测的一致性。如图 4.6 中黄色框所示,出 ICE-Unet 和本文研究方法外, 几乎其他的方法

都在连续的细粒度的边界出现了漏检情况, ICE-Unet 虽然检测出了部分的狭长边界,但出现了程度较深的误检,其余方法在此区域尽管没有检测出连续的细长边界,但是并没有明显的误, DDPM-CD 通过扩散模型生成变化掩码,在一定程度上减少了噪声干扰,使得检测结果更加平滑,在一定程度上少降低了误检率,

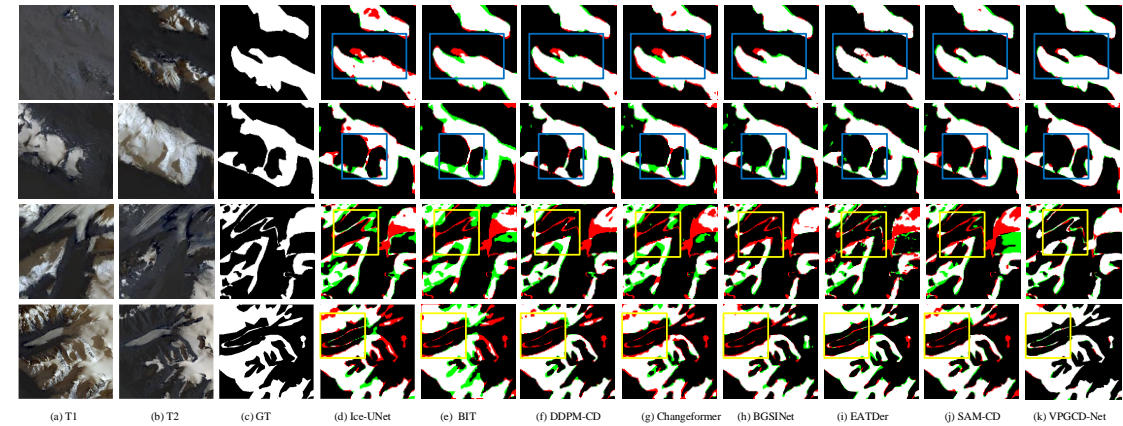


图 4.6 EACD-Net 与深度学习方法的实验结果定性分析对比
Fig. 4.6 Comparison of qualitative analysis of experimental results
between EACD-Net and deep learning methods

如图 4.6 的(f)部分中黄色框所示。Changeformer 结合 Transformer 结构进行时空特征提取,在变化区域的识别能力上有所提升,尤其是对于较大尺度的变化区域表现较为稳定,如图 4.6 的(g)部分中黄色框所示。然而,在复杂地形场景下,Changeformer 仍然存在边界断裂或变化区域不完整的问题。这主要是由于 Transformer 结构对全局信息的关注较多,而对局部边缘特征的捕捉能力相对较弱,导致最终检测结果的边界细节仍然不够精确,如图 4.6 的(g)部分中蓝色框所示。BGSINet 在边界检测方面有所提升,相比于前述方法,其检测结果在整体上

表 4.2 EACD-Net 与深度学习变化检测方法的定量对比

Table. 4.2 Quantitative comparison of EACD-Net and deep learning change detection				
方法	精确度 (%)	召回率 (%)	交并比 (%)	F1 (×100)
Ice-UNet	89.5	87.6	81.0	88.1
BIT	91.0	90.5	83.1	90.4
BGSINet	89.7	89.4	85.2	89.5
DDPM-CD	90.4	91.5	84.0	91.0
ChangeFormer	91.2	90.1	83.1	90.6
EATDer	93.1	90.7	84.5	91.1
SAM-CD	93.3	93.1	85.4	93.2
EACD-Net	94.1	92.5	86.7	93.2

