毕业论文(设计)

论文（设计）题目：

多模态图像分割研究

姓　　名 赵一然

学　　号 202000800656

学　　院 机电与信息工程学院

专　　业 计算机科学与技术

年　　级 2020 级

指导教师 杨飞

2024 年 5 月 17 日

摘　　要

道路裂缝是道路质量的重要标志之一，其严重性预示了道路的使用寿命和安全状况。由于道路裂纹的隐蔽性以及不易被发现等特点，单纯使用人工来对道路裂纹进行识别效率较为低下，并且容易产生疏忽。近年以来，以深度学习为基础的单模态图像分割算法取得了比较大的进展，尤其是 U-Net 架构在图像分割领域产生了巨大的影响。

本文提出了 MultiCrack-Net 架构，旨在将 U-Net 架构扩展到多模态分割领域，应用于道路裂缝识别，期望能够帮助人们对道路裂纹进行自动化、智能化识别，从而能够及时发现道路裂纹，维护城市道路安全。我们收集了有关道路裂纹的多模态数据集，包括可见光图像和红外图像两种不同模态的数据集，并对其进行了预处理和标注。模型采用层级融合策略，将不同模态图像通过不同路径输入到包含编码器-解码器的神经网络当中，并进行同一路径和不同路径的紧密连接。对于每一层神经网络，模型采用扩展的 InceptionNet，使用不同大小的卷积核更好地挖掘图像信息。通过对比试验与消融实验，证明 MultiCrack-Net 模型在多模态道路裂纹图像分割领域的良好效果。

关键词：深度学习；多模态图像分割；U-Net；层级融合策略；道路裂纹

ABSTRACT

Road cracks are one of the important indicators of road quality, and their severity predicts the longevity and safety of the road. Due to the concealment of road cracks and the characteristics of not easy to be found, the simple use of manual identification of road cracks is relatively inefficient, and it is easy to cause negligence. In recent years, great progress has been made in single-modal image segmentation algorithms based on deep learning, especially the U-Net architecture has had a great impact in the field of image segmentation.

This article proposes the MultiCrack-Net architecture, which aims to extend the U-Net architecture to the field of multi-modal segmentation and apply it to road crack identification, hoping to help people identify road cracks automatically and intelligently, so as to detect road cracks in time and maintain urban road safety. We collected multimodal datasets about road cracks, including visible light images and infrared images, and preprocessed and annotated them. The model adopts a hierarchical fusion strategy, which inputs images of different modalities into a neural network containing encoder-decoder through different paths, and closely connects the same path and different paths. For each layer of neural network, the model uses an extended InceptionNet to better mine image information using convolutional kernels of different sizes. Through comparative experiments and ablation experiments, it is proved that the MultiCrack-Net model has a good effect in the field of multi-modal road crack image segmentation.

Key Words: deep learning(DL), multimodal image segmentation(MIS), U-Net, fusion strategy(FS)

目　录

1 绪　　论

1.1 研究背景及研究意义

1.1.1 研究背景

道路裂纹是道路缺陷的早期表现形式的一种，也是评价一条道路是否健康的重要参考标准，其预示了道路可能会面临的安全问题和性能问题，需要人们定期对道路裂纹进行调查，从而判断道路状况，为以后的政策制定做好相必要的准备。道路裂纹的危害程度会直接影响城市对于交通安全的判断和对道路维护方案的制定。但是现有的沥青道路裂纹检测技术无法满足人们对于效率和精度的需要。单纯使用人工来进行道路裂纹的检测通常准确率较为低下，非常耗费人工，并且单纯的人工检测容易对细小的道路裂纹进行忽略，从而造成非常严重的交通事故。现有的道路裂纹检测算法大多都是单一模态，同样存在着疏漏问题。为了方便检测人员能够更好地对道路裂纹进行检测，同时为了解决在传统道路裂纹检测方法中存在的精度地下、噪声较高、计算量大，并且容易丢失图像细节方面等多个问题，采用多模态技术对道路裂纹进行检测便显得尤为必要。

1.1.2 研究目的

通过研究多模态图像分割在道路裂纹方面的应用，本文期待能够通过综合利用多模态图像的信息，更加全面地捕捉道路裂纹的特征，从而提高图像分割的效率，帮助人们提高对道路裂纹检测的准确性和效率。通过实现多模态图像分割在道路裂纹方面的应用，本文还期待能够实现道路维护与管理的自动化，通过自动检测和定位道路裂纹，使得人们能够及时发现和维修道路裂纹，从而提高道路安全性和道路使用寿命。

1.1.3 研究意义

通过对以往单一图像分割技术进行比较，本文对传统的道路裂纹分割模型进行了改进，采用了同样基于深度学习的处理方法，使用了多模态技术来对道路裂纹进行检测，促进了对道路裂纹的自动识别，提升了分割准确性 0。通过分析多模态模型在道路裂纹分割上的构建与优势，采用了大量的数据集来进行训练，这些数据集包括了普通图片和不同时间段的红外线图片。在道路裂纹检测时，对比经典的图像分割技术能够获得更好的性能与效率，能够忍受道路裂纹中存在的如光照不平衡、阴影不均等问题。从而能够更好地识别出各种形态的道路裂纹，能

够提高道路裂纹识别的检测精度，使道路裂纹分割更加明显、完整。通过正确识别道路裂纹，有助于人们及时发现和修复道路裂纹，从而提高道路的安全性和可靠性，减少交通事故发生的概率。并且还能够通过利用图像分割技术进而实现对道路裂纹识别的自动化、智能化，减少人工成本，并提高检测的准确性和覆盖范围。

1.2 研究现状

1.2.1 国内外研究现状

图像分割问题一直是计算机视觉研究当中的一个基本组成部分。图像分割在医学影像分析、自动驾驶、无人机图像处理、工业质检、道路裂纹识别等领域都有广泛的应用。图像分割是众多视觉理解系统中不可缺少的部分，它是指将数字图像划分成多个子区域或者是分割出图像中对象的过程。图像分割问题通常包括区域划分、边界检查测、对象识别等方面。区域划分是指将图像中的像素组织成具有相似特征的区域或者连通区域，这些特征可以是颜色、亮度、纹理、形状等。边界检测是指在分割过程中，通常需要检测出不同区域之间的边界，以便准确地将它们分开。对象识别是指分割的目的通常是为了识别图像中的对象或区域。因此，一种常见的分割任务是将图像中的不同对象或物体分开。因此，在国内外的研究现状中，图像分割需要使用到各种各样的算法和不同的技术，这其中包括传统的阈值分割、边缘检测、区域增长、分水岭算法等方法，以及现代的深度学习方法，如卷积神经网络（CNN）和语义分割模型等。但这些图像分割算法通常都是通过单一模态来进行训练，可能因为数据集受到污染，如光照不足，阴影覆盖不均等原因而造成图像分割不精准、出现错误。这时便需要多模态数据来进行融合分割。

1.2.2 国内外研究现状分析

关于道路裂纹图像分割问题需考虑很多实际挑战，道路图像可能包含复杂的结构、变化的光照条件、噪声以及对象之间的相互、阴影不均等因素。这给传统的图像分割算法造成了不小的困难。在道路裂纹分割领域当中，基于卷积神经网络 CNN 和 Transformer 的模型架构都得到了人们的广泛探索与应用，但这两种算法在关于道路裂纹方面的图像分割都存在这不同的优劣之处。基于 CNN 的图像分割算法用于较好的适用性，应用范围比较广泛，并且能够学习到图像中的高级

特征，对于复杂的图像结构也能拥有较好的识别能力。但基于 CNN 的图像分割算法中，CNN 卷积核的感受视野通常是固定大小，在进行特征提取时只能关注固定大小的局部区域，可能会忽略全局信息。当图像中存在尺度变化较大的目标或结构时，固定大小的感受野可能无法适应不同尺度的目标，导致提取的特征不够准确或完整。在基于 Transformer 的图像分割算法中，Transformer 模型相较于 CNN 模型能够捕捉到图像的全局信息，从而有利于理解图像的整体结构，Transformer 中的自注意力机制能够很好地获取图像内部各个像素之间的关系，也能够提高分割的准确性。但 Transformer 在图像分割领域同样具有缺点，相比于 CNN，Transformer 模型的计算复杂度更高，特别是在处理大规模图像时会消耗更多的资源。

