

# 一种多分辨率特征提取红外图像语义分割算法

徐慧琳<sup>1</sup>, 赵鑫<sup>1,2</sup>, 于波<sup>1</sup>, 韦小牙<sup>1</sup>, 胡鹏<sup>1,2</sup>

(1. 安徽理工大学 人工智能学院, 安徽 淮南 232000;

2. 安徽理工大学 深部煤矿采动响应与灾害防控国家重点实验室, 安徽 淮南 232000)

**摘要:** 针对现有图像语义分割算法在对低分辨率红外图像进行分割时存在准确率不高的问题, 提出了一种多分辨率特征提取算法。该算法以 DeepLabv3+ 为基准网络, 添加了一组对偶分辨率模块, 该模块包含低分辨率分支和高分辨率分支, 以进一步聚合红外图像特征。低分辨率分支采用 GPU 友好的注意力模块捕获高层全局上下文信息, 同时引入一个多轴门控感知机模块并行提取红外图像局部信息和全局信息; 高分辨率分支采用跨分辨率注意力模块将低分辨率分支上学习到的全局特征传播扩散到高分辨率分支上以获取更强的语义信息。实验结果表明, 该算法在数据集 DNDS 和 MSRS 上的分割精度优于现有语义分割算法, 证明了提出算法的有效性。

**关键词:** 对偶分辨率模块; 语义分割; DeepLabv3+; 红外图像; 注意力模块

中图分类号: TP391.41 文献标识码: A 文章编号: 1001-8891(2024)05-0556-09

## Multi-resolution Feature Extraction Algorithm for Semantic Segmentation of Infrared Images

XU Huilin<sup>1</sup>, ZHAO Xin<sup>1,2</sup>, YU Bo<sup>1</sup>, WEI Xiaoya<sup>1</sup>, HU Peng<sup>1,2</sup>

(1. School of Artificial Intelligence, College of Anhui University Of Science and Technology, Huainan 232000, China; 2. State Key Laboratory of Mining Response and Disaster Prevention and Control in Deep Coal Mines, Huainan 232000, China)

**Abstract:** A multi-resolution feature extraction convolution neural network is proposed for the problem of inaccurate edge segmentation when existing image semantic segmentation algorithms process low-resolution infrared images. DeepLabv3+ is used as the baseline network and adds a multi-resolution block, which contains both high and low resolution branches, to further aggregate the features in infrared images. In the low-resolution branch, a GPU friendly attention module is used to capture high-level global context information, and a multi-axis-gated multilayer perceptron module is added in this branch to extract the local and global information of infrared images in parallel. In the high resolution branch, the cross-attention module is used to propagate the global features learned on the low resolution branch to the high resolution branch, hence the high resolution branch can obtain stronger semantic information. The experimental results indicate that the segmentation accuracy of the algorithm on the dataset DNDS is better than that of the existing semantic segmentation algorithm, demonstrating the superiority of the proposed method.

**Key words:** multi resolution block, semantic segmentation, deepLabv3+, infrared image, attention module

## 0 引言

红外成像相较于可见光成像可以保存更多的信息<sup>[1]</sup>, 有较强的穿透力<sup>[2]</sup>, 具有不受恶劣环境(雨、雪、雾天气等)和光照的影响等优点<sup>[3]</sup>, 因此, 针对

红外图像语义分割的研究在自动驾驶<sup>[4]</sup>、姿态估计<sup>[5]</sup>、图像搜索<sup>[6]</sup>、医学诊断<sup>[7]</sup>等领域具有重要的应用价值。相较于传统的图像处理算法, 基于深度学习的算法因其具有自动特征提取、高级语义理解、端到端训练以及更好的泛化性能等优势已被用于图像语义分割<sup>[8]</sup>。

收稿日期: 2023-08-11; 修订日期: 2023-09-22。

作者简介: 徐慧琳(1999-), 女, 安徽安庆人, 硕士研究生, 研究方向: 目标分割、目标检测。E-mail: 2021201730@aust.edu.cn。

通信作者: 赵鑫(1991-), 男, 山西运城人, 讲师, 博士, 研究方向: 机器视觉。E-mail: zhaoxin@aust.edu.cn。

基金项目: 安徽省教育厅重点项目(KJ2020A0289); 淮南市科技计划项目(2020186); 安徽省教育厅重点项目(2022AH050801); 安徽理工大学青年教师科学基金(13200390)。

2017年, Jonathan Long 等人<sup>[9]</sup>提出了用于图像语义分割的全卷积神经网络(Fully Convolutional Networks, FCN)。之后,为了实现更高效且精细的语义分割,在FCN的基础上相继提出了SegNet<sup>[10]</sup>、U-Net<sup>[11]</sup>、PSPNet<sup>[12]</sup>、RefineNet<sup>[13]</sup>等模型,同期提出的DeepLabv1<sup>[14]</sup>使用条件随机场(Conditional random field, CRF)预测了更为精确的目标边界信息,但上述模型得到的语义分割结果比较粗糙,忽略了像素与像素之间的空间一致性关系。针对以上问题,DeepLabv2<sup>[15]</sup>在DeepLabv1的基础上增加了空洞空间金字塔池化(Atrous spatial pyramid pooling, ASPP)结构,达到较好的分割效果。DeepLabv3+<sup>[16]</sup>使用基于深度分离卷积的Xception改进DeepLabv2原有的基网络,同时加入简单高效的解码器模块,达到了目前最先进的分割效果。但红外图像的分辨率较低且目标边缘模糊<sup>[17]</sup>,采用目前的语义分割算法在处理红外图像相关任务时存在分割准确率不高的情况。

