

# 基于四通道不可分加性小波与DeepLabv3+结合的语义分割模型

刘 斌, 潘 蜜

湖北大学计算机与信息工程学院, 湖北 武汉

收稿日期: 2023年6月18日; 录用日期: 2023年7月9日; 发布日期: 2023年7月20日

## 摘 要

为了改善传统语义分割模型中因丢失细节, 从而导致信息下降的问题, 我们提出了一种改进的DeepLabv3+网络分割模型。首先将主干网络替换为MobileNetV2网络; 其次通过构造四通道不可分小波低通滤波器, 对源图像进行分解, 提取源图像的高频子图; 再次, 将普通卷积更换为深度可分离卷积并且引入卷积注意力模块(CBAM)自适应细化特征, 从而提高网络模型的分割效果。实验结果表明, 改进后的模型在VOC数据集上均交并比(mean intersection over union, MIoU)比原始的DeepLabv3+模型提高0.94%, 平均像素精度(mean pixel accuracy, MPA)比原始DeepLabv3+模型提高了1.34%, 准确度比原始DeepLabv3+模型提高0.19%。在BDD100K数据集上均交并比比原始的DeepLabv3+模型提高0.53%, 平均像素精度比原始DeepLabv3+模型提高了0.15%, 准确率比原始DeepLabv3+模型提高0.13%。在主观和客观结果上均显示我们的模型优于原模型。

## 关键词

语义分割, 空洞空间卷积池化金字塔, 不可分小波, 加性小波, 卷积注意力模块

## Semantic Segmentation Model Based on Four Channel Non-Separable Additive Wavelet Combined with DeepLabv3+

Bin Liu, Mi Pan

School of Computer Science, Hubei University, Wuhan Hubei

Received: Jun. 18<sup>th</sup>, 2023; accepted: Jul. 9<sup>th</sup>, 2023; published: Jul. 20<sup>th</sup>, 2023

文章引用: 刘斌, 潘蜜. 基于四通道不可分加性小波与 DeepLabv3+结合的语义分割模型[J]. 图像与信号处理, 2023, 12(3): 279-289. DOI: 10.12677/jisp.2023.123028

## Abstract

In order to improve the loss of details in the traditional semantic segmentation model, which leads to the decline of information, we propose an improved DeepLabv3+ network segmentation model. Firstly, replace the backbone network with the MobileNetV2 network. Secondly, the source image is decomposed by constructing a four-channel non-separable wavelet low-pass filter, and the high-frequency subimage of the source image is extracted. Thirdly, the common convolution is replaced by deep separable convolution and the adaptive refinement feature of convolutional attention module (CBAM) is introduced to improve the segmentation effect of the network model. The experimental results show that on the VOC data set, the mean intersection over union (MIOU) of the improved model is 0.94% higher than that of the original DeepLabv3+ model, the mean pixel accuracy (MPA) is 1.34% higher than the original DeepLabv3+ model, and the accuracy is 0.19% higher than the original DeepLabv3+ model. On the BDD100K data set, mean intersection over union is 0.53% higher than the original DeepLabv3+ model. The DeepLabv3+ mean pixel accuracy is 0.15% higher than the original DeepLabv3+ model, and the accuracy is 0.13% higher than the original DeepLabv3+ model. Both subjective and objective results show that our model is better than the original model.

## Keywords

Semantic Segmentation, Atrous Spatial Pyramid Pooling, Non-Separable Wavelet, Additive Wavelet, Attention Mechanism, Convolutional Attention Module

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

语义分割作为计算机视觉中的一个重要分支, 通过为每个像素进行密集的预测推断标签来实现细粒度推理, 从而实现图像像素级的分类。语义分割任务就是从低层语义向高层语义推理的过程。目前语义分割在现实生活, 如自动驾驶[1] [2] [3], 人机交互[4], 计算摄影[5], 图像搜索引擎[6]以及增强现实技术等[7]具有广泛的应用。