更加完整,边界信息的表达能力有所增强,如图(h)中黄色框所示。然而,该方法在光谱特征变化较大的区域容易受到误导,导致误检现象较多,特别是在高反差场景下,模型易将非变化区域误判为变化区域。此外,该方法在部分细节区域仍然存在边缘模糊的问题,影响了最终检测结果的精度,如图 4.6 的(h)部分中蓝色框所示。EATDer 在精细化边缘方面表现较好,相较于其他方法,其检测结果在一定程度提高了边界的检测精度,如图 4.6 的(i)部分中蓝色框所示。然而,该方法在边缘提取能力上仍然存在不足,特别是在微小变化区域的检测上,部分边界信息未能被准确识别,并且出现了程度较高的噪声,导致检测结果的精度仍然受到影响。此外,在部分复杂地形区域,EATDer 仍然存在一定程度的误检,影响了最终检测结果的一致性。

SAM-CD 在边缘细节方面表现较好,相比于其他方法,其检测结果在误检、漏检方面有所提升。然而,在提取细小狭长边界方面仍然存在缺陷,影响了最终的整体检测效果。此外,SAM-CD 仍然存在一定的误检现象,在某些复杂场景下,模型无法完全抑制光谱变化导致的误判如图 4.6 的(j)部分中蓝色框所示。

相比于上述方法,EACD-Net 在边缘提取、变化区域完整性和噪声抑制方面均表现更优。首先,在边缘提取方面,EACD-Net 通过专门的边缘增强模块,使得检测结果的边界更加清晰,能够精确刻画变化区域的轮廓如图(k)中蓝色框所示。其次,在变化区域的完整性方面,EACD-Net 通过多尺度特征融合,有效提

升了对变化区域的整体感知能力,使得检测结果更加稳定,不会出现断裂或不连续的情况。此外,在噪声抑制方面,EACD-Net 结合了多尺度信息和特征自适应机制,使得模型能够有效区分真实变化区域和噪声干扰,从而显著减少误检和漏检现象。

综合来看,EACD-Net 在变化检测任务中表现出色,尤其是在复杂场景下展现出了更强的鲁棒性。其在边缘提取、变化区域完整性和噪声抑制等方面均优于现有方法,为冰川变化检测任务提供了更高精度的解决方案。实验结果进一步验证了 EACD-Net 的有效性,证明其在变化检测任务中具有较高的应用价值。

4.3.3 消融实验分析

如表 4.4 所示,我们对 EACD-Net 中各组件的作用进行了以下分析:

去除 EAM 后,交并比 (IoU) 从 86.7%下降至 82.3%,降幅为 4.4%,F1 分数下降 1.9%。这表明 EAM 在增强模型对冰川边界区域的检测能力方面至关重要,尤其在边界模糊或复杂地形条件下,能够显著提升定位精度。去除 CAM (w/o CAM) 后,交并比下降 4.6%,F1 分数下降 1.8%。这表明跨时相特征交互对变化区域的精确定位起到关键作用。尤其在低对比度场景下,CAM 增强了真实变化区域的对比度,提高了检测的稳定性。不使用 MAB (w/o MAB) 时,IoU 下降 2.2%,F1-score 下降 0.8%。MAB 通过多尺度信息融合增强了特征表达能力,使得模型在微小变化区域的表现更优。实验结果表明,多尺度特征交互能够提升整体检测性能。

在实验中,我们尝试采用采用 Canny 算子替代 Sobel,其 mIoU 较 EACD-Net (完整)下降 5.5%。分析发现 Canny 的二值化阈值处理导致高频细节丢失,而 Sobel 的梯度幅值保留能更好适应冰川表面的纹理渐变特性。此外我们在实验中尝试了单尺度 sobel 算子的效果,mIoU 分数下降 2.6%,F1 分数下降 2.1%,表明使用所尺度算子操作能够增强模型对于不同尺度边界特征的获取能力