为进一步提高红外图像语义分割准确度,本文提出一种用于红外图像语义分割的多分辨率特征提取卷积神经网络算法(Multi-resolution feature extraction convolution neural network, MRFE-CNN)。该算法以DeepLabv3+为基准网络,通过引入一组对偶分辨率模块(Multi resolution block, MRBlock)对输入图像高分辨率和低分辨率的特征进行聚合,以更好地提取红外图像边缘信息,同时在该模块中加入一个多轴门控多层感知机模块(Multi-axis gated multi-layer perceptron block, MAGBlock)以并行捕获红外图像局部信息和全局信息。同时,本文建立了一个标准的红外图像语义分割数据集:昼夜分割数据集(Day and Night Dataset, DNDS)。

## 1 网络结构

### 1.1 DeepLabv3+网络架构

DeepLabv3+模型的网络结构如图1所示。首先,输入的图片通过卷积和ResNet网络进行特征提取。之后,利用ASPP模块提取多尺度的卷积特征,在该模块中多个空洞速率的空洞卷积通过采用级联和并行的方式聚合不同层次的上下文信息和保留更多的空间信息解决了多尺度分割对象的问题。由ASPP层输出的高级特征经过上采样后与经过第一个ResNetBlock处理的低级特征进行融合,最后,经过一个普通卷积层后再经过上采样输出语义分割结果。

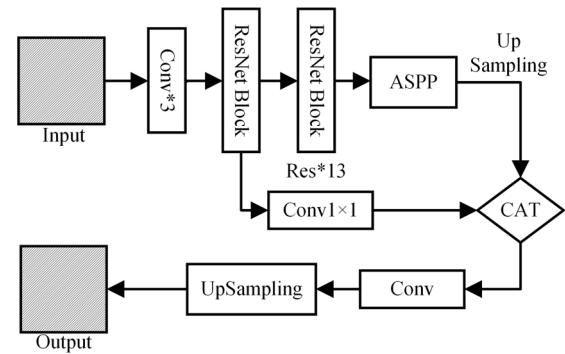


图1 DeepLabv3+网络结构

Fig.1 DeepLabv3+ network structure

ASPP模块网络结构如图2所示,该模块以主干网络提取的高级特征作为输入,经过具有不同空洞速率的空洞卷积,最后将结果融合,达到覆盖多尺度感受野的目的。假设 $H_k^r$ 表示卷积核大小为 $k$ 、空洞速率为 $r$ 的卷积操作,则ASPP模块用公式可表示为:

$$y = H_3^6(x) + H_3^{12}(x) + H_3^{18}(x) \quad (1)$$

式中:  $x$ 表示ASPP模块的输入;  $y$ 表示ASPP模块的输出。

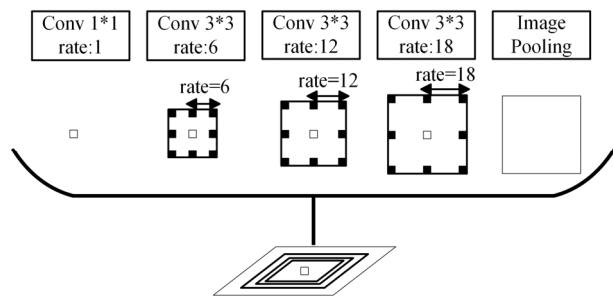


图2 ASPP模块网络结构

Fig.2 ASPP network structure

### 1.2 MRFE-CNN 网络架构

MRFE-CNN网络结构图如图3所示,该网络基于DeepLabv3+语义分割网络模型。原始DeepLabv3+网络结构在分割红外图像时,经过ASPP层的输出特征仅包含多尺度的卷积特征,无法有效提取低分辨率图像边缘信息。在此基础上,本文设计了MRBlock模块,该模块在保留输入红外图像多尺度信息的同时对高分辨率特征和低分辨率特征进行聚合,利用MAGBlock模块并行捕获局部信息和全局信息,实现对低分辨率红外图像边缘信息的提取。该模块将由ASPP模块输出再上采样后的高级特征与经过第一个ResNetBlock处理的低级特征进行融合,其输出作为高级特征经过上采样再与低级特征进行融合,最后经过一个普通卷积后再上采样输出。

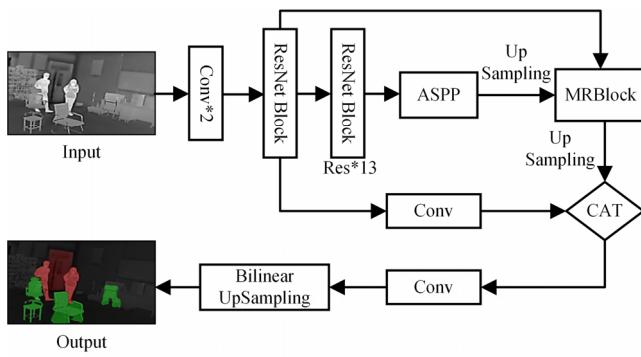


图3 MRFE-CNN 网络结构

Fig.3 MRFE-CNN network structure

### 1.3 对偶分辨率模块 MRBlock

为解决现有算法在分割低分辨率红外图像时无法准确分割其边缘的问题,本文借鉴 RTFormer<sup>[18]</sup>的设计思想构建了一种对偶分辨率模块 MRBlock, MRBlock 结构如图 4 所示,其由双分辨率分支组成,分别为低分辨率 (Low\_level\_feature) 分支与高分辨率 (High\_level\_feature) 分支,同时包含两种类型注意力模块。

经过第一个 ResNetBlock 处理的低级特征保留了低分辨率图像较多的边缘信息,将其作为低分辨率分支的输入,为保留图像边缘信息同时提取整体信息,在该分支采用了一个多轴门控 MAGBlock 以并行地捕获局部和全局信息,同时在该分支采用了 GPU 友好注意力模块 (GPU-Friendly Attention, GFA),使得网络可以保持多头机制中学习多样特征的能力。

经过 ASPP 模块后输出的高级特征作为高分辨率分支的输入,在该分支中采用跨分辨率注意力机制 (Cross-Attention, CRA) 对高层全局上下文信息进行传播扩散,使模型在提取全局信息的同时不丢失图像边缘信息。

