分类号：TP391　　学校代码：10363
密　级：公开　　学　　号：2210330180

安徽工程大学
Anhui Polytechnic University

硕士学位论文

题目 面向道路场景的夜间红外图像语义分割研究

论 文 作 者　苏世林
校 内 导 师　许钢 教授
校 外 导 师　朱标 高级工程师
专业学位类别　电子信息
研究方向（领域）　智能信息处理及应用

论文提交日期：2024 年 5 月 31 日

分类号：TP391　　单位代码：10363

密类级：公开　　学　　号：2210330180

题　目　面向道路场景的夜间红外图像语义分割研究

题　目　Research on Semantic Segmentation of Night Infrared Image for Road Scene

学 生 姓 名：苏世林

校 内 导 师：许钢 教授

校 外 导 师：朱标 高级工程师

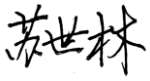
专　　业：电子信息

研 究 方 向：智能信息处理及应用

论文答辩日期：2024 年 5 月 23 日

安徽工程大学学位论文原创性声明

本人郑重声明：我恪守学术道德，崇尚严谨学风。所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已明确注明和引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品及成果的内容。论文为本人亲自撰写，我对所写的内容负责，并完全意识到本声明的法律后果由本人承担。

学位论文作者签名： 

日期：2024 年 5 月 31 日

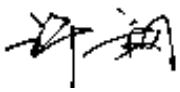
安徽工程大学学位论文版权使用授权书

学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅或借阅。本人授权安徽工程大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保密 □，在 ____ 年解密后适用本版权书。

本学位论文属于

不保密☑。

学位论文作者签名：  　　指导教师签名： 

日期：2024 年 5 月 31 日　　日期：2024 年 5 月 31 日

面向道路场景的夜间红外图像语义分割研究

摘 要

近年来，随着汽车及智能交通系统的快速发展，语义分割在道路场景中的应用也越来越广泛，较多的应用于汽车的自动驾驶技术。目前，针对可见光图像的语义分割算法已经较为成熟，然而汽车在道路场景下自动驾驶，如果遇到暗光、眩光、大雾等夜间复杂环境，可见光相机在采集数据过程中将受到很大影响，自动驾驶系统的感知精度和决策能力随之降低，将造成很大的行车安全隐患。而使用红外热像仪来应对上述夜间环境比较有优势，能有效避免感知精度下降的问题。因此，本文针对道路场景下的夜间红外图像语义分割算法进行了以下研究：

（1）当前开源夜间道路场景红外图像数据集非常匮乏，且红外图像数据具有低信噪比、低对比度等特性，不能很好的满足本文算法训练的需求。因此，从道路驾驶场景较为全面的 FLIR 红外目标识别数据集中筛选出部分夜间红外图像数据作为本文算法训练的数据集，并人工制作标签，从而建立适合本文算法训练的红外数据集；使用 CLAHE 算法对红外图像进行增强处理，并通过对图像质量各项指标增强前后的对比，验证了增强处理后红外图像质量的改善，使其利于标签制作和算法训练。

（2）针对道路场景语义分割的实时性与精确性都有较高要求，本文基于 DeepLabV3+的网络结构进行改进，解决其上下文特征信息不丰富、细节特征丢失较多以及网络计算量较大的问题。将骨干网络改为轻量化的 MobileNetV2，提高网络的实时性能；替换 DeepLabV3+

网络结构的空洞卷积空间金字塔模块（ASPP），引入基于深度可分离卷积的密集连接方式的空间金字塔模块(SPP)，即 D-ASPP，使得DeepLabV3+算法在获得更大更密集的特征尺度范围和更丰富的上下文信息的同时减少了网络的计算量，进一步提高实时性能；在编解码的特征提取阶段分别引入通道和空间注意力机制，通过对每个特征值赋值不同的权重，有效减少算法细节特征的丢失。改进算法在红外数据集和 Pascal VOC 2012 的 MIoU 分别达到了 80.84%和 73.83%。

（3）针对图像语义分割关键性难题之一进行研究，即语义分割模型在边缘域分割效果不佳的问题，改进的 DeepLabV3+算法仍然出现类别边缘细节特征丢失的问题。因此，本文通过将基于深度学习的语义分割算法（改进的 DeepLabV3+算法和 UNet 算法）与传统图像处理方法 SLIC 超像素进行结合，将高层语义特征与超像素的类别边缘细节信息相融合，从而构建红外图像融合分割算法，实现端到端的对夜间红外图像训练和学习，从而达到优化图像边缘分割的目的。融合算法（SLIC+改进 DeepLabV3+）和（SLIC+UNet）在红外数据集的 MIoU 分别达到了 81.82%和 83.16%。

关键词：语义分割, 道路场景, 自动驾驶, 红外图像, DeepLabV3+, 实时性

RESEARCH ON SEMANTIC SEGMENTATION OF NIGHT INFRARED IMAGE FOR ROAD SCENE

ABSTRACT

In recent years, with the rapid development of automobiles and intelligent transportation systems, semantic segmentation is more and more widely used in the road scene, and it is more used in automobile self-driving technology. At present, the semantic segmentation algorithm for visible images has been more mature, but if the car drives automatically in a road scene, if it encounters dim light, glare, fog, and other complex environment at night, the visible light camera will be greatly affected in the process of data acquisition, and the perception accuracy and decision-making ability of the autopilot system will be reduced, which will cause great driving safety risks. The use of an infrared thermal imager to deal with the above-night environment has an advantage, which can effectively avoid the decline of perceptual accuracy. Therefore, this paper studies the semantic segmentation algorithm of night infrared images in road scenes as follows:

The main contributions of this paper are as follows:

(1) The current open source night road scene infrared image data set is very scarce, and the infrared image data has the characteristics of low signal-to-noise ratio and low contrast, which can not well meet the training needs of this algorithm. Therefore, from the FLIR infrared target recognition data set with a more comprehensive road driving scene, part of the night infrared image data is selected as the data set for the training of this algorithm, and the label is made manually, to establish the infrared data set suitable for the training of this algorithm. CLAHE algorithm is used to

enhance the infrared image, and through the comparison of various image quality indicators before and after enhancement, the improvement of infrared image quality after enhancement processing is verified, which makes it convenient for label-making and algorithm training.

(2) Given the high requirements of real-time and accuracy of road scene semantic segmentation, this paper improves the network structure based on DeepLabV3+ to solve the problems of insufficient context feature information, large loss of detail features, and large amount of network computation. Change the backbone network to lightweight MobileNetV2 to improve the real-time performance of the network. The hollow convolution spatial pyramid module (ASPP) is replaced by the hollow convolution spatial pyramid module (ASPP), and the spatial pyramid module (SPP) based on deep separable convolution is introduced, namely D-ASPP, which enables the DeepLabV3+ algorithm to reduce the amount of network computation while obtaining a larger and denser range of feature scales and richer context information, and further improve the real-time performance. In the feature extraction stage of coding and decoding, channel and spatial attention mechanisms are introduced respectively to effectively reduce the loss of detailed features of the algorithm by assigning different weights to each feature value. The MIoU of the improved algorithm in the infrared data set and Pascal VOC 2012 is 80.84% and 73.83%, respectively.

(3) One of the key problems of image semantic segmentation is studied, that is, the segmentation effect of the semantic segmentation model is not good in the edge domain, and the improved DeepLabV3+ algorithm still has the problem of loss of category edge detail features. Therefore, by combining the semantic segmentation algorithm based on deep learning (improved DeepLabV3+ algorithm and UNet algorithm)

with the traditional image processing method SLIC super-pixel, and combining the high-level semantic features with the category edge detail information of super-pixel, this paper constructs the infrared image fusion segmentation algorithm to achieve end-to-end training and learning of night infrared image, to achieve the purpose of optimizing image edge segmentation. The MIoU of the fusion algorithm (SLIC+improved DeepLabV3+) and (SLIC+UNet) in the infrared data set reached 81.82% and 83.16% respectively.

Su Shilin (Electronic Information)

Supervised by Professor Xu Gang

KEY WORDS: semantic segmentation, road scene, autopilot, in frared image, DeepLabV3+, real-time

目录

摘 要..... I

ABSTRACT.....III

第 1 章 绪论.....1

1.1 研究背景与意义.....1

1.2 国内外研究现状.....3

1.2.1 可见光图像语义分割研究现状.....3

1.2.2 热红外图像语义分割研究现状.....5

1.3 论文主要研究内容及创新点.....7

1.4 本文章节安排.....8

第 2 章 语义分割与热成像技术理论.....10

2.1 引言.....10

2.2 语义分割基础理论.....10

2.2.1 卷积神经网络.....10

2.2.2 语义分割基本原理.....12

2.2.3 语义分割评价指标.....14

2.3 热红外成像技术理论.....15

2.3.1 热红外技术基础.....15

2.3.2 热红外成像原理.....19

2.4 本章小结.....21

第 3 章 红外数据集的建立与预处理.....22

3.1 引言.....22

3.2 语义分割数据集介绍.....22

3.2.1 Microsoft COCO 数据集.....22

3.2.2 Pascal VOC 数据集.....23

3.2.3 Cityscapes 数据集.....23

3.3 红外图像的增强处理.....24
3.3.1 CLAHE 图像增强算法.....24
3.3.2 红外图像预处理结果.....26
3.4 建立红外数据集.....28
3.4.1 数据类别与标注.....28
3.4.2 数据集扩充.....29
3.4.3 数据集划分与统计.....30
3.5 本章小结.....31
第 4 章 基于 DeepLabV3+的轻量化语义分割算法.....32
4.1 引言.....32
4.2 DeepLabV3+算法原理.....32
4.3 DeepLabV3+算法的改进.....33
4.3.1 骨干网络 MobileNetV2.....33
4.3.2 基于空洞可分离卷积的 D-ASPP.....36
4.3.3 注意力机制.....38
4.4 实验结果与分析.....40
4.4.1 实验配置与参数.....41
4.4.2 损失函数.....41
4.4.3 D-ASPP 模块对比实验.....41
4.4.4 红外数据集的实验结果与分析.....42
4.4.5 可见光数据集的实验结果与分析.....46
4.5 本章小结.....48
第 5 章 融合 SLIC 超像素的红外图像语义分割算法.....49
5.1 引言.....49
5.2 融合 SLIC 的边缘分割优化算法.....49
5.2.1 SLIC 超像素分割.....49
5.2.2 融合 SLIC 的 DeepLabV3+算法.....51
5.2.3 融合 SLIC 的 UNet 算法.....52

5.3 实验结果与分析.....53
5.3.1 实验配置与参数.....54
5.3.2 基于改进的 DeepLabV3+融合算法实验.....54
5.3.3 基于 UNet 的融合算法实验.....59
5.4 本章小结.....63
第 6 章 总结与展望.....65
6.1 总结.....65
6.2 展望.....66
参考文献.....68
攻读硕士学位期间主要研究成果.....75
致谢.....76

第 1 章 绪论

1.1 研究背景与意义

提到面向道路场景的语义分割技术，自然想到目前的研究热门，即自动驾驶技术。自动驾驶技术是从 1980 年起就被很多专家及研究机构所看重，被重视的原因就在于发展自动驾驶技术可以改善行车安全、缓解驾驶者负担、节约不可再生能源的消耗等等。自 21 世纪开始，全球举办各类自动驾驶领域的比赛[1]，使得自动驾驶越来越受到研究者们重视，成为非常具有发展前景的科技研究领域。随着科技的快速发展，包括人工智能、机器视觉等，无疑会加快自动驾驶技术发展的步伐，市场化也慢慢的成为现实。

自动驾驶汽车的感知系统和决策系统[2]是其核心部分。感知系统通过与计算机视觉技术的相结合，实现多任务的应用，包括车辆定位、静态景物的地图建立、动态物体的识别和跟踪[3]等。在这个过程中，语义分割技术起着重要的作用。

语义分割是一种像素级的图像分类技术，它将图像中的每个像素分配到预定义的类别中。这种技术可以帮助自动驾驶系统理解道路场景，例如将道路、车辆、行人等类别实现像素级的区域划分。图 1-1 的两张图中，分别为道路场景图像（Cityscapes 数据集样张）以及对应的语义分割预测结果。

一个准确的、实时的语义分割结果对于自动驾驶系统的场景理解[4]和后续的决策具有重要的意义。例如，它可以帮助系统识别出道路的边界，预测其他车辆和行人的行为，以及理解交通信号和标志。这些信息可以帮助系统做出更准确的决策，例如避开障碍物，遵守交通规则，以及安全驾驶。

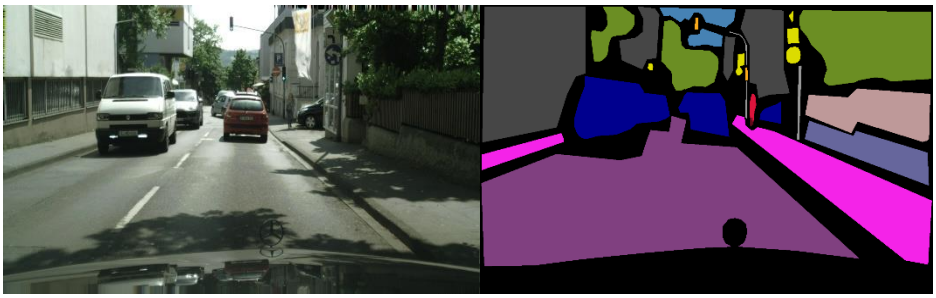


图 1-1 道路场景图与对应语义分割结果

目前，自动驾驶技术所用到的语义分割多用于可见光图像，但是在特殊环境下，比如黑夜、眩光和浓雾中，可见光相机的探测性能受限，这会影响语义分割的精度，对自动驾驶系统的可靠性和安全性构成挑战。

为了解决这一问题，研究人员已经开始探索结合其他传感器以及不同成像技术来提高自动驾驶系统的感知能力。其中，红外热成像技术因其在特殊环境下的优势而备受关注。通过结合可见光图像和红外图像，可以弥补可见光图像在特殊环境下的局限性，从而提高语义分割的精度和自动驾驶系统的整体性能。

这种多模态传感器融合[5]的方法有望为自动驾驶系统带来更全面的感知能力，同时也提高了在各种复杂环境下的安全性和鲁棒性。

随着汽车技术的发展，部分汽车配置了热红外成像仪，比如谷歌、比亚迪等一些公司在其研发的部分车型上既配置了可见光探测器用来获取可见光图像，又装备了红外热像仪用来辅助驾驶，有效提高了夜间以及特殊环境下驾车的安全性，如图 1-2 所示。

图 1-2 谷歌和比亚迪红外辅助驾驶汽车

红外技术因其独特性能，在军事侦察、夜间监控、无损检测、医疗诊断、环境监测等诸多领域发挥着重要作用。红外图像能够反映物体的热辐射特性，不受光照条件限制，尤其在黑暗或恶劣天气条件下具有明显优势。

另外，红外图像的低对比度、低信噪比以及缺乏色彩信息等特点，给语义分割、目标检测、目标识别等任务带来了挑战。针对这些难点，计算机视觉和人工智能领域的深度学习技术为红外图像分析提供了新的解决思路。通过构建深层神经网络模型，可以有效地从原始红外图像中提取高层次的语义特征，进而实现对场景的精确理解。例如，使用卷积神经网络（Convolutional Neural Networks，CNN）进行特征提取，结合全连接条件随机场（Dense Conditional Random Field，

DenseCRF）、图神经网络（Graph Neural Networks，GNN）或者 Transformer 架构来考虑像素间的上下文关系[6]，以提升分割和识别精度。

此外，通过数据增强、迁移学习以及自适应特征融合等策略优化深度学习模型，可以在有限的红外图像样本基础上改善模型泛化能力，并且在复杂的红外场景下实现更准确的目标检测与场景解析。随着研究的深入和技术的不断进步，红外图像场景理解这一研究方向将持续取得突破性进展。

1.2 国内外研究现状

目前，随着语义分割技术的研究与发展，出现了很多的算法模型，本文将语义分割国内外研究情况分为基于可见光图像的语义分割方法和基于热红外图像语义分割方法来具体阐述，能够较为全面的了解目前语义分割技术的发展动态与趋势。

1.2.1 可见光图像语义分割研究现状

在可见光图像语义分割方面，出现了很多优秀的算法模型。J.Long等人[7]首次详细介绍了全卷积网络（Fully Convolutional Networks，FCN）应用于图像语义分割任务的方法。提出利用全卷积神经网络来处理任意尺寸的输入图像，并通过反卷积层（也称为转置卷积层）将高层特征映射恢复到与原始输入相同的空间分辨率，从而输出同尺寸的像素级分割结果。Ronneberger等人[8]提出了UNet这一深度学习架构，专门用于生物医学图像的分割任务。UNet的结构包括下采样路径（编码器）和上采样路径（解码器）。下采样路径用于捕获输入图像的上下文信息，而上采样路径则用于实现精确定位，并生成与输入图像相同尺寸的分割图。这两条路径之间存在跳跃连接，有助于保留更多的细节信息。

Badrinarayanan等人[9]提出了SegNet网络，是一种编解码结构的语义分割算法。SegNet的设计理念在于利用已知的池化索引来执行上采样操作。在编码过程中，最大池化层会记录下每个像素的最大池化索引，这些索引在解码阶段被用来恢复原始空间维度，即进行非线性上采样（也称为“Unpooling”）操作，这种方法使得SegNet无需额外学习上采样的权重参数，同时保留了图像的空间结构信息。与

其它编码器-解码器结构相比，SegNet的独特之处在于其对池化索引的有效利用以及轻量级的设计，如图 1-3[9]所示。

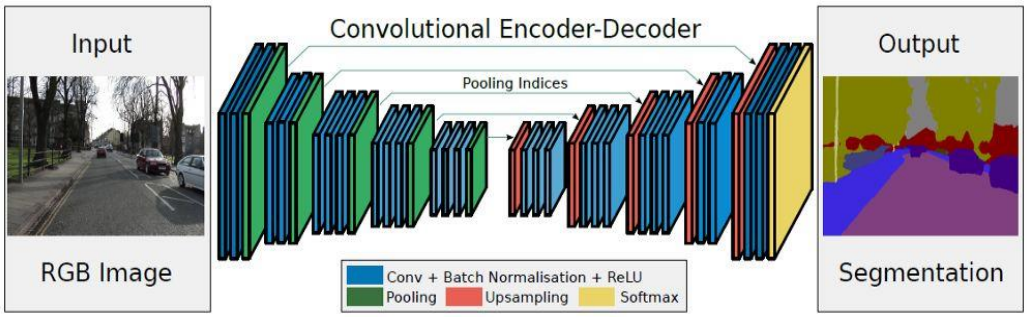


图 1-3 SegNet 网络结构

Liu等人[10]在FCN之后应用条件随机场（Conditional Random Fields, CRFs）模型，可以在深度学习预测的基础上，利用像素之间的成对势能函数来优化标签分配，使得相邻像素具有更高的概率分配相同的类别标签，从而有效改善了分割边界的质量。LC Chen等人[11-13]提出的DeepLab系列模型中使用空洞卷积（Atrous Convolution）和空洞空间金字塔池化（Atrous Spatial Pyramid Pooling，ASPP）。利用了这些技术克服了传统CNN在处理高分辨率输出时容易丢失边缘细节的问题，并且有效提高了对图像中目标的多尺度识别能力，从而在图像语义分割领域取得了突破性进展。

DeepLabV3+[14]算法展现出了卓越的性能，并在PASCAL VOC 2012 和Citys capes等重要数据集上取得了当时最佳的结果。DeepLabV3+结合了编码器-解码器架构的优点，其中采用了空洞卷积以及自适应的空洞空间金字塔池化（ASPP）模块，有效地捕获了不同尺度下的上下文信息，从而提高了分割精度，网络结构如图 1-4[14]所示。PSPNet[15]算法精度相对DeepLabV3+算法也较为优秀，通过引入金字塔池化模块（Pyramid Pooling Module）来获取不同尺度下的上下文信息，从而有助于提高对整个场景的理解能力。

Huang Z等人[16]提出CCNet（Criss-Cross Network）网络，通过引入Criss-Cross Attention模块来改进传统注意力机制处理图像信息的方式，能够沿着垂直和水平方向形成交叉路径，从而收集并整合远距离的相关特征。Li H等人[17]提出的DFANet（Deep Feature Aggregation Network）网络，利用轻量级Xception骨干网络提取特征并聚合子网络中区分度强的特征，以多尺度特征传递的形式增大感受野

并加强模型的学习能力，从而保持分割速度和分割精度的均衡。

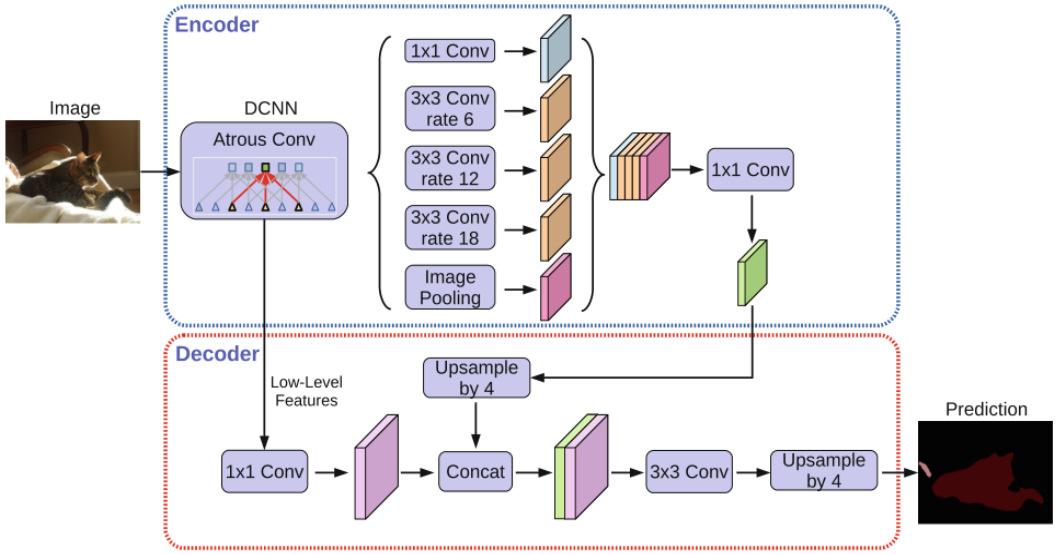


图 1-4 DeepLabV3+网络结构

1.2.2 热红外图像语义分割研究现状

考虑到可见光传感器在低光照等特殊条件下的探测精确度较低，近年来出现了利用多光谱融合手段来提高分割精度的方法。Hazirbas 等人[18]提出了一个名为 FuseNet 的深度学习模型，用于同时利用可见光图像和深度信息进行场景语义分割。FuseNet 网络结构采用了编码器-解码器（Encoder-Decoder）架构，并针对双模态输入进行了优化设计。在该模型中，可见光图像和深度图像分别通过各自的卷积路径进行特征提取，在编码器阶段，深度分支提取出的高级语义特征被融合到可见光分支中，实现了多模态信息的有效整合，如图 1-5[18]所示。

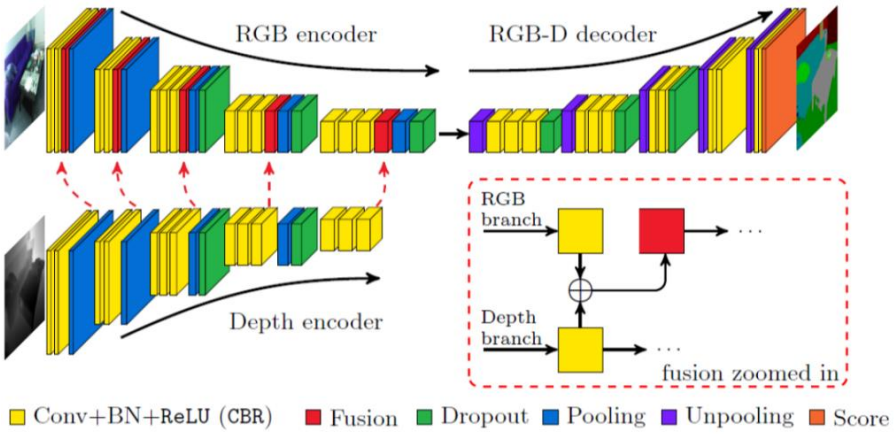


图 1-5 FuseNet 网络结构

Ha 等人[19]研究的 MFNet（Multi-modal Fusion Network）是一种针对可见光与热红外信息融合进行图像语义分割的深度学习模型。该网络采用了双编码器结构，其中一个通道专门处理可见光图像，另一个通道则负责处理热红外图像。在解码器阶段，两个通道提取的高级特征被有效地结合在一起，以便更好地利用两种不同类型的传感器提供的互补信息，从而提高图像分割的精度和准确性。