1.3 论文主要研究内容

本文将基于深度学习的单模态图像分割模型 U-Net 架构扩展为多模态图像分割算法 MultiCrack-Net，数据集采用多模态数据集，包括了不同传感器拍摄的图像，例如可见光图像和红外图像，并对数据进行预处理，以确保数据输入模型时的一致性。MultiCrack-Net 架构由编码器、解码器和跳跃连接构成，能够有效地处理、捕捉图像的细节。接下来通过各种性能指标来评估模型的准确性，如 MioU、MPA 等。最终探讨模型的优缺点，以及未来可能改进的空间。

1.4 论文组织结构

本论文具体的章节安排如下：本文第一章先多模态图像分割在道路裂纹方面研究的背景、目的、意义，对多模图像分割在道路裂纹研究做简要概述。之后提出了国内外研究现状，并对其进行简要分析。第二章便对文中主要概念：多模态图像分割、U-Net 进行阐述和概括，便于人们对其理解。第三章介绍基于 U-Net 的多模态图像分割算法 MultiCrack-Net 的思想，并对其架构进行分析，并介绍数据集、实验设计以及模型对比。第四章对本文进行总结与展望。

2 相关概念及理论基础

2.1 多模态图像分割

2.1.1 概念

多模态图像分割是指利用多种类型的图像数据，从而完成对图像的分割，使用多模态图像分割需要将多种信息进行融合，从而进行分割。在传统的图像分割领域中，不管是采用什么类型的方法，通常都是基于单一类型的图像数据，例如单一的灰度图像或者是单一的彩色图像。然而，在某些图像分割应用中，例如医学图像分割或者是道路裂纹的图像分割中，仅仅依靠单一类型的图像数据集可能无法提供足够的信息来完成准确的分割。多模态图像分割可以通过结合多种类型的图像，从而获得更加准确的分割结果。多模态图像的数据集通常来自不同传感器或者成像工具的图像，如可见光图像、红外线图像，或者是 CT 扫描图像。每种类型数据都可以对最终的分割结果提供不同的信息。多模态图像分割的过程通常包括数据获取、图像配准、特征提取、特征融合、模型构建、分割、后处理、结果评估等过程。多模态图像分割的融合策略具有多样化，常见的有概率论法[2]、机器学习方法[3][4][5]，还有基于深度学习的方法。但是对于前两种方法，采取浅层模型建模比较困难，因为不同的模态的统计特性之间存在差异。当下，对于多模态图像融合多采用基于深度学习的方法，例如 AlexNet[6]、ZFNet[7]、VGG[8]、DenseNet[9]、GoogleNet[10]、FCN[11]、ResidualNet[12]、U-Net[13]。

2.1.2 多模态图像分割优点

多模态图像分割相较于单一模态图像分割拥有众多优点，例如信息丰富、互补性强、鲁棒性强以及适用性广泛等优点。多模态图像融合了来自不同传感器或成像方式的图像数据，提供了更丰富、更全面的信息。不同类型的图像数据可以捕获目标的多个方面特征，如形状、纹理、颜色等，有助于提高分割的准确性；不同类型的图像数据通常具有互补的特点，可以相互弥补单一图像数据的不足。例如，可见光图像通常可以提供目标的形状和颜色信息，而红外图像则更适合在夜间或低光条件下对目标进行图像分割。通过融合这些不同类型的信息，可以增强分割算法的效果，提高准确性；多模态图像分割对于噪声、光照变化、遮挡等因素具有较强的鲁棒性。由于融合了多种类型的信息，分割算法可以更好地适应复杂环境和不确定性因素，从而提高算法的稳定性和可靠性；多模态图像分割可

以应用于多种领域和应用场景，包括医学影像、地质勘探、军事侦察、环境监测、交通管理等。不同领域和应用场景可能需要不同类型的信息来实现准确的分割，而多模态图像分割可以根据需要选择适合的图像数据类型进行处理，具有较强的通用性和灵活性。

2.1.3 一般原理

深度学习也是神经网络中的一种[14]，其有多层非线性处理单元。每一个连续层的输入都为前面一层的输出。深度学习可以利用这些连续的层从大量数据中捕获相对复杂的结构特征。在当今人们常用的深度学习算法包括桨深度玻尔兹曼机[15]、卷积神经网络[16]等。图 2-1 描述了基于深度学习的多模态图像分割流程。该管道由四部分组成：数据准备、网络架构、融合策略和数据后处理。在数据准备阶段，首先选择数据维度，并使用预处理来减少图像之间的变异，还可以使用数据增强策略来增加训练数据以避免过拟合问题。在网络架构和融合策略阶段，提出了基本网络和详细的多模态图像融合策略来训练分割网络。在数据后处理阶段，植入形态学技术、条件随机场等一些后压技术，以细化最终的分割结果。

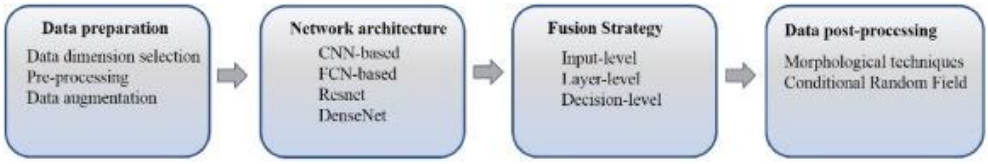

图 2-1 多模态图像分割流程

2.1.4 融合策略

多模态图像分割的融合策略大致分为三种不同的方法：输入级融合网络、层级融合网络、决策级融合网络。

输入级融合网络中，采取将不同模态的图像输入到多通道中，然后多通道逐步融合，最后以学习融合的特征来表示结果，之后再进行训练网络分割。在当下，大多数多模态图像分割网络中大部分都会采取输入级融合网络策略，对输入到空间中的不同模态图像直接进行融合。图 2-2 是对输入级融合网络通用架构的描述。采用可见光图像和红外图像作为两种不同模态图像作为输入。[17]

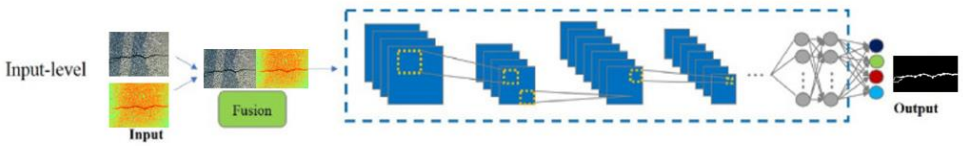

图 2-2 输入级融合网络

在层级融合策略中，用单个或两个模态图像作为单个输入来训练个体分割网络，然后这些学习到的个体特征表示将在网络的各层中融合，最后将融合结果馈送到决策层得到最终的分割结果。层级融合网络可以有效地集成和充分利用多模态图像。图 2-3 描述了层级融合分割工作的通用网络架构。[17]

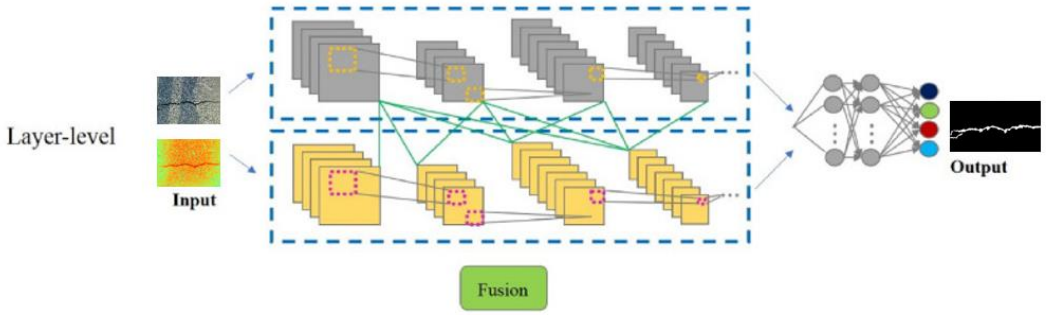

图 2-3 层级融合网络

在决策级融合分割网络中，与层级融合一样，每个模态图像作为单个分割网络的单个输入。单一网络能更好利用相应模态的独特信息。然后将整合各个网络的输出以获得最终的分割结果。由于图像采集技术的不同，多模态图像在原始图像空间中几乎没有直接的互补信息，因此决策级融合分割网络被设计为独立学习来自不同模态的互补信息。图2-4描述了层级融合分割工作的通用网络架构。[17]