在图像处理的早期, 传统图像分割方式包括基于阈值的图像分割方法、基于边缘的图像分割方法[8] [9] [10]、基于区域的图像分割方法[11] [12] [13]、基于聚类的图像分割方法[14] [15] [16]和基于图论的图像分割方法[17] [18] [19]。这些传统图像分割方法大多数是利用图像的表层信息进行分析处理, 因此对于需要“处理”大量语义信息的图像分割任务并不适用。而基于小波变换的图像分割技术具有较强的抗噪声性能, 并且能够有效地保留边缘信息, 以达到分割的目的。

随着深度学习技术的不断发展, 涌现出了大量基于深度学习的高效语义分割方法。其中 Long 等人首次将卷积神经网络(CNN)的全连接改为卷积操作, 得到全卷积神经网络(FCN) [20]。FCN 对图像进行像素级的分类, 从而解决了语义级别的图像分割问题, 被称为深度学习用于语义分割的开创之作。随后 SegNet 网络[21], UNet [22], PSPNet [23], DeepLab 系列[24]等经典网络相继被提出。然而 FCN 和 SegNet 模型不够精细, 没有充分利用上下文关系, UNet 网络冗余太大, PSPNet 网络得到的结果不够精细。DeepLab 系列相比这些网络具有准确度高、速度快、参数量小和感受野大等优点。其中, DeepLabv3+利用空间金

字塔模块和 encoder-decoder 来实现语义分割。为了增强分割结果, Xu 等人提出加入通道注意力机制模块以增强分割结果的方法[25]; Zhen 等人采用更密集的特征池化模块有效聚合多尺度特征,同时使用深度可分离卷积降低网络计算复杂度[26]; Ma 等人提出了使用 DeepLabv3+网络识别时频分布中故障特征的方法,对采集到的滚动轴承振动信号使用短时傅里叶变换得到时频分布[27]; Zhang 等人提出了一种基于 DeepLabv3 架构的小波域 DeepLabv3-MRF (Markov random field, MRF)算法,从而获得更为清晰的边缘细节信息[28]。

由于经典方法和深度学习方法不足,为了提高 DeepLabv3+网络分割图像效果,本文提出一种基于改进 DeepLabv3+模型的语义分割网络,将轻量级网络 MobileNetV2 [29]作为主干网络,利用不可分小波具有各向同性的特点,将加性小波加入模型中进行各个方向的特征提取。实验结果显示,相比于原模型,我们的模型在主观视觉分割效果以及客观指标上都取得了更好的结果。

## 2. 相关理论介绍

### 2.1. 二维不可分小波

Liu 等人[30] [31]对不可分小波进行了广泛的研究,并成功地将其应用于图像处理的不同领域,如图像分割、图像融合和图像增强等。Chen 等人以高维小波的多尺度分析为基础,提出了构造具有正交性和紧支撑的高维非张量积小波滤波器组[32],其构造的高维低通滤波器频域形式如下:

$$m_0(\xi) = \frac{1}{s} X(\xi) \prod_{j \in Z_n} (U_j D(A^T \xi)) \left( \prod_{j \in Z_n} U_{N-1-j}^T \right) V_0, \quad \xi \in R^d \quad (1)$$

对应的  $s-1$  个相应的正交共轭滤波器(CQF)的形式如下:

$$m_j(\xi) = \frac{1}{s} X(\xi) \prod_{l \in Z_n} (U_l D(A^T \xi)) \left( \prod_{l \in Z_n} U_{N-1-l}^T \right) V_j, \quad \xi \in R^d, (j=1, 2, \dots, s-1) \quad (2)$$

Liu 等人在此基础上构建了二维四通道滤波器组[30],假设小波变换的伸缩矩阵为 $[2, 0; 0, 2]$ ,则滤波器组的形式可构造为:

$$[m_0(x, y), m_1(x, y), m_2(x, y), m_3(x, y)] = \frac{1}{4} X(1, x, y, xy) \prod_{j=1}^k [U_j D(x^2, y^2) U_j^T] V \quad (3)$$