MRBlock 模块将包含边缘信息最多的低级特征与包含全局信息最多的高低特征多次进行聚合,有效提升了网络提取低分辨率红外图像特征能力。

#### 1.3.1 GPU 友好注意力模块 GFA

首先,MRBlock 模块中的 GFA 通过使用普通矩阵乘操作替换多头机制中的分组矩阵乘操作,实现了更适合在 GPU 上运行的注意力计算方式。同时 GFA 引入了组归一化,这使得网络可以保持多头机制中学习多样特征的能力。此外,由于普通矩阵乘更适合 GPU 推理, GFA 可以一定程度扩大参数数量,提升网络容量。

设  $X \in \mathbb{R}^{N \times d}$  表示输入特征,其中  $N$  是元素 (或图像中的像素) 的数量,  $d$  是特征维数,原始的外部注意力模块 (External Attention, EA) 计算公式如下:

$$EA(X, K, V) = DN(X \cdot K^T) \cdot V \quad (2)$$

式中:  $K, V \in \mathbb{R}^{M \times d}$  为可学习参数;  $M$  是参数维度, DN 是双重归一化运算。多头 EA (Multi-head EA) 计算公式如下:

$$MHEA(X) = \text{Concat}(h_1, h_2, \dots, h_H) \quad (3)$$

$$h_i = EA(X_i, K', V'), i \in [1, H] \quad (4)$$

式中:  $K', V' \in \mathbb{R}^{M \times d'}$ ,  $d' = d/H$ ,  $H$  是头的数量;  $X_i$  是  $X$  的第  $i$  个头, Concat 为拼接函数。

多头注意力模块 Multi-head EA 和 GFA 模块如图 5 所示。

High\_level\_feature Low\_level\_feature

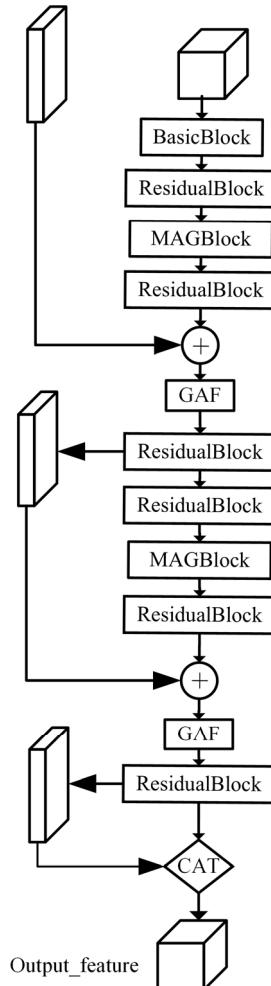


图4 MRBlock 结构

Fig.4 MRBlock structure

标准的自注意力<sup>[19]</sup>有 2 次矩阵乘法,在输入时通过通道切片得到多个头计算后再拼接到一起即多头。GFA 不切片直接分别与  $K$  矩阵计算后求取注意力,简化了操作,同时很好地利用了 GPU 大型矩阵并行速度快的优势。GFA 计算公式如下,  $K$  矩阵和  $V$  矩阵为可学习矩阵:

$$GFA(X, K_g, V_g) = GDN(X \cdot K_g^T) \cdot V_g \quad (5)$$

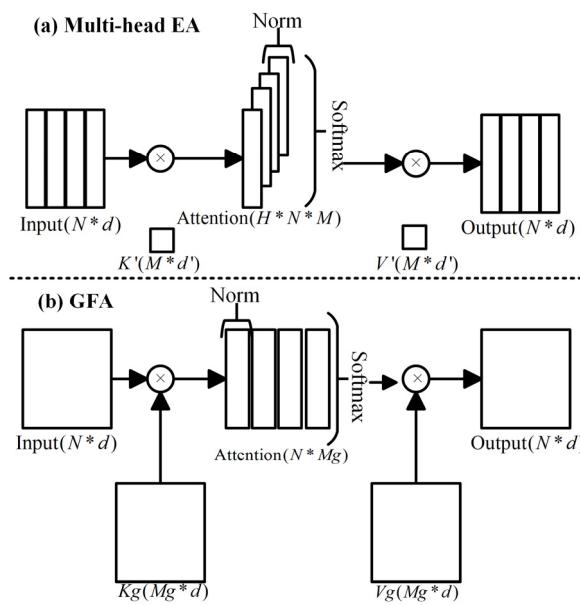


图5 多头注意力模块 Multi-head EA(a)和 GFA(b)

Fig.5 Multi-head EA (a) and GFA (b)

式中:  $K_g, V_g \in \mathbb{R}^{mg \times d}$ ,  $M_g = M \times H$ , GDN 表示双组重归一化, 它将原始双重归一化的第二次归一化拆分为  $H$  组。

### 1.3.2 跨分辨率注意力机制 CRA

在高分辨率分支上, 通过跨分辨率的跨注意力机制将低分辨率分支上学习到的全局特征传播到高分辨率分支上, 从而使得高分辨率分支获得更强的语义信息。该过程用公式可描述如下:

$$CA(X_h, K_c, V_c) = \text{Soft max}\left(\frac{X_h \cdot K_c^T}{\sqrt{d_h}}\right) \cdot V_c \quad (6)$$

$$K_c, V_c = \phi(X_c) \quad X_c = \theta(X_l) \quad (7)$$

式中:  $X_h$ 、 $X_l$  分别表示高分辨率分支和低分辨率分支上的特征映射;  $\phi$  为一组矩阵操作, 包括分裂, 置换和重塑;  $d_h$  为高分辨率分支的特征维数。 $\theta$  是由池化层和卷积层组成的函数,  $X_c$  是  $X_l$  通过  $\theta$  函数计算出来的交叉特征,  $K_c$  和  $V_c$  是两组可学习矩阵。