吴骏逸等人[20]针对夜间道路场景解析的挑战，在全卷积神经网络（FCN）的基础上提出了一种改进方法。他们设计了一种并行全卷积神经网络结构，该结构能够同时处理来自可见光和热红外等多模态传感器的数据，以充分利用不同频谱信息间的互补性。不仅增强了对夜间复杂光照条件下道路场景的理解能力，而且在实际应用中取得了显著的效果，为自动驾驶、智能交通监控等领域提供了有力的技术支持。Y Sun 等人[21]提出 RTFNet（RGB-Thermal Fusion Network）这一新型深度学习网络，此网络使用 ResNet（Residual Neural Network）作为编码器的一部分，有助于提高网络对输入图像的特征提取能力。另外，在解码部分还使用了一种创新结构的上采样模块，使得网络在还原分辨率的同时能够更好地保留细节信息，从而提高了分割的准确性。

Azarbad等人[22]在研究热红外图像分割时，采用了蜜蜂算法这一启发式优化方法。蜜蜂算法模拟自然界中蜜蜂的觅食行为，通过搜索空间中的多个潜在解（即不同的阈值），最终确定出使图像信息熵达到最大化的阈值。这种方法被证明在处理无偏优化问题上具有高效性和稳定性，尤其适用于快速处理大量数据且需要精确分割的场景，因此在热红外图像分割领域具有较高的应用价值。Yu等人[23]在热红外图像分割领域提出了一种创新方法，其融合了免疫场理论和克隆选择算法，该方法利用了免疫系统自我调节和学习适应的特点，结合克隆选择算法的全局优化能力，在热红外图像分割任务中取得了良好的效果。

Vertens 等人[24]在多模态语义分割的研究中，针对自动驾驶场景下的全天候图像理解问题，设计了一种能够在白天和黑夜条件下有效进行语义分割的模型。鉴于自动驾驶研究领域内热红外图像数据的匮乏，作者还构建了一个新的数据集，该数据集包含 20000 对时间同步且空间对齐的可见光图像与热红外图像配对样本。Chen 等人[25]在研究低分辨率热红外图像的语义分割任务时，创新性提出了

一种全卷积神经网络架构。为了优化模型性能和精确度，引入了加权交叉熵损失函数，该函数能够根据类别不平衡问题对各类别像素的重要性进行不同的权重分配，从而更准确地衡量预测值与真实标签之间的差异，并有效地指导模型训练。

1.3 论文主要研究内容及创新点

通过对国内外研究现状的了解，当前针对道路场景的语义分割算法现有的研究方法中，基于热红外图像的语义分割相对较少，存在开源红外数据集匮乏的问题，也存在实时性能与分割精度失衡的问题等主要问题。因此，本文针对这些问题进行研究并进行改善。

本文旨在基于深度学习的语义分割算法，使得自动驾驶汽车在面向夜间复杂道路场景时，在热红外成像仪的加持下，感知系统能够兼具良好的实时性和优良的语义分割精度。

由于目前完全公开的夜间道路场景下的红外图像语义分割数据集匮乏，而且基本不适合本课题的研究，因此需首先解决这一关键问题；另外，由于在面向道路场景时所出现的复杂的道路状况，对象往往显示出非常大的尺度变化，而且本文研究对象为红外图像，因此实现精确预测较为困难。考虑到这个问题本文研究的语义分割模型选择了 DeepLabV3+模型。其空间空洞金字塔模块（ASPP）在输入特征上应用多采样率扩张卷积，探索多尺度上下文信息，结构通过逐渐恢复空间信息来捕捉清晰的目标边界。但是该模型在面对道路驾驶场景中的对象尺度变化非常大时，其理解能力仍不足，上下文信息不够丰富，过程中丢失了较多的细节信息，而且模型的参数量较大不利于实时处理，本文研究中会对该模型进行改进，从而改善这些不足；另外，考虑图像语义分割模型在分割中的难题，即边缘信息并不能得到很好的保留，导致边缘分割效果普遍较差，因此本文对以上研究难点展开了以下研究来改善不足：

（1）筛选出道路驾驶场景较为全面的 FLIR 红外目标识别数据集中的部分夜间红外图像数据，作为本文算法训练验证的数据集，并人工制作标签。针对红外图像具有低信噪比、低对比度等特性，使用 CLAHE（Contrast Limited Adaptive Histogram Equalization）算法增强处理，并与其他直方图均衡化算法对比说明。

另外验证了图像各项指标的改善有效性。

（2）基于 DeepLabV3+算法结构的改进，以满足本文所需道路驾驶场景下的实时性能与精确性能。将骨干网络改为轻量化的 MobileNetV2，提高网络的是实时性能；替换 DeepLabv3+网络结构的空洞卷积空间金字塔模块（ASPP），引入基于深度可分离卷积的密集连接方式的空间金字塔模块(Spatial Pyramid Pooling，SPP)，即 D−ASPP（Dense Atrous Spatial Pyramid Pooling），改善算法上下文特征信息还不够丰富的不足；另外在特征提取阶段分别引入通道和空间注意力机制，提高算法对细节特征信息的提取能力。

（3）在改进的 DeepLabV3+算法结构基础上，与 SLIC（Simple Linear Iterative Cluster）超像素分割算法构建融合算法，实现端到端的对夜间红外图像训练和学习，从而达到优化图像边缘分割的目的。另外，将 UNet 算法结合 SLIC 超像素算法进行对比实验，用以验证融合 SLIC 超像素后的语义分割模型是可行的、有效的。

本文通过以上主要三方面进行了研究与改进，改进后的算法模型相对于当前前的研究方法来说，分割精度与实时性能都有所更高，实现两者的平衡。当然，与轻量化的语义分割网络相比，实时性能还有待提高，但分割精度却有较大优势。因此，本文的研究内容基于当前研究现状所展现的问题，进行有效地改善，能够验证本文改进方法的有效性与必要性。

1.4 本文章节安排

本文共有六个章节，具体的章节安排如下：

第一章 绪论。阐述了本研究课题的背景及意义，然后概述了语义分割技术的国内外研究现状，具体从可见光图像语义分割和热红外图像语义分割两个方面进行阐述。另外，介绍了本文的主要研究内容及章节结构安排。

第二章 语义分割与热红外成像技术理论。简要介绍了基于深度学习的语义分割的理论知识，包括图像语义分割神经网络基础知识等等，还介绍了热红外成像技术相关理论。

第三章 夜间红外数据集的建立与预处理。首先介绍了一些语义分割广泛使

用的开源数据集。然后介绍了建立本文适用的夜间红外数据集的流程以及关键技术，其中包括：红外图像数据的筛选、红外图像数据的增强预处理、人工标注标签文件、采用数据增强方法对数据集进行扩充以及数据集的划分统计。

第四章 基于DeepLabV3+的轻量化语义分割算法。主要介绍了在DeepLabV3+算法的基础上，对其结构进行改进优化，使其满足道路驾驶场景的需求，即良好的实时性和语义分割精度。因此对其骨干网络模块、空洞空间金字塔池化模块（ASPP）进行改进，并引入了通道注意力和空间注意力机制。在红外数据集上进行训练，并在Pascal VOC 2012 可见光数据集进行对比实验。另外，还进行了相关消融实验以及与经典语义分割算法的对比实验等。

第五章 融合 SLIC 超像素的红外图像语义分割算法。介绍了 SLIC 超像素算法等相关理论知识；进行了 SLIC 超像素算法与改进的 DeepLabV3+算法融合训练，以及与 UNet 算法融合训练，从而验证结合 SLIC 算法能够优化红外图像边缘分割的有效性。

第六章 总结与展望。综合性的将本文工作内容做出了梳理与总结，指出了本文研究内容的创新之处以及不足之处，并对面向道路场景的夜间红外图像语义分割算法研究方向进行了展望。

第 2 章 语义分割与热成像技术理论

2.1 引言

本文针对夜间道路场景的语义分割研究是计算机视觉领域中的重要课题。这种技术旨在将图像分割成像素级别的区域，并对这些区域进行标记和分类，以便理解图像中不同物体的角色和位置。与目标检测不同，语义分割需要对每个像素进行预测，以实现对整个场景的理解。在道路场景语义分割中，这项技术可以应用于识别道路上的不同元素，比如汽车、摩托车等，并确定它们的边界，这对于自动驾驶和交通管理等领域具有重要意义。

另外，本文针对的夜间红外图像是通过热红外成像仪器采集的，热红外成像技术凭借其特有的优势已经广泛应用于各个领域，本文研究内容体现出了其有效弥补可见光相机的不足之处。

2.2 语义分割基础理论

2.2.1 卷积神经网络

语义分割中去掉了用于分类的传统全连接层，取而代之的是将卷积操作应用于网络结构，将全卷积神经网络作为网络的编码部分，从而构建逐像素语义分割的基础。它通过逐像素的方式提取高层语义特征，为后续的解码部分提供必要的信息输入。因此，卷积层、池化层以及全连接层成为卷积神经网络的三大组成部分。

卷积层是卷积神经网络（CNN）的核心组件之一，它负责提取输入数据的特征。在图像处理中，卷积层可以理解为一个滤波器，它在输入图像上滑动，提取出不同位置的局部特征。

卷积层的主要参数[26]包括：

（1）卷积核（Filter）：卷积核是卷积层的参数，它是一个小的矩阵，可以理

解为一个滤波器。卷积核的大小通常是 3×3 或 5×5。

（2）步长（Stride）：步长是卷积核在输入数据上滑动的距离。步长越大，输出的特征图的尺寸就越小。

（3）填充（Padding）：填充是在输入数据的边缘上添加额外的像素，以便卷积核可以滑动到边缘。常用的填充方式包括“valid”（不填充）和“same”（填充后输出的特征图与输入数据的尺寸相同）。

卷积层的输出通常称为特征图（Feature Map），它是输入数据的局部特征的映射。特征图的深度等于卷积核的数量，每个深度对应一个卷积核提取的特征。卷积层的工作原理[27]是通过卷积操作来提取输入数据的特征，而卷积操作是将卷积核与输入数据的局部区域进行元素级的乘法，然后将所有乘积的结果相加，得到输出的特征图，如图 2-1 所示。

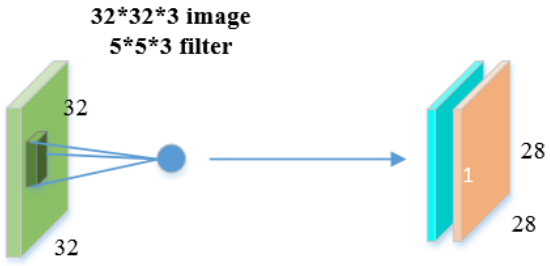


图 2-1 卷积层使用卷积核输出特征

池化层（Pooling Layer）是 CNN 中的重要层，用于减少特征图的维度，从而降低计算量，减少过拟合，以及提高网络的鲁棒性。

池化层的主要作用[28]是对特征图进行下采样，通常通过取最大池化（Max Pooling）或者取平均池化（Average Pooling）来实现，如图 2-2 所示。例如，对于一个 2×2 的池化层，最大池化操作会在每个 2×2 的区域内取最大值，从而将特征图的大小减半。

池化层通常在卷积层之后使用，以便在保留重要特征的同时，减少特征图的维度。这有助于提高网络的计算效率，同时也有助于防止过拟合。总的来说，池化层是卷积神经网络中的一个重要组成部分，它通过对特征图进行下采样，从而减少特征维度，提高网络实时性能。

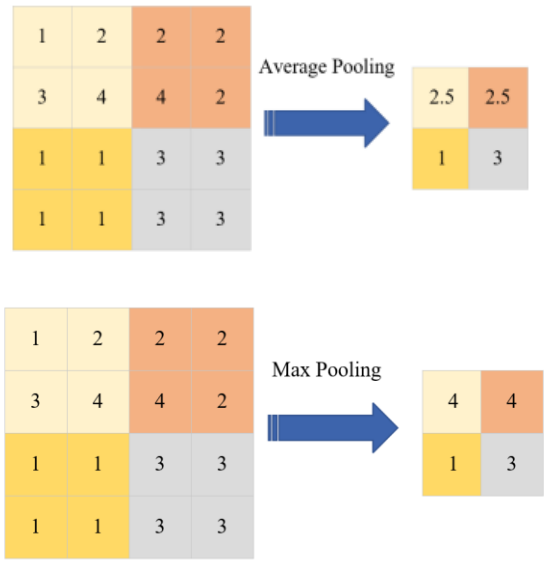


图 2-2 平均池化与最大池化

全连接层（Fully Connected Layer）也被称为密集层[29]（Dense Layer）。全连接层的每个神经元都与上一层的所有神经元相连，这意味着全连接层的输出是上一层所有特征的线性组合。

全连接层通常在卷积层和池化层之后使用，用于将卷积层和池化层提取的特征进行分类或回归。例如，对于图像分类任务，全连接层可以将图像的特征映射到不同的类别。全连接层的输出可以通过激活函数进行非线性变换，如ReLU（Rectified Linear Unit）或Sigmoid函数，这有助于网络学习非线性的特征。总的来说，全连接层是卷积神经网络的重要组成，它通过将卷积层和池化层提取的特征进行分类或回归，从而完成网络的最终输出。

2.2.2 语义分割基本原理

在语义分割中，通常会将RGB图像（H×W×3）或者灰度图像（H×W×1）作为输入，然后输出预测分割图像（H×W×1），其中每个像素包含了其对应的类别标签。输入图像可以是彩色图像，其中每个像素由红、绿、蓝三个通道的数值表示。同样也可以是灰度图像，其中每个像素只包含一个单独的数值，本文使用的则是单通道的红外灰度图像。

输出预测分割图的大小应该与输入图像的大小相同，其中每个像素的值表示了该像素所属的类别。一般来说，类别标签使用整数编码，每个整数对应一种类

别。例如，可以使用 0 表示背景类别，1 表示汽车类别，2 表示行人类别，以此类推。为了清晰起见，简化了输出分割图的分辨率，实际上分割图的分辨率应与原始输入的分辨率相匹配。语义分割输入输出如图 2-3 所示。

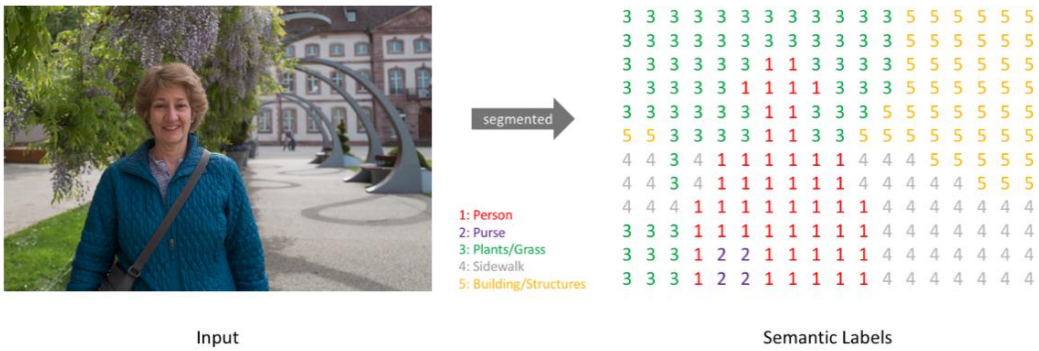


图 2-3 语义分割输入与输出

语义分割中，通常使用多分类Softmax函数来计算每个像素属于每个类别的概率，比如本文数据集包含 5 个类别（包括背景类），那么算法模型的输出将是一个H×W×5 的张量，其中每个元素代表了对应像素属于每个类别的概率。

为了得到最终的分割结果，可以使用 Argmax 函数[30]来选择每个像素概率最高的类别。这将得到一个 H×W 的矩阵，其中的每个元素代表了对应像素的类别，这个过程使用的是独热编码[31]（One-hot），如图 2-4 所示。

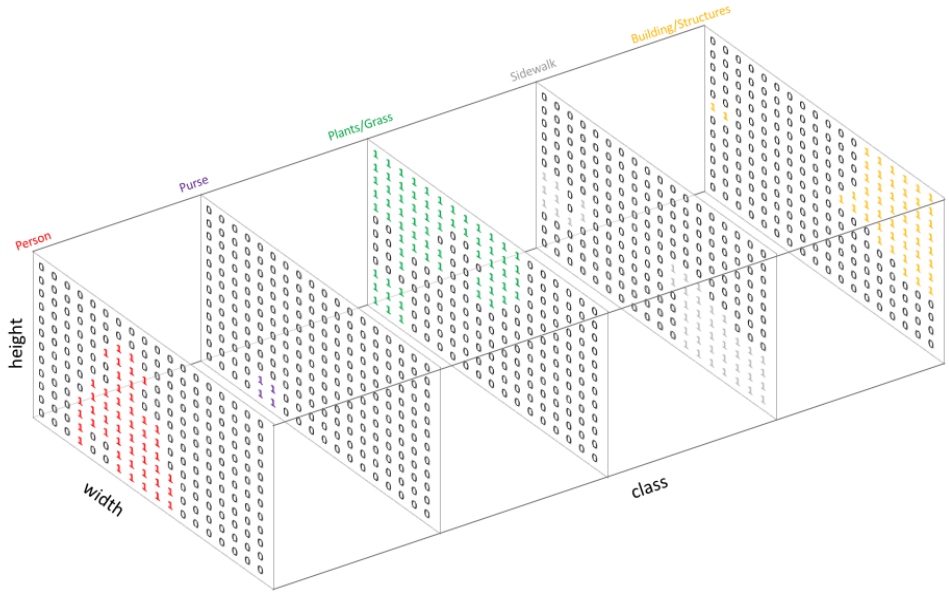


图 2-4 语义分割独热编码结果

整个过程也可以看作是对每个可能的类别创建一个输出通道，然后选择每个像素概率最高的类别，这与处理标准分类值的方式类似，只是在语义分割中，需

要逐像素进行处理，而不是整幅图像。最终得到的分割结果是一个与原始图像大小相同的矩阵，这个矩阵可以用来进行图像分割，或者其他需要像素级分类的任务。语义分割分类结果如图 2-5 所示。

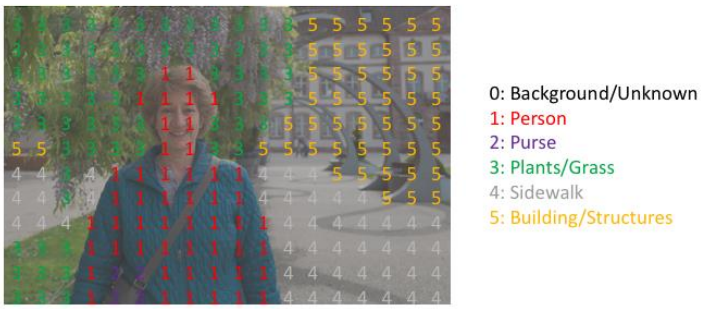


图 2-5 语义分割逐像素的分类结果

2.2.3 语义分割评价指标

在语义分割任务中，一个可靠且稳定的评价方法是非常重要的。平均像素精度、像素精度、平均交并比等都是评价算法性能的常见指标。本文则主要以平均交并比（MIoU）作为算法的评价指标。

首先介绍一下混淆矩阵[32]，如表 2−1 所示。

表 2-1 混淆矩阵表

混淆矩阵		真实值	
		Positive	Negative
预测值	Positive	True Positive （*TP*） 真阳性	False Positive （*FP*） 假阳性
	Negative	False Negative （*FN*） 假阴性	True Negative （*TN*） 真阴性

为了方便解释，先设定出数据集中类别的总数 $k+1$（0...k），背景通常用 0 来表示。p_{ii} 表示模型正确地预测出正类别以及负类别，即原本是 i 类别，预测的结果也是 i 类别，也就是表中提到的真阳性（*TP*）以及真阴性（*TN*），p_{ij} 表示模型错误地将负类别预测为正类别或将正类别预测为负类别，即原本是 i 类别被错误预测成了 j 类别，也就是表中提到的假阳性（*FP*）以及假阴性（*FN*）。若

第 i 类别为正类别，那么在 i 与 j 不相等时，那么 p_{ii}、p_{jj}、p_{ji} 和 p_{ij} 则分别表示为 TP、TN、FP 和 FN。

（1）像素精度（Pixel Accuracy，PA），其衡量了图像分割模型在整个图像上的分类准确性。该指标并不考虑不同类别之间的权重或平衡性，只是简单地计算正确分类的像素所占的比例。如公式(2-1)所示。

$$PA=\frac{TP+TN}{TP+TN+FP+FN} \tag{2-1}$$

（2）平均交并比（Mean Intersection over Union，$MIoU$），是用于评价算法性能的重要指标之一。$MIoU$ 的计算方法[33]是对每个类别计算其交并比（Intersection over Union，IoU），然后取所有类别的平均值。通常情况下，$MIoU$ 值越高，则表示模型的分割结果与实际情况的吻合程度越好。计算公式如式(2-2)和(2-3)所示。

$$IoU=\frac{TP}{TP+FP+FN} \tag{2-2}$$

$$MIoU=\frac{1}{k+1}\sum_{i=0}^{k}\frac{TP}{TP+FP+FN} \tag{2-3}$$

2.3 热红外成像技术理论

目前汽车行业正在开发更安全、更有效的 ADAS 系统和自动驾驶系统，常搭载毫米波雷达、激光雷达、高感知摄像机等等，而在漆黑、眩光、烟雾等特殊环境下则考虑结合使用热成像仪来探测汽车周围物体，其感知热红外辐射或热量的能力为现有的传感器技术提供了独特的互补优势。常见的热成像仪红外辐射的波长多为 8~14 μm，属于远红外或长波红外。

2.3.1 热红外技术基础

任何温度高于绝对零度(−273.15℃)的物体，都会根据其温度发出红外能量，即物体的热特征[34]。通常来讲，目标物体温度越高，其产生的热辐射就会越多。红外热成像仪实质是一种高度敏感的热传感器，物体产生的非常小的温度差，也能够检测和捕捉到。因此，其将物体产生的红外辐射进行收集，同时依据温差信息来建立像素，然后组成图像。红外彩色热图示例如图 2−6 所示。

图 2-6 红外彩色热图示例

（1）辐射理论：

根据物体的温度，任何物体都会以一定频率辐射电磁波，这种辐射称为热辐射。根据普朗克辐射定律，热辐射的强度与物体的温度有关。同时，物体也会吸收和发射电磁波，这些特性被广泛应用于许多领域，包括红外线成像技术、激光技术和通信技术等。

物理学家在研究热辐射时引入了黑体（Black Body）[35]这一理想物体，黑体是一个完美的吸收体和发射体，它能够完全吸收所有照射在其上的任何波长的电磁辐射，同时按照其自身温度无选择性地向外发出辐射。这意味着黑体不反射也不透射任何光线，因此在所有方向上观察到的辐射强度只与黑体本身的温度有关。通过研究黑体辐射，科学家不仅了解了物质热辐射的基本原理，还发现了许多自然界中的基本物理常数，如普朗克常数等，并且对现代科技领域如红外成像技术、光源设计、天体物理学等领域产生了深远影响。

基尔霍夫在 1860 年发现了一项重要的规律，即物体对于某一特定波长辐射的吸收能力与其发射能力之比是一个恒定值，这一比值不依赖于物体的具体性质，而是一个普适性的常数。这一发现被称为基尔霍夫定律[36]，它揭示了物体的辐射特性与其材质无关，如式(2-4)所示：

$$W_\lambda(T)=\alpha(\lambda,T)\cdot E_\lambda(T) \tag{2-4}$$

式(2-4)中，$W_\lambda(T)$ 是辐射出射度，$\alpha(\lambda,T)$ 为辐射吸收率，$E_\lambda(T)$ 是黑体对相同波长的辐射出射度。因此，还定义了比辐射率，其描述了特定波长和温度下的辐射吸收率与辐射出射度之间的关系，如式(2-5)所示：

$$\varepsilon=\frac{W_\lambda(T)}{E_\lambda(T)}=\alpha(\lambda,T) \tag{2-5}$$

德国物理学家马克斯·普朗克于 1900 年提出了著名的普朗克辐射定律[37]，

用以解释黑体辐射的光谱分布问题。这一定律描述了黑体辐射的光谱能量密度与波长和温度之间的关系，对于不同波长的辐射能量分布进行了准确的描述。如式(2-6)所示：

$$W_{\lambda}(T)=\frac{2\pi hc^{2}}{\lambda^{5}}\times\frac{1}{e^{\frac{ch}{\lambda kT}}-1} \tag{2-6}$$