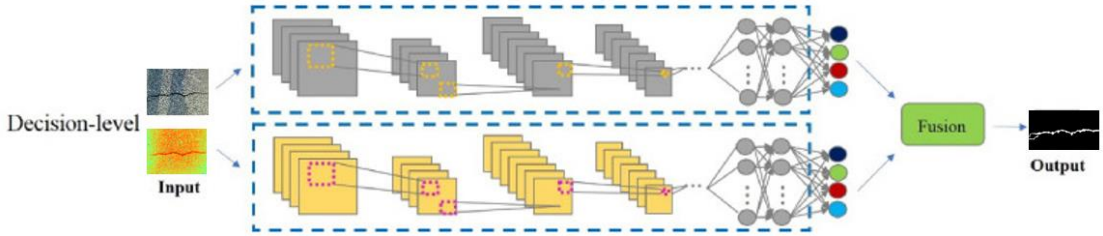

图 2-4 决策级融合网络

2.2 U-Net 架构

2.2.1 U-Net 架构概念

U-Net 架构是一种深度学习框架，最初为了解决生物医学图像分割问题而产生的。由于其良好效果，后来被广泛应用于语义分割的各个方向，比如卫星图像分割等。U-Net 的提出使得原本需要多达数千个带有注释的数据才能进行训练的深度学习神经网络大大减少了训练所需的任务量。U-Net 并不需要很多数据来进行训练，但是 U-Net 仍能取得比较好的训练效果。U-Net 架构是由 FCN 衍生而来的，采用了编码器-解码器（Encoder-Decoder）结构，并且在中间添加了跳跃连接

部分。

图 2-5 描述了 U-Net 的基本架构。左边部分是编码器，也被称为收缩路径。U-Net 的编码器由一系列卷积层和池化层组成，用于提取输入图像的特征，编码器可以有效地提取不同尺度和抽象级别的特征。右边部分为解码器，也被称为扩展路径。解码器负责将编码器提取的特征图进行上采样，并恢复到原始图像尺寸的分辨率。U-Net 中的解码器采用了反卷积操作，从而实现上述采样。此外，中间的灰色箭头即为跳跃连接，用于将编码器中相应层的特征图与解码器中的特征图进行连接。这些跳跃连接有助于将低级别和高级别的特征信息进行融合，从而提高了分割结果的准确性。

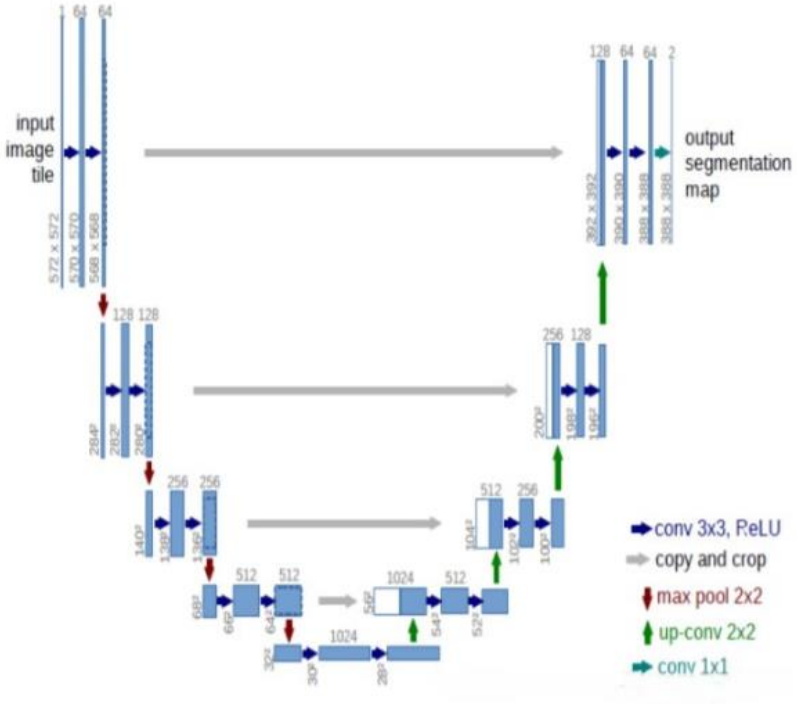

图 2-5 U-Net 基本架构

2.2.2 U-Net 架构优势

U-net 在小样本数据集上表现出色，这使得它特别适用于道路裂纹分割等领域，因为这些领域的数据往往稀缺且成本高昂。由于 U-Net 具有较少的参数量和简单的结构，它能够在小样本数据上快速收敛并取得良好的性能；U-Net 可以生成与输入图像相同分辨率的分割结果，这对于需要高精度分割的任务非常重要，如医学图像中的器官分割和道路裂纹分割方面。通过有效地将特征映射恢复到原始图像的尺寸，U-Net 能够保留细节信息并减少分割误差；U-Net 的结构相对简单，易于实现和训练。这使得它成为深度学习图像分割领域的一个流行选择，尤

其是对于那些初学者或需要快速迭代的项目来说；U-Net 的结构灵活，可以根据具体任务进行调整和扩展。例如，可以通过调整编码器和解码器的深度或宽度，或者通过增加或减少跳跃连接，来适应不同的图像分割任务。

3 基于多模态 U-Net 的道路裂缝定位和分割

3.1 算法思想

与单张图像相比，多模态数据带来了互补的信息，有助于更好的数据表示和判别能力。但由于不同模态图像之间的关系多种多样，极具复杂性，并且不能直接通过单一层次进行建模。为了解决不同模态数据之间建模的非线性问题，本文提出了 MultiCrack-Net 算法。

MultiCrack-Net 算法是一个扩展了 U-Net 神经网络算法，将不同模态图像通过不同路径输入到包含编码器-解码器的神经网络当中，采用的是多模态图像处理的层级融合策略。在编码器中，将各个模态图像分别进行同一路径和不同路径的紧密连接，充分挖掘及整合多模态图像的特征。在解码器中，从编码器的跳跃连接获得相应层的特征图，同时采用反卷积操作，融合编码器提取的特征图和相应层的特征图，进行上采样，最终得到与原始图像大小相同的分割掩膜。

该算法主要有三方面创新与贡献：（1）MultiCrack-Net 算法将每种模态图像以不同的路径进行处理，每种模型图像单独作为一个路径的输入，以更好地利用其独特的信息。（2）MultiCrack-Net 网络不仅仅包括了在同一路径中的层与层之间的连接，而且包括了在不同路径上的层与层之间的密集连接，使模型能够自由地学习单一模态的处理和组合不同模态的特征。（3）MultiCrack-Net 改进了标准的 U-Net 模块，将初始模块扩展为两个具有不同尺度的膨胀卷积的卷积块，这有助于处理多尺度上下文。

该算法架构不仅仅包括了在同一路径中的层于层之间的连接，进而包括了在不同路径上的层与层之间的密集链接。为了对不同尺度上下文信息进行捕捉，并处理在过程中图片尺度的变化，采用了多种不同的路径处理以不同尺度但是从同一位置提取的区域。并且在多模态图像分割中间，还会采用随机模态图像体素丢弃策略，此策略能够有效地减少不同类型模态之间的特征过于依赖性性，能够有效协调不同模态图像，并且促进不同的单一模态之间独立学习并且判断信息。

3.2 网络架构介绍

MultiCrack-Net 算法由两部分路径组装而成，图 3-1 是算法的网络架构图。第一部分是编码器（Encoder）,用于收缩图像信息，提取图像特征；第二部分是解码器（Decoder），用于将图像特征扩张到原始图像大小，得到图像分割掩膜。

前一部分路径由多条输入路径组合而成，每条输入路径对应于一种模态图像，将输入的图像放入四层卷积神经网络中，进行特征提取，由此生成一组高级特征，形成紧凑的图像特征中间表示形态。

中间部分路径是连接前后两部分路径的跳跃连接[18]，即将编码器每一层生成的图像特征，输送到解码器同一层当中，作为解码器该层的一部分输入，从而能够结合深层信息和浅层信息，缓解卷积操作时的边界数据丢失问题，保证给分割提供更加精细的特征。