### 1.3.3 多轴门控模块 MAGBlock

MRBlock 中的多轴门控<sup>[20]</sup>模块 MAGBlock 如图6 所示, 该模块可以并行地捕获局部和全局信息。在局部分支中, 大小为( $H, W, C/2$ )的特征的半头被阻塞为形状张量( $H/b \times W/b, b \times b, C/2$ ), 在全局分支中, 使用固定的( $d \times d$ )网格将另一半头部网格化为形状( $d \times d, H/d \times W/d, C/2$ ), 每个窗口具有大小( $H/d \times W/d$ )。同时为了使其完全卷积, 只在每个分支的单个轴上应用门控 MLP (Gate MLP, gMLP) 模块, 同时在其它空间轴上共享参数。直观地说, 并行应用多轴 gMLP 分别对应于空间信息的局部和全局混合。

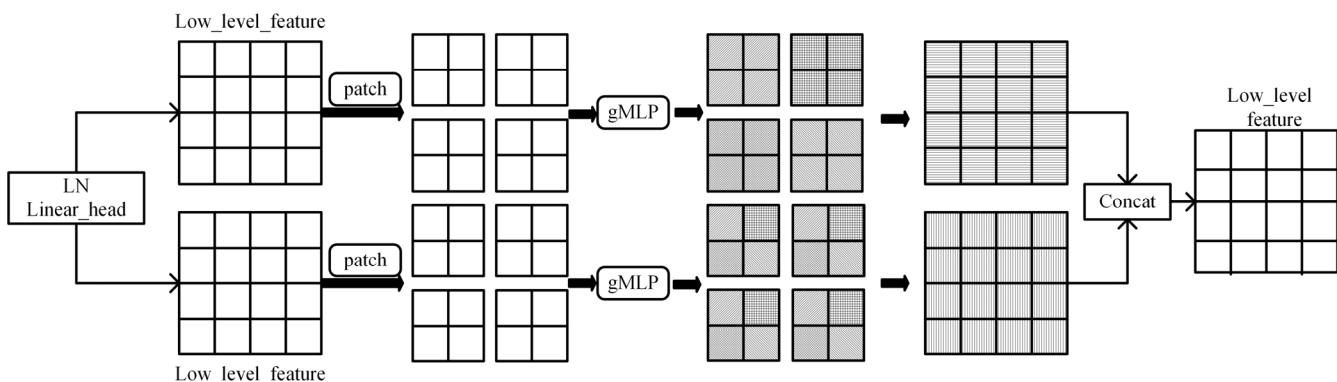


图6 MAGBlock 结构

Fig.6 MAGBlock structure

在 MRBlock 低分辨率分支引入 MAGBlock, 有效提取低分辨率红外图像局部信息和全局信息, 增强了提取红外图像边缘特征能力。

### 1.4 损失函数

本文采用交叉熵函数作为损失函数<sup>[21]</sup>, 其公式表示为:

$$L = -\frac{1}{n} \sum_i^n [y_i \ln G_\theta(X_i | Z) + (1 + y_i) \ln (1 - G_\theta(X_i | Z))] \quad (8)$$

式中:  $G_\theta(X_i | Z)$  是本文网络中参数为  $\theta$  的像素  $i$  处的标签分配概率;  $y_i$  是真实值标签。

## 2 数据集

本文建立了一个标准的红外图像语义分割数据集: 昼夜分割数据集(Day and Night Dataset, DNDS)。DNDS 数据集包含 3264 张红外图像, 由 HIKVISION 热成像双光谱半球摄像机(DS-2TD1217-3/PA)获取, 包含 15 个语义标签, 包括: 人、椅子、建筑物、树、

交通锥、窗户、门、桌子、显示器、道路、汽车、空调、自行车、路灯和背景标签。所获取的图像从现实场景中的不同视角以及不同遮挡度等条件下采集。标注工具为开源标注工具 labelme。在标注完所有图片