式(2-6)中，$W_{\lambda}(T)$ 表示光谱辐射通量密度 $(w\cdot cm^{-1}\cdot\mu m^{-1})$，而 h 和 k 分别表示普朗克常数和玻尔兹曼常数，λ 表示波长，T 表示绝对温度，c 表示光速。光谱辐射通量密度如图 2−7 所示。

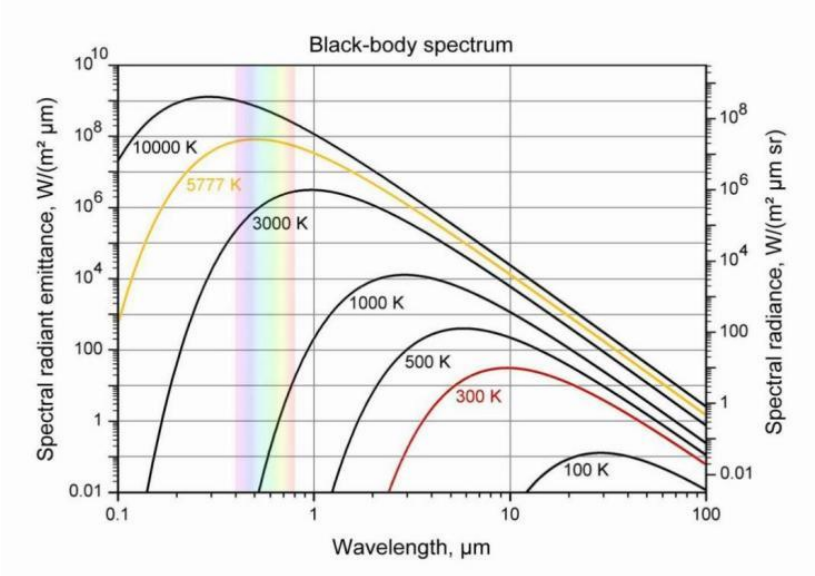


图 2-7 光谱辐射通量密度随温度和波长的变化

从图 2-7 不难发现，黑体辐射光谱分布随着温度的升高会发生变化，当黑体的温度上升时，其辐射能量最强的光谱位置会向短波方向移动，从红外波段逐渐过渡到可见光甚至紫外波段[38]。在室温（大约 300K）下，黑体辐射的峰值位于红外区域。通过对光谱辐射通量密度曲线积分则可得到辐射通量密度，如式(2-7)所示：

$$W(T)=\int_{0}^{\infty}W_{\lambda}(T)d\lambda \tag{2-7}$$

同理，计算某一波段 $[\lambda_1,\lambda_2]$ 的辐射通量密度，如式(2-8)所示：

$$W(T)=\int_{\lambda_1}^{\lambda_2}W_{\lambda}(T)d\lambda \tag{2-8}$$

而斯蒂芬、玻尔兹曼在他们的实验结果中发现了黑体辐射的总能量与其绝对

温度的四次方之间的联系是正比例的[39]，从新的角度解析了普朗克定律，简化了黑体辐射通量密度的计算，如式(2-9)所示：

$$W(T)=\int_0^{\infty}W_{\lambda}(T)d\lambda=\frac{2\pi^5k^4}{15c^2h^3}T^4=\sigma T^4 \tag{2-9}$$

式(2-9)中，斯蒂芬−玻尔兹曼常数$\sigma=5.6697\times10^{-12}(W\cdot cm^{-2}\cdot K^{-4})$。

另外，维恩位移定律[40]通过求黑体光谱辐射通量密度分布函数对波长的偏微分导数，并令其等于零，可以找到辐射强度峰值所对应的波长，计算得到峰值波长λ_m和黑体绝对温度T之间的关系，如式 (2-10)所示。

$$\lambda_m T=2897.8\mu m \tag{2-10}$$

根据光谱辐射通量密度分布关于波长的偏微分，可知温度的 5 次方和峰值波长的谱能量密度成正比，如式 (2-11)所示。

$$W_{\lambda_m}(T)=bT^5 \tag{2-11}$$

式(2-11)中，$b=1.2862\times10^{-15}(W\cdot cm^{-2}\cdot Sr^{-1}\cdot\mu^{-1}\cdot K^{-5})$。

（2）红外辐射大气窗口：

目标物体发出的红外辐射往往需要经过大气以及光学设计系统的镜头才能到达探测器，此处只考虑大气对于红外辐射的衰减。日常生活中，红外辐射往往除了透过大气还需要通过一些遮挡物才能到达探测器，由于遮挡物性质不同对红外辐射的透过率是不同的，本文仅阐述红外辐射在大气中的透射情况，空气中不同成分对不同波段的红外辐射的吸收率是不同的[41]。其中，水蒸气、臭氧、二氧化碳的影响最大，相关实验表明，红外辐射波段在大气中的透射率（未被吸收衰减的辐射能量与总能量的比值）如图 2-8 所示。

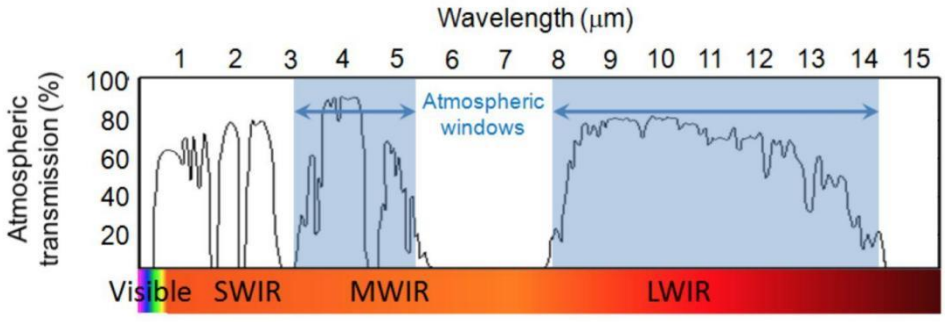


图 2-8 红外波段的大气窗口

其中，红外波段的大气窗口分为三个波段：1至2.5μm SWIR、3至5μm MWIR、8至14μm LWIR，如图 2-9 所示。不同波段因为受到不同成分的吸收干扰，在不同的场景中也各具特点：

（1）SWIR：可使用常规可见光镜头，可透过玻璃成像；可探测1.06μm及1.55μm的激光；可复现可见光图像细节。

（2）MWIR：在高温、潮湿的海洋大气条件下，中波红外的传输优于长波红外[42]，如舰船发动机等高温目标中波红外特征更明显。

（3）LWIR：长波红外在地面大气环境的传输最好[43]；长波红外与室温目标的红外辐射光谱的匹配最好；在烟雾、黑夜等环境下适应性好；非制冷长波红外成像成本较低。

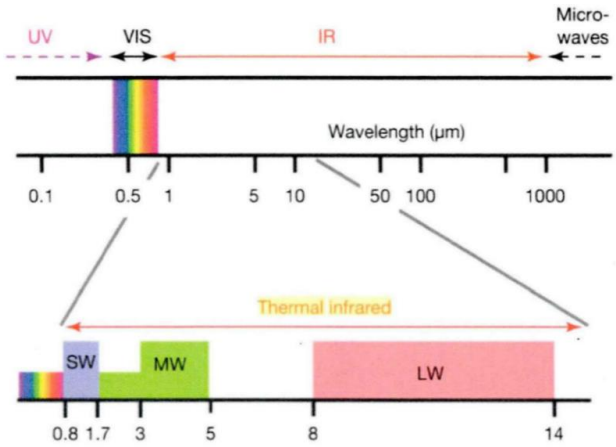


图 2-9 红外的三个大气窗口对应波段

2.3.2 热红外成像原理

一般的红外热成像原理是利用光电设备探测和测量红外辐射，并建立红外辐射与表面温度之间的关系。首先，通过红外探测器、光学热成像透镜等光学装置将被测物的红外辐射能量进行接收[44]，进而会生成红外辐射能量的分布图。然后，能量分布图会进行反射，红外探测器上的光敏元会进行接收，从而得到与物体表面的热分布场相对应的红外热图。红外热图像的一般生成过程如图 2-10 所示。

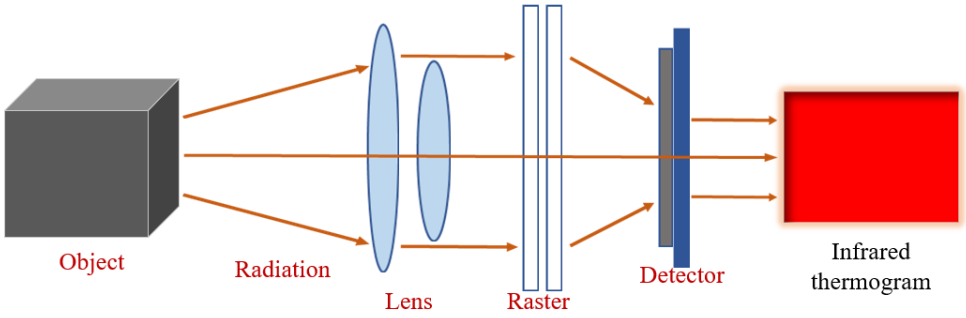


图 2-10 红外热图像的一般生成过程

本文所用的红外图像数据则是使用 FLIR 红外热成像仪[45]采集的。常见的热红外成像系统具体由红外光学系统（包括镜头、光学滤波器等）、探测器、信号处理系统、显示系统、控制系统等组成。图 2-11 为 FLIR 公司的一款长波红外热像仪模块，图 2-12 为 FLIR 热像仪采集的彩色热图像及对应的灰度红外图像。

图 2-11 FLIR 红外热像仪模块

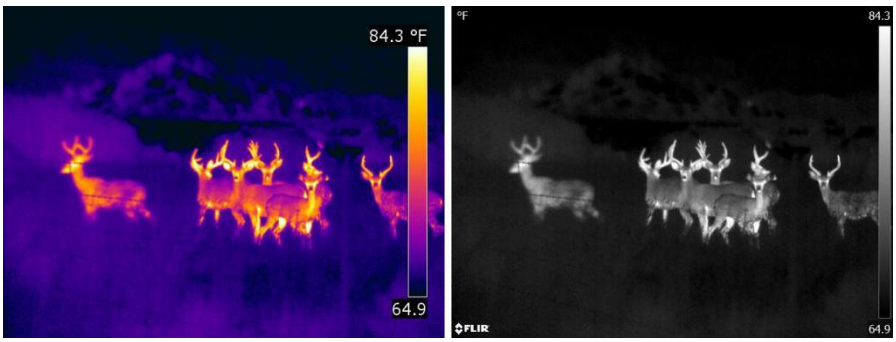


图 2-12 热图像及对应灰度图像

热成像系统是一种利用物体发出的红外辐射来生成图像或者视频的技术。其工作原理[46]如下：

（1）辐射捕捉：热成像系统使用红外传感器来捕捉物体发出的红外辐射，这种辐射是由物体本身的热量产生的。

（2）信号处理：捕捉到的红外辐射信号被转换成电信号，然后将电信号进行模数信号转换（A/D），并经过放大和处理，以便进行后续的分析和图像或视

频生成。

（3）热图像或视频生成：通过信号处理，包括数字信号处理（DSP）算法处理以及数模信号转换（D/A）等，系统会根据不同的红外辐射强度将其转换成对应的亮度或颜色，从而形成热成像图像或者形成标准的视频信号。

（4）显示与分析：最终的热成像图像或视频会被显示在监视器上供人们观察。通过分析，可以得知物体表面的温度分布情况，从而用于识别问题区域或进行热量分析。

总的来说，热成像系统利用物体发出的红外辐射来获取温度信息，并将其转换成可视化的图像或者视频，使人们能够直观地观察物体的温度分布情况。红外成像系统结构如图 2-13 所示。

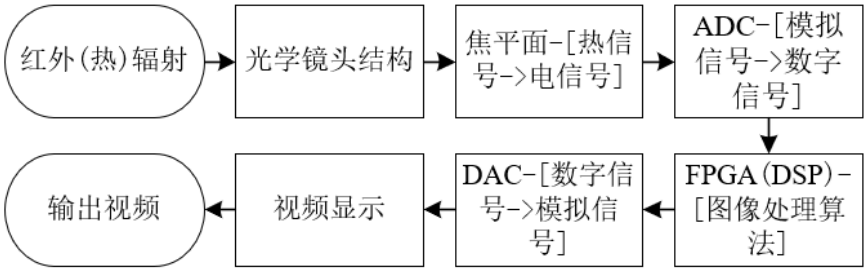


图 2-13 红外成像系统结构

2.4 本章小结

本章主要是阐述了基于深度学习的语义分割技术以及热红外成像技术的两方面的理论内容。先介绍的是语义分割技术理论，其中有语义分割卷积神经网络的基本内容，包括卷积层、池化层、全连接层等几个主要部分；另外，详细阐述了语义分割的基本原理；然后介绍了语义分割常用的评价指标，也是本文中评估语义分割算法性能的主要指标。之后介绍了热红外成像技术相关理论内容，先阐述了热红外技术的基本理论，包括红外辐射理论、红外辐射大气窗口等；然后详细介绍了热红外成像的工作原理，包括对红外热像仪的介绍，以及热红外成像系统的构成。

第 3 章 红外数据集的建立与预处理

3.1 引言

目前开源的语义分割数据集有 Cityscapes、Pascal VOC、Microsoft COCO 等较为常见，不过它们基本都是可见光数据集，而开源的语义分割红外数据集则比较匮乏。本文的研究应用于夜间道路场景，需要夜间道路场景红外数据集。因此，本文选择从 10000 多张 FLIR 目标识别红外数据集中一一筛选出较为清晰且场景差异性较大的夜间道路红外图像，共 1200 张，并自动划分了训练集与验证集。另外，考虑到红外图像低信噪比、低对比度、特征信息不够丰富等特性，会对后续的人工标注标签以及算法训练有一定的影响。因此，本文采用限制对比度自适应直方图均衡化算法（CLAHE）对其批量化增强处理，并与常规的直方图均衡算法作对比试验，通过常见的图像质量评价指标来验证 CLAHE 的有效性。

3.2 语义分割数据集介绍

3.2.1 Microsoft COCO 数据集

Microsoft COCO 数据集是微软开发维护的大型图像数据集，该数据集的任务包括识别（Recognition）、分割（Segmentation）及检测（Detection）。数据集包含目标分割的属性，也常用于语义分割。该数据集有超过 30 万张图，超过 200 万个实例，涵盖 80 种类别，平均每张图含有 5 个目标。图 3-1 为 Microsoft COCO 数据集样张。

图 3-1 Microsoft COCO 数据集样张

3.2.2 Pascal VOC 数据集

Pascal VOC 对于目标检测或分割类型来说属于先驱者的地位，对于现在的研究者来说比较重要的两个年份的数据集是 Pascal VOC 2007 与 Pascal VOC 2012。目前常用到的 Pascal VOC 2012 数据集在语义分割部分涵盖了 20 种物体类别，类别中包含人类、交通工具、生活物品等大类，语义标签图有 2913 张，其中训练集有 1464 张，验证集有 1449 张，共有 3422 个物体[47]。图 3-2 为 Pascal VOC 2012 数据集示例。

图 3-2 Pascal VOC 2012 数据集示例

3.2.3 Cityscapes 数据集

Cityscapes 数据集是一个专为自动驾驶和计算机视觉领域设计的大型语义分割数据集。它包括 8 个类别（平面、人类、车辆、构造、对象、自然、天空和虚空）的 30 个类提供语义实例和像素注释。

数据集由大约 5000 个精细注释图像和 20000 个粗糙注释图像组成，分别在 50 个城市中采集。其中，涵盖多种天气条件、时间（白天和黄昏）、季节变化和地理区域，能够确保模型对各种环境有较好的泛化能力。如图 3-3 则是 Cityscapes 数据集的部分图像示例。

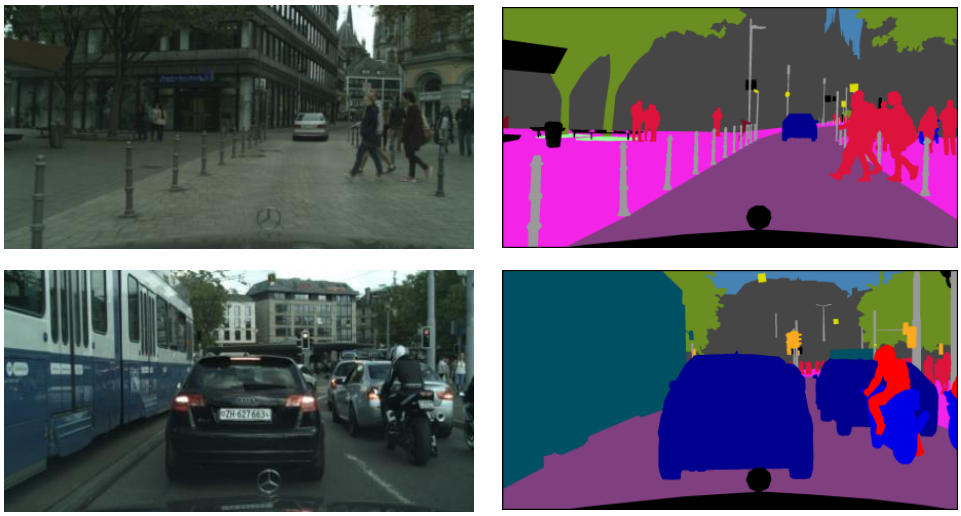


图 3-3 Cityscapes 数据集示例

3.3 红外图像的增强处理

可见光图像增强算法较为常见，算法增强可以有效提高图像质量，以满足使用要求。红外图像具有的低信噪比、低对比度、特征信息不丰富等特性，不利于手工制作标签和后续的算法训练。因此，使用合适的增强算法很有必要，能够提高红外图像质量的同时抑制噪声的引入。

3.3.1 CLAHE 图像增强算法

常规的直方图均衡化算法（Histogram Equalization，HE）[48]是采用基于整幅图像的增强算法，通过对原始图像的灰度直方图进行变换，使得处理后图像的灰度级分布更加均匀。但是会造成两个主要问题：一是由于全局性导致图像的整体灰度过暗或过亮，局部细节会被淹没；二是增加了背景噪声，图像质量不太理想。为了解决问题一，相关研究者提出了局部直方图均衡化的方法，即将图像被划分为多个小块或窗口，每个子区域独立进行直方图均衡化处理，这样可以更好地保留图像的局部对比度和细节；为了解决问题二，相关研究者提出对图像对比度进行限制的相关方法，能够有效克服全局或局部直方图均衡化过度增强背景噪声的问题。

本文采用对比度受限的自适应直方图均衡化（Contrast Limited Adaptive Histogram Equalization，CLAHE）对红外图像进行增强处理，它是结合了以上两

个关键问题的解决方法，适用于局部对比度提升和噪声抑制，而且该图像增强算法非常快速高效。

CLAHE 算法图像处理流程[49]：

（1）首先将输入图像分割成多个小块。使用4×4或更大尺寸的矩形区域对图像进行分块处理；

（2）对每个小块独立地计算其灰度直方图，并应用常规的直方图均衡化方法来增加该小块内部的对比度；

（3）在进行直方图拉伸时，为了避免由于过度拉伸而导致的噪声放大，CLAHE 会对直方图的累积分布函数（Cumulative Distribution Function，CDF）进行饱和处理，即设定一个上限值限制像素增益的最大幅度；

（4）为了消除相邻块之间的边界效应，通常在合并各个块的结果之前，会采用重叠区域并对其进行邻域平均操作，确保过渡平滑；

（5）分块后的结果通常需要通过双线性插值上采样到原始图像大小，以保持图像分辨率不变且确保边缘细节自然连续；

逐个访问图像中的每一个像素以及其邻域内的所有像素，并计算对应的直方图以及累积分布函数（CDF），这本身就涉及到大量的循环和内存访问操作，这种方式非常耗时。CLAHE 通过使用线性插值方法[50]，在处理效果上得到保证的同时节约了时间成本。图 3-4 为双线性插值方法的具体展现，将 4 个角的位置（红色阴影处）进行直接计算直方图以及累积分布函数，边缘位置（绿色阴影处）的图像块进行线性插值，而蓝色阴影处的像素则采用双线性插值。使用这种双线性插值方法能提供更平滑、更精确的结果，而且计算量较小。具体的双线性插值计算如下：

已知点$Q_{11}=(x_1,y_1)$，$Q_{12}=(x_1,y_2)$，$Q_{21}=(x_2,y_1)$，$Q_{22}=(x_2,y_2)$这四个点处的灰度值$f(Q)$，而$f(P)$为双线性插值求解$P=(x,y)$的灰度值：

$$f(R_1)\approx\frac{x_2-x}{x_2-x_1}f(Q_{11})+\frac{x-x_1}{x_2-x_1}f(Q_{21})\quad where\quad R_1=(x,y_1) \tag{3-1}$$

$$f(R_1)\approx\frac{x_2-x}{x_2-x_1}f(Q_{12})+\frac{x-x_1}{x_2-x_1}f(Q_{22})\quad where\quad R_2=(x,y_2) \tag{3-2}$$

$$f(Q) \approx \frac{y_2 - y}{y_2 - y_1} f(R_1) + \frac{y - y_1}{y_2 - y_1} f(R_2) \tag{3-3}$$

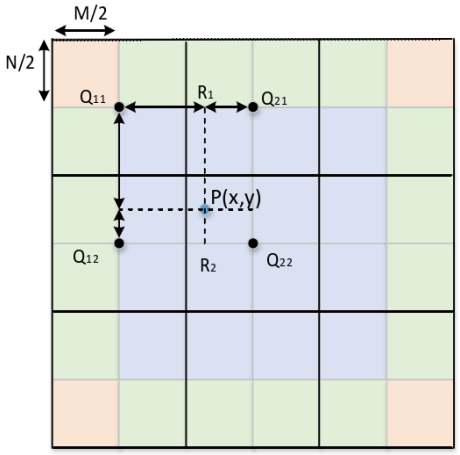


图 3-4 双线性插值方法

3.3.2 红外图像预处理结果

本文对通过 CLAHE 算法增强处理后的红外图像做了一些对比实验以及主、客观评价。将常见的几个图像评价指标作为客观评价的主要参考，其中包括标准差（Standard Deviation，SD）、峰值信噪比（PSNR）、平均梯度（Average Gradient，AG）[51]、信息熵（Information Entropy，IE）、结构相似性（Structural Similarity，SSIM）。

PSNR 用于衡量原始图像与压缩或处理后图像之间差异的一个量化指标。PSNR 值是基于均方误差（MSE）计算[52]得出的，值越大，说明处理后的图像相对于原始图像的失真越小，图像质量越好。AG 和 SD 都能够评估图像的清晰程度，前者通常用来反映图像边缘信息的丰富程度以及图像的整体对比度，较高的平均梯度值意味着图像具有更多的边缘细节，图像会更清晰。后者是对图像像素值分布离散程度的度量，一定程度上能够体现图像的变化剧烈程度或者灰度级的多样性。标准差较大意味着图像有更高的动态范围和更多视觉细节，图像也就会更清晰。

IE 则用来衡量图像所含信息量的多少，信息熵越高则说明图像包含的信息种类越多，这在一定程度上代表图像包含的信息量更大。SSIM 是衡量两张图像相似性的指标，越接近 1，结构性失真就越小[53]。下图 3-5 由左到右每一列分别