如图 4.7 所示展示了消融实验的定性分析结果,以不同的方式分析了模型在去除关键模块后的表现。如图 4.7 中 (d) 部分蓝色框和黄色框所示,可以观察到去除 EAM 后,误检区域明显增多,边界检测能力下降,尤其是在变化区域的边缘部分,绿色 (误检) 和红色 (漏检) 部分增多,说明 EAM 在增强边界信息方面起到了重要作用。如图 4.7 中 (e) 部分黄色框所示,去掉 MAB 后,结果中的细节丢失较多,说明 MAB 在捕捉多尺度特征方面起到了关键作用,缺少该模块可能导致对不同尺度变化的适应能力下降。如图 4.7 中 (f) 部分黄色框所示,采用单尺度的 Sobel 算子进行边缘增强,可以看到部分边界得到加强,但由于 Sobel 仅能捕捉局部梯度变化,整体检测结果仍然较为粗糙,且误检较多。相比 Sobel, Canny 算子在细节部分处理不够到位,变化边缘仍然存在断裂或不连续

表 4.3 EACD-Net 中关键模块消融实验定量分析

Table. 4.3 Quantitative analysis of key module ablation experiments in EACD-Net

方法	精确度 (%)	召回率 (%)	平均交并比 (%)	F1 (%)
w/o EAM	91.3	88.9	82.3	89.9
w/o CAM	90.5	89.2	82.1	90.0
w/o MAB	91.8	90.1	84.5	91.0
单尺度 sobel	91.5	90.8	84.2	91.1
多尺度 Canny	89.3	85.2	81.2	73.8
EACD-Net (完整)	94.1	92.5	86.7	93.2

现象如图 4.7 中 (i) 部分中蓝色框所示。总体来看 EACD-Net (完整模型) 的检测结果明显优于消融版本, 变化区域的边界更加清晰, 误检和漏检减少, 说明完整的 EACD-Net 在边缘感知和多尺度信息融合方面的有效性。

综上所述, 定性分析结果和定量分析结果表明, EACD-Net 的每个组件均对模型性能有正向贡献, 只有边缘感知模块 (EAM) 和多尺度特征融合 (MAB) 等组件协同工作, 才能使得 EACD-Net 在冰川变化检测任务中表现出色, 尤其在边界定位精度、复杂场景适应性和变化区域检测能力上表现出显著优势。

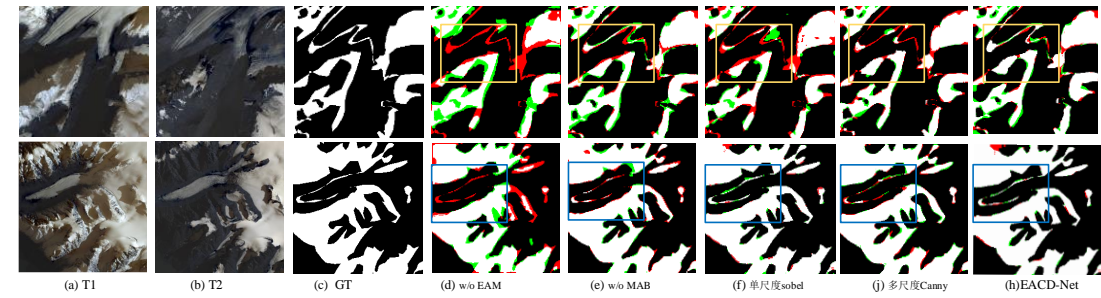


图 4.7 EACD-Net 中关键模块消融实验定性分析

Fig. 4.7 Qualitative analysis of key module ablation experiments in EACD-Net

4.4 本章小结 (Chapter Summary)

本章首先详细阐述了基于边缘信息感知的变化检测网络方法的研究流程, 该流程涵盖了数据预处理、深度神经网络的训练及测试等关键步骤。随后, 本章深入介绍了 EACD-Net, 该网络结构由主干网络、边缘感知模块以及多尺度聚合模块组成。通过接入边缘感知模块, EACD-Net 能够精确地捕捉到由地形因素导致的冰川的复杂形状特征, 如模糊的边缘。此外, 边缘感知模块的引入, 能够保留并强化来自编码器的多层次特征, 进而增强网络对边缘区域微小变化的识别能力, 以提升网络整体的性能表现。随后, 本章针对各种场景下的冰川遥感影像设计了实验, 以实际评估所提出方法的有效性。实验部分包括了对比分析, 其中将