后一部分路径则融合跳跃连接获得的编码器相应层的特征图（即浅层信息），和编码器最终输出的紧凑高级特征（即深层信息），进行四层反卷积操作，最终生成与原始图像相同的图像分割掩模结果。

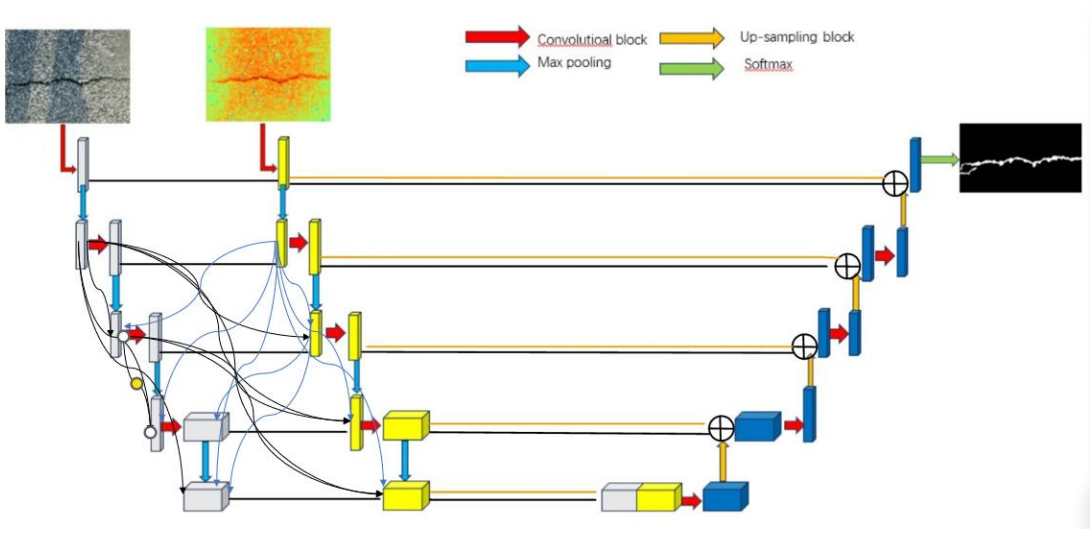

图 3-1 MultiCrack-Net 算法网络架构图

3.2.1 多模态分别处理

为了充分利用多模态数据，MultiCrack-Net 算法选择使用多模态图像分割中的层级融合策略。算法首先创建一个由多个路径流组成的编码器，每个路径流处理不同的图像模态。也就是说，每种模态的图像都作为单独的一条路径流的输入，顺着该路径流利用神经网络进行特征提取。

通过对不同的模态图像采用单独的路径流可以将不同模态的信息单独解开与提取特征，由此能够防止多模态图像信息在早期阶段过早地融合，并且能够促进网络捕捉不同模态之间地复杂信息的能力。

3.2.2 扩展的初始模块

由于数据的多样性与差异性，多模态图像中有分割意义的区域大小可能会发生较大变化。由于变化的不确定性，选择一个较为通用并且分割准确的内核大小会比较困难。过大或过小的内核都会造成分割的不确定性。较小的内核可以捕捉到更加充分、细致的信息，较大的内核可以拥有更加广泛的视野。

因此 MultiCrack-Net 算法加入了 InceptionNet，采用了多个不同大小的卷积核在同一层神经网络上进行图像特征的提取。同时，为了获得更加充分的上下文信息，算法在 InceptionNet 模块中设置了两个与现有模块并行运行的扩张卷积模块，并且这些模块拥有着不同的扩张率，因此有助于从不同的视野中进行信息的捕捉，增加了对初始模块上下文的学习。InceptionNet 模块中还包括了两个空洞卷积块，空洞卷积相比于传统的卷积模块拥有更加广大的感受视野。在传统的标准卷积模块中，每一层的输出都是仅仅依赖于前一层的局部区域，通过引入空洞，空洞卷积卷积核能够覆盖更加广泛的输入区域，所以网络可以捕捉更大范围的上下文信息，从而帮助网络进行语义分割。空洞卷积相较于传统的标准卷积还可以保持分辨率。在传统的卷积操作中，输入图像和输出图像的分辨率可能会产生变化，输出图像的分辨率通常会降低。通过引入空洞卷积，我们可以在增加感受视野的同时保持输入图像和输出图像的分辨率不变。同时，相比于传统的常规卷积操作，空洞卷积还可以减少参数数量以及计算量。空洞卷积不需要增加额外的参数，因此，空洞卷积可在不增加模型大小的情况下提高网络的感知能力。本文所提出的扩展初始模块如图 3-2 所示。

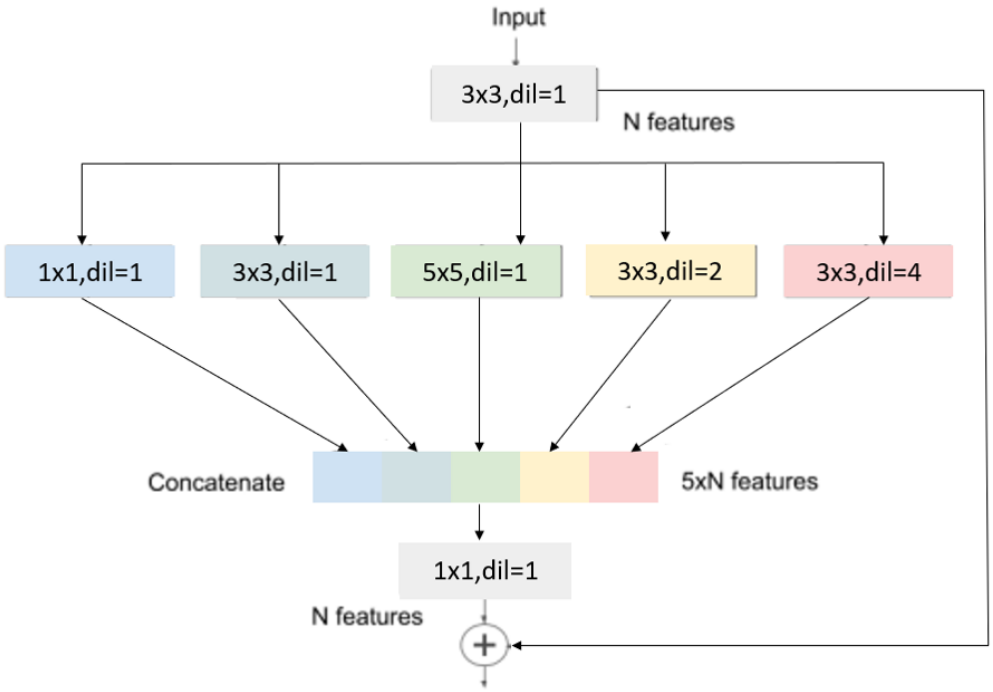

图 3-2 算法的扩展初始模块

3.2.3 超密集连接网络

为了充分挖掘多模态的特征，MultiCrack-Net 参考[9]，采用了超密集连接方法，将密集网络 DenseNet 融合到 U-Net 模型中，将 U-Net 算法拓展到多模态图像分割问题。

采用超密集连接在多模态图像的特征挖掘上主要有四方面的优势：第一，将不同路径流组合成同一条编码路径能够更好的提取不同模态之间的相互关系。第二，在对所有层之间使用直接连接能够更好地促进了整个网络的信息流和梯度，缓解了梯度消失。第三，在网络中包含到所有特征的短路径，引入了一种隐式的深度监督。第四，超密集连接能够起到正则化的效应，从而能够降低了算法模型在较小训练集上容易产生的过拟合问题。

如图 3-1 所示，为了实现这种密集的连接模式，不仅单个模态图像特征在同一路径流的不同神经网络层之间进行连接，而且不同的模态图像也在不同路径流的层之间发生密集连接。值得注意的是，这两种连接均只对当前层神经网络的所有下层神经网络进行特征的传输。