其中  $x = e^{-i\omega_1}$ ,  $y = e^{-i\omega_2}$ ,  $U_j (j=1, 2, \dots, k)$  以及  $V/2 = (V_0, V_1, V_2, V_3)/2$  为正交阵,  $V_1, V_2, V_3$  为  $4 \times 1$  向量,  $V_0 = (1, 1, 1, 1)^T$ 。

### 2.2. 加性小波

近些年来小波分析迅速发展,在信号处理、图像识别、计算机视觉和数据压缩等方面具有广泛的应用。二维小波可以分为张量积小波和不可分小波,张量积小波并不具有多方向性,无法处理图像中多方向的信息。与张量积小波相比,不可分小波可以提取到图像各个方向的信息,具有良好的适应性。

加性小波是利用低通滤波器对图像进行 atrous 分解,具体过程如下:

$$\begin{aligned} H_0(p_0) &= p_1, w_1 = p_0 - p_1, \\ H_0(p_1) &= p_2, w_2 = p_1 - p_2, \\ H_0(p_2) &= p_3, w_3 = p_2 - p_3, \\ &\vdots \end{aligned} \quad (4)$$

其中  $w_i$  为小波平面,也就是对应图像的高频信息;  $p_i$  为低频分量,也就是对应图像的低频信息;  $p_r$  为

残余图像。重构公式如下:

$$p_0 = \sum_{i=1}^n w_i + p_r \tag{5}$$

### 2.3. 注意力机制

CBAM [33]是一种轻量级通用的模块, 同时也是一个即插即用模块, 可以嵌套在任何 CNN 架构中, 它结合了空间与通道注意力机制模块。CBAM 包含 CAM 和 SAM 两个子模块。输入特征  $F \in R^{C \times H \times W}$ , 再将通道注意力模块进行一维卷积  $M_c \in R^{C \times 1 \times 1}$ , 然后将卷积结果与原图相乘, 以 CAM 的输出结果作为输入, 进行空间注意力模块的二维卷积  $M_s \in R^{1 \times H \times W}$ , 再将输出结果与原图相乘, 具体过程如下:

$$F' = M_c(F) \otimes F \tag{6}$$

$$F'' = M_s(F') \otimes F' \tag{7}$$

通道注意力机制类似于 SENet [34], 其主要思想是增加有效通道的权重, 降低无效通道的权重。公式表示为:

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) = \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c))) \tag{8}$$

其中  $F_{avg}^c$  表示平均池特征,  $F_{max}^c$  表示最大池特征,  $W_0 \in R^{c/r \times c}$ ,  $W_1 \in R^{c \times c/r}$  表示共享网络层的两层神经网络。在网络层  $W_0$  后面, 使用函数 ReLU 作为激活函数,  $\sigma$  表示 Sigmoid 函数。为了减少参数开销, 隐藏的激活大小设置为  $R^{c/r \times 1 \times 1}$ , 其中  $r$  是压缩率, 本文实验中  $r$  取值为 16。

通道注意机制关注的是通道层面的哪些层具有更强的反馈能力, 空间注意机制则是增强特定区域的特征。空间注意机制的表述如下:

$$M_s(F) = \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) = \sigma(f^{7 \times 7}([F_{avg}^s; F_{max}^s])) \tag{9}$$