EACD-Net 与其他现有的冰川变化检测方法进行了比较,以突出 EACD-Net 的优越性。最后通过一系列消融实验,本章验证了 EACD-Net 中各个组成部分对于提升网络性能的具体贡献。

5 总结与展望

5 Summary and Prospects

5.1 总结 (Summary)

本研究在全球气候变化背景下,深入探讨了冰川遥感变化检测中的挑战,并且深度探究了深度学习方法在遥感变化检测领域的应用,目前冰川遥感变化检测主要有两个难题,第一,由于成像时间、光照条件及地形复杂性使得冰川遥感影像中存在很多的阴影,这些阴影的存在导致影像间的光谱差距变小,造成了模型的误判;第二,由于重力、温度等因素的影响以及冰川自身存在流动性,冰川存在这复杂的边缘结构,这对模型精细化定位冰川边界带来了不小的困难。为了解决这两个问题,本研究提出了两个创新性的网络模型:VPGCD-Net 和 EACD-Net。

VPGCD-Net 通过引入视觉提示工程以增强双时相影像的语义一致性,并结合基于 Transformer 的变化检测架构及 FiLM 模块来提升变化区域的特征表达能力,从而有效减少了阴影等成像因素对检测精度的影响,有效的减少了误判。实验结果表明, VPGCD-Net 能够更精准地识别冰川的变化区域,并减少误检和漏检的情况,在双时相冰川数据集上有着很好的表现。

面对复杂边缘结构的挑战,提出了 EACD-Net,一种基于边缘信息增强的冰川变化检测网络,采用孪生网络架构独立提取双时相遥感影像特征,通过显式建模多尺度边缘信息,提高模型对变化区域边界的感知能力,并在深层引入跨时相注意力交互与对比损失约束,以提升变化区域的对比度和特征聚焦能力。在 Landsat-8 遥感影像上的实验验证了边缘信息增强在处理复杂冰川形状和边界信息方面的高效性和准确性。

综上所述,本研究针对全球气候变化背景下冰川遥感变化检测的两大挑战——阴影干扰导致的误判问题与复杂模糊边缘结构的精细化检测难题,提出了新的解决方案 VPGCD-Net 和 EACD-Net,取得了一定的技术突破和进步。本研究为冰川遥感变化监测提供了可解释性强、适应性优的解决方案,也为遥感变化检测研究贡献了一份力量,这一研究为准确监测冰川变化以及对全球气候变化的定量评估提供了有力的技术支持。

5.2 展望 (Prospects)

在全球气候变化加剧的背景下,冰川动态监测的精准性与时效性需求愈发迫切。本研究提出的 VPGCD-Net 与 EACD-Net 模型虽然有效缓解了阴影干扰与复杂边缘检测的难题,取得了一定的阶段性成果,但冰川遥感变化检测仍面临多维度技术挑战,未来研究将在以下方向深化探索:

（1）多模态数据融合与跨域泛化能力提升

当前模型主要基于光学遥感影像（如 Landsat-8）构建，但冰川环境受云雾、季节性积雪等因素影响显著，单一数据源的局限性可能制约模型鲁棒性。未来可探索多模态数据协同分析框架，例如结合合成孔径雷达（SAR）的穿透性优势、高光谱影像的精细光谱特征以及无人机近地观测的高分辨率空间信息，构建多源异构数据互补机制。技术上，可引入图神经网络（GNN）建模多模态数据间的拓扑关联，或利用知识蒸馏技术实现跨传感器域自适应特征迁移。此外，需进一步验证模型在极地、高山等不同冰川类型的泛化能力，开发针对小样本冰川区域的元学习策略，以应对全球冰川分布异质性带来的挑战。