为红外原始图像、HE 算法处理图像、CLAHE 算法处理图像以及对应的直方图。

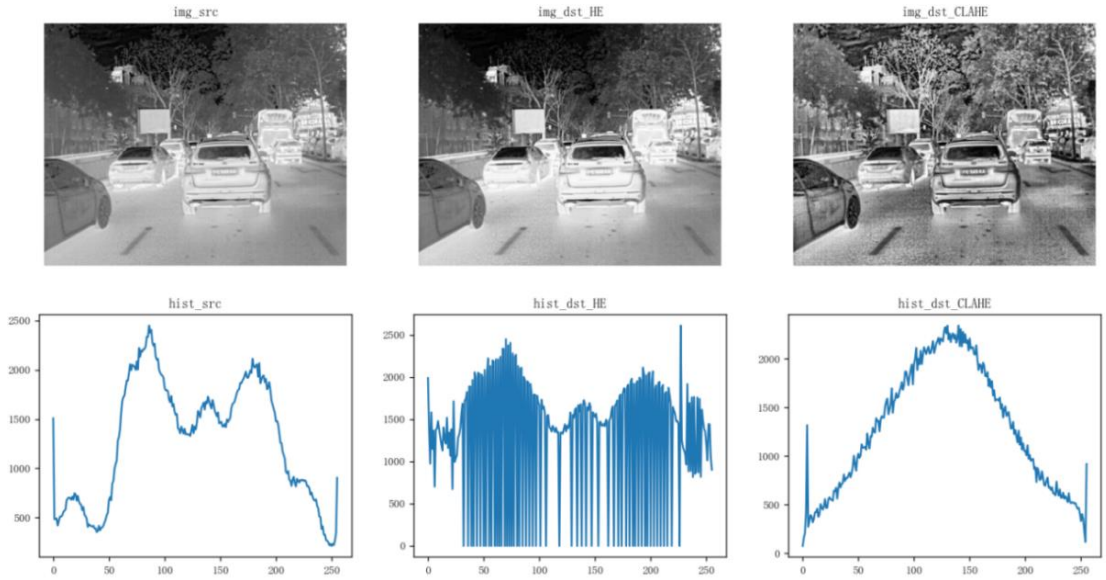


图 3-5 原始红外图、HE 处理图、CLAHE 处理图及对应直方图

主观评价：明显的看到图 3-5 中红外原始图像的直方图生成了多个峰值，这就表明像素在这些灰度区间的分布更加集中，不均匀的像素分布导致图像的对比度下降，图像中的各类别景物细节特征不太明显。使用全局直方图均衡化（HE）算法处理红外图像，确实能够使图像的灰度级分布更加均匀，从而提高整体对比度。然而，这种全局处理方法可能导致图像局部区域的细节信息丢失，因为所有像素都受到相同程度的增强，无法针对图像中的不同特征区域进行自适应调整。相比之下，本文采用 CLAHE 算法处理的红外图像，直方图的峰值变得更加平滑，整体灰度级分布更加均匀，同时保留了更多局部细节信息，这非常利于手工标注标签，提高效率的同时减少了标注失误，精细的标签文件也有利于后续算法的训练。

表 3-1 红外图像各评价指标数据对比

评价指标	红外原图	HE 处理图	CLAHE 算法处理图
PSNR	/	19.23	26.19
AG	8.24	13.40	12.81
IE	7.15	7.56	7.99
SD	0.14	0.18	0.20
SSIM	/	0.74	0.82

客观评价：表 3-1 中各个图像评价指标数据是从 300 组红外图像计算所得到的均值。从表中数据对比来看，CLAHE 算法处理的红外图像的 PSNR 指标是要优于 HE 算法，说明图像失真较少。SSIM 指标也是优于 HE 算法，说明相对于原始红外图像在结构上相似性更高，结构失真较少。另外，AG、IE 和 SD 指标均优于原始红外图像，说明了图像清晰度更高，图像特征信息更丰富。因此，也验证了原始红外图像在本文算法（CLAHE）预处理后的可行性与有效性，图像更清晰，视觉效果得到了很大改善。

3.4 建立红外数据集

3.4.1 数据类别与标注

本文使用的红外图像数据是从 FLIR 目标识别红外道路数据集中筛选出来的，筛选出 1200 张夜间红外图像。通过 CLAHE 算法的增强预处理后，图像质量得到改善。将预处理后的夜间道路场景红外图像中最常见到的 4 类景物（建筑、汽车、人和植被）进行手工标注，使用的是 Labelme 软件来标注。首先，将图像中 4 种类别的景物轮廓标画出来，为了生成彩色标签图，于是把所标画的 4 类景物以及背景类给予不同颜色。在手工标注时，这 4 个类别都不区分子类别，例如植被这一类别包括树木、草坪等子类别，其他景物统一标记为背景类。另外，背景和 4 个类别的类号用数字 0~4 来分别表示，更加利于区分。表 3-2 为各个类别具体标注信息。

表 3-2 红外数据集的标注信息

类别	子类别	类号	颜色	RGB 值
车辆	轿车、卡车	1	红	255 0 0
植被	树木、草坪	2	绿	0 255 0
人类	人类	3	黄	255 255 0
建筑物	房屋、墙壁	4	蓝	0 0 255
背景	其他背景	0	黑	255 255 255

人工精细标注标签是非常耗时的，但是精细标注的标签图利于算法的训练，图 3-6 为本文红外数据集的示例。

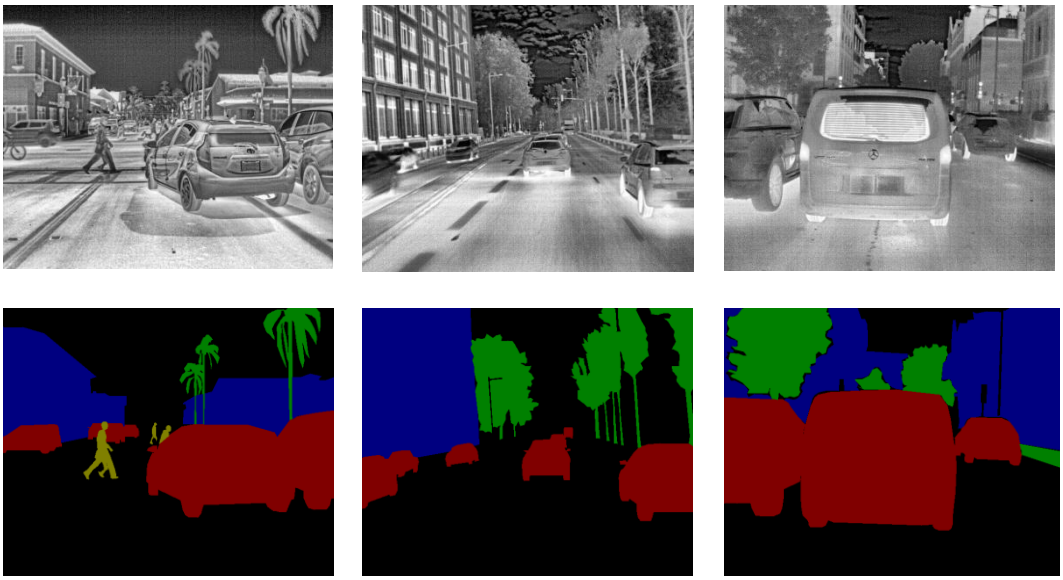


图 3-6 红外数据集示例

3.4.2 数据集扩充

本文使用的红外数据集包含的图像数量为 1200 张，在深度学习算法中进行训练数量上可能还不够充分，可能会出现训练过拟合现象。因此，本文采取数据增强的方法对红外数据集进行扩充处理，其中包含旋转、翻转、平移、对比度变换等数据增强方法，通过这些方法实现了对红外数据集的扩充，图像数量是原数据集的 5 倍。本文数据增强方法有以下几种：

（1）旋转：自由选择角度，对图像进行旋转；

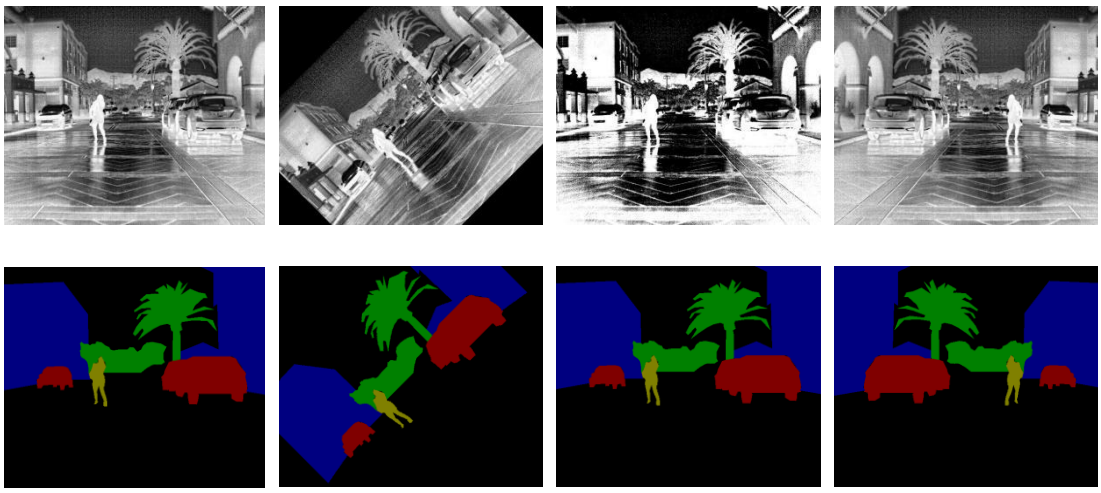
（2）翻转：水平或垂直方向进行翻转图像；

（3）对比度变换：设定对比度因子，改变图像的对比度；

（4）平移：通过随机或指定平移的方向和步长，沿着水平或竖直方向平移，改变图像内容位置；

（5）随机噪声：对图像的每个像素值进行随机扰动，导致原始图像信息失真。

本文使用的几种数据增强方法都是较为常见的，在数据集图像数量匮乏的情况下，数据增强则是较好的处理方式。图 3-7 为数据增强示例。



(a)原数据　(b)旋转　(c)对比度变换　(d)翻转

图 3-7 数据增强示例

3.4.3 数据集划分与统计

红外数据集共有 1200 张图像，在设定训练集与验证集的比例后，采用随机划分方法将数据集划分为 900 张图像的训练集与 300 张图像的验证集。在数据集扩充之后，图像数量达到了原有的 5 倍，在这里验证集图像数量没有扩充，仍然保持 300 张，有效保证训练集部分的扩充。

为了较直观的了解本文红外数据集（Our Infrared）的类别分布情况，本文从 1200 张图像中统计出了每个类别数据集所能占到的比例，同时也统计了 Cityscapes 数据集中各个类别的占比情况，与本文数据集进行对比。从表 3-3 可以得知，两个数据集中的人、汽车、建筑和植被这 4 个类别分布较为类似，“人”的占有比例都是最少的，其次就是“汽车”，“建筑”和“植被”占比都较大。因此。通过对比可以看出本文的红外数据集是比较符合道路场景的客观性。

表 3-3 本文数据集和 Cityscapes 数据集的类别像素占比

数据集/类别	人类	车辆	建筑物	植被	背景
Our Infrared	1.10%	5.67%	11.58%	16.68%	65.42%
Cityscapes	1.21%	6.63%	20.68%	14.09%	57.38%

3.5 本章小结

本章节主要阐述了本文红外数据集的建立及预处理。首先简单介绍了目前比较常见的语义分割数据集；然后针对红外图像的特性，通过对比度受限的自适应直方图均衡算法（CLAHE）实现了对红外图像的增强处理，与全局直方图均衡化算法（HE）做了实验对比，并使用图像评价指标来验证有效性；之后，介绍了红外数据集建立的规则、划分，以及通过数据增强对训练集进行扩充，最后统计出数据集中各个类别的像素占比情况。

第 4 章 基于 DeepLabV3+的轻量化语义分割算法

4.1 引言

在第三章中建立了本文必需的红外数据集，为后续的算法实验打下基础。目前，面向道路场景的语义分割算法基本用于自动驾驶领域。常见的语义分割算法在进行场景理解时，由于计算及参数量很大，运算时间就随之较长，而汽车行驶过程中速度较快，如果遇到复杂的交通路况时，汽车自动驾驶决策系统就无法及时的识别判断景物类别，将造成一定的安全隐患。因此，本文选择在性能良好的DeepLabV3+算法的基础上进行优化，达到精确性和实时性兼顾的目的，以满足汽车自动驾驶的需求。针对 DeepLabV3+算法的计算量较大，细节信息丢失较多，获取上下文特征信息以及多尺度特征信息的能力不足的缺点，对其骨干网络、特征提取以及空洞空间金字塔池化（ASPP）三点进行改进。

4.2 DeepLabV3+算法原理

DeepLabV3+算法是编码-解码结构。DeepLabV3+网络结构如图 4-1 所示。

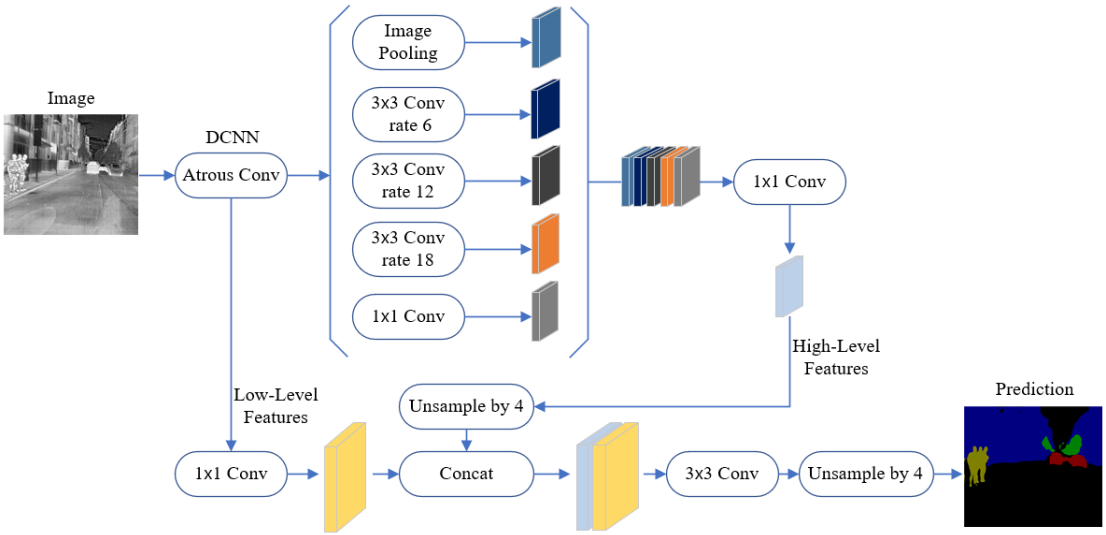


图 4-1 DeepLabV3+网络结构

编码器部分通过骨干网络 Xception 提取图像特征信息，接着利用 ASPP 模块中不同速率的并行空洞卷积获取高层语义信息，其中空洞速率分别为 6、12、

18，并通过1×1卷积进行通道压缩；在解码器部分，将主干网络所获取的低级特征与经过 4 倍双线性插值上采样之后的高级特征进行融合。然后，利用3×3卷积恢复空间信息，最后通过上采样恢复到原图大小，并输出最终的预测结果。

4.3 DeepLabV3+算法的改进

DeepLabV3+算法性能较为优秀，是目前众多研究者们使用的经典语义分割算法，但是应用于自动驾驶场景时，还是需要相关改进：首先，为了降低网络参数量，将骨干网络 Xception 替换为更加轻量化的 MobileNetV2；然后，为了网络特征提取到更加丰富的上下文特征信息，使用密集连接方式的空洞空间金字塔池化（ASPP）模块，即本文的 D-ASPP 模块，并且考虑到密集连接会导致计算量增加，于是将空洞卷积替换为空洞可分离卷积；最后，为了网络在特征提取中减少细节特征信息的丢失，分别在编码和解码阶段的特征提取中引入通道注意力模块和空间注意力模块。改进后的 DeepLabV3+网络结构如图 4-2 所示。

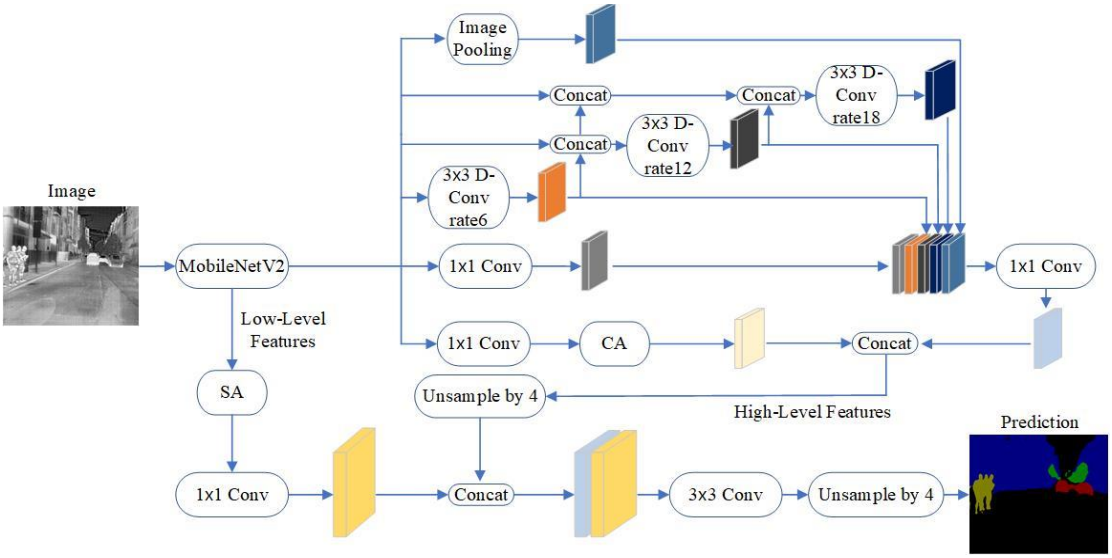


图 4-2 改进后的 DeepLabV3+网络结构

4.3.1 骨干网络 MobileNetV2

MobileNetV2 相较于V1 主要有两个改进点，分别为线性瓶颈（Linear Bottleneck）和倒残差结构（Inverted Residuals）[54]。在传统的瓶颈结构中，通常会先通过1×1卷积进行降维（通道数减少），然后执行3×3卷积或者其他形式的卷积

操作以捕获特征，最后再通过1×1卷积进行升维（恢复通道数）。而MobileNetV2提出的“Linear Bottleneck”则有所不同：它首先通过1×1卷积进行通道扩展（即升维）然后应用Relu6 激活函数来限制输出范围，这有助于防止激活值过大或过小的问题，并且有利于硬件加速；然后，使用深度可分离卷积进行特征提取，包括深度卷积和逐点卷积两部分，进一步降低计算量；最后，不同于传统瓶颈结构在此处的1×1卷积升维，MobileNetV2 的“Linear Bottleneck”选择直接将经过Relu6 激活后的结果传递给下一层，也就是说，在这个模块的末尾没有采用非线性激活函数。

这样的结构设计能够保留更多的信息流经网络，同时保持较低的计算成本和参数数量，有助于提高模型的效率和性能。模块结构如图 4-3 所示。

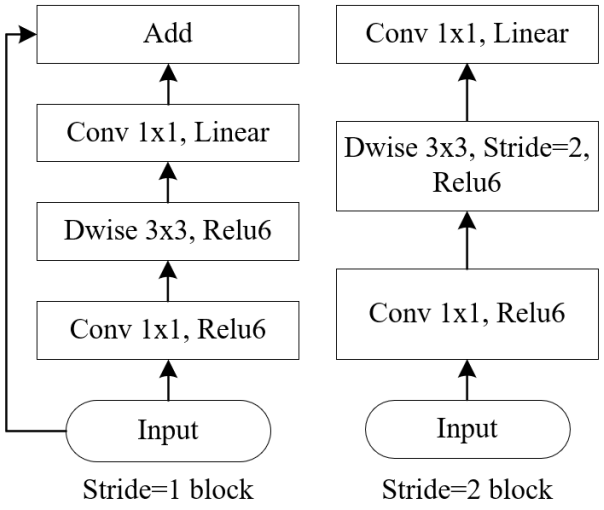


图 4-3 MobileNetV2 的 Linear Bottleneck 结构

Bottleneck Residual Block（ResNet论文中的）是中间窄两头胖，残差网络ResNet有两种基本结构如图 4-4，左图是常规的ResNet层级中的网络结构组件，右图是更深的称为瓶颈（Bottleneck）结构。传统的残差块中，通常采用先降维、执行卷积操作、再升维的设计方式。而MobileNetV2 中的倒残差结构[55]则是反其道而行之。首先，通过一个1×1的卷积层将输入通道数进行扩展（乘以扩张因子k），这一过程增加了模型的表达能力，但不增加计算复杂度的空间维度；接着，对经过宽度扩展后的特征图采用深度可分离卷积，即先进行逐通道的深度卷积操作，然后是一个1×1的逐点卷积。这种结构能够极大地减少计算量，同时保持有效的特征提取能力；然后，并没有像常规残差块那样立即恢复到原始通道数，而是直

接将经过Relu激活函数处理过的深度可分离卷积结果传递给下一层；为了与下一层匹配，还需要一个最后的1×1卷积层来调整通道数。

通过这种倒残差结构设计，MobileNetV2 能够在保持较高性能的同时，大幅度降低模型的计算复杂性和参数数量，非常适合部署在资源有限的移动设备和嵌入式系统上。因此，MobileNetV2 网络非常符合道路场景语义分割的实时性要求。MobileNetV2 网络的Bottleneck Residual Block如图 4−5 所示。

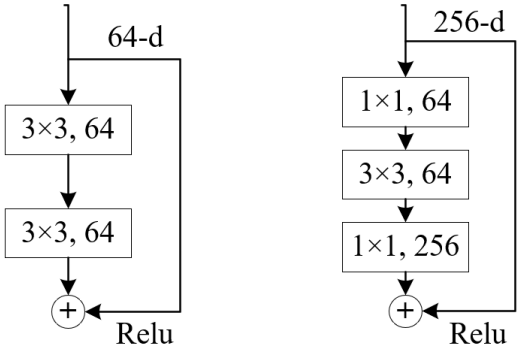


图 4-4 ResNet 的两种基本结构

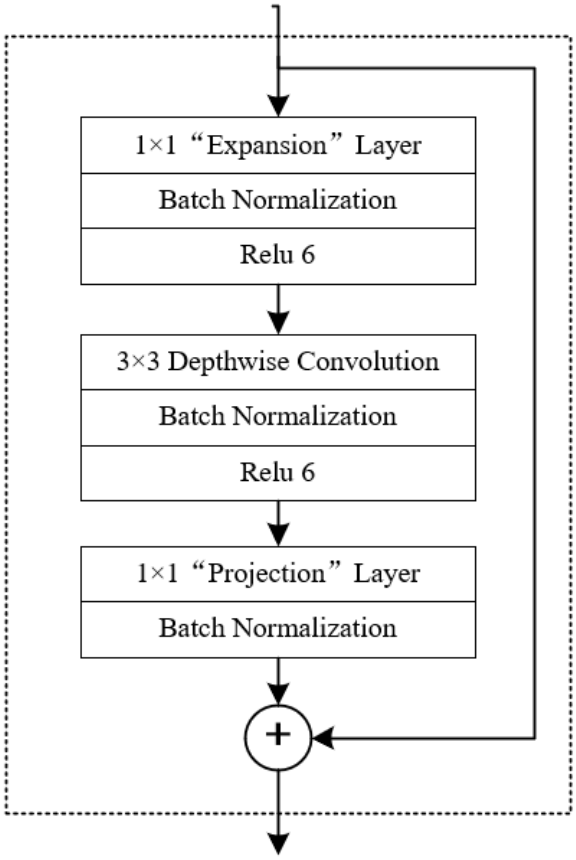


图 4-5 MobileNetV2 网络的 Bottleneck Residual Block

4.3.2 基于空洞可分离卷积的 D-ASPP

DeepLabV3+网络结构中的 ASPP 模块，主要思路是通过不同速率的并行空洞卷积来获取特征信息，实现覆盖多尺度的感受域。本文引入密集连接方式的 ASPP[56]，即 D-ASPP，并将卷积替换为深度可分离卷积，在获取更密集的特征金字塔和更大感受野的同时，又减少了 ASPP 模块的计算量与参数量，更加轻量化。

所谓空洞，即在卷积核中注入空洞进行膨胀卷积，根据需求设置不同的空洞速率（Dilation Rate）来进行卷积计算[57]。常规标准卷积因为没有空洞，即一个 3×3 的标准卷积核在应用空洞速率为 1 的情况下，其实际覆盖的区域为 3×3；若将空洞速率设置为 2，则卷积核在移动过程中每两个元素间都会跳过一个像素，从而实际上覆盖了 5×5 的区域，感受野得到扩大。以此类推，随着空洞率的增大，卷积核能够感知到更大范围的输入特征，而无需增加卷积核的实际大小或减少输出特征图的分辨率，这对于图像分割、目标检测等任务具有重要意义，可以捕捉到更远距离的上下文信息。图 4-6 中，左图则为标准卷积，右图为空洞率为 2 的空洞卷积，卷积后的感受野为 5。