使用平均池化和最大池化来评估信息, 并使用卷积进行提取。最后, 用 sigmoid 层进行归一化。通道注意机制和空间注意机制示意图见图 1。

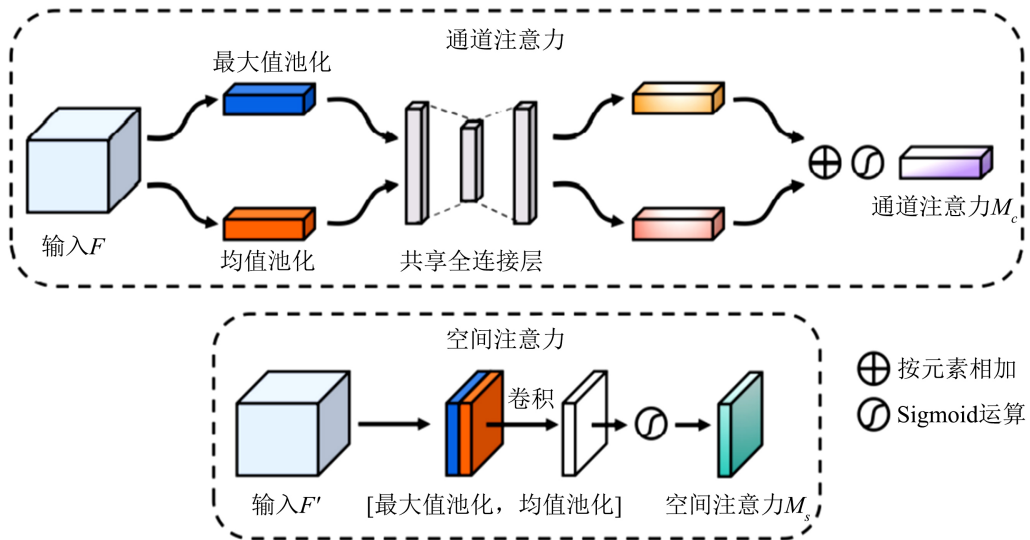


Figure 1. Channel attention and Spatial attention structure in CBAM  
 图 1. CBAM 中通道注意力和空间注意力结构

## 2.4. DeepLabv3+

DeepLabv3+是一个经典的语义分割模型, 是经典的编码器-解码器结构, 为了使得模型具有轻量级特点, 且提高模型训练速度, 这里我们选用 MobileNetV2 为主干网络。经过主干网络进行特征提取后, 进入 ASPP (空洞空间卷积池化金字塔), 其中 ASPP 包括  $1 \times 1$  卷积, 空洞率分别为 6、12、18 的空洞卷积和全局平均池化[35]。然后将其进行堆叠, 再经过  $1 \times 1$  卷积从而实现降低特征图的通道数。在 decoder 部分, 主干网络会将低级特征(Low-Level Features)经过  $1 \times 1$  卷积后的特征与 decoder 部分得到的特征图进行 4 倍上采样, 再进行拼接融合, 融合得到的特征图进行  $3 \times 3$  卷积后进行 4 倍上采样, 最终得到语义分割后的图像。

## 3. 方法

### 3.1. 二维四通道不可分小波滤波器构造

在公式(3)的基础上, 取  $k=2$ , 构造四通道  $6 \times 6$  滤波器组:  
取:

$$V = \begin{pmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \end{pmatrix} \quad (10)$$

$$P = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & -1 & 0 \end{bmatrix} \quad (11)$$

$$A_j = \begin{bmatrix} -\cos \alpha_j & \sin \alpha_j & 0 & 0 \\ \sin \alpha_j & \cos \alpha_j & 0 & 0 \\ 0 & 0 & -\cos \beta_j & \sin \beta_j \\ 0 & 0 & \sin \beta_j & \cos \beta_j \end{bmatrix} (j=1,2,\dots,k) \quad (12)$$

按照  $U_j = 1/2 \times P \times A_j \times P^T$  构造  $U_j (j=1,2,\dots,k)$ ,  $U_j(\alpha_j, \beta_j)$  是中心对称矩阵。