在单模态的垂直密集连接中，配方，让x_1表示l^{th}层，以及H_l是一个映射函数，它对应于一个卷积层，然后是非线性激活。在标准 CNN 中，输出l^{th}层通常是从前一层的输出中获得的x_{l-1}，见式（3-1）

$$x_l = H_l(x_{l-1}) \quad (3-1)$$

然而，在密集连接的网络中，所有特征输出都以前反馈方式连接起来，见式（3-2）

$$x_l = H_l([x_{l-1}, x_{l-2}, \dots, x_0]) \quad (3-2)$$

在多模态的水平密集连接中，与 HyperDenseNet 一样，不同流中前几层的输出也被连接起来，形成后续层的输入。在多模态上下文中，这种连通性产生了比早期或晚期融合策略更加强大的特征表示，因为网络能够学习所有抽象级别内部和之间的不同模态之间的更复杂的关系。为简单起见，让我们仅考虑两种模式的方案。让x_l^1和x_l^2表示l^{th}分别在流 1 和 2 中图层。然后，输出l^{th}给定流中的层s可以定义为下式（3-3）

$$x_l^s = H_l^s([x_{l-1}^1, x_{l-1}^2, x_{l-2}^1, x_{l-2}^2, \dots, x_0^1, x_0^2]) \quad (3-3)$$

此外，最近的研究发现，在 CNN 中随机和交错完整的特征图（或单个特征图元素）可以提高其性能，因为它可以作为强大的正则化器。受此启发，我们可以以不同的顺序为每个分支和层连接特征图，其中l^{th}图层现在变为下式（3-4）

$$x_l^s = H_l^s(\pi_l^s[x_{l-1}^1, x_{l-1}^2, x_{l-2}^1, x_{l-2}^2, \dots, x_0^1, x_0^2]) \quad (3-4)$$

跟π_l^s是一个排列作为输入给出的特征图函数。因此，在两种图像模态的情况下，输出l^{th}两个流中的层都可以定义为式(3-5)和式(3-6)

$$x_l^1 = H_l^1([x_{l-1}^1, x_{l-1}^2, x_{l-2}^1, x_{l-2}^2, \dots, x_0^1, x_0^2]) \quad (3-5)$$

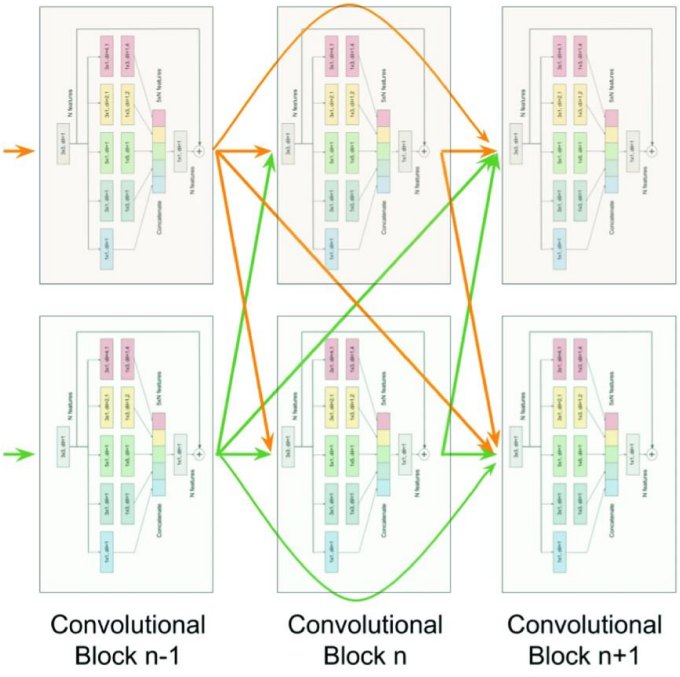

图 3-3 两种模态图像的超密集连接的详细示例

图 3-3 描述了在两种模态图像的情况下采用的超密集连接的详细示例，展示了三层神经网络的模块交互，是整个算法架构的一部分。两种模态图像在单独的路径中进行特征提取，箭头表示模块之间的连接模式。

3.3 实验设计

3.3.1 数据集

本文的数据集来自：Asphalt Pavement Crack Detection Based on Convolutional Neural Network and Infrared Thermography 此篇文章。该数据集中有四种类型的图像，包括可见光图像、红外图像、融合图像和地面实况图像。可见光图像和红外图像分别是完全可见光和红外光。融合图像是通过 IR-Fusion™技术实现的可见光和红外图像的组合，分别约为 50%和 50%，使用 Photoshop 软件在像素级别手动标记数据集中的 Ground Truth。图 3-3 描述了此数据集的基本信息。在三个时段获取数据（图像），包括早上（8:00 am）、中午（12:00 pm）和黄昏（5:00 pm）。路面最高温度与日最高气温基本一致，黄昏温度略高于早晨。图 3 给出了三个时段同一位置的可见光图像、红外图像和融合图像的示例。尽管这些可见光图像非常相似甚至相同，但它们的红外图像和融合图像却有显着不同。首先，它们的温度不同。中午的图像（图 8b）的温度最高，范围为 17.1 至 31.0 °C，黄昏的图像（图 8c）的温度第二高（5.0 至 15.9 °C），早晨的图像具有最低温度（0.5 至 7.5 °C）。这意味着这些图像中的相同颜色代表不同的温度。其次，裂缝区域温度最高，但背景图案不同。中午的图像背景大约具有其温度范围的中间温度。黄昏时的图像背景温度高于中间温度，早晨的图像背景温度也高于中间温度，但颜色更接近红色（远高于中间温度）。第三，裂纹的区分不同。中午图像中的一些裂缝区域与背景温度相似，导致裂缝的一些误识别。黄昏和早晨图像的裂纹区分更加清晰和明显。此外，护栏和树木造成的阴影可能会影响可见光图像和红外图像。为了消除此类环境因素，仅在没有护栏和树木的人行道上拍摄图像。每种类型共有 448 张图像（4 种类型，包括可见光图像、红外图像、融合图像和地面实况图像）。早上拍摄了 186 张图像，中午拍摄了 142 张图像，黄昏拍摄了 120 张图像。为了训练和评估分割模型，整个数据集分为两个子集：训练集和测试集。训练集和测试集的比例为 8：2。[19]

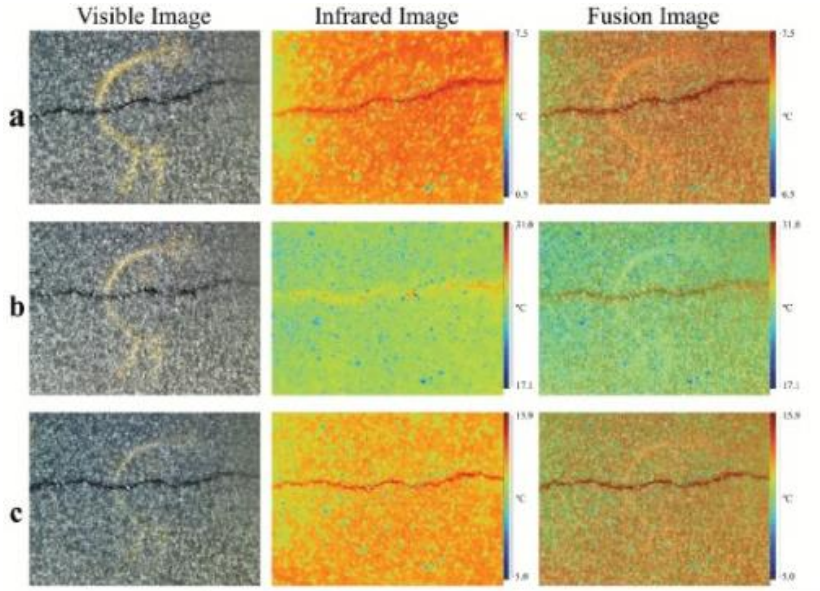

图 3-4 三段实践同一位置的可见光图像、红外图像、融合图像示例

表 3-1 说明了该数据集的摘要。每种类型共有 448 张图像（4 种类型，包括可见光图像、红外图像、融合图像和地面实况图像）。早上拍摄了 186 张图像，中午拍摄了 142 张图像，黄昏拍摄了 120 张图像。为了训练和评估分割模型，整个数据集分为两个子集：训练集和测试集。训练集有 382 张图像，测试集有 66 张图像。此外，裂纹像素和非裂纹像素的百分比也总结在表 3-1 中。裂纹区域仅占整个图像的一小部分（小于 4%）。[19]