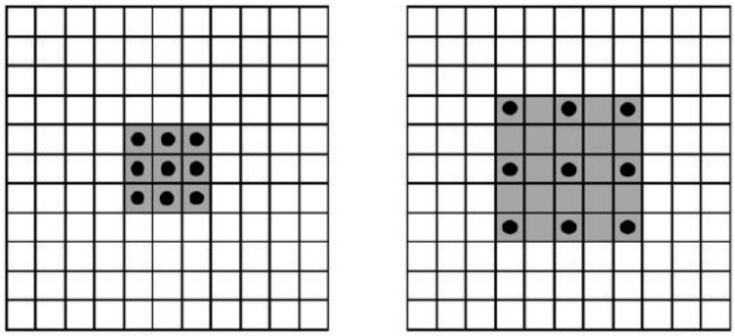


图 4-6 标准卷积和空洞率为 2 的空洞卷积

本文使用深度可分离卷积替换标准卷积来进行卷积核的膨胀，而深度可分离卷积[58]的实现分两步：第一步，则是用三个卷积对三个通道分别做卷积，即逐层卷积（Depthwise Convolution），并输出三个通道的属性；第二步，用 $1\times1\times3$ 的卷积核，即逐点卷积（Pointwise Convolution），对三个通道继续做卷积。此时的输出就和标准卷积一样，为 $8\times8\times1$。具体实现过程如图 4-7 所示。

如果要提取更多的属性，那么只要设计更多的 $1\times1\times3$ 卷积核就可以实现。如果仅仅是提取一个属性，深度可分离卷积的方法不如标准卷积，但是随着要提取

的属性越来越多，深度可分离卷积就能够节省更多的参数。

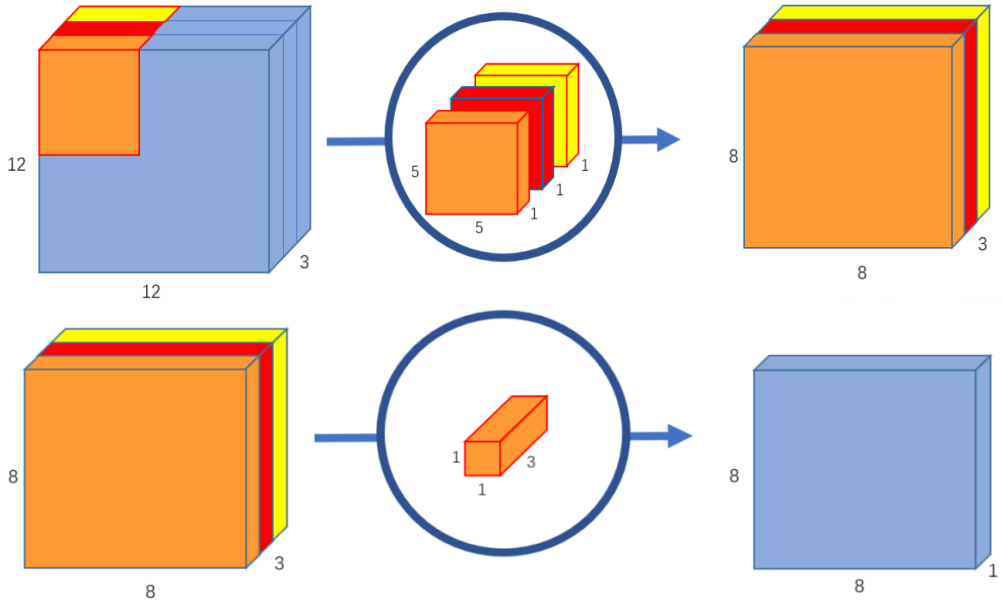


图 4-7 深度可分离卷积的实现过程

依靠空洞可分离卷积的特性，即在保持特征图分辨率不变的情况下大幅增加网络的感受野。那么，在输出信号 y 和输入信号 x 都属于一维向量时，空洞卷积的处理具体为：

$$y(i)=\sum_{h=1}^{H}x(i+r\cdot h)w(h) \tag{4-1}$$

式(4-1)中，r 是空洞速率；$w(h)$ 是滤波器处于第 i 个位置的参数；H 是滤波器的尺寸。空洞卷积相当于卷积核元素之间插入一定数量 $(r-1)$ 的零值，随着空洞速率的增大，卷积核在处理输入特征时跳跃的距离也会增加，即不连续地采样输入特征。那么，卷积核大小为 h、空洞速率为 r 的空洞卷积的感受野大小具体为：

$$M=(r-1)\times(h-1)+h \tag{4-2}$$

另外，通过上下层级联方式的空洞卷积层，其感受野范围大小具体为：

$$M=M_1+M_2-1 \tag{4-3}$$

式(4-3)中，M_1、M_2 分别为上下级联的两个空洞卷积层感受野。因此，D-ASPP 模块中各个空洞可分离卷积支路感受野的变化如图 4-8 所示。

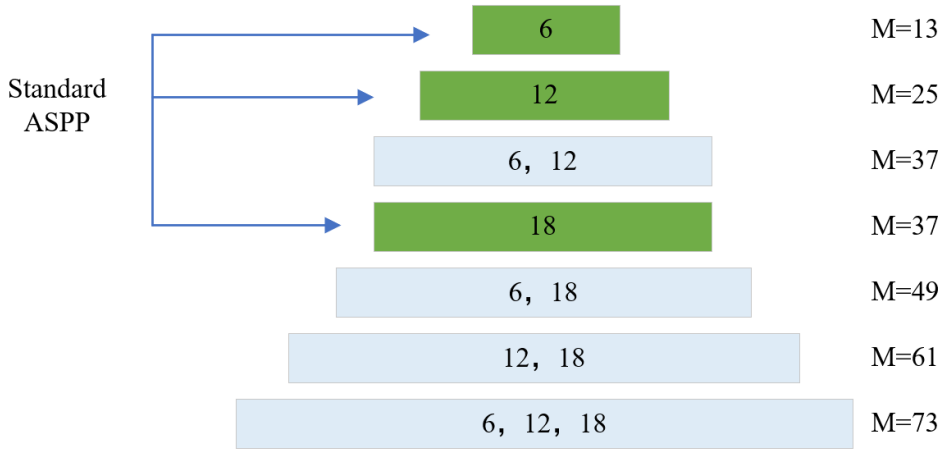


图 4-8 D-ASPP 感受野的大小

以空洞速率（6，12，18）为例，根据式 4-3，若设 M_h^r 是卷积核大小为 h、空洞速率为 r 的卷积感受野，那么原有的 ASPP 的感受野最大值具体为：

$$M_{\max} = \max(M_3^6, M_3^{12}, M_3^{18}) = M_3^{18} = 37 \tag{4-4}$$

D-ASPP 的感受野最大值具体为:

$$M_{\max} = M_3^6 + M_3^{12} + M_3^{18} - 2 = 73 \tag{4-5}$$

本文则是采取不同的空洞速率（6，12，18）来构建 D-ASPP。通过密集连接的方式，加强了各支路的相关性。由于卷积空洞速率的层层增加，上级联层的卷积特征能够和下级联层的卷积特征相互补充，实现更加密集化的像素采样，高感受域的特征也能拥有低感受域的特征信息，从而获得更密集的特征金字塔[59]。

4.3.3 注意力机制

DeepLabV3+算法在语义分割中的综合表现较为优秀，但是在提取细节特征信息以及图像类别之间的边缘分割效果还有待提升。因此，本文在算法的编码阶段和解码阶段分别引入通道注意力模块以及空间注意力模块。

通道注意力（Channel Attention，CA）是通过考虑输入特征图中不同通道之间的相关性来自适应地调整每个通道的权重。它的目标是将更多的注意力放在对当前任务更有贡献的通道上。相比于普通的卷积操作，通道注意力机制可以使模型更加关注重要的特征，提高模型的表征能力。通道注意力机制[60]通常通过以下步骤实现：

（1）输入特征图通过全局池化操作（如平均池化 *AvgPool* 和最大池化 *MaxPool* ）得到全局特征描述；

（2）全局特征通过一个或多个全连接层进行映射，即将两个特征描述送入多层感知（Multilayer Perceptron，*MLP*）中，得到注意力向量；

（3）注意力向量通过激活函数（如 Sigmoid）进行归一化，以控制每个通道的权重；

（4）权重与输入特征图逐元素相乘，得到加权后的特征图。具体表示为：

$$F_{avg} = MLP(AvgPool(F)) \tag{4-6}$$

$$F_{\max} = MLP(MaxPool(F)) \tag{4-7}$$

$$F_m = F_{avg} + F_{\max} \tag{4-8}$$

$$F_c = \sigma(F_m) \cdot F \tag{4-9}$$

其中，F 为输入；F_{avg} 为平均池化后进行多层映射的特征图；$F_{\max}$ 为最大池化后进行多层映射的特征图；F_m 为将多层映射特征进行加权操作；F_c 表示加入通道注意力后的特征图；图 4-9 则为 CA 模块结构，其中 F_{cs} 为步骤（3）处理的特征图。

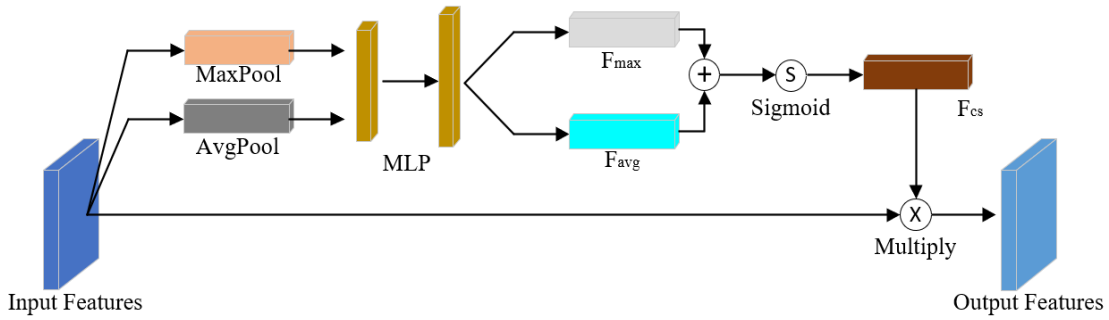


图 4-9 通道注意力模块

DeepLabV3+算法解码阶段提取的低级特征直接作为解码的输入，会引入较多的背景特征，影响分割效果。空间注意力模块能够过滤背景信息，关注于前景信息，聚焦目标特征信息，一定程度改善边缘分割效果。

空间注意力（Spatial Attention，SA）更加关注空间位置信息，是通过考虑输入特征图中不同空间位置的重要性来自适应地调整每个位置的权重。它的目标是放大对关键目标或区域的关注，以提高模型对空间位置的感知能力。空间注意力

机制[61]通常通过以下步骤实现：

（1）输入特征图在通道维度上进行平均池化和最大池化操作，分别得到对应特征图；

（2）使用 Concat 操作将两个不同表示的特征图进行拼接；

（3）通过卷积操作和 Sigmoid 函数得到注意力权重图；

（4）注意力权重图与输入特征图进行逐元素相乘，得到加权后的特征图。

具体如下式所示：

$$F_{avg}' = Avg(F') \tag{4-10}$$

$$F_{\max}' = Avg(F') \tag{4-11}$$

$$F_{ss} = \sigma(f[F_{avg}'; F_{\max}']) \tag{4-12}$$

其中，F'为输入特征图；F_{avg}'为平均池化后的特征图；$F_{\max}'$为最大池化后的特征图；F_{ss}为注意力权重图；$f[F_{avg}'; F_{\max}']$为卷积操作。SA 模块结构如图 4-10 所示。

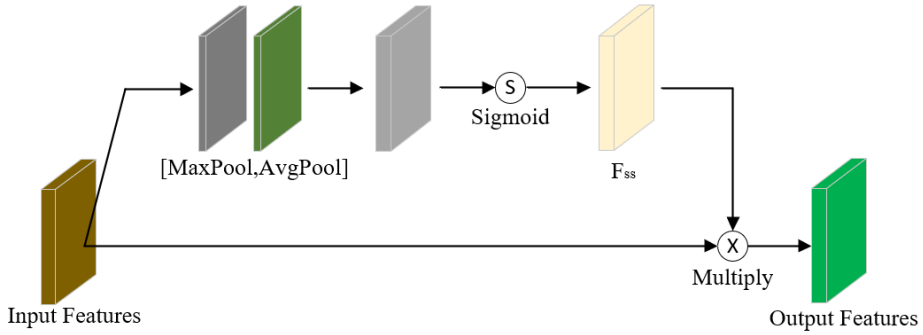


图 4-10 空间注意力模块

4.4 实验结果与分析

本文旨在实现面向道路场景的语义分割算法的优化，基于对 DeepLabV3+算法结构的改进，以达到算法的实时性与精确性。本节则是对本文提出的算法进行实验与结果分析，使用本文建立的夜间红外数据集以及 Pascal VOC 2012 数据集进行训练与验证，并与常见的语义分割算法作对比实验及结果分析。然后对几个改进点进行了消融实验，从而验证改进点的有效性和可行性。另外，针对了 D-ASPP 模块进行相关实验，包括分别基于空洞卷积和空洞可分离卷积的 D-ASPP

模块对比实验、基于不同空洞速率构建 D-ASPP 的对比实验等。

4.4.1 实验配置与参数

本实验是在 Windows10 系统上基于 Pytorch 框架来实现的，采用 CUDA 进行处理，GPU 型号：NVIDIA RTX2080Ti，显卡内存 11G。数据集为本文建立的红外数据集，训练集图像统一为 640×512 分辨率。在对比实验中同时使用了 Pascal VOC 2012 数据集。

在算法训练中常规的参数设置：数据批量大小（Batch Size，BS）设置为 4，学习率初始化为 0.001，学习率下降方式为余弦退火（Cosine Annealing，COS）。选择随机梯度下降法（Stochastic Gradient Descent，SGD）作为算法的优化器，优化器参数（Momentum）设为 0.9。损失函数为交叉熵函数。

4.4.2 损失函数

整个网络采用逐像素的交叉熵损失作为损失函数[62]，交叉熵的值越小，模型预测效果就越好。交叉熵涉及到计算每个类别的概率，所以交叉熵几乎每次都和 Sigmoid 函数一起出现，Sigmoid 函数将线性回归的结果（$-\infty,+\infty$）映射到（$0,1$）范围内。分类器（Softmax）的输出如式(4-13)：

$$p_k(x)=\frac{\exp(\alpha_k(x))}{\sum_{k=1}^{K}\exp(\alpha_k(x))} \tag{4-13}$$

其中，x 表示像素位置；K 为总的类别个数；$\alpha_k(x)$ 为分类器输出的像素 x 对应的第 k 个通道的值；$p_k(x)$ 为像素 x 属于第 k 类的概率。因此，算法的损失如式(4-14)：

$$E=\sum\nolimits_x w_n\log(p_n(x)) \tag{4-14}$$

其中，w_n 为类别 n 的损失权重；$p_n(x)$ 为像素 x 属于真实类别 n 的概率。

4.4.3 D-ASPP 模块对比实验

不同空洞速率构建 D-ASPP 的对比实验。此实验是将标准卷积替换为深度可分离卷积后所作的对比。由表 4-1 中可以看出密集连接的 ASPP 在以（6，12，

18，24）不同空洞速率构建时，算法的平均交并比是相对最好，但是在预测单幅图像的速度较慢，不利于实时性场景的应用。本文选择以（6，12，18）不同空洞速率构建 ASPP，是均衡算法精确度和实时性之后的最佳选择。

表 4-1 不同结构的 D-ASPP 指标数据对比

结构	最大尺度	MIoU/%	时间/ms
ASPP(6，12，18)	37	78.71	31.56
D-ASPP（6，12）	37	76.88	21.71
D-ASPP（6，12，18）	**73**	**80.01**	**32.68**
D-ASPP（6，12，18，24）	122	80.13	48.21
D-ASPP（6，12，18，24，30）	173	79.90	63.10

本文的 D-ASPP 模块是基于空洞可分离卷积，因此，本文将两种不同类型卷积下的 D-ASPP 的算法性能作了对比实验。如表 4-2 所示，可以看出深度可分离卷积的 D-ASPP 的处理耗时降低了 39.3%。基于标准卷积的 D-ASPP，其平均交并比较好一些，算法精确性更好，但是由于计算量相比基于深度可分离卷积的 D-ASPP 要大，单幅图像预测处理时间比较长，满足不了道路场景的实时性语义分割的要求。

表 4-2 基于不同卷积类型的 D-ASPP 数据对比

卷积类型	MIoU/%	时间/ms
标准卷积	80.23	54.12
深度可分离卷积	**80.01**	**32.86**

4.4.4 红外数据集的实验结果与分析

将本文建立的红外数据集在改进的 DeepLabV3+算法上进行训练与验证，针对骨干网络、ASPP 模块及编解码特征提取阶段三点进行优化改进。分别在改进前后的 DeepLabV3+算法训练与验证。

训练损失曲线和验证损失曲线如图 4-11、4-12，由曲线对比可以看出改进后算法的训练损失和验证损失均小于改进之前，损失值分别为 0.245 和 0.359，曲线也较为平滑，说明改进后的算法精度有所提高，其分割结果与真实标签更加相

似，算法性能有所改善。从图 4-13 和 4-14 的平均交并比曲线也可以看出算法改进前后的精度区别，改进后的算法平均交并比达到 80.84%，改进前则为 78.71%。

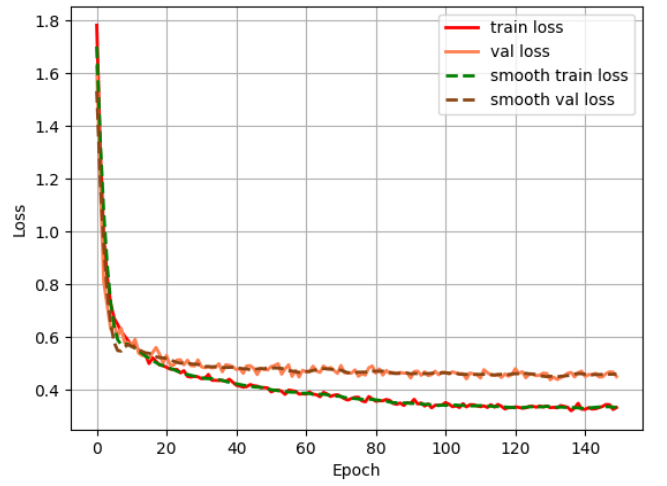


图 4-11 改进前 DeepLabV3+算法损失曲线

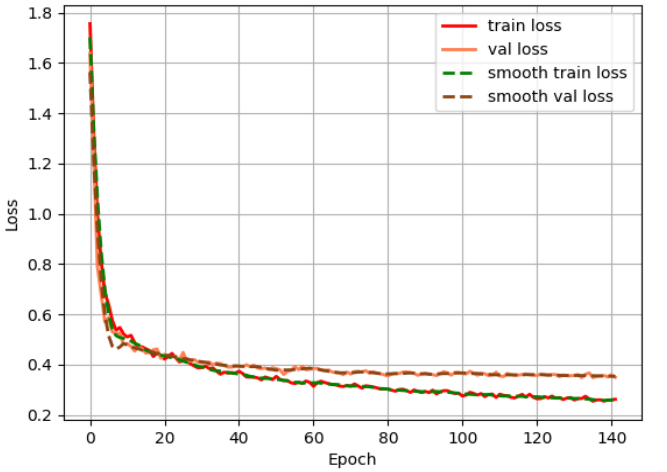


图 4-12 改进后 DeepLabV3+算法损失曲线

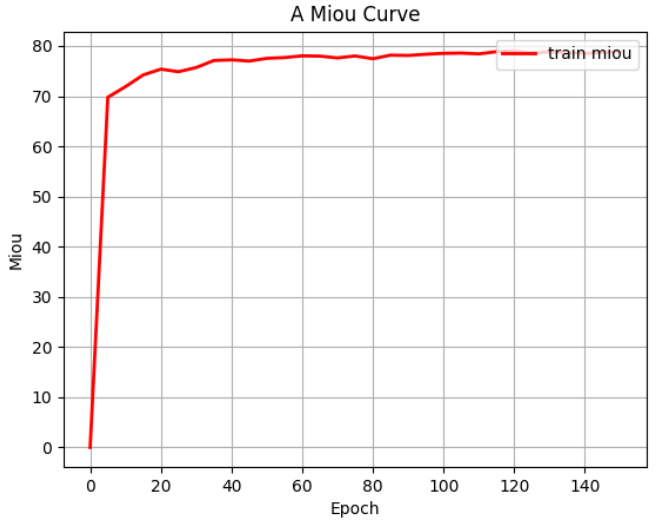


图 4-13 改进前 DeepLabV3+算法的 MIoU 曲线

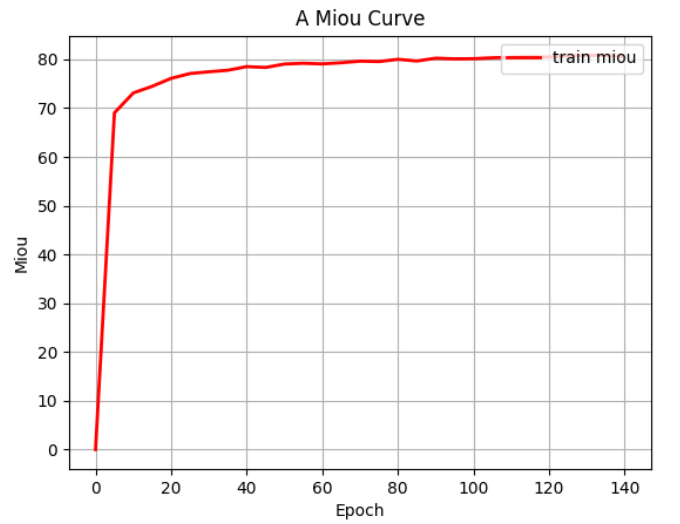


图 4-14 改进后 DeepLabV3+算法的 MIoU 曲线

为了验证本文方法的有效性，与其他语义分割方法进行了实验结果对比。其中，常见的语义分割指标的对比则包括平均交并比（MIoU）、平均像素准确率（MPA）、分类准确率（ACC）等。另外，重点关注的指标还有单幅图像预测处理时间，实时性也是本章节研究重点之一，验证本文算法实时性是否满足道路场景的实时语义分割，如表 4-3 所示。通过主、客观评价来对比说明实验结果。

表 4-3 不同算法实验数据对比

算法	MIoU/%	MPA/%	ACC/%	时间/ms
ERFNet	70.76	80.90	84.83	19.33
SegNet	71.37	83.12	89.10	38.62
PSPNet	74.58	84.37	90.21	24.36
DeepLabV3+	78.71	88.05	92.43	27.29
本文算法	**80.84**	**89.19**	**93.24**	**18.80**

客观评价：表 4-3 中的DeepLabV3+和PSPNet算法，其骨干网络均为MobileNetV2。从表中数据可以看出DeepLabV3+算法相比其他语义分割算法有一定的优势，但是本文算法在DeepLabV3+的基础上更加具有竞争力，MIoU值达到了 80.84%，相比原算法提高了 2.13%，单幅图像预测时间也比较大的减少，只有 18.80ms。本文算法在各项评价指标上均是最佳的。其中ERFNet算法是轻量化语义分割算法，其实时性能也较为优秀，但是在红外数据集上的精度较低。因此，本文算法具备良好的分割精度和实时性，一定程度上提高了工程实用性。

图 4-15 中，由上到下的每一行分别是红外原图、真实标签图、SegNet 预测图、PSPNet 预测图、DeepLabV3+预测图以及本文算法预测图。通过各个算法的预测结果图对比能够更直观的去观察到不同算法的图像分割效果，进一步验证本文算法的有效性与可行性。