取  $\alpha_1 = \pi$ ,  $\beta_1 = \pi/30$ ,  $\alpha_2 = \pi/15$ ,  $\beta_2 = \pi/24$ , 可得四通道不可分小波滤波器组的低通滤波器为:

$$H_0 = \begin{bmatrix} -0.0000056724 & 0.0001082364 & -0.001082868 & -0.001875582 & -0.0001666694 & -0.0000087348 \\ -0.0001082364 & -0.002065273 & -0.02066235 & 0.03578824 & -0.003180241 & 0.0001666694 \\ 0.0009882841 & -0.01885758 & 0.190422 & 0.2932242 & 0.02595524 & 0.001360256 \\ 0.001360256 & 0.02595524 & 0.2932242 & 0.190422 & -0.01885758 & 0.0009882841 \\ 0.0001666694 & -0.003180241 & 0.03578824 & -0.02066235 & -0.002065273 & -0.0001082364 \\ -0.0000087348 & -0.0001666694 & -0.001875582 & -0.001082868 & 0.0001082364 & -0.0000056724 \end{bmatrix} \quad (13)$$

### 3.2. 提取图像高频子图

根据我们所构造出的滤波器, 基于加性小波的分解原理, 我们将图像经过所构造的低通滤波器进行卷积提取到低频  $p_i$ , 因为加性小波是完全无损的, 保存边缘信息较好。因此我们将原图利用所构造的低通滤波器进行不可分小波分解, 然后把源图像与上次分解的低频相减得到高频  $w_i$ , 取到图像高频部分。提取到图像如图 2(a)的低频及高频信息如图 2(b), 图 2(c)所示。



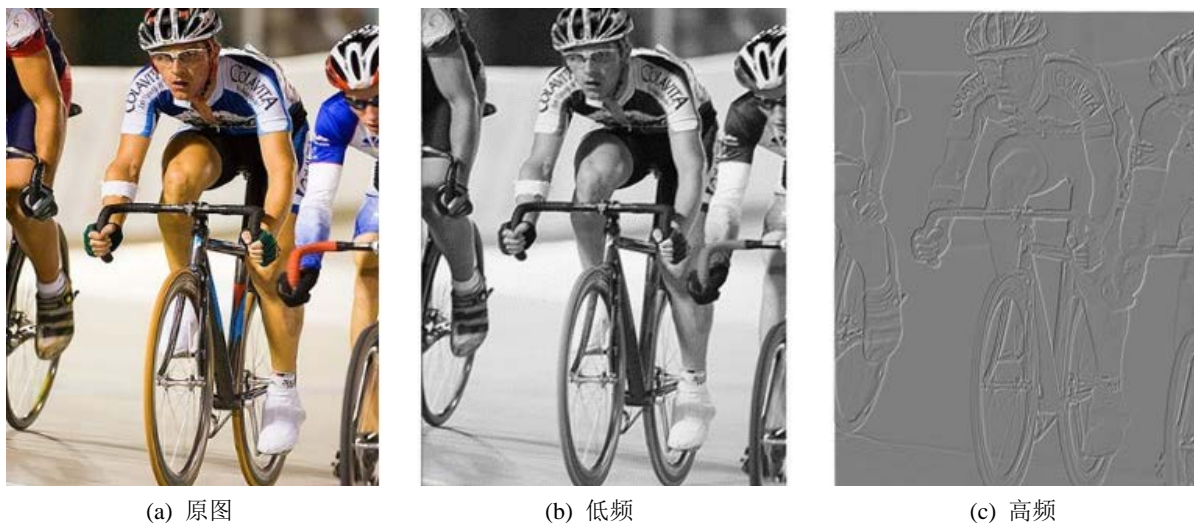


Figure 2. Low frequency and high frequency of image  
图 2. 图像的低频高频

### 3.3. 对 DeepLabv3+模型进行修改

语义分割最终结果正是图像的轮廓信息，而图像的轮廓信息是由边缘信息所构成，因此提取图像中的高频信息对语义分割而言具有重要作用。不可分小波具有各向同性特征，可以提取到图像当中各个方向的信息，具有不丢失源图像信息的特点[36]。因此可以很好提取图像轮廓信息。基于此，我们利用加性小波的分解原理，将不可分小波的高频信息送入到网络中。