表 3-1 本文数据集摘要

Dataset	Number	Crack Pixel (%)	Non-crack Pixel (%)
Morning	186	3.85	96.15
Noon	142	3.97	96.03
Dusk	120	3.20	96.80
Train	382	3.86	96.14
Test	66	2.88	97.12
Total	448	3.71	96.29

3.3.2 数据预处理

在对图像的预处理中，我们采用了图像增强技术和图像均衡化技术，通过使用图像增强和图像均衡化，能够使得模型拥有更高的性能和更高的鲁棒性，从而更好地处理数据。

图像增强可以增加数据的多样性，通过随机生成一个-20 到+20 的角度来对图像进行旋转，可以生成更加多样性的训练样本，增加数据的多样性，降低过拟合风险，从而能够帮助模型更好地泛化。

通过图像均衡化，可以调整图像的像素值分布，使其像素值分布更加均匀，增强像素的对比度，从而使得图像的细节更加清晰。并且，通过图像均衡化技术，可以使得图像中的局部特征更加突出，能够使得模型更好地识别和利用这些特征进行分割。

3.3.3 评估指标

像素精度（PA），见式（3-6），是计算正确分类的像素数量与其总数之间的比率。

$$PA = \frac{\sum_{i=0}^{k} P_{ii}}{\sum_{i=0}^{k}\sum_{j=0}^{k} P_{ij}} \tag{3-6}$$

平均像素精度（MPA），见式（3-7），是一种改进的 PA，考虑了每个类别的正确率，基于每个类别正确的像素总数与每个类别总数比率求和得到的均值。

$$MPA = \frac{1}{k+1}\sum_{i=0}^{k}\frac{P_{ii}}{\sum_{j=0}^{k} P_{ij}} \tag{3-7}$$

频率权重并交比（MIoU），见式（3-8），是语义分割网络的一个标准评价指标，它通过计算两个集合（ground truth 和预测分割结果）的交集和并集之间的比率来度量，是重新计算 TP 跟（TP+FN+FP）之间的比率。IoU 是基于每个类别计算，然后再求均值。

$$MIoU = \frac{1}{k+1}\sum_{i=0}^{k}\frac{P_{ii}}{\sum_{j=0}^{k} P_{ij} + \sum_{j=0}^{k} P_{ji} - P_{ii}} \tag{3-8}$$

F1 score，见式（3-9），F1 分数是二元分类准确性的衡量标准。

$$F_1 score = 2 \times \frac{Precision \cdot Recall}{Precision + Recall} = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \tag{3-9}$$

3.3.4 实现细节

模型整个架构是用 PyTorch 实现的，并使用 Adam 优化器来训练模型，使用$\beta_1 = 0.9$和$\beta_2 = 0.99$。训练在 200 个周期后收敛，初始学习率为1×10^{-4}，100 个周期后减少为一半。

图像在 0 和 1 之间归一化，没有使用其他预处理或后处理步骤。此外，没有采用数据增强来提高网络的性能。

3.3.5 对比模型

为了验证我们的方法的性能，我们将此方法的结果与其他多模态图像分割模型的分割结果进行了比较。

RFBNet[20]是一个输入级融合策略的多模态分割模型，RFBNet 架构如图 3-5 所示。它提出了一种具有高效融合机制的方案，该机制探索了编码器之间的相互依赖性，以残余融合块（RFB）为基础模块，以自下而上的方式实现交互式融合，由两个模态特定残余单元（RU）和一个门控融合单元（GFU）组成。

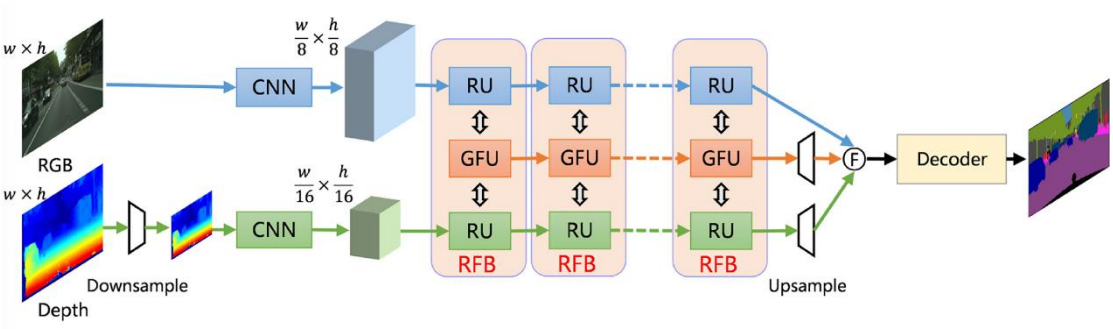

图 3-5 RFBNet 网络架构

LFC[21]由 Valada 等人提出，是一个基于决策级融合策略的多模态图像分割方法，架构如图 3-6 所示。此融合架构在相应的分支上分别提取多模态特征。将计算出的特征图相加以进行联合表示，然后将其输入一系列卷积层中，得到最终的输出掩膜。