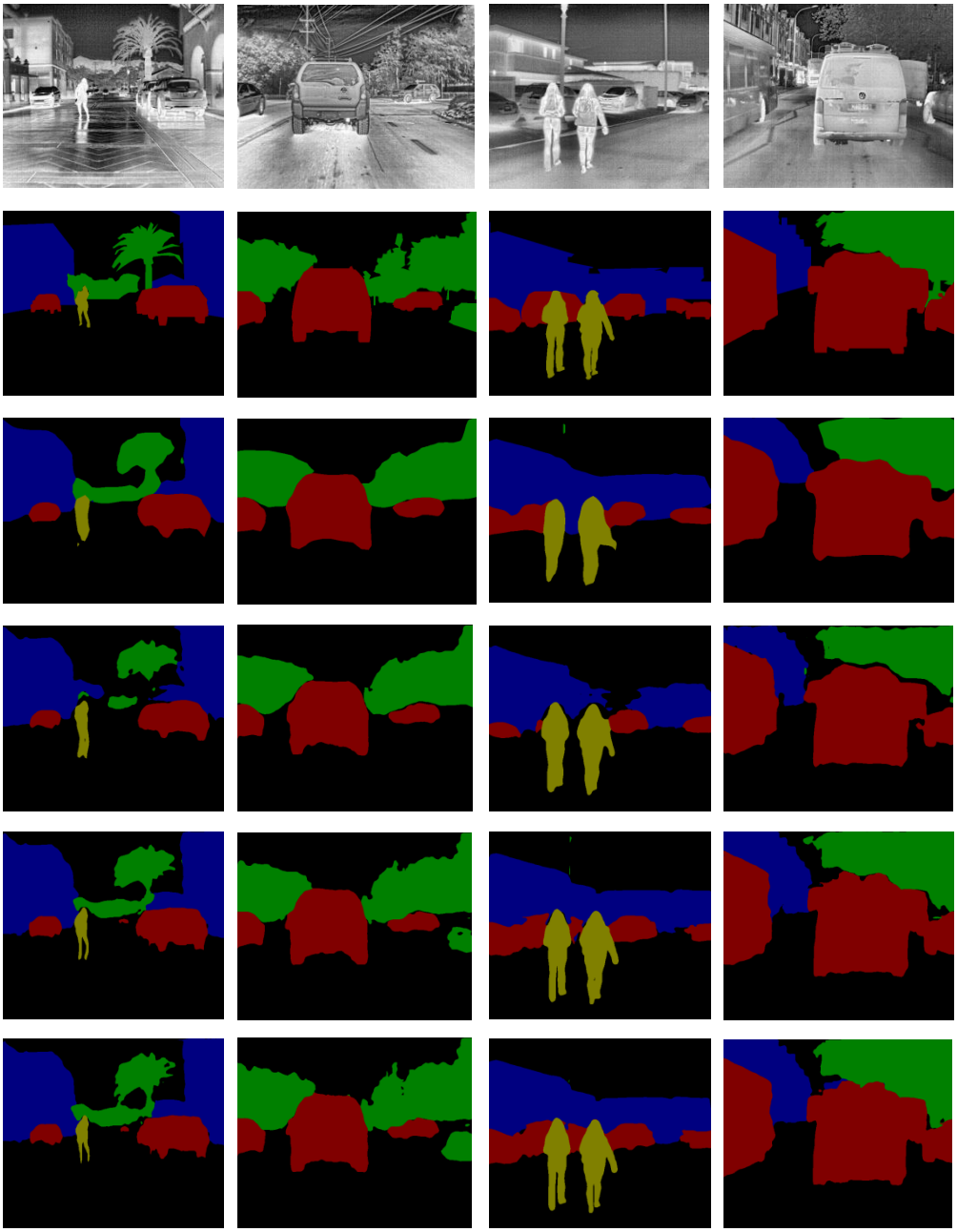


图 4-15 不同算法的预测结果对比

主观评价：观察图 4-15 可看出，本文算法分割效果是最好的，基本可以正确识别出各个类别的位置，包括较远距离的小形态物体也可以识别出来，并且可以较为准确的进行分割，比如远处的树干、汽车等景物，可以较好的适应景物尺

度变化较大的道路场景。另外，本文算法的边缘分割效果也是相对最好的，保留了很多细节信息。而算法 PSPNet 和 SegNet 都会出现目标识别不准确的问题，边缘分割效果也比较差。DeepLabV3+方法相对 PSPNet 方法分割效果要好一些，但也存在类似的分割问题。本文算法之所以具有良好的分割精度，是因为使用了注意力模块和密集连接的 ASPP 模块。综上分析，可以验证本文算法的分割性能得到改善。为了进一步验证本文改进点注意力机制模块和 D-ASPP 模块的有效性，做了消融实验，主要以 MIoU 评价指标来进行对比，具体实验数据如表 4-4 所示。

表 4-4 消融实验结果

组别	注意力机制模块	D-ASPP 模块	MIoU/%
1	—	—	78.71
2	√	—	79.53
3	—	√	80.02
4	√	√	80.84

由表 4-4 可知，不同组合方法的图像语义分割结果有较为明显的分割精度差距。其中，1，2 两组的结果对比可以看出，在 DeepLabV3+算法的基础上引入注意力机制模块，MIoU 提高了 0.82%，因此可以验证注意力模块能够提高特征提取的准确性；1，3 两组的结果对比可以看出，引入 D-ASPP 模块来获取更密集的特征金字塔和更大的感受野，使得 MIoU 值提高了 1.31%，算法的分割精度有了明显提高，进一步得到验证改进点的有效性。

4.4.5 可见光数据集的实验结果与分析

上述实验均使用的是夜间道路场景红外数据集，而且已经验证了本文算法在训练红外数据集时表现出了良好的性能。因此，使用白天场景的可见光数据集 Pascal VOC 2012 来进一步验证本文算法的效果。表 4-5 为不同算法在该可见光数据集上的实验结果对比，主要以 MIoU 指标做数据对比。本文算法是分割精度最佳的。

表 4-5 不同算法数据对比

算法	MIoU/%
SegNet	68.40
PSPNet	70.21
DeepLabV3+	72.25
本文算法	**73.83**

另外，将不同算法在 Pascal VOC 2012 数据集上的预测结果图进行对比，更加直观的看出不同的分割效果。如图 4-16 所示，由上到下每一行分别是可见光图、真实标签图、PSPNet 预测图、SegNet 预测图、DeepLabV3+预测图以及本文算法预测图。

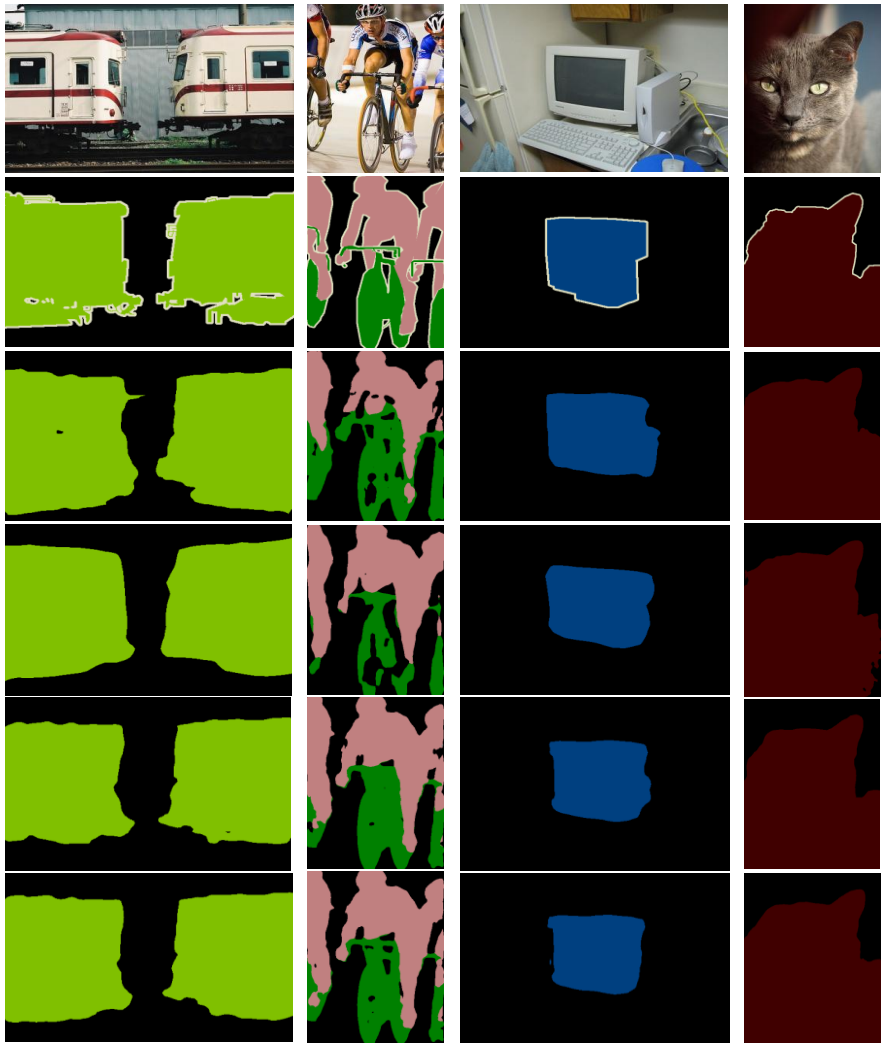


图 4-16 不同算法预测结果对比

由图 4-16 对比可以看出，本文算法的分割效果最好，景物识别更加精确，

像素分类也更准确，类别边界也相对较好的分割，如图中列车车厢和自行车的分割效果可以看出，车厢细节特征较为丰富，边缘部分像素分类较为准确。从自行车的车轮可以明显看出本文算法的像素精度是最好的，保留很多的细节信息。从而验证了本文算法在白天场景的可见光数据集上也能很好的进行分割。

4.5 本章小结

本章主要阐述了在 DeepLabV3+算法的基础上进行的三个改进点，包括换用骨干网络 MobileNetV2、引入基于空洞可分离卷积的密集连接方式的 ASPP 模块以及引入通道注意力和空间注意力模块。分别在红外数据集和可见光数据集进行训练与验证，并做了系列的对比实验、消融实验等。另外，对 D-ASPP 模块也做了不同空洞速率构建 D-ASPP 的对比实验。通过实验的结果与分析，主、客观评价了本文算法的分割效果，验证了本文算法的可行性与有效性，也为第五章的实验奠定了基础。

第 5 章 融合 SLIC 超像素的红外图像语义分割算法

5.1 引言

在第四章中，基于 DeepLabV3+算法进行了改进，提高了网络的实时性能以及分割精度，但是边缘分割精度明显不足，这也是图像语义分割模型在分割中十分常见的难题之一。边缘信息没有得到很好的保留，最根本的原因还是因为相邻类别的像素对应感受野内的图像信息太过相似导致。如果采用更加多尺度和多层次的网络结构去进行边缘预测，无疑让算法增加很多的计算量，这在道路场景的实时性语义分割中是难以接受的，本文的初衷是将分割精度与实时性兼顾优化，在实时性能良好的前提下，分割精度自然越高越好。而 SLIC 超像素算法是线性迭代聚类算法，采用均值聚类方法高效地生成超像素，相比以往的超像素算法可以更好地获取边界的同时，具有更快的速度，更高的内存效率，并且能提高图像分割性能，非常适合用于本文轻量化算法。因此，本文将语义分割算法结合 SLIC 超像素来构建融合算法，能够实现端到端的对夜间红外图像训练和学习，从而达到优化图像边缘预测精度的目的。

本文采用两种语义分割算法（即：本文第四章改进的 DeepLabV3+算法和经典算法 UNet）分别结合 SLIC 算法之后，分别在红外数据集和 Pascal VOC 2012 数据集上进行训练，并且与其他语义分割算法进行实验对比，从而验证融合 SLIC 超像素的语义分割算法的可行性与有效性。

5.2 融合 SLIC 的边缘分割优化算法

5.2.1 SLIC 超像素分割

SLIC（Simple Linear Iterative Clustering）是一种用于图像分割的超像素算法[63]。超像素是将图像像素划分为具有相似特征的紧凑区域的方法，它可以将图像分割为更高级别的语义区域，而不是仅仅将图像划分为单个像素。SLIC 算法通

过将图像分割为紧凑且相似的区域来生成超像素。它的基本思想是将图像彩色空间和空间距离结合起来，以确定最佳的初始超像素中心点，然后通过迭代过程将像素分配到最近的超像素中心点。

SLIC 算法的优点是简单而高效，能够在保持图像边缘信息的同时生成具有紧凑形状的超像素，也可以对灰度图进行分割。它在计算成本和超像素质量之间提供了一个权衡，因此常被用于图像分割、目标检测和图像处理等应用领域。

SLIC 算法常用来分割可见光 RGB 图像，而本文则是针对红外灰度图像，于是利用 SLIC 能兼容分割灰度图像的特点，应用于本文当中对红外图像进行处理。

SLIC 算法分割红外灰度图的步骤[64]如下：

（1）初始化超像素中心点：根据指定的超像素数量 K，均匀地在图像中选择初始中心点。图像中若具有 N 个像素点，那么每一个超像素的大小则为 N/K。相邻点的网格间距则为 $S=\sqrt{N/K}$。

（2）聚合像素：对于每个超像素中心点，将其周围的像素聚集起来，形成一个紧凑的超像素。

（3）更新中心点：计算每个超像素的重心，并将其作为新的超像素中心点。

（4）重复迭代：重复步骤（2）和（3），直到超像素中心点的位置不再发生显著变化。

（5）后处理：根据需要，可以对生成的超像素进行后处理，例如合并或分割。

具体计算公式为下式(5-1)、(5-2)及(5-3)。

$$d_c=\sqrt{(l_i-l_j)^2} \tag{5-1}$$

$$d_s=\sqrt{(x_i-x_j)^2+(y_i-y_j)^2} \tag{5-2}$$

$$D'=\sqrt{(\frac{d_c}{N_c})^2+(\frac{d_s}{N_s})^2} \tag{5-3}$$

其中，d_c 表示颜色距离，d_s 表示空间距离。N_s 为类内最大空间距离，定义为 $N_s=S=\sqrt{N/K}$，颜色距离 N_c 一般取固定常数值。

SLIC 算法中设置分割超像素个数 K 是重要的参数，其大小将会直接影响分割效果。因此，为了提高边缘分割的精确性，本文将 K 值设为 1000。如图 5−1 所

示，从左到右分别为红外图像和 K 值为 200、500、1000 的 SLIC 超像素分割图。

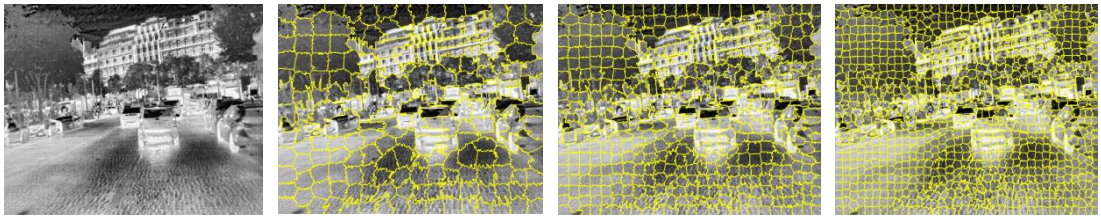


图 5-1 红外灰度图和不同K值下的分割图

5.2.2 融合 SLIC 的 DeepLabV3+算法

这里的 DeeLabV3+算法为第四章改进后的算法。虽然是改进后的算法，但是针对 DeeLabV3+边缘预测精度方面没有优化，即使原 DeeLabV3+增加了解码阶段来对边缘特征进行恢复，但是在上采样恢复特征图大小的过程中使用了两次双线性插值的方式，这就会导致图像类别之间的邻域特征分割效果不太理想。因此，本文将改进后的 DeeLabV3+算法结合 SLIC 超像素来改善边缘域的分割效果。融合算法的网络结构如图 5-2 所示。

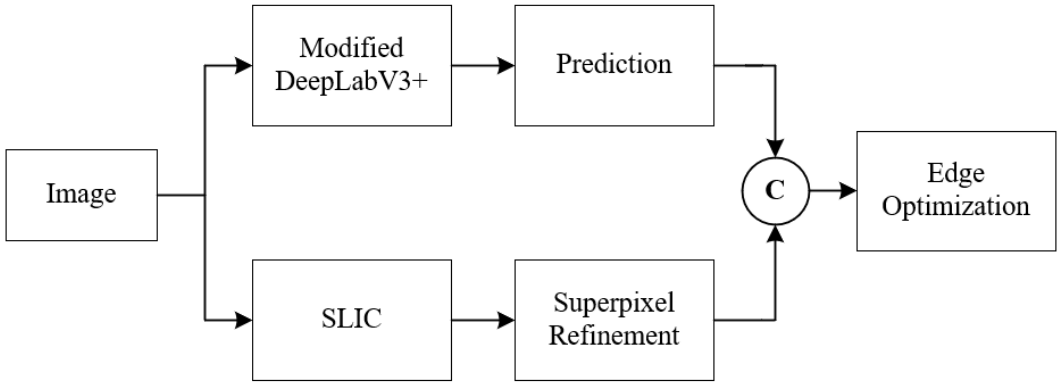


图 5-2 改进的 DeeLabV3+融合 SLIC 的算法结构

本文将卷积神经网络提取的高层语义信息与超像素类别边缘信息进行融合，实现改善边缘分割的效果，进而提高整体算法的语义分割精度。首先通过改进的 DeeLabV3+网络得到粗糙的语义预测结果，随后会统计每个超像素内各个类别所占像素数量，最后选择像素总数最多的类别并将其赋给该超像素[65]。

融合算法的具体实现流程：

（1）红外图像作为输入设为 I，通过改进的 DeeLabV3+得到较粗糙的语义预测图 L；

（2）通过 SLIC 获得 K 数量的超像素，$R=\{R_1,R_2,...,R_K\}$，R_i 表示第 i 个超像素；

（3）外部：for $i=1:K$

（a）用 $M=\{C_1,C_2,...,C_K\}$ 表示 R_i 中的全部像素，其中 C_j 为被类别 j 所标记的像素点。

（b）通过卷积神经网络来一一提取 C 中像素的特征，进而初始化权重 $W_C=0$。

（c）内部：*for* $j=1:N$

将 C_j 的特征标签保存为 L_{C_j}，然后更新超像素内一切标签的权重 $W_{C_j}=W_{C_j}'+1/N$，其中 W_{C_j}' 为 W_{C_j} 的上一个值。若 $W_{C_j}>0.8$，则退出内部循环，否则继续循环。

end

（d）搜索所有的 W_C 并判断是否具有某一 W_{C_j} 值大于 0.8。若有，执行下一步；否则，继续寻觅最大的 $W_{\max}$ 和第二大的 W_{sub}，然后判断它们之间的插值是否大于 0.2。如果是，则进行下一步；否则继续外部循环。

（e）使用目前超像素内几率最大的分类 $L_{C_{\max}}$，然后重新标注目前超像素的语义标签。

end

（4）最终得到超像素优化后的语义预测图，并且输出 $\tilde{L}$。

5.2.3 融合 SLIC 的 UNet 算法

为了验证 SLIC 超像素算法可以有效结合深度学习语义分割算法，能够实现语义类别边缘预测的精度提高，本文将 UNet 算法结合 SLIC 超像素，和第 5.2.2 节中将改进的轻量化 DeepLabV3+算法结合 SLIC 超像素类似，构建可以端到端的语义分割优化算法，使用红外数据集进行训练与学习，输出边缘优化分割的预测结果，另外使用可见光数据集 Pascal VOC 2012 作对比验证。

UNet 网络因其独特的架构而得名，其特点是具有对称的编码器-解码器结构和跳跃连接[66]。编码器部分是通过一系列卷积层、下采样操作（如最大池化或步长为 2 的卷积），逐渐减小特征图的空间尺寸，同时增加深度（通道数），以捕获更高级别的语义信息。解码器部分则是通过上采样操作（如转置卷积或双线性插值）恢复空间维度，同时进行特征融合，将高维特征映射回原始输入大小。

UNet 的关键创新之一是引入了跳跃连接机制，它直接将编码器中对应层次的特征图与解码器中的特征图拼接起来，这样可以将低级的细节信息传递给解码器，使得网络能够更好地保留边缘和细节信息，从而提高分割精度。最终，经过多层解码和特征融合后的特征图被送到一个1×1卷积层，用于生成最终的像素级分类结果，即每个像素点所属类别。UNet 融合 SLIC 超像素的算法整体结构并不复杂，如图 5-3 所示。

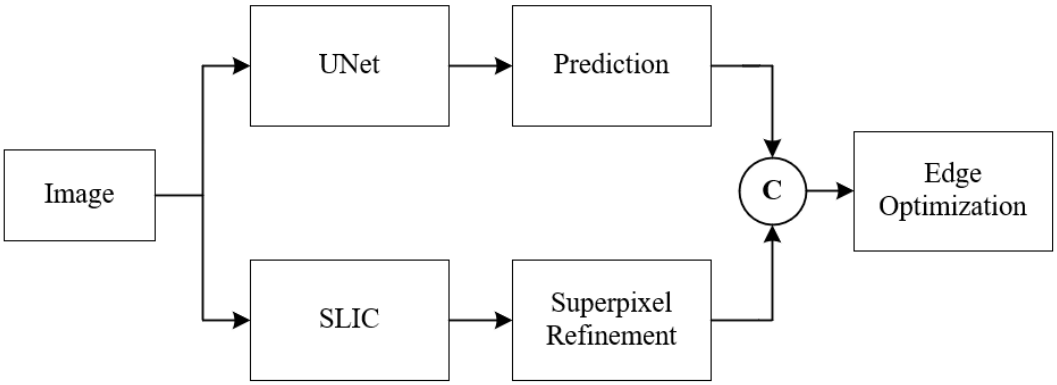


图 5-3 UNet 融合 SLIC 超像素的算法结构

5.3 实验结果与分析

本章节的实验是针对语义分割常见难题之一，语义类别边缘分割效果不太理想所进行的。基于深度学习的边缘分割的优化，一般是加深特征提取层次以及多尺度特征提取，这会使算法满足不了轻量化的语义分割，无法应用于道路场景。因此，本文将深度学习语义分割模型（本文改进的 DeepLabV3+和 UNet 算法）与传统的图像处理算法 SLIC 相结合，通过对比实验来验证该融合算法的可行性与有效性。

5.3.1 实验配置与参数

本实验是在 Windows10 系统上基于 Pytorch 框架来实现的，采用 CUDA 进行处理，GPU 型号：NVIDIA RTX2080Ti，显卡内存 11G。数据集主要为本文建立的红外数据集，在对比实验中同时使用了 Pascal VOC 2012 数据集。

在算法训练中常规的参数设置：基于改进的 DeepLabV3+融合算法和基于 UNet 的融合算法的数据批量大小（Batch Size，BS）设置均为 4，学习率初始化均为 0.001，学习率下降方式均为余弦退火（Cosine Annealing，COS）。前者的优化器种类选择随机梯度下降（Stochastic Gradient Descent，SGD），后者的内优化器选择自适应矩估计（Adaptive Moment Estimation，ADAM），优化器内部参数均设置为 0.9，损失函数均为交叉熵函数。

5.3.2 基于改进的 DeepLabV3+融合算法实验

针对该融合算法在红外数据集进行了训练与验证，图 5-4、5-5 为算法融合前后的训练损失曲线和验证曲线。通过对比可以看出融合算法的训练损失与验证损失都更小，分别为 0.223 和 0.270，损失值也较为接近，这说明其分割结果与真实标签更加相似，验证了融合算法的可行性与有效性。

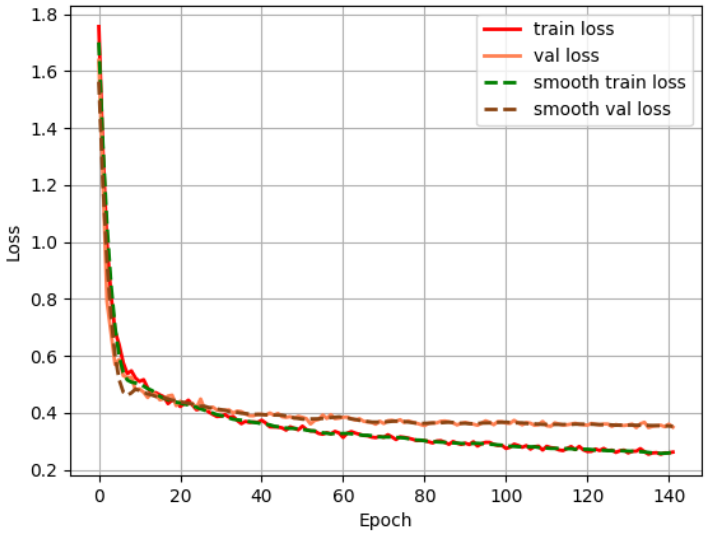


图 5-4 算法融合前的训练与验证损失曲线

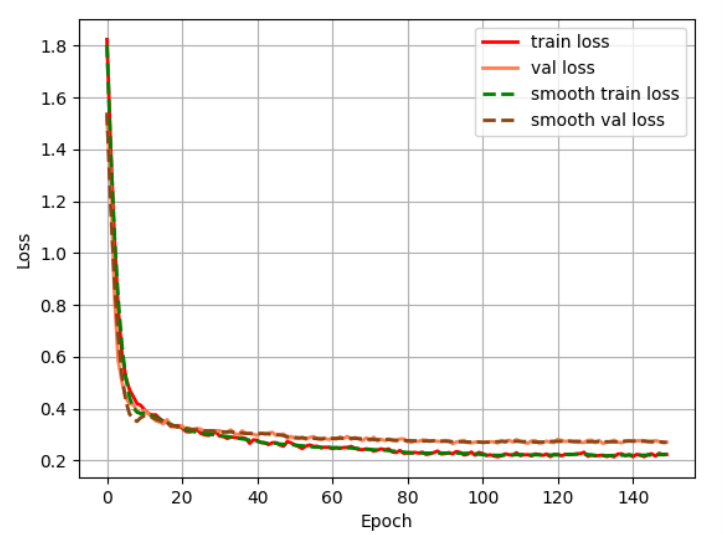


图 5-5 算法融合后的训练与验证损失曲线

根据图 5-4 和 5-5 训练与验证曲线的对比，融合算法的性能得到了进一步增强，那么其语义分割重要指标平均交并比（MIoU）也对应的会提高，如图 5-6、5-7 则分别为融合前后的平均交并比曲线，由图可以看出，融合后的算法其 MIoU 指标最终要高于融合之前的算法，达到了 81.82%，而融合前的算法指标为 80.84%，说明了融合算法的分割精度得到了改善，这主要在于提高了边缘域的预测精度。