后,进行第二轮检查,剔除不标准的图片,最终筛选出3264张可用且标注良好的图片。部分数据集图片展示如图7所示。

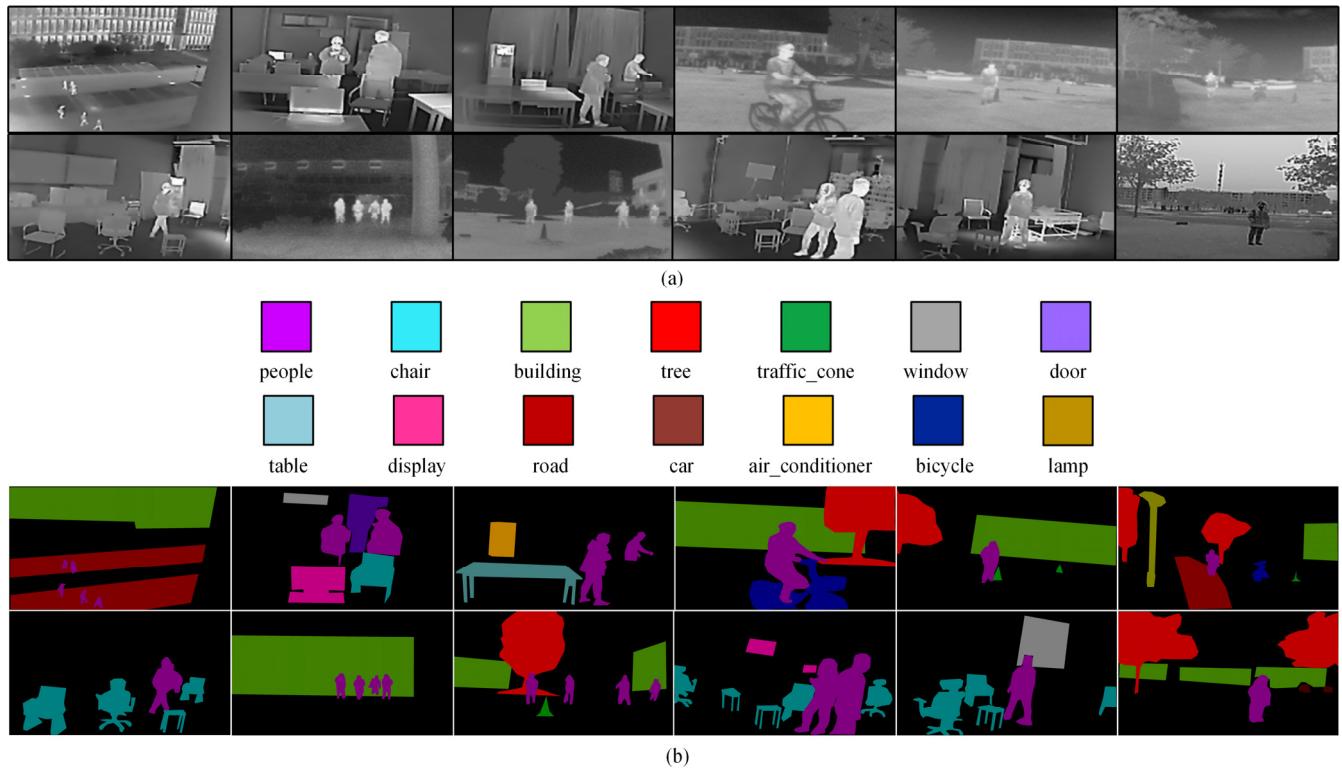


图7 DNDS数据集部分图片展示。(a)为红外图像, (b)为真实语义标签

Fig.7 Some pictures of the DNDS. (a) infrared image, (b) real semantic label

DNDS数据集各类标签数量如图8所示。

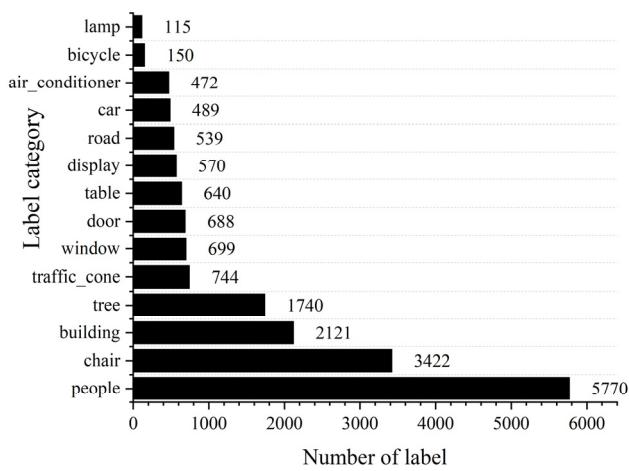


图8 DNDS数据集各类标签数量

Fig.8 The number of labels in the DNDS

本文所使用的DNDS数据集划分为训练集,验证集和测试集,比例为1:0.8:0.2,训练集包含1632张图像,验证集包含1305张图像,测试集包含327张图像。

### 3 实验内容

本文算法基于PyTorch框架,实验硬件配置如表1所示。

表1 实验硬件配置

Table 1 Experimental hardware configuration

CPU	GPU	Memory	System
2.60GHz Intel Xeon Platinum 8350C CPU	NVIDIA RTX 3090	24GB	Linux

#### 3.1 参数设置

本文网络预训练权重为mobilenetv2<sup>[22]</sup>,初始学习率设置为0.007,训练批次设置为200、动量和权重衰减分别设置为0.9和0.0001。批处理大小设置为4,输入图像的大小缩放为1024×1024。

#### 3.2 评价指标

本文采用的语义分割评价指标为类别平均像素准确率(Mean Pixel Accuracy, MPA)和平均交并比

(Mean Intersection over Union, MIOU) [23]。

MPA 是指语义分割中所有类别像素准确率之和的平均。计算公式如下:

$$\text{MPA} = \frac{1}{n_c} \sum_i \frac{n_{ii}}{N_i} \quad (9)$$

MIOU 是指分别对每个类别计算真实标签和预测结果的交并比, 然后再对所有类别的交并比求均值。计算公式如下所示:

$$\text{MIOU} = \frac{1}{n_c} \sum_i \frac{n_{ii}}{N_i + \sum_j n_{ji} - n_{ii}} \quad (10)$$

表 2 MRBlock 模块消融实验

Table 2 Ablation experiment of MRBlock

Models	Residual Block	Basic Block	MAGBlock	ASPP	Input_feature	MPA%	MIOU%
Model_1	√	-	-	-	x_l	90.8	81.07
Model_2	√	√	-	-	x_l+x_h	91.13	81.7
Model_3	√	√	*1	-	x_l+x_h	91.91	82.6
Model_4	√	√	*2	-	x_l+x_h	92.36	83.43
Model_6	√	√	*2	√	x_l+x_h	92.72	84.1
Model_7	√	√	*2	√	x_l	92.31	82.1
Model_8	√	√	*2	√	x_h	91.1	83.7

### 3.4 整体性能

表 3 显示了本文算法与目前常用的 6 种算法在数据集 DNDS 上两个客观评价指标数值。包括 FCN-8s, FCN-16s, FCN-32s, U-Net, DUC<sup>[24]</sup>, DeepLabv3+。可以看出, 本文提出的方法所有指标值均取得了最好值, 相较于 DeepLabv3+ 算法 MIOU 提升了 1.8%, MPA 提升了 2.32%, 证明了提出 MRBlock 的有效性。

表 3 流行算法性能比较

Table 3 Performance comparison of popular algorithms

Methods	MPA%	MIOU%
FCN-8s	8.16	5.74
FCN-16s	63.32	48.3
FCN-32s	58.37	42.32
U-Net	78.12	72.35
DUC	78.35	69.61
DeepLabv3+	90.4	82.3
MRFE-CNN	92.72(+2.32)	84.1(+1.8)

本文提出方法与 DeepLabv3+ 算法在 DNDS 训练

式中:  $n_c$  为总类别数;  $n_{ij}$  表示实际类别为  $i$ 、被预测类别为  $j$  的像素数量;  $N_i$  为类  $i$  的像素总数,  $N$  为总像素数量。MIOU 通常为语义分割最终的评价标准。