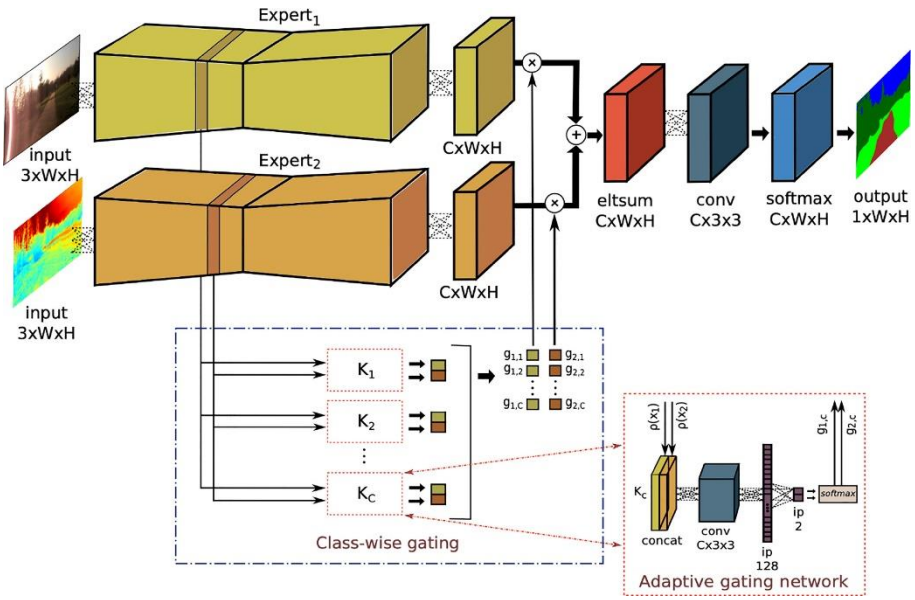

图 3-6 LFC 网络架构

S-M Fusion[22]是基于层级融合的多模态图像分割模型，考虑语义引导的多级融合，以自下而上的方式学习特征表示（架构见图 3-7）。这种融合策略采用级联

语义引导融合块 SFB 来按顺序融合跨模态的低级特征。

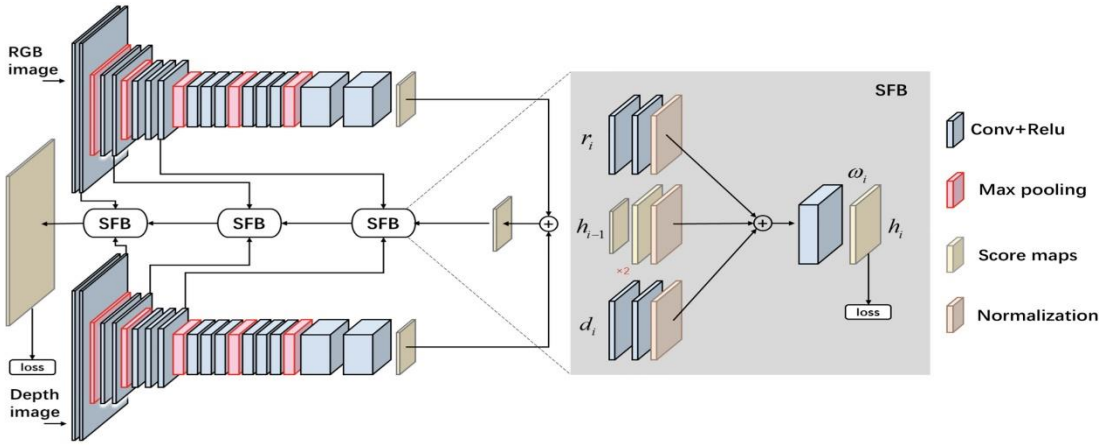

图 3-7 S-M 网络架构

3.4 实验结果和分析

3.4.1 分割预测

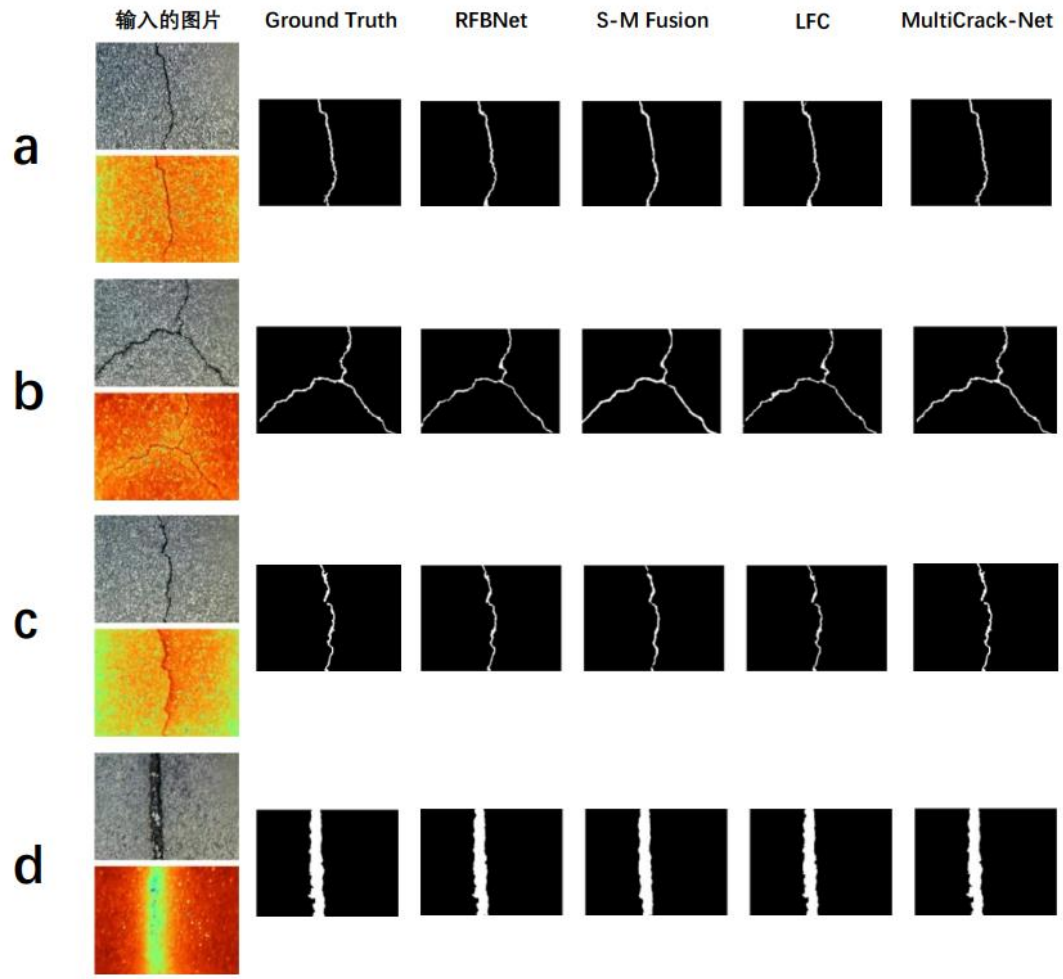

图 3-8 不同多模态图像分割模型预测对比

上图显示了 RFBNet、S-M Fusion、LFC 和 MultiCrack-Net 多模态图像分割模型对可见光图像和红外图像进行计算后得到的不同裂缝的预测。

总之，所有多模态分割模型的预测结果几乎相同。仔细对比之下，基于混合融合策略的多模态模型 S-M Fusion 和 MultiCrack-Net 生成的结果更好，对于多裂缝和细裂缝的分割更为准确。其中，MultiCrack-Net 生成的掩膜最接近真实情况，无需后处理即可用于准确的裂纹检测和识别。

3.4.2 计算精度

表 3-2 各种算法精度对比

模型	F1	PA	MPA	MioU
RFBNet	0.751	0.984	0.922	0.792
LFC	0.756	0.984	0.923	0.796
S-M Fusion	0.805	0.987	0.953	0.818
MultiCrack-Net	**0.808**	**0.989**	**0.965**	**0.833**

表3-2显示了各个多模态图像分割模型的精度指标，包括F1分数、PA、MPA和 MIoU。MultiCrack-Net在四个准确度指标中均获得最高值，这证明了多输入流编码器和超密集连接网络的有效性及先进性。

其次，S-M Fusion模型在四个准确度指标中效果仅次于MultiCrack-Net，每个指标仅有微小差距。S-M Fusion模型与MultiCrack-Net模型同为基于层级融合策略的多模态图像分割模型。相比之下，基于输入级融合策略的RFBNet模型在所有准确性指标中均具有最低值。这表明了对于道路裂缝的可见光图像和红外图像，层级融合更能挖掘和融合这两种模态图像中的低阶信息。基于输入级融合策略和决策级融合策略的模型可能不足以将得到两种模态中的特征进行融合。

3.4.3 消融实验

表 3-1 消融实验计算精度对比

模型	F1	PA	MPA	MIoU
MultiCrack-Net（仅可见光图像）	0.750	0.983	0.910	0.793
MultiCrack-Net（仅红外图像）	0.735	0.977	0.886	0.782
MultiCrack-Net（仅融合图像）	0.751	0.984	0.910	0.792
MultiCrack-Net	**0.808**	**0.989**	**0.965**	**0.833**

为了证明多模态图像分割相对于单模态的改进程度，我们对 MultiCrack-Net

模型和道路裂缝数据集做消融实验。表 3-3 是消融实验在四个评估指标的效果对。

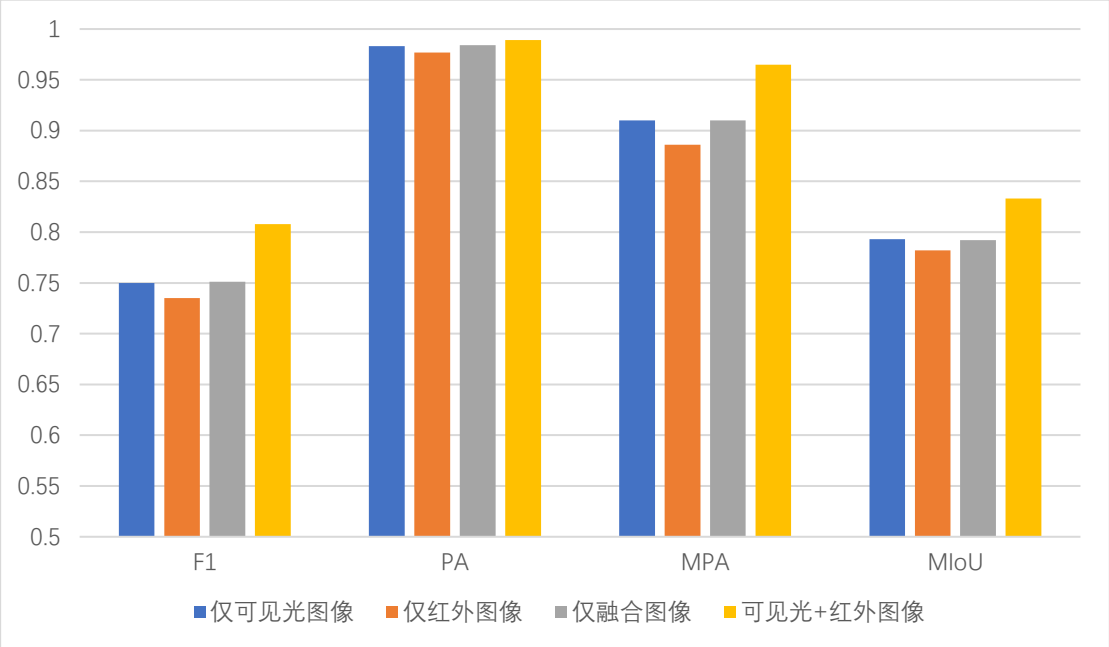

图 3-9 多模态图像分割效果

如图 3-9 所示，处理两个模态图像（可见光图像+红外图像）的 MultiCrack-Net 模型在四个准确度指标中均获得最高值，这证明了多模态图像分割比但模态有着不小的提高。