（2）模型轻量化与边缘计算部署

尽管 VPGCD-Net 与 EACD-Net 在精度上表现优异，但其基于 Transformer 与孪生网络的设计可能带来较高计算成本，难以满足冰川实时监测需求。未来需探索模型压缩与加速技术，例如通过神经架构搜索（NAS）优化网络深度与宽度，或采用动态稀疏化机制减少冗余特征计算。同时，可结合边缘计算架构（如 FPGA 或专用 AI 芯片）实现星载/机载端实时处理，构建“空-天-地”一体化智能监测系统。此外，联邦学习框架的引入可解决冰川数据隐私性与分散性问题，在保护区域数据主权的前提下实现分布式模型协同训练。

（3）动态连续监测与物理机制耦合

现有研究集中于双时相变化检测，但冰川动态具有时空连续性特征。未来需发展时序密集连续变化监测方法，例如构建基于神经辐射场（NeRF）的冰川三维运动场重建模型，或利用时空 Transformer 捕捉冰川消融的长程依赖关系。此外，当前深度学习模型与冰川物理运动机制（如应力-应变模型、热力学方程）的耦合尚不充分，可探索物理信息神经网络（PINN）框架，将质量守恒方程、流速约束等先验知识嵌入网络训练过程，提升模型对冰川动力学的物理可解释性。

（4）全球尺度应用与数字孪生平台构建

为支撑 IPCC（政府间气候变化专门委员会）等国际机构的评估需求，需推动模型向全球尺度的拓展。可通过迁移学习将模型适配至 Sentinel、MODIS 等不同分辨率卫星数据，并构建冰川变化知识图谱整合历史观测与实时数据。在此基础上，发展冰川数字孪生平台，集成多模型预测结果与气候情景模拟（如 CMIP6），实现冰川消融趋势的可视化推演与风险预警。该平台还可嵌入碳汇计算、海平面上升影响评估等模块，服务于碳中和目标的科学决策。

总之，冰川遥感变化检测正从“静态双时相分析”向“动态智能感知”跨越式发展。通过多学科交叉创新、物理可解释模型构建及全球协作网络完善，下一代技术有望实现冰川消融过程的毫米级精度监测与多尺度影响评估，为人类应对

气候变化提供坚实的技术基座。本研究提出的框架可作为这一演进过程中的一环，但其全面落地仍需学术界与产业界的持续协同攻关。

参考文献

- [1] 李开明, 李忠勤, 高闻宇, 等. 近期新疆东天山冰川退缩及其对水资源影响[J]. 科学通报, 2011, 56(32): 2708-2716.
- [2] 段建平, 王丽丽, 任贾文, 等. 近百年来中国冰川变化及其对气候变化的敏感性研究进展[J]. 地理科学进展, 2009 (2): 231-237.
- [3] 曹泊. 祁连山东段冷龙岭现代冰川变化研究[D]. 兰州: 兰州大学, 2013.
- [4] 曲建升, 张志强, 曾静静. 气候变化科学国际发展态势分析[J]. 2008.
- [5] 樊羿, 郝卫峰, 李斐, 等. 基于多源空间测量技术的南极冰下湖研究进展[J]. 地球物理学进展, 2017, 32(3): 939-945.
- [6] 邢鹏, 李彪, 韩一萱, 等. 淡水生态系统对全球变化的响应: 研究进展与展望[J]. 植物生态学报, 2020, 44(5): 565-574.
- [7] 效存德, 丁永建. 冰冻圈变化及其影响研究的主要科学问题概论[J]. 地球科学进展, 2013, 28(10): 1067-1076.
- [8] 夏军, 刘春蓁, 任国玉. 气候变化对我国水资源影响研究面临的机遇与挑战[J]. 地球科学进展, 2011, 26(1): 1.
- [9] Paul F, Winsvold S H, Kääb A, et al. Glacier remote sensing using Sentinel-2. Part II: Mapping glacier extents and surface facies, and comparison to Landsat 8[J]. Remote Sensing, 2016, 8(7): 575.
- [10] SUI H, FENG W, LI W, et al. Review of change detection methods for multi-temporal remote sensing imagery[J]. Geomatics and Information Science of Wuhan University, 2018, 43(12): 1885-1898.
- [11] 王宁练, 刘时银, 李海涛, 等. (2006). 中国西部冰川变化的遥感监测及其气候变化响应. 遥感学报, 10(5), 831-836.
- [12] 李忠勤, 刘时银, 高磊. (2011). 基于代数运算的遥感图像变化检测方法及其在冰川监测中的应用. 遥感学报, 15(6), 1021-1027.
- [13] 睦海刚, 冯文卿, 李文卓, 等. 多时相遥感影像变化检测方法综述[J]. Geomatics & Information Science of Wuhan University, 2018, 43(12).
- [14] 聂勇, 等. (2016). 珠穆朗玛峰地区冰川变化监测研究. 遥感学报, 20(7), 1245-1253.
- [15] Mouginit J, Rignot E, Scheuchl B, et al. Comprehensive annual ice sheet velocity mapping using Landsat-8, Sentinel-1, and RADARSAT-2 data[J]. Remote Sensing, 2017, 9(4): 364.
- [16] Zhang C, Yao X, Li S, et al. Glacier Change in the West Kunlun Main Peak Area from 2000 to 2020[J]. Remote Sensing, 2023, 15(17): 4236.