### 3.3 消融实验

为验证对偶分辨率模块 MRBlock 的有效性, 本文通过在原 DeepLabv3+ 的基础上构建不同结构的 MRBlock 对 DNDS 红外数据集进行分割, 结果见表 2。可以看出 model\_6 中, 在 MRBlock 中加入 BasicBlock、MAGBlock, 同时将高级特征  $x_h$  与低级特征  $x_l$  共同作为该模块的输入时, 评价指标 MPA、MIOU, 由此可验证 MRBlock 的有效性。

集上训练的 LOSS 变化曲线图与 MIOU 变化曲线图如图 9 所示。

可以看出, 相较于 DeepLabv3+ 算法, 本文提出的方法在训练时 LOSS 收敛速度更快, MIOU 均为领先。

MRFE-CNN 与 DeepLabv3+ 在 DNDS 测试集上的预测效果如图 10 所示。通过对比测试集图片分割结果, 可以看出, DeepLabv3+ 分割图片结果存在边缘模糊、纹理信息丢失严重的问题, 而 MRFE-CNN 分割图片结果对应位置相对 DeepLabv3+ 分割效果更佳, 明显观察到分割目标更加准确, 边缘分割更加完整, 因此本文所提出的算法相较于 DeepLabv3+ 可以输出更精准的分割结果。

为进一步验证本文提出算法的有效性, 在公共红外图像数据集 MSRS 上与 DeepLabv3+ 进行比较, 训练配置与 DNDS 数据集相同, 其中划分训练集图片 722 张, 验证集图片 577 张与测试集图片 145 张。获得的相关评价指标数据如表 4 所示。

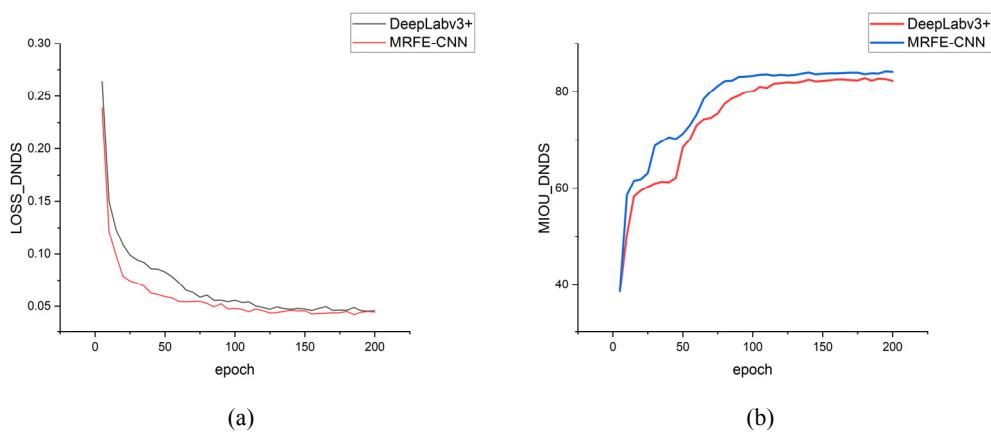


图9 训练指标比较。(a)LOSS 变化曲线图, (b)MIOU 变化曲线图

Fig.9 Comparison of training indicators . (a) LOSS curves graph, (b) MIOU curves graph