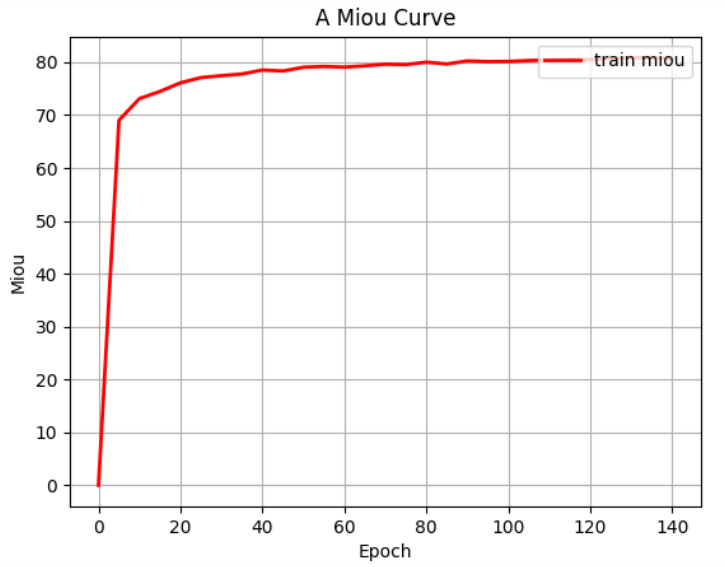


图 5-6 融合前的平均交并比曲线

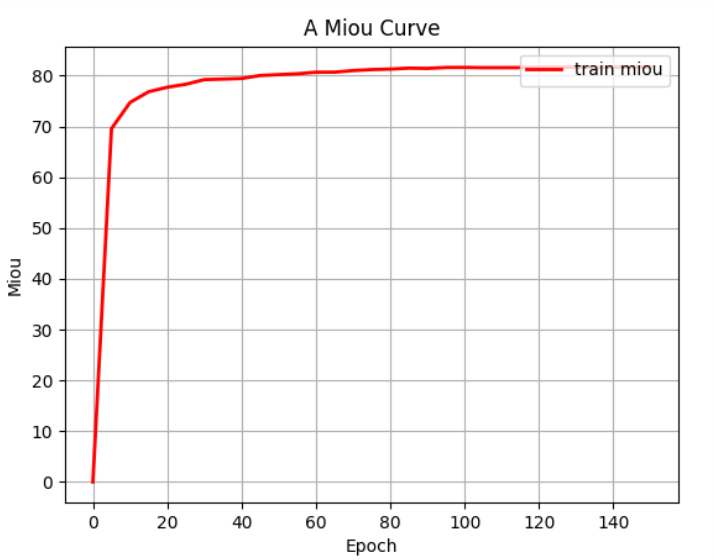


图 5-7 融合后的平均交并比曲线

通过训练与验证曲线以及MIoU曲线能够大致知道本文的融合算法性能得到一定改善，但是还不够直观。因此，本文以主、客观评价的方式来更加充分的展现融合算法分割性能的表现。如表 5-1 所示为不同算法各项评价指标的客观数据对比。

表 5-1 不同算法各项评价指标的客观数据对比

算法	MIoU/%	MPA/%	ACC/%	时间/ms
PSPNet	74.58	84.37	90.21	24.36
DeepLabV3+	78.71	88.05	92.43	27.29
改进 DeepLabV3+	80.84	89.19	93.24	**18.80**
本文融合算法	**81.82**	**89.66**	**93.58**	20.31

客观评价：对比实验中各算法骨干网络均为 MobileNetV2。通过客观的指标数据对比本文的融合算法在单幅图像预测时间上略微比改进的 DeepLabV3+算法多，但是在平均交并比、平均像素精度及整体精确度各个指标均优于改进的 DeepLabV3+算法，分别为 81.82%、89.66%以及 92.43%，单幅图像预测时间在 18.80ms 的基础上只有很小的增加，为 20.31ms，很好的均衡了精确性与实时性能，其中关键指标 MIoU 提高了 0.98 个百分点。因此，该融合算法从客观角度看，其分割性能有效得到改善，仍可以满足在道路场景下的语义分割要求。

为了更加直观的看出融合算法的性能改善效果，本文将不同算法的预测分割结果进行了四组对比，数据集为本文建立的夜间道路场景红外数据集。对比图如

图 5-8 所示，由上到下的每一行分别是红外图、真实标签、PSPNet 预测图、DeepLabV3+预测图、本文改进的 DeepLabV3+预测图以及本文融合算法预测图。

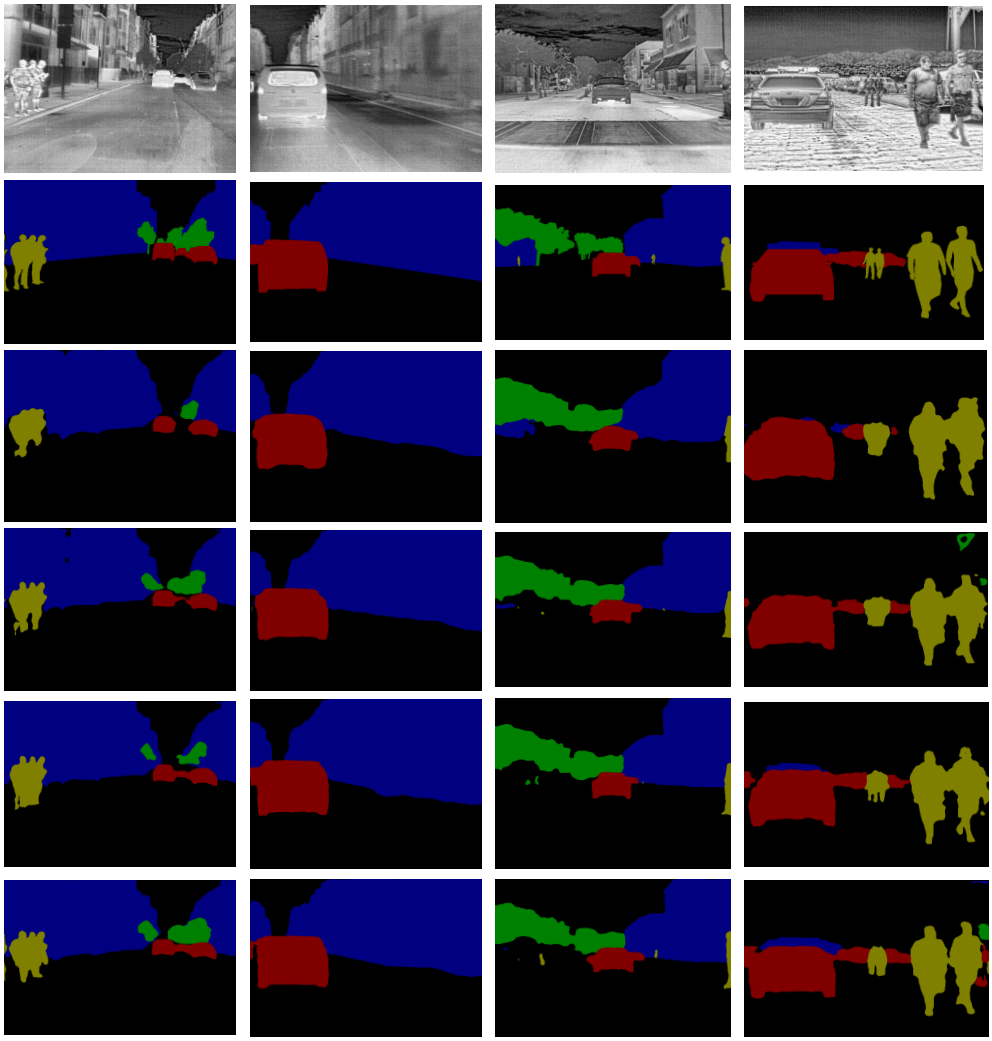


图 5-8 不同算法预测图对比

主观评价：从图 5-8 中各个算法的预测分割图可以更加直观的观察到分割效果的差异性。PSPNet预测图中，目标类别像素精确度较低，边缘域细节信息丢失较严重。改进的DeepLabV3+预测图中，细节信息有较多的保留，小目标类别物体也可以较准确分割，这归功于注意力机制以及D-ASPP模块，但是边缘域细节信息相对融合算法预测图仍然不够丰富，融合算法预测图可以观察到远、近景物的特征信息保留的最丰富，尤其是近景的细节信息，预测图中行人与建筑物的边缘域分割更加精确，与真实标签最相似，验证了将语义分割高层语义特征与超像素信息结合能够一定程度改善边缘分割精度。

另外，本文使用可见光数据集 Pascal VOC 2012 进一步验证本文融合算法的性能，将不同算法训练预测结果进行对比。表 5-2 为各算法使用该可见光数据集

进行训练的结果对比，主要以 MIoU 指标数据为参照。

表 5-2 各算法在可见光数据集训练指标对比

算法	MIoU/%
PSPNet	70.21
DeepLabV3+	72.25
改进 DeepLabV3+	73.83
本文融合算法	**74.75**

由表 5-2 可以看出，本文融合算法在改进的 DeepLabV3+算法的精度基础上又提高了约 1%的指标数值。为了直观的认识精度差异性，将各算法预测图进行了四组对比。如图 5-9 所示，由上到下每一行分别是可见光图、真实标签、PSPNet 预测图、DeepLabV3+预测图、改进的 DeepLabV3+预测图及本文融合算法预测图。

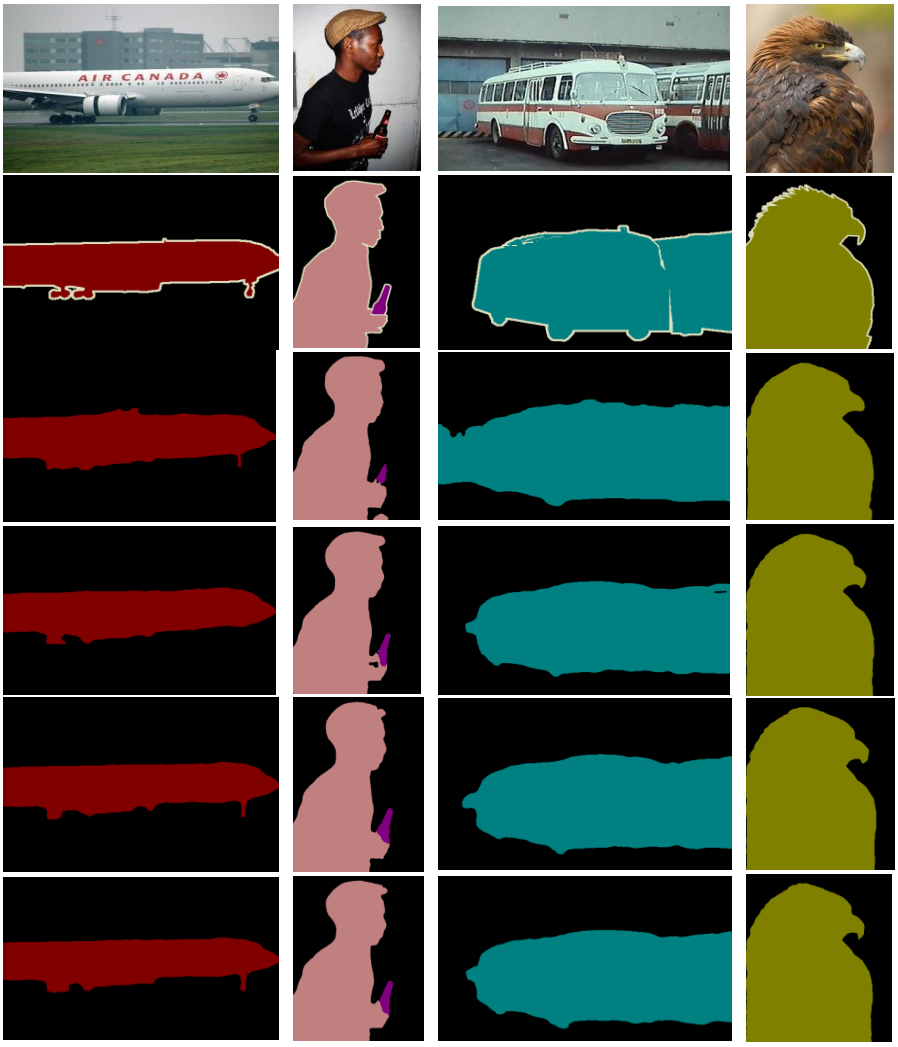


图 5-9 不同算法预测图对比

图 5-9 中预测图的对比可以观察到，本文融合算法的边缘域预测准确度是最佳的，预测图中的飞机腹部和老鹰的喙可以更加明显的看出边缘分割的优化效果得到了改善，从而验证本文融合算法（SLIC+改进 DeepLabV3+）在可见光数据集训练时的有效性。

5.3.3 基于 UNet 的融合算法实验

本文将UNet算法融合SLIC超像素来构建端到端的训练模型，所使用的UNet算法是基于VGG16 骨干网络，其模型计算量相对较大，此次实验是为了再次验证SLIC超像素分割可以有效结合深度学习语义分割算法，能够优化边缘分割效果，从而提高语义分割的整体精度。图 5-10 和 5-11 分别为该算法融合前后的训练损失曲线与验证损失曲线。

根据损失曲线的前后对比可以看出，融合算法的最终损失值更小，包括训练损失和验证损失，损失值分别为 0.146 和 0.275，而且训练损失值和验证损失值的差值也更接近，这说明在训练与学习过程中，融合算法使得预测精确度提高，像素精度也更高，预测值与真实值更加接近。

图 5-12 和 5-13 分别为算法融合前后，其主要评价指标平均交并比（MIoU）曲线，从曲线对比可以看出，融合算法的 MIoU 值更高，实际达到了 83.16%，而融合前算法的 MIoU 值为 82.21%，提高了 0.95 个百分点。

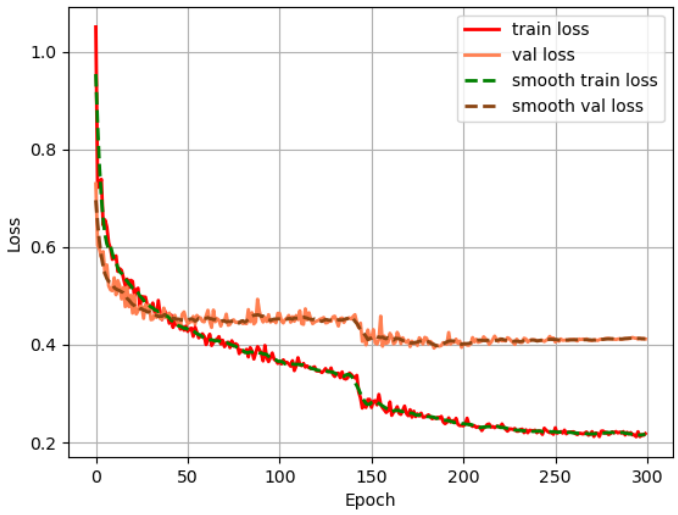


图 5-10 融合前算法的训练损失及验证损失曲线

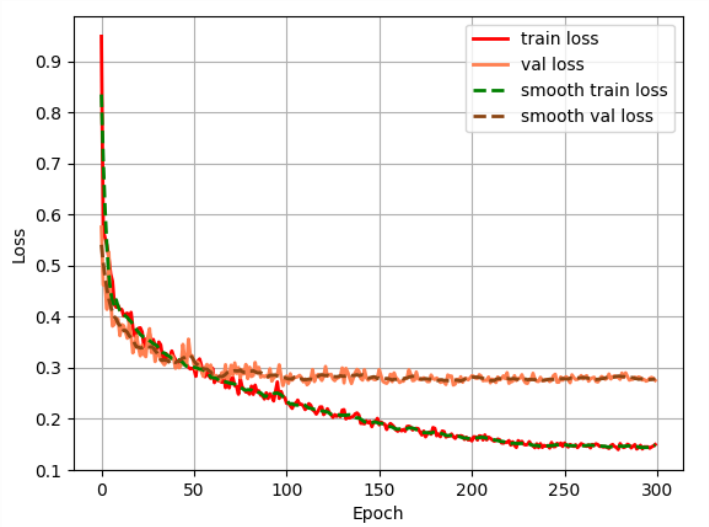


图 5-11 融合后算法的训练损失及验证损失曲线

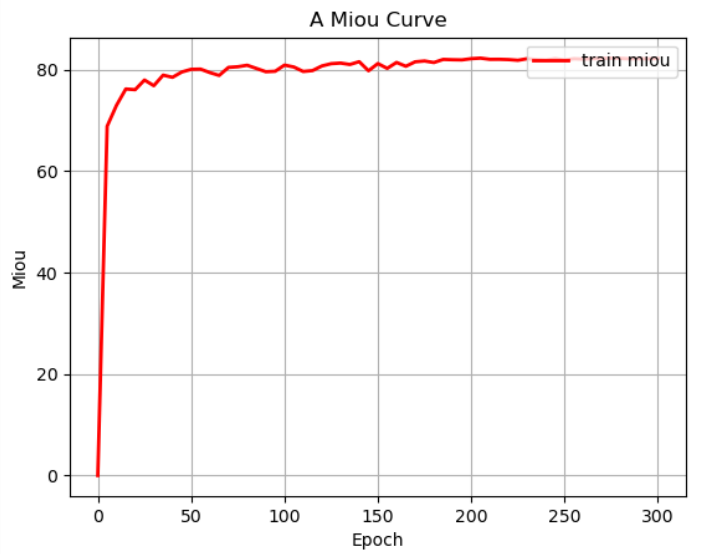


图 5-12 融合前算法的平均交并比曲线

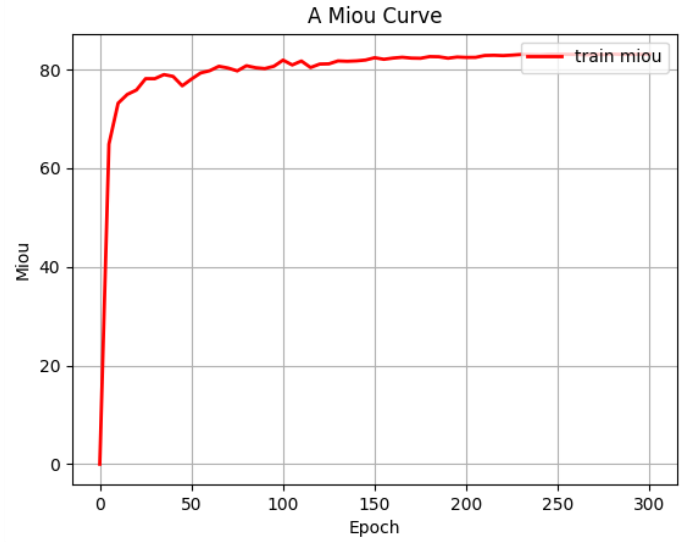


图 5-13 融合后算法的平均交并比曲线

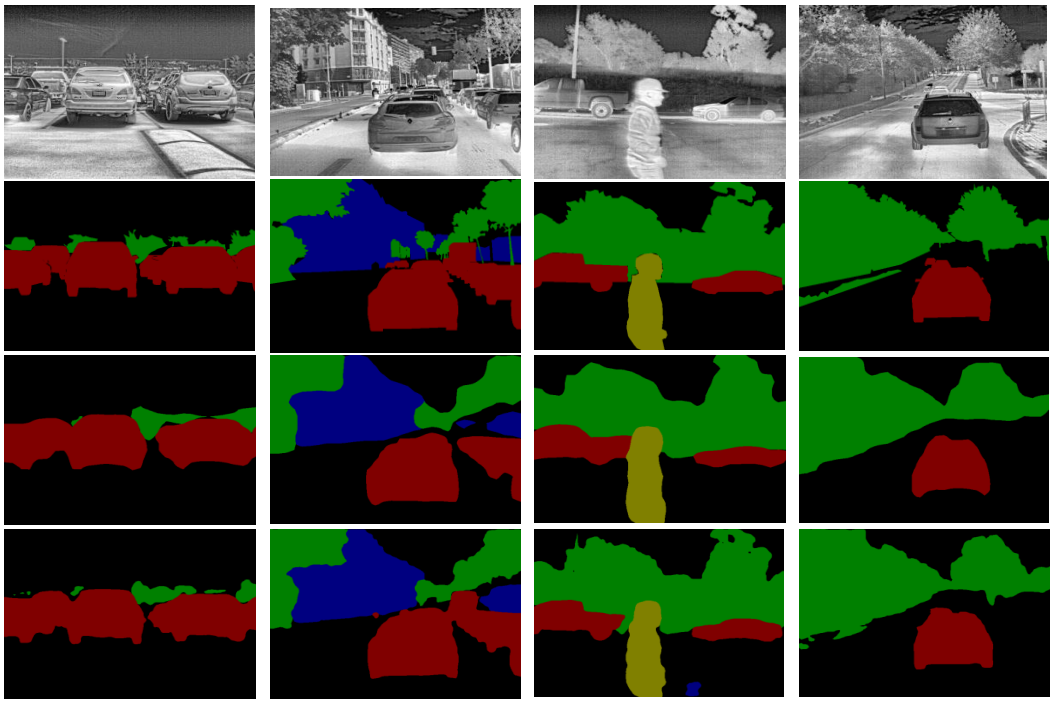
通过上述中的曲线图对比可以大致了解该融合算法的分割性能。同样的，为了更加直观、全面的认识本文融合算法的表现，本文从主、客观评价两个角度去分析算法的整体表现。表 5-3 为各算法在红外数据集训练的指标数据对比。

表 5-3 各算法在红外数据集训练指标对比

算法	MIoU/%	MPA/%	ACC/%	时间/ms
SegNet	71.37	83.12	89.10	38.62
SLIC+改进 DeepLabV3+	81.82	89.66	93.58	**20.31**
UNet	82.21	90.15	93.64	88.35
SLIC + UNet	**83.16**	**90.59**	**93.85**	90.47

客观评价：这里 UNet 算法是基于参数量较大的 VGG16 骨干网络。从上表可以看出，由于网络差异性，UNet 算法的几个常见指标比之前算法具有一定的优势，尤其在融合 SLIC 算法之后的评价指标均得到了有效提高，提高了 0.95%，但高计算量和高参数量带来的则是实时性能大幅下降，单幅图像预测时间达到了 90.47ms，是本文 SLIC+改进 DeepLabV3+融合算法的约 4.5 倍。

图 5-14 为各算法在使用红外数据集得到的预测图像对比，由上到下每一行分别是红外图、真实标签、SegNet 预测图、SLIC+改进 DeepLabV3+预测图、UNet 预测图、SLIC+UNet 预测图。