-
- [17]李海, 杨成生, 惠文华, et al. 基于遥感技术的高山极高山区冰川冰湖变化动态监测——以西藏藏南希夏邦玛峰地区为例 [J]. 中国地质灾害与防治学报, 2021, 32(05): 10-7.
- [18]Winsvold S H, Kb A, Nuth C. Regional Glacier Mapping Using Optical Satellite Data Time Series[J]. IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing, 2017, 9(8):3698-3711.
- [19]Sood V, Tiwari R K, Singh S, et al. Glacier boundary mapping using deep learning classification over Bara Shigri Glacier in Western Himalayas[J]. Sustainability, 2022, 14(20): 13485.
- [20]范吉延, 柯长青, 姚国慧, 等. 基于深度学习的全极化 SAR 影像冰川边界识别[J]. Journal of Remote Sensing, 2023, 27(9).
- [21]李志伟, 许文斌, 胡俊, 等. InSAR 部分地学参数反演[J]. Cehui Xuebao, 2022, 51(7): 1458.
- [22]边金虎, 李爱农, 王少楠, 等. 基于 MODIS NDVI 的 Landsat TM 影像地形阴影区光谱信息恢复方法研究[J]. 遥感技术与应用, 2016, 31(1): 12-22.
- [23]顾菊, 张勇, 刘时银, 等. 青藏高原冰川底部滑动估算方法研究: 进展, 问题与展望[J]. 地球科学进展, 2021, 36(3): 307-316.
- [24]鄔光剑, 姚檀栋, 王伟财, 等. 青藏高原及周边地区的冰川灾害[J]. 中国科学院院刊, 2019, 34(11): 1285-1292.
- [25]Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.
- [26]祁宣豪, 智敏. 图像处理中注意力机制综述[J]. Journal of Frontiers of Computer Science & Technology, 2024, 18(2).
- [27]徐春晖, 徐向东. 前馈型神经网络新学习算法的研究[J]. 清华大学学报: 自然科学版, 1999, 39(3): 1-3.
- [28]He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [29]刘文婷, 卢新明. 基于计算机视觉的 Transformer 研究进展[J]. Journal of Computer Engineering & Applications, 2022, 58(6).
- [30]周飞燕, 金林鹏, 董军. 卷积神经网络研究综述[J]. 计算机学报, 2017, 40(6): 1229-1251.
- [31]Dai J, Qi H, Xiong Y, et al. Deformable convolutional networks[C]//Proceedings of the IEEE international conference on computer vision. 2017: 764-773.
- [32]张顺, 龚怡宏, 王进军. 深度卷积神经网络的发展及其在计算机视觉领域的应用[J]. 计算机学报, 2019, 42(3): 453-482.
- [33]左宗成, 张文, 张东映. 融合可变形卷积与条件随机场的遥感影像语义分割方法[J]. 2019.