在模型的 Decoder 部分以及 ASPP 部分加入 CBAM 模块，注意力机制可以告知 Feature Map 关注的对象，提高了特征感兴趣的表现，提高了模型自适应细化特征。改进后的网络结构见图 3。

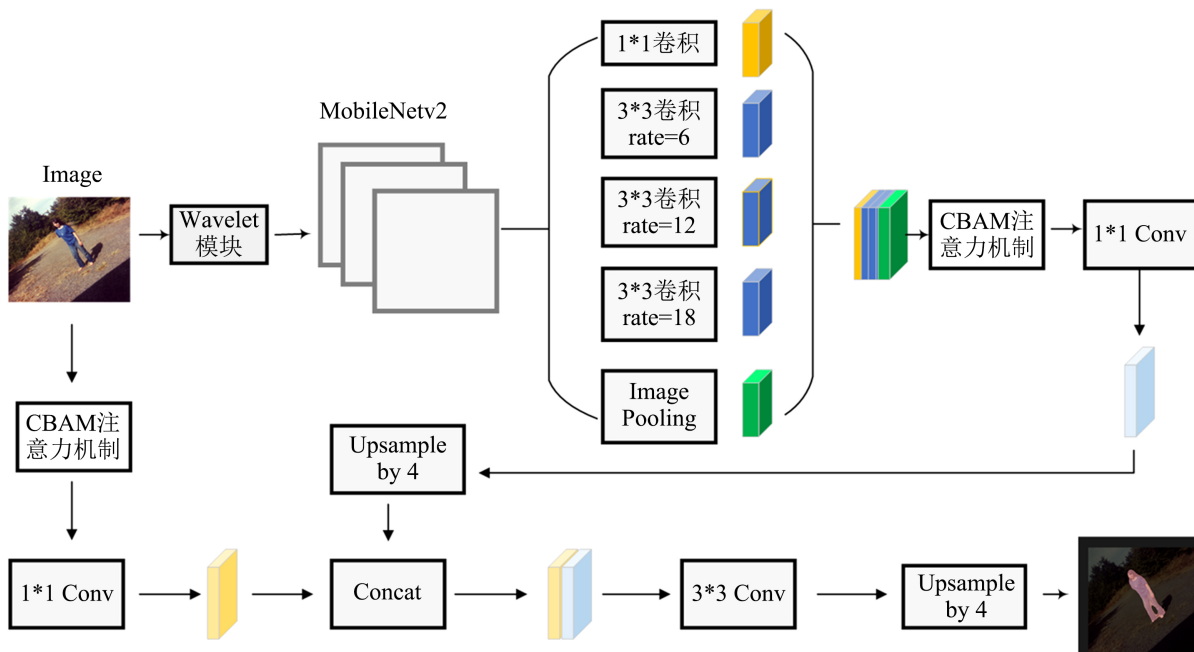


Figure 3. Overall framework of our model  
图 3. 本文模型框架

## 4. 结果

### 4.1. 数据集

我们采用 PASCAL VOC 以及 BDD100K 数据集, 将数据集按照 9:1 划分训练集和测试集, VOC 数据集以及 BDD100K 数据集均为语义分割领域中常用的数据集。其中 VOC 数据集包含 20 个类别以及一个背景类别, 共有 10828 张图像作为训练集图像, 共有 7200 张图像作为 BDD100K 训练集图像。

### 4.2. 实验环境

实验的系统环境在处理器为 Intel(R) Xeon(R) Silver 4210 CPU@2.20GHz, 运行内存为 24G, 显卡为 NVIDIA Quadro RTX 5000 的主机上进行的, 程序是基于 CUDA11.0 以及 CUDNN10.1 的深度学习框架上实现的。训练分为两个阶段, 利用迁移学习思想将模型分为冻结和解冻阶段。冻结阶段会将模型的主干部分冻结, 特征提取网络不发生变化, 占用的显存较小, 仅仅只是对网络进行微调, 提升速度的同时保留训练网络的参数。在解冻阶段, 此时模型的主干不再被冻结, 特征提取网络发生变化, 模型的所有参数都会发生变化。冻结阶段将 Epoch 设置为 50, Batch size 设置为 8, 模型的最小学习率设置为 0.00007; 解冻阶段, Epoch 设置为 100, Batch size 设置为 4, 学习率设置为 0.00007。