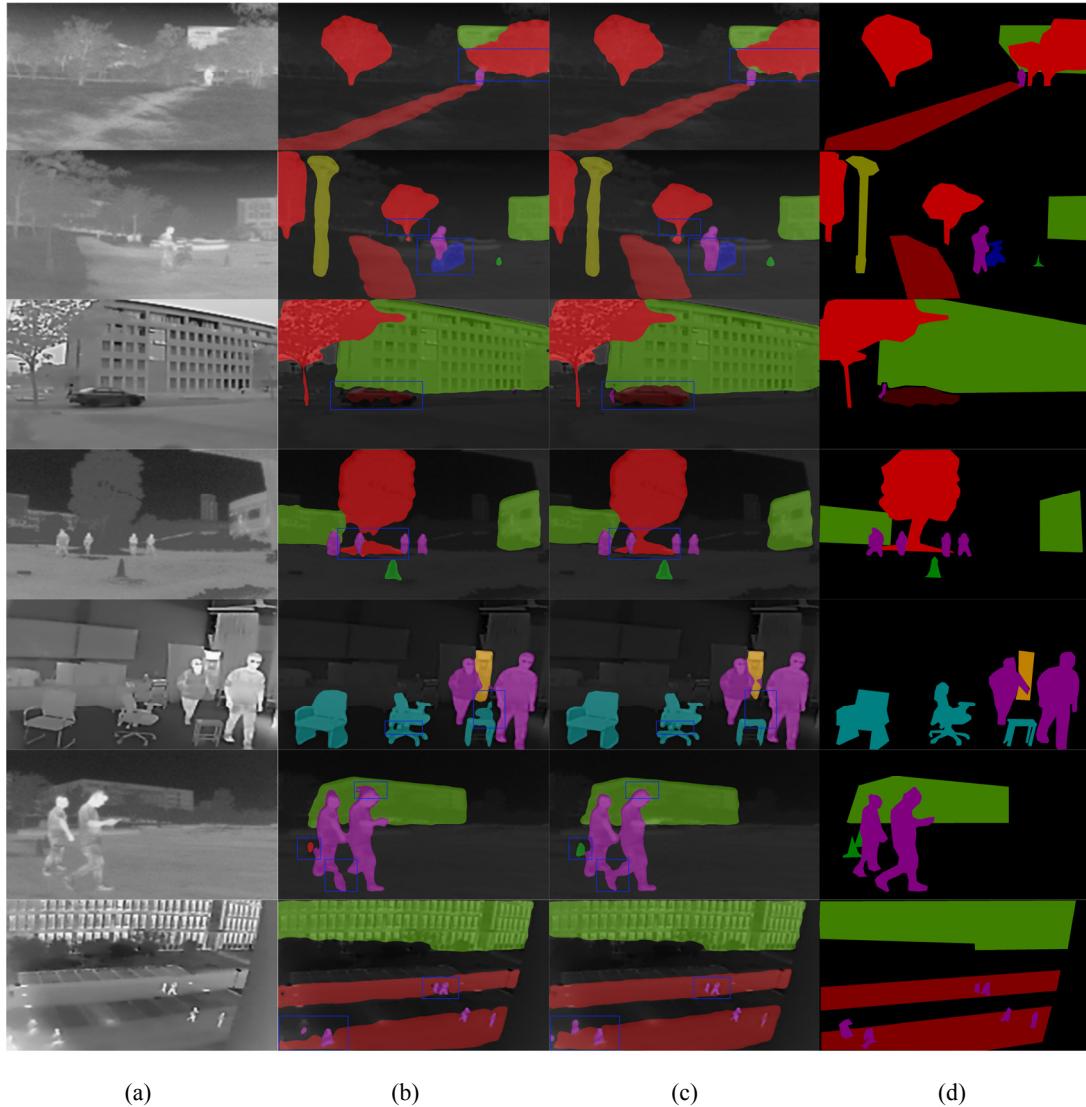


图10 测试集结果比较。(a)原图; (b)DeepLabv3+; (c)MRFE-CNN; (d)真实标签

Fig.10 Comparison of results. (a) Original image, (b) DeepLabv3+, (c) MRFE-CNN, (d) Real label

表 4 公共数据集 MSRS 性能比较

Table 4 Performance comparison of MSRS

Methods	MPA%	MIOU%
DeepLabv3+	68.43	56.37
MRFE-CNN	71.82(+3.39)	58.14(+1.77)

可以看出, 在公共红外图像数据集 MSRS 上, 本文提出的方法相较于 DeepLabv3+ 算法 MIOU 提升了 1.77%, MPA 提升了 3.39%, 证明了提出算法的有效性。

本文提出方法与 DeepLabv3+ 算法在 MSRS 数据

集训练集上训练的 LOSS 变化曲线图与 MIOU 变化曲线图如图 11 所示。可以看出, 在该数据集上, 相较于 DeepLabv3+ 算法, 本文提出方法在训练时 LOSS 收敛速度更快, MIOU 更高。

MRFE-CNN 与 DeepLabv3+ 在公共红外数据集 MSRS 上的预测效果如图 12 所示。通过对比图片分割结果, 可以看出, 而 MRFE-CNN 分割图片结果对应位置相对 DeepLabv3+ 分割效果更佳, 进一步验证了本文提出算法的有效性。

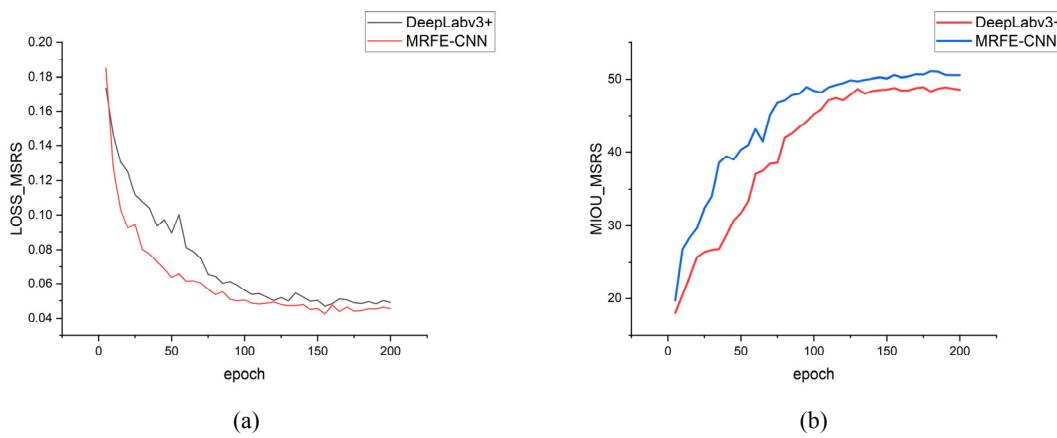


图 11 训练指标比较。(a)LOSS 变化曲线图, (b)MIOU 变化曲线图

Fig.11 Comparison of training indicators . (a) LOSS curves graph, (b) MIOU curves graph