-
- [34]Tang X, Zhang T, Ma J, et al. Wnet: WNet: W-Shaped Hierarchical Network for Remote-Sensing Image Change Detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 60: 1-14.
- [35]Feng Y, Xu H, Jiang J, et al. ICIF-Net: Intra-Scale Cross-Interaction and Inter-Scale Feature Fusion Network for Bitemporal Remote Sensing Images Change Detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-13.
- [36]Fu Z, Li J, Ren L, et al. SLDDNet: Stage-wise Short and Long Distance Dependency Network for Remote Sensing Change Detection [J]. IEEE Transactions on Geoscience and Remote Sensing, 2023. 61: 1-19.
- [37]Li Y, Gupta A. Beyond grids: Learning graph representations for visual recognition[J]. Advances in neural information processing systems, 2018, 31: 56-73.
- [38]Chen Y, Dai X, Liu M, et al. Dynamic convolution: Attention over convolution kernels[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 11030-11039.
- [39]Zhao H, Jia J, Koltun V. Exploring self-attention for image recognition[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 10076-10085.
- [40]Peng X, Zhong R, Li Z, et al. Optical remote sensing image change detection based on attention mechanism and image difference[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59(9): 7296-7307.
- [41]Chen H, Shi Z. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection[J]. Remote sensing, 2020, 12(10): 1662.
- [42]Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.
- [43]Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.
- [44]Perez E, Strub F, De Vries H, et al. Film: Visual reasoning with a general conditioning layer[C]//Proceedings of the AAAI conference on artificial intelligence. 2018, 32(1).
- [45]崔政, 胡永利, 孙艳丰, 等. 面向跨模态数据协同分析的视觉问答方法综述[J]. 北京工业大学学报, 2022, 48(10): 1088-1099.
- [46]López V, Casteleiro-Roca J L, Gato F Z, et al. Intelligent model hotel energy demand forecasting by means of LSTM and GRU neural networks[C]//International Symposium on Distributed Computing and Artificial Intelligence. Cham: Springer International Publishing, 2022: 81-90.

-
- [47]Shtedritski A, Rupprecht C, Vedaldi A. What does clip know about a red circle? visual prompt engineering for vlms[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 11987-11997.
- [48]Wang J, Liu Z, Zhao L, et al. Review of large vision models and visual prompt engineering[J]. Meta-Radiology, 2023, 1(3): 100047.
- [49]潘梦洋, 杨航, 范祥晖. 基于 CNN-Transformer 结构的遥感影像变化检测[J]. Chinese Journal of Liquid Crystals and Displays, 2024, 39(10): 1361-1379.
- [50]Journal of Liquid Crystals and Displays, 2024, 39(10): 1361-1379.
- [51]Zhang C, Yang Z, He X, et al. Multimodal intelligence: Representation learning, information fusion, and applications[J]. IEEE Journal of Selected Topics in Signal Processing, 2020, 14(3): 478-493.
- [52]Gao Y, Wei L, Li J, et al. MCCSeg: Morphological embedding causal constraint network for medical image segmentation[J]. Expert Systems with Applications, 2024, 249: 123666.
- [53]Li K M, Li Z Q, Gao W Y, et al. Recent glacial retreat and its effect on water resources in eastern Xinjiang[J]. Chinese Science Bulletin, 2011, 56: 3596-3604.
- [54]朱节中, 陈永, 柯福阳, 等. 基于 Siam-UNet++ 的高分辨率遥感影像建筑物变化检测[J]. Application Research of Computers/Jisuanji Yingyong Yanjiu, 2021, 38(11).
- [55]Wang Q, Yuan Z, Du Q, et al. GETNET: A general end-to-end 2-D CNN framework for hyperspectral image change detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 57(1): 3-13.
- [56]Li X, Du Z, Huang Y, et al. A deep translation (GAN) based change detection network for optical and SAR remote sensing images[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 179: 14-34.
- [57]Chen H, Qi Z, Shi Z. Remote sensing image change detection with transformers[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1-14.
- [58]Cui B, Liu C, Yu J. BGSINet-CD: Bitemporal Graph Semantic Interaction Network for Remote Sensing Images Change Detection[J]. IEEE Geoscience and Remote Sensing Letters, 2024.
- [59]Lüddecke T, Ecker A. Image segmentation using text and image prompts[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 7086-7096.
- [60]Bandara W G C, Patel V M. A transformer-based siamese network for change detection[C]//IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2022: 207-210.
- [61]Bandara W G C, Nair N G, Patel V M. DDPM-CD: Denoising diffusion probabilistic models as feature extractors for change detection[J]. arXiv preprint arXiv:2206.11892, 2022.