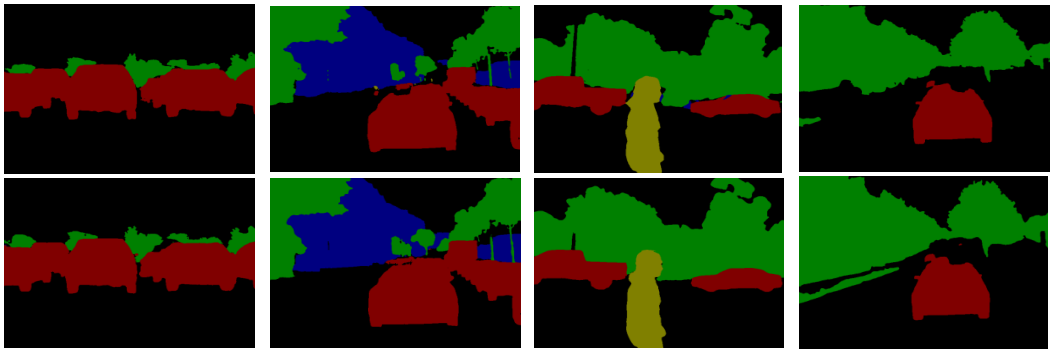


图 5-14 不同算法预测图对比

主观评价：根据上图对比观察可以得知，本节的融合算法（SLIC+UNet算法）借助着VGG16 骨干网络以及本身网络优势，无论是边缘域分割效果还是远近景物识别准确度都是最佳的，比如预测图中建筑与树木相邻区域以及车辆与树木相邻区域的分割更精确，景物像素识别也更精准，相比原UNet算法的表现更加优秀，验证了该融合算法的有效性。而上节提到的融合算法（SLIC+改进DeepLabV3+）也较为准确的进行了对景物类别的预测，包括在边缘域的预测效果也比较好，由于本文旨在研究满足道路场景的语义分割算法，应兼顾良好的实时性与精确性，因此，SLIC结合改进DeepLabV3+的融合算法会更加适用。

对于本节的融合算法，也是使用可见光数据集 Pascal VOC 2012 进一步验证该融合算法的性能，将不同算法训练预测结果进行对比。表 5-4 为各算法使用该可见光数据集进行训练的结果对比，主要以 MIoU 指标数据为参照。

表 5-4 不同算法在可见光数据集训练结果对比

算法	MIoU/%
SegNet	68.40
SLIC + 改进 DeepLabV3+	74.75
UNet	76.62
SLIC + UNet	**77.54**

从上表可以看出本节融合算法在使用可见光数据集预测时，分割精度也得到一定提升，提高了 0.92 个百分点。图 5-15 则为各算法在使用 Pascal VOC 2012 数据集得到的预测图像对比，由上到下每一行分别是红外图、真实标签、SegNet 预测图、SLIC+改进 DeepLabV3+预测图、UNet 预测图、SLIC+UNet 预测图。

根据图 5-15 中的预测结果对比可以看出本节融合算法（SLIC+UNet）边缘

域分割效果相比原 UNet 算法更好，相比上节融合算法（SLIC+改进 DeepLabV3+）也要更优秀，比如图中汽车底盘部分以及人与酒瓶相邻部分，像素准确度明显较高，更加接近真实标签。因此，验证了本节融合算法使用可见光数据集预测也可以有效改善边缘分割精度，融合算法的可行性与有效性也进一步得到了验证。

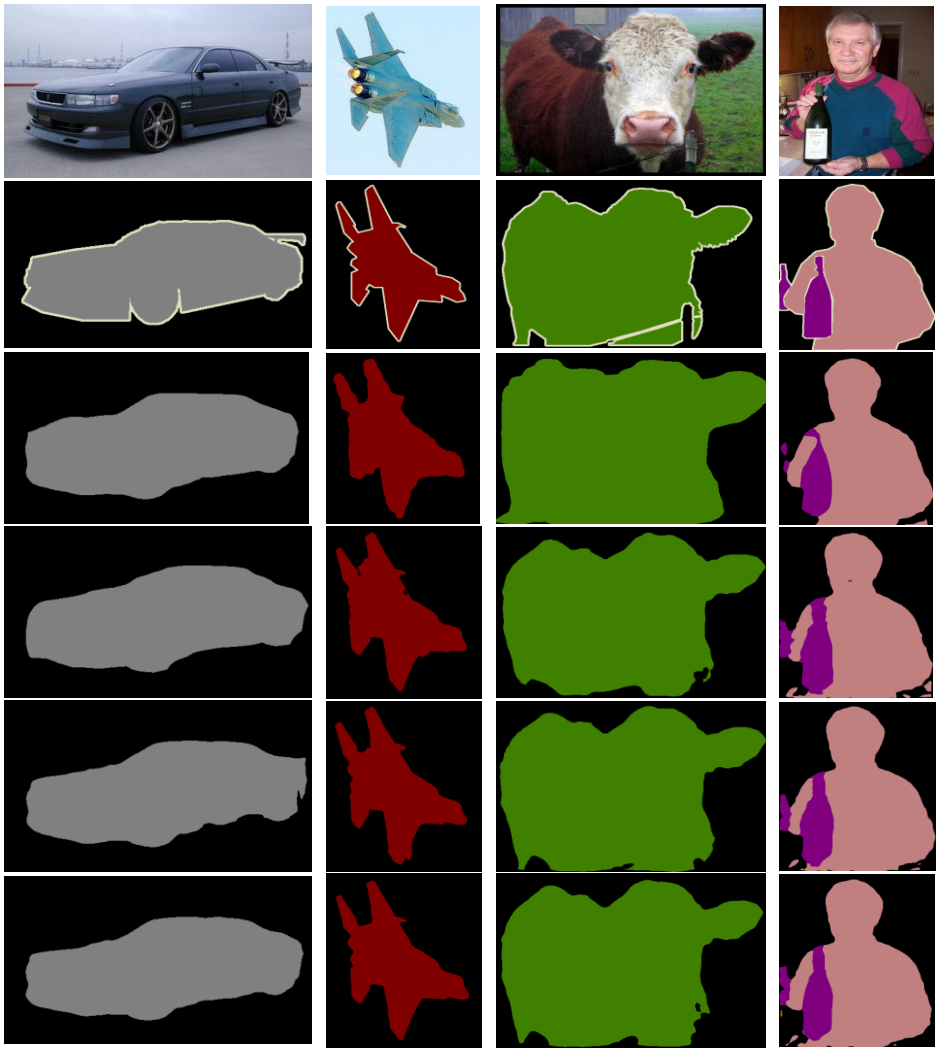


图 5-15 不同算法预测图对比

5.4 本章小结

本章主要研究融合SLIC超像素的语义分割算法，将SLIC超像素分别与本文改进的DeepLabV3+算法和UNet算法融合，分别使用红外数据集和Pascal VOC 2012 数据集进行训练与验证。本章首先阐述了SLIC超像素算法处理红外灰度图像的基本原理；然后，介绍了两个融合算法（SLIC+UNet和SLIC+改进DeepLabV3+）的结构，详细说明了SLIC超像素与高层语义特征融合并输出优化预测图的过

程；最后，针对两个融合算法与其他语义分割算法进行实验数据对比以及预测结果对比。总之，基于深度学习的红外图像语义分割算法与传统的图像处理方法的相结合策略，其可行性与有效性在本章节得到了验证。

第 6 章 总结与展望

6.1 总结

目前，面向道路场景的语义分割技术多应用于自动驾驶、无人车等领域，而这些领域在实际运用语义分割技术过程中，常受限于特殊环境影响（黑暗、大雾、眩光等）导致智能感知系统精度严重下降，语义分割精度也大大降低。因此，本文基于夜间道路红外场景来进行语义分割算法研究，通过红外成像仪进行数据采集不会受到外界特殊环境影响，能够实现自动驾驶系统全天候的精准感知。

通过对红外图像语义分割的研究，基于深度学习的可见光语义分割算法能够实现红外图像像素级的分类与分割得到了验证，这取决于在语义分割模型中红外图像可以获得每个像素的类别标签。但由于红外图像属于灰度图像，为单通道，语义分割的效率会明显提升。也正因为其为单通道图像，图像分辨率与清晰度也较低，图像质量不如可见光图像，将会影响语义分割的精确度。

语义分割不仅要求模型识别图像中存在哪些不同的对象（分类），还必须清楚地描绘出每个对象在图像中的具体位置和形状（分割），初步的实现对图像场景的语义理解，为更深层次的图像理解打下基础。本文将基于深度学习的语义分割技术应用在红外图像上，实现对红外场景的语义理解，并通过改进的语义分割算法进行训练与预测，在满足道路场景实时性分割的前提下，预测精度也比较可观。另外，将语义分割模型结合传统图像处理方法来改善边缘域的预测精度，可以进一步提高模型整体的语义分割精度。获得具体成果如下：

（1）对夜间红外数据增强预处理，建立本文夜间红外数据集。针对红外图像特性，使用 CLAHE 算法进行增强处理，包括提高图像对比度与清晰度、降低图像噪声放大等，更利于人工标签制作与算法训练；人工制作标签文件，并且对数据集进行增强扩充，自动划分训练集与验证集。

（2）改进 DeepLabV3+算法结构，以兼顾良好的预测精度与实时性能。使用 MobileNetV2 作为算法的基础网络；引入密集链接方式的 ASPP 模块，考虑实时性能，将模块标准卷积替换为深度可分离卷积；另外，在编码和解码阶段，分别

引入通道注意力和空间注意力模块。改进后，在保证较高的预测精度的同时，有效改善原算法的实时性能。

（3）构建融合 SLIC 超像素的语义分割算法，验证了该融合算法的可行性与有效性。本文将改进的 DeepLabV3+算法和 UNet 算法分别与 SLIC 超像素进行融合，并且使用夜间红外数据集和 Pascal VOC 2012 数据集分别进行训练与验证，最终实验结果都验证了本文融合算法的有效性，边缘预测精度得到有效提高。

6.2 展望

随着自动驾驶等技术的不断发展，道路场景的语义分割算法也随之成为现在的热门研究之一，如何改善实时性与精确性也一直是该领域的主要研究内容，而针对夜间道路场景红外图像的语义分割研究相对不多。另外，边缘域分割效果不佳是语义分割关键难题之一，本文结合传统图像处理方法来优化边缘分割，并起到了一定的改善效果，但是这仍然无法彻底解决该难题。因此，还有很多方面需要去研究、去探索。

（1）建立大规模的开源的夜间道路场景红外数据集

本文建立的夜间红外数据集的图像是从 FLIR 目标识别数据集中筛选出来的，并且人工制作标签文件，比较耗时费力，而且数据集规模较小，不利于进行算法研究，这是由于目前开源的夜间道路红外数据集非常匮乏的原因。因此，建立一个大规模、多任务、综合性的道路场景红外数据集，包括白天与夜晚场景的红外数据，这是很有必要的，能够满足语义分割、目标检测、目标识别等方向的使用，便于研究者们的工作开展。

（2）实时性语义分割算法的进一步研究

道路场景语义分割的应用需要的关键性能就是实时性，盲目的追求分割精度的提高而忽略实时性对自动驾驶等领域的发展是不利的。本文针对这一点进行提高算法实时性的研究，但也只是起到有限的作用，如果想要道路场景语义分割发展的更长远，需要研究更加实时性、轻量化的语义分割算法。

（3）边缘域分割问题的进一步研究

目前，边缘域分割问题仍然得不到有效解决，常见的边缘优化算法也只能做到一定程度的改善。因此，在以后的研究中，要不断地针对该问题进行探索，包括对现有的优化算法进行不断改进以及研究更优秀的优化算法。

总之，基于深度学习的各种应用算法是人工智能领域的研究重点，未来也会面临各种难题与挑战，但仍然相信通过研究者们的不懈探索与努力下，这些挑战与难题都将解决，人类社会也将更加的智能化。

参考文献

[1] 王晓, 张翔宇, 周锐, 等. 基于平行测试的认知自动驾驶智能架构研究[J]. 自动化学报, 2024, 50(02): 356-371.

[2] 蒋彪, 吴楷, 蒋炜等. 边缘计算下整车驾驶辅助系统动态优化[J]. 智能网联汽车, 2022, (01): 54-58.

[3] 成欢, 王硕, 李孟等. 面向自动驾驶场景的神经辐射场综述[J]. 图学学报, 2023, 44(06): 1091-1103.

[4] 毕阳阳, 郑远帆, 史彩娟, 等. 基于深度学习的图像全景分割综述[J]. 计算机科学与探索, 2023, 17(11): 2605-2619.

[5] Huang Z, Lv C, Xing Y, etc. Multi-modal sensor fusion-based deep neural network for end-to-end autonomous driving with scene understanding[J]. IEEE Sensors Journal, 2020, PP(99): 1-1.

[6] Hu K, Xie Z, Hu Q. Dual-resolution transformer combined with multi-layer separable convolution fusion network for real-time semantic segmentation [J]. Computers Graphics, 2024, 118220-232.

[7] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.

[8] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//Proceedings of the medical image computing and computer-assisted intervention. 2015: 234-241.

[9] Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(12): 2481-2495.

[10] Liu W, Rabinovich A, Berg A C. Parsenet: Looking wider to see better[J]. arXiv preprint arXiv: 1506.04579, 2015.

[11] Chen L C, Papandreou G, Kokkinos I, etc. Semantic image segmentation with deep convolutional nets and fully connected crfs[J]. arXiv preprint ar Xiv: 1412.7062, 2014.

[12] Chen L C, Papandreou G, Kokkinos I, etc. Deeplab: Semantic image segm entation with deep convolutional nets, atrous convolution, and fully connec ted crfs[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 40(4): 834-848.

[13] Chen L C, Papandreou G, Schroff F, etc. Rethinking atrous convolution fo r semantic image segmentation[J]. arXiv preprint arXiv: 1706.05587, 2017.

[14] Chen L C, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separ able convolution for semantic image segmentation[C]//Proceedings of the E uropean conference on computer vision (ECCV). 2018: 801-818.

[15] Zhao H, Shi J, Qi X, etc. Pyramid scene parsing network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2881-2890.

[16] Huang Z, Wang X, Huang L, etc. Ccnet: Criss-cross attention for semantic segmentation[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 603-612.

[17] Li H, Xiong P, Fan H, etc. Dfanet: Deep feature aggregation for real-time semantic segmentation[C]//Proceedings of the IEEE/CVF conference on co mputer vision and pattern recognition. 2019: 9522-9531.

[18] Hazirbas C, Ma L, Domokos C, etc. Fusenet: Incorporating depth into sem antic segmentation via fusion-based cnn architecture[C]//Proceedings of the computer vision. 2017: 213-228.

[19] Ha Q, Watanabe K, Karasawa T, etc. MFNet: Towards real-time semantic segmentation for autonomous vehicles with multi-spectral scenes[C]//Procee dings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2017: 5108-5115.

[20] 吴骏逸, 谷小婧, 顾幸生. 基于可见光/红外图像的夜间道路场景语义分割[J]. 华东理工大学学报(自然科学版), 2019, 45(02): 301-309.

[21] Sun Y, Zuo W, Liu M. Rtfnet: Rgb-thermal fusion network for semantic s egmentation of urban scenes[J]. IEEE Robotics and Automation Letters, 20 19, 4(3): 2576-2583.

[22] Azarbad M, Ebrahimzade A, Izadian V. Segmentation of infrared images a nd objectives detection using maximum entropy method based on the bee algorithm[J]. International Journal of Computer Information Systems and In dustrial Management Applications, 2011, 3: 26-33.

[23] Yu X, Zhou Z, Gao Q, etc. Infrared image segmentation using growing i mmune field and clone threshold[J]. Infrared physics & technology, 2018, 88: 184-193.

[24] Vertens J, Zürn J, Burgard W. Heatnet: Bridging the day-night domain gap in semantic segmentation with thermal images[C]//Proceedings of the IEE E/RSJ International Conference on Intelligent Robots and Systems (IROS). 2020: 8461-8468.

[25] Chen S, Chen Z, Xu X, etc. Nv-Net: Efficient infrared image segmentatio n with convolutional neural networks in the low illumination environment [J]. Infrared Physics & Technology, 2020, 105: 103184.

[26] 李敬兆, 何娜, 张金伟, 等. 基于VMD和CNN-BiLSTM的矿井提升电动机故障诊断方法[J]. 工矿自动化, 2023, 49(07): 49-59.

[27] C Junbin, H Ruyi, Z Kun, etc. Multiscale convolutional neural network wi th feature alignment for bearing fault diagnosis[J]. IEEE Transactions on I nstrumentation and Measurement, 2021, 70.

[28] 王婷, 王其兵, 何志方, 等. 基于改进的一维级联神经网络的异常流量检测[J]. 计算机应用与软件, 2023, 40(09): 320-326.

[29] Anand V, Gupta S, Koundal D, etc. Deep learning based automated diagnosis of skin diseases using dermoscopy[J]. Computers, Materials Continua, 2022, 71(2): 3145-3160.

[30] 鹿静雅, 伍志发, 刘梦秋, 等. 基于卷积神经网络的胸部薄层CT图像肋骨骨折分割与检测研究[J]. 临床放射学杂志, 2023, 42(06): 996-1002.

[31] 臧颖. 基于边缘感知的图像语义分割算法的研究[D]. 浙江大学, 2021.

[32] H J C, Gyang E R. Evaluating prediction model performance[J]. Surgery, 2023, 174(3): 723-726.

[33] 沈川贵. 基于Deeplab V3+框架的图像语义分割技术研究[D]. 贵州财经大学, 2022.

[34] 李晶. 温度传感器在真空冻干设备中的应用与研究[J]. 低温与特气, 2009, 27(03): 42-45.

[35] Hamidi S Z, Shariff M N N, Ibrahim A Z, etc. Magnetic reconnection of solar flare detected by solar radio burst type III[J]. Journal of Physics: Conference Series, 2014, 539(1): 012006-012006.

[36] J. K S, Souvik B, Bo Z, etc. Direct observation of the violation of Kirchhoff's law of thermal radiation[J]. Nature Photonics, 2023, 17(10): 891-896.

[37] 林向礼. 观念的转变——纪念普朗克量子理论发表 100 年[J]. 科学与文化, 2001, (03):32.

[38] 郑娜娥, 吕品品, 任修坤等. 目标特性与传感原理[M]. 电子工业出版社: 20 2011.161.

[39] 银浩, 陈志军, 马永吉, 等. 循环流化床锅炉微熔电磁辐射技术研究与应用[J]. 仪器仪表用户, 2023, 30(03): 70-73.

[40] 裴训, 谢敏, 李瑞一, 等. (Gd_(1-x)Er_x)_2Zr_2O_7 红外发射率对高温导热性能的影响规律研究[J]. 中国稀土学报, 2023, 41(06): 1111-1118.

[41] 王闪. 基于深度学习的红外图像真彩转换算法研究[D]. 哈尔滨工程大学, 2023.

[42] 李延彬, 康为民, 戴景民. 红外景象模拟器波长相关问题的研究[C]//中国宇航学会光电技术专业委员会, 2008: 4.

[43] 李延彬, 康为民, 戴景民. 红外景象模拟器波长相关问题的研究[J]. 红外与激光工程, 2008, 37(S2): 381-384.

[44] 肖学文, 王亚淑, 刘康林. 基于红外热像技术的过程设备无损检测[J]. 化工机械, 2020, 47(06): 742-746.

[45] Weiwen Z, Yi Y, Jiyong J, etc. Value of the combination of a smartphone -compatible infrared camera and a hand-held doppler ultrasound in preoper ative localization of perforators in flaps[J]. Heliyon, 2023, 9(6): 17372-173 72.

[46] Peeters J, Louarroudi E, Greef D D, etc. Time calibration of thermal rolli ng shutter infrared cameras[J]. Infrared Physics and Technology, 2017, 801 45-152.

[47] 王梦瑶, 宋薇. 动态场景下基于自适应语义分割的RGB-D SLAM算法[J]. 机器人, 2023, 45(01): 16-27.

[48] 李习习. 基于YOLOX的雾天交通目标检测方法研究[D]. 安徽工程大学, 20 23.

[49] 李洲, 吴晗, 闫满等. 基于改进ViBe算法以及改进CLAHE算法的烟雾信息增强技术[J]. 激光杂志, 2024, 45(01): 121-128.

[50] 王晨. 基于深度学习的红外图像语义分割技术研究[D]. 中国科学院大学(中国科学院上海技术物理研究所), 2017.

[51] Asem K. Smoke removal technique of industrial scene images based on se cond-generation wavelets and dark channel prior model[J]. Soft Computing, 2023, 27(23): 17505-17514.

[52] S. S. R. G. Determining the quality of compression in high resolution sate llite images using different compression methods[J]. International Journal o f Engineering Research in Africa, 2015, 20(20-20): 202-217.

[53] 徐开丽, 张乾, 何剑. 一种用于人脸图像修复的TPDCU-Net算法[J]. 湖北民族大学学报(自然科学版), 2024, 42(01): 105-110+150.

[54] 葛振强. 基于改进DeepLabv3+的道路分割算法[J]. 现代信息科技, 2024, 8(04): 132-135+141.

[55] 王兴涛, 单慧琳, 孙佳琪等. 基于改进YOLOv3 的遥感目标检测算法[J]. 兵器装备工程学报, 2023, 44(11): 279-286.

[56] Ahmad P, Jin H, Qamar S, etc. RD 2 A: Densely connected residual networks using ASPP for brain tumor segmentation[J]. Multimedia Tools and Applications, 2021, 80: 27069-27094.

[57] 李秋晓, 刘建明, 杨晓冬. 基于多模型融合的遥感房屋提取[J]. 测绘与空间地理信息, 2023, 46(10): 81-84.

[58] 张云佐, 李文博, 郭威. 融合多尺度信息和跨维特征引导的轻量行人检测[J]. 光电子·激光, 2024, 35(04): 344-350.

[59] 刘致驿. 无人车场景的图像语义分割与语义SLAM研究[D]. 东华大学, 2020.

[60] 张晓倩, 罗建, 杨梅, 等. 基于改进SE-Net网络与多注意力的脑肿瘤分类方法[J]. 西华师范大学学报(自然科学版), 2024, 45(01): 93-101.

[61] 王囡, 侯志强, 蒲磊等. 空洞可分离卷积和注意力机制的实时语义分割[J]. 中国图象图形学报, 2022, 27(04): 1216-1225.

[62] Yandong H, Zhengbo W, Xinghua R, etc. BFFNet: a bidirectional feature fusion network for semantic segmentation of remote sensing objects[J]. International Journal of Intelligent Computing and Cybernetics, 2024, 17(1): 20-37.

[63] Qiu W, Gao X, Han B. A superpixel-based CRF saliency detection approach[J]. Neurocomputing, 2017, 24419-32.

[64] Shifei D, Lijuan W, Lin C. Super-pixel image segmentation algorithm based on adaptive equalisation feature parameters[J]. IET Image Processing, 2020, 14(17): 4461-4467.

[65] 任凤雷. 基于智能车辆视觉导航的环境感知技术研究[D]. 中国科学院大学(中国科学院长春光学精密机械与物理研究所), 2020.

[66] Wenbo Z, Weidong L, Le L, etc. A framework for the efficient enhancement of non-uniform illumination underwater image using convolution neural network[J]. Computers Graphics, 2023, 11260-71.

攻读硕士学位期间主要研究成果

[1] Su S, Xu G, Wang B. Research on semantic segmentation of night infrared image for road scene[C]//Third International Conference on Signal Image Processing and Communication (ICSIPC 2023). SPIE, 2023, 12916: 314-320.（第一作者）

[2] 王博,许钢,苏世林.一种基于 SwiftNet 面向室内 RGBD 场景的高效语义分割算法[J/OL].重庆工商大学学报(自然科学版),1-10[2024-03-26].（第三作者）

致谢

时光飞逝，转眼间，三年的硕士学习生活即将落下帷幕，回顾这段珍贵的学习历程，我受益良多。到了临别院校的时刻，我内心充满了不舍与感激。在此，感谢所有在我的学习过程中给予支持与帮助的老师和同学。

首先，要特别感谢我的导师许钢教授。许老师工作认真负责，学识广博，性格随和友善。无论我遇到什么难题，他都会耐心指导，解答我疑惑，指明我前进的方向。在学术研究上，许老师用他渊博的学识为我的学术研究指明了方向，把我培养成可以独立进行学术研究的硕士生。许老师认真地带领我完成了从选题、理论研究到具体实验，再到毕业论文的撰写、修改过程。在生活上，许老师教会我如何待人接物，并在就业问题上给予了我宝贵的建议。在此，我向许老师表示最衷心的感谢！

其次，感谢师兄师姐们，他们在我课题研究的路上一路陪伴着，帮助我解决了很多课题研究中的困难，并且提供了毕业论文的格式参考。感谢课题组的所有同学在学术和生活上对我的关心和帮助。另外，感谢我的室友，他们营造了良好的休息环境，使我有充沛的精力投入学习与科研中。

最后，感谢一路以来一直陪伴我成长的亲人、朋友。向百忙之中审阅论文的各位评审老师表示诚挚的感谢！