仅处理可见光图像和仅处理混合图像的 MultiCrack-Net 模型效果几乎一致，而仅处理红外图像的模型效果最差。这说明了可见光图像和混合图像包含更多有用的信息，可以提取更多有效的特征。

4 总结与展望

4.1 工作总结

通过阅读大量国内外相关文献，本文提出了 MultiCrack-Net 的多模态图像分割架构，这是一种基于多模态 U-Net 的道路裂缝定位和分割算法。本文所提出的算法采用的数据集包含了可见光图像和红外线图像这两种模态数据，经由两条路径分别捕捉图像信息后再对模态信息进行融合后分割，并在确立的评估指标中进行评估。此外，还与 RFBNet、S-M Fusion 和 LFC 这三个先进的多模态图像分割方法的分割结果进行比较，证明本文所提出的 MultiCrack-Net 具有更加优异的表现，能够较为准确的识别出道路裂缝。

4.2 进一步研究设想与展望

尽管基于深度学习的 U-Net 架构在多模态分割任务中表现较以往算法更加出色，但仍有进一步改进的空间。例如，我们可以进一步探索不同的网络架构，引入更加优异的损失函数，从而进一步提高模型的性能和稳定性。还可以引入更加广泛的跨模态信息融合进制，以便更好地利用多模态图像的信息，方便各种模态信息进行融合。目前的研究大多集中在特定数据集上对基于 U-Net 的多模态分割算法进行训练和评估，在未来的研究中，我们我可以提高模型的泛化能力，使得模型能够适用于不同领域的多模态图像数据集。并且，在很多应用时，实时性也是一个非常重要的考虑因素，因此，在接下的研究中，我们可以考虑在保持准确率的同时进一步提高模型的实时推断速度，这可以通过模型剪枝、高性能硬件等技术来实现。

参考文献

[1] Zeiler D M,Fergus R. Visualizing and Understanding Convolutional Networks.[J]. CoRR, 2013, abs/1311.2901.

[2] Khalil A,Turki T. Automatic Classification of Melanoma Skin Cancer with Deep Convolutional Neural Networks[J]. AI, 2022, 3(2).

[3] S S,Juliet S. Deep Medical Image Reconstruction with Autoencoders using Deep Boltzmann Machine Training[J]. EAI Endorsed Transactions on Pervasive Health and Technology, 2020, 6(24).

[4] Pan M,Shi Y,Song Z. Segmentation of Gliomas Based on a Double-Pathway Residual Convolution Neural Network Using Multi-Modality Information[J]. Journal of Medical Imaging and Health Informatics, 2020, 10(11).

[5] Shanchen P,Yaqin Z,Mao D, et al. A Deep Model for Lung Cancer Type Identification by Densely Connected Convolutional Networks and Adaptive Boosting[J]. IEEE Access, 2020, 8.

[6] Zhou T,Ruan S,Canu S. A review: Deep learning for medical image segmentation using multi-modality fusion[J]. Array, 2019, 3-4(C).

[7] Guo Z,Li X,Huang H, et al. Deep Learning-Based Image Segmentation on Multimodal Medical Imaging[J]. IEEE Transactions on Radiation and Plasma Medical Sciences, 2019, 3(2).

[8] Jerome L,Jing-Hao X,Su R. Segmenting Multi-Source Images Using Hidden Markov Fields With Copula-Based Multivariate Statistical Distributions.[J]. IEEE transactions on image processing : a publication of the IEEE Signal Processing Society, 2017, 26(7).

[9] He K,Zhang X,Ren S, et al. Identity Mappings in Deep Residual Networks.[J]. CoRR, 2016, abs/1603.05027.

[10] Bilwaj G,David H,Neil M, et al. Deep learning in the small sample size setting: cascaded feed forward neural networks for medical image segmentation[J]. Univ. of California, Los Angeles (United States);Oak Ridge National Lab. (United States);The Univ. of Chicago (United States), 2016, 9785.

[11] Ronneberger O,Fischer P,Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation.[J]. CoRR, 2015, abs/1505.04597.

[12] Shelhamer E,Long J,Darrell T. Fully Convolutional Networks for Semantic Segmentation.[J]. CoRR, 2016, abs/1605.06211.

[13] Srivastava N,Salakhutdinov R. Multimodal learning with deep Boltzmann machines.[J]. Journal of Machine Learning Research, 2014, 15(1).

[14] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. nature, 2015, 521(7553): 436-444.

[15] Zhang N,Ruan S,Lebonvallet S, et al. Kernel feature selection to fuse multi-spectral MRI images for brain tumor segmentation[J]. Computer Vision and Image Understanding, 2010, 115(2).

[16] LeCun Y,Boser B,Denker S J, et al. Backpropagation Applied to Handwritten Zip Code Recognition[J]. Neural Computation, 1989, 1(4).

[17] Junzhi Z,Zhaoyun S,Ju H, et al. Automatic pavement crack detection using multimodal features fusion deep neural network[J]. International Journal of Pavement Engineering, 2023, 24(2).

[18] Dolz J, Desrosiers C, Ben Ayed I. IVD-Net: Intervertebral disc localization and segmentation in MRI with a multi-modal UNet[C]//International workshop and challenge on computational methods and clinical applications for spine imaging. Cham: Springer International Publishing, 2018: 130-143.

[19] Liu F, Liu J, Wang L. Asphalt pavement crack detection based on convolutional neural network and infrared thermography[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(11): 22145-22155.

[20] L. Deng, M. Yang, T. Li, Y. He, C. Wang, RFBNet: Deep Multimodal Networks With Residual Fusion Blocks for RGB-D Semantic Segmentation, arXiv, 2019 , preprint arXiv:1907.00135. Liu H, Miao X, Mertz C, et al (2021) CrackFormer: transformer network for fine-grained crack detection. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, pp 3763–3772.

[21] Valada, G.L. Oliveira, T. Brox, W. Burgard, Deep multispectral semantic scene understanding of forested environments using multimodal fusion, International Symposium on Experimental Robotics, Springer 2017, pp. 465–477.

[22] Y. Li, J. Zhang, Y. Cheng, K. Huang, T. Tan, Semantics-guided multi-level RGB-D feature fusion for indoor semantic segmentation, Proceedings - International Conference on Image Processing, ICIP, volume 2017-Septe, IEEE 2018, pp. 1262–1266.

致　　谢

在完成这篇关于多模态图像分割在道路裂纹方面的研究论文的过程中，我深深感激并想要表达我最诚挚的谢意。在这段旅程中，有许多人给予了我宝贵的帮助、支持和鼓励，他们的贡献是我取得成就的重要原因。

首先，我要感谢我的指导老师杨飞。他的专业知识、悉心指导和宝贵建议为我提供了坚实的学术支持。他不仅在学术上给予我指引，还在生活和职业规划方面给予了我宝贵的建议。我要特别感谢他对我的耐心和信任，他的指导使我能够不断探索、学习和成长。

其次，我要衷心感谢我的女朋友刘绮琪。在我整个研究过程中，她给予我无微不至的支持和理解，鼓励我克服困难，始终保持信心。她的支持让我能够全身心地投入到研究中，我会永远感激她的陪伴和支持。

除此之外，我要感谢我的舍友陈骋。在繁忙的学术生活中，他为我营造了轻松愉快的生活氛围。他的支持和理解使我能够在学业和生活之间取得平衡，同时享受大学生活的乐趣。

并且，我要感谢我的朋友刘展。在学术交流和合作中，他与我分享了宝贵的经验和见解。我共同探讨问题、解决困难，相互激励和促进，这让我的研究之路更加丰富和有趣。我要特别感谢他对我的支持和鼓励，与他一起学习和成长是我人生中宝贵的经历。

在这个特殊的时刻，我想对所有曾经支持、帮助和鼓励过我的人表示最诚挚的感谢。没有你们的支持和鼓励，我将无法完成这篇论文。你们的付出和支持将永远铭记在我心中。