-
- [62]Ma J, Duan J, Tang X, et al. Eatder: Edge-assisted adaptive transformer detector for remote sensing change detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 62: 1-15.
- [63]Ding L, Zhu K, Peng D, et al. Adapting segment anything model for change detection in VHR remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024.
- [64]Fang S, Li K, Li Z. Changer: Feature interaction is what you need for change detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-11.
- [65]Yu X, Wu X, Luo C, et al. Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework[J]. GIScience & Remote Sensing, 2017, 54(5): 741-758.

作者简历

一、基本情况

姓名：石智深 性别：男 民族：汉 出生年月：1999.07.27 籍贯：山东省潍坊市

1. 2018.09-2022.06 石家庄铁道大学学士
2. 2022.09-2025.06 山东科技大学硕士学习阶段

二、攻读学位期间发表学术论文情况

- 1.

三、攻读学位期间获奖情况

1. 2023 年 10 月获华为昇腾 AI 创新大赛青岛区域决赛优胜奖；
2. 2023 年 11 月获山东省数字应用创新创业大赛省级二等奖；
3. 2024 年 10 月获山东科技大学二等学业奖学金；

四、攻读学位期间研究项目情况

1. 自然资源部第一海洋研究所技术委托项目：高分海洋资源环境遥感信息处理与业务应用示范系统（二期），2021—2023 年，功能模块开发人员；
2. 青岛市西海岸新区智慧大脑视觉中枢项目子课题：城市违章违法建筑监测系统，2021-2022 年，算法开发人员。
3. 国家自然科学基金（面上项目）：融合地学知识图谱的多生态型滨海湿地植被精细分类方法研究，项目批准号：42276185，2023—2026 年，参与人员。

致谢

行笔至此，感悟良多。

学位论文数据集

关键词*	密级*	中图分类号*	UDC	论文资助
冰川变化检测；遥感影像；深度学习；语义引导；边缘增强；	公开			
学位授予单位名称*	学位授予单位代码*	学位类别*	学位级别*	
山东科技大学	10424		硕士	
论文题名*		并列题名*		论文语种*
		t		中文
作者姓名*	石智深	学号*	202282060078	
培养单位名称*	培养单位代码*	培养单位地址	邮编	
山东科技大学	10424	山东省青岛市	266590	
学科专业*	研究方向*	学制*	学位授予年*	
计算机应用技术	人工智能	三年	2025	
论文提交日期*		2025 年 6 月 1 日		
导师姓名*	崔宾阁	职称*	教授	
评阅人		答辩委员会主席*	答辩委员会成员	
电子版论文提交格式 文本（√） 图像（ ） 视频（ ） 音频（ ） 多媒体（ ） 其他（ ）				
推荐格式：application/msword; application/pdf				
电子版论文出版（发布）者	电子版论文出版（发布）地		权限声明	
论文总页数*		82		
注：共 33 项，其中带*为必填数据，共 22 项。				