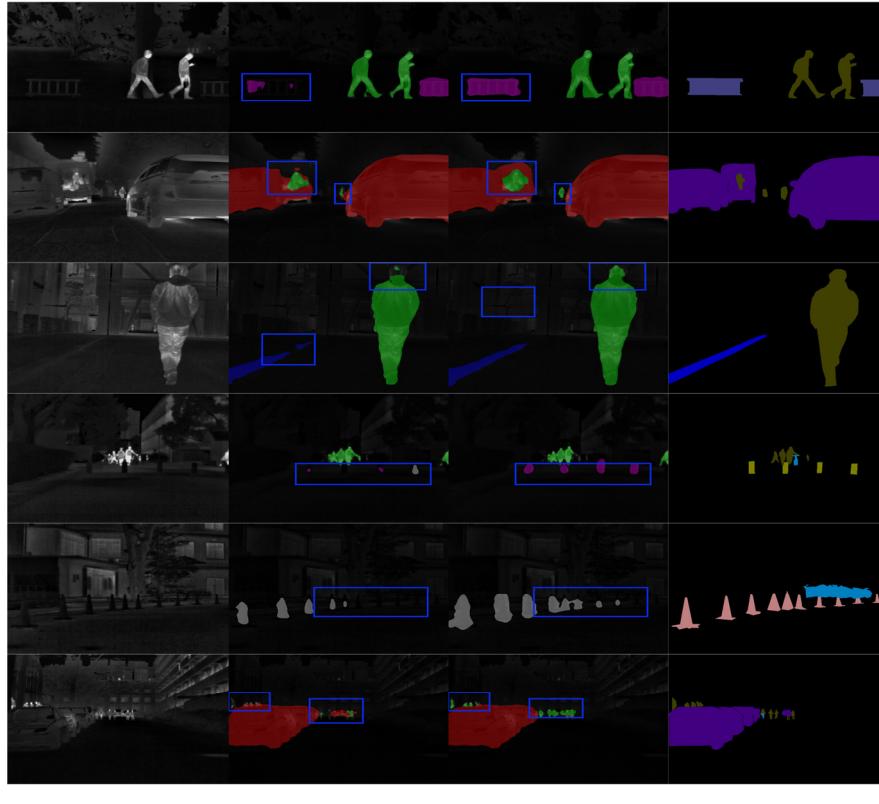


图 12 测试集结果比较。(a)原图; (b)DeepLabv3+; (c)MRFE-CNN; (d)真实标签

Fig.12 Comparison of results. (a) Original image, (b) DeepLabv3+, (c) MRFE-CNN, (d) Real label

## 4 结语

针对现有算法无法准确分割低分辨率红外图像、容易丢失边缘信息的问题，本文以 DeepLabv3+为基准网络，提出了一种多分辨率特征提取算法 MRFE-CNN，其中采用对偶分辨率模块 MRBlock 有效提取红外图像边缘信息，提升了现有算法精确度，达到比 DeepLabv3+更好的分割效果。

未来将致力于解决红外图像语义分割在实际生产应用中存在实时性不高的问题，进而推动红外技术在电力领域、智能制造业等行业以及在夜晚及恶劣天气下自动驾驶与智能环境监控等方面的应用和发展。

## 参考文献：

- [1] 刘致驿, 孙韶媛, 任正云, 等. 基于改进 DeepLabv3+的无人车夜间红外图像语义分割[J]. 应用光学, 2020, 41(1): 180-185.  
LIU Zhiyi, SUN Shaoyuan, REN Zhengyun, et al. Semantic segmentation of nocturnal infrared images of unmanned vehicles based on improved DeepLabv3+[J]. *Journal of Applied Optics*, 2020, 41(1): 180-185.
- [2] 夏威. 基于卷积神经网络的热红外图像语义分割研究[D]. 合肥: 安徽大学, 2020.  
XIA Wei. Thermal Image Semantic Segmentation Based on Convolutional Neural Networks[D]. Hefei: Anhui University, 2020.
- [3] 景庄伟, 管海燕, 彭代峰, 等. 基于深度神经网络的图像语义分割研究综述[J]. 计算机工程, 2020, 46(10): 1-17.  
JING Zhuangwei, GUAN Haiyan, PENG Daifeng, et al. Survey of research in image semantic segmentation based on deep neural network[J] *Computer Engineering*, 2020, 46(10): 1-17.
- [4] ZHAO L, WANG M, YUE Y. Sem-aug: improving camera-lidar feature fusion with semantic augmentation for 3d vehicle detection[J]. *IEEE Robotics and Automation Letters*, 2022, 7(4): 9358-9365.
- [5] WANG J, LIU L, LU M, et al. The estimation of broiler respiration rate based on the semantic segmentation and video amplification[J]. *Frontiers in Physics*, 2022, 10: 1-13.
- [6] XUE Z, MAO W, ZHENG L. Learning to simulate complex scenes for street scene segmentation[J]. *IEEE Transactions on Multimedia*, 2021, 24: 1253-1265.
- [7] WANG Y, TIAN S, YU L, et al. FSOU-Net: Feature supplement and optimization U-Net for 2D medical image segmentation[J]. *Technology and Health Care*, 2023, 31(1): 181-195.
- [8] 郭尹. 基于深度学习的电力设备热红外图像语义分割方法研究[D]. 合肥: 安徽大学, 2022.  
GUO Yin. Research on Electrical Thermal Image Semantic Segmentation Method Based on Deep Learning[D]. Hefei: Anhui University, 2022.
- [9] 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.
- [10] Adrinarayanan 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.
- [11] OLAF R, PHILIPP F, THOMAS B. U-Net: Convolutional networks for biomedical image segmentation[J]. *CoRR*, 2015, abs/1505.04597.
- [12] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2017: 2881-2890.
- [13] LIN G, MILAN A, SHEN C, et al. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1925-1934.
- [14] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFS[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 40(4): 834-848.
- [15] CHEN L C, ZHU Y, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 801-818.
- [16] Chollet F. Xception: Deep learning with depthwise separable convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1251-1258.
- [17] 练铮, 张宝辉, 江云峰, 等. 基于语义分割的红外图像增强方法[J]. 红外技术, 2023, 45(4): 394-401.  
LIAN Zheng, ZHANG Baohui, JIANG Yunfeng, et al. An infrared image enhancement method based on semantic segmentation[J]. *Infrared Technology*, 2023, 45(4): 394-401.
- [18] WANG J, GOU C, WU Q, et al. RTFormer: efficient design for real-time semantic segmentation with transformer[J]. arXiv e-prints, 2022: arXiv: 2210.07124.
- [19] VASWANI A, SHAZER N, PARMAR N, et al. Attention is all you need[J]. arXiv, 2017. DOI:10.48550/arXiv.1706.03762.
- [20] TU Z, TALEBI H, ZHANG H, et al. Maxim: Multi-axis MLP for image processing[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 5769-5780.
- [21] JADON S. A survey of loss functions for semantic segmentation[C]// IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). IEEE, 2020: 1-7.
- [22] Sandler M, Howard A, ZHU M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4510-4520.
- [23] 于营, 王春平, 付强, 等. 语义分割评价指标和评价方法综述[J]. 计算机工程与应用, 2023, 59(6):13.  
YU Ying, WANG Chunping, FU Qiang, et al. Survey of evaluation metrics and methods for semantic segmentation[J]. *Journal of Computer Engineering & Applications*, 2023, 59(6): 13.
- [24] WANG P, CHEN P, YUAN Y, et al. Understanding convolution for semantic segmentation[C]//IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2018: 1451-1460.