### 4.3. 分析与评价

为了客观评价改进后的模型性能, 本文选用 MIoU (Mean Intersection over Union), MPA (Mean Pixel Accuracy, 类别平均像素准确率), Accuracy (准确度)三个指标来评价语义分割算法的性能。

以 VOC 数据集为例, 该数据集包含 20 个类以及一个背景类,  $k$  表示类别, 因此共有  $k+1$  个类别,  $i$  表示真实值,  $j$  表示预测值,  $p_{ij}$  表示将  $i$  预测为  $j$ 。

1) MIoU 即均交并比, 计算所有类别交集和并集之比的平均值。某一个类别的 MIoU 计算公式如下:

$$\text{MIoU} = \frac{1}{k+1} \frac{\sum_{i=0}^k p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}}$$

2) MPA 即计算每一类分类正确的像素点数和该类的所有像素点数的比例, 然后求取平均值, 计算公式如下:

$$\text{MPA} = \frac{1}{k+1} \frac{\sum_{i=0}^k p_{ii}}{\sum_{j=0}^k p_{ij}}$$

3) Accuracy 即全部预测正确的概率。计算公式如下:

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

其中: TP (True Positive 真正例)指预测的结果为正例, 并且预测正确。FN (False Negative 假反例)表示实际为真, 预测为假。FP (False Positive)指假正例, 模型预测为正例, 实际是反例。FN (False Negative)表示假反例, 模型预测为反例, 实际是正例。TN (True Negative)表示真反例, 模型预测为反例, 实际是反例。

### 4.4. 不同网络在 VOC 数据集上的测试结果

基于以上所提的评价指标, 将本文所提出的新的模型与目前几种比较先进的语义分割模型在 VOC 数据集上进行对比, 结果见表 1。

**Table 1.** Comparison of objective indicators of different Classical semantic Segmentation methods on VOC datasets  
**表 1.** 不同经典语义分割方法在 VOC 数据集上客观指标比较

训练网络	MIoU%	MPA%	Accuracy%
DeepLabv3+	76.29	86.51	93.39
UNet	61.68	72.44	91.28
PSPNet	70.31	83.04	92.83
<b>Our Methods</b>	<b>77.23</b>	<b>87.85</b>	<b>93.58</b>

#### 4.5. 不同网络在 BDD100K 数据集上的测试结果

将本文所提出的新的模型与目前几种比较先进的语义分割模型在 BDD100K 数据集上进行对比, 结果见表 2。

**Table 2.** Comparison of objective indicators of different Classical semantic Segmentation methods on BDD100K datasets  
**表 2.** 不同经典语义分割方法在 BDD100K 数据集上客观指标比较

训练网络	MIoU%	MPA%	Accuracy%
DeepLabv3+	51.75	59.52	90.09
UNet	48.35	56.31	91.35
PSPNet	33.28	39.57	86.62
<b>Our Methods</b>	<b>52.28</b>	<b>59.67</b>	<b>90.22</b>

#### 4.6. Wavelet 模块性能验证

为了进一步验证 Wavelet 模块的性能, 我们将加入 Wavelet 模块的网络模型与原模型在 VOC 数据集上进行对比试验, 实验结果见表 3。加入 Wavelet 模块后, MIoU 提高 0.79%, MPA 提高 0.31%, Accuracy 提高了 0.11%。证明 Wavelet 模块对于分割任务有明显提升效果。

**Table 3.** Wavelet module performance  
**表 3.** Wavelet 模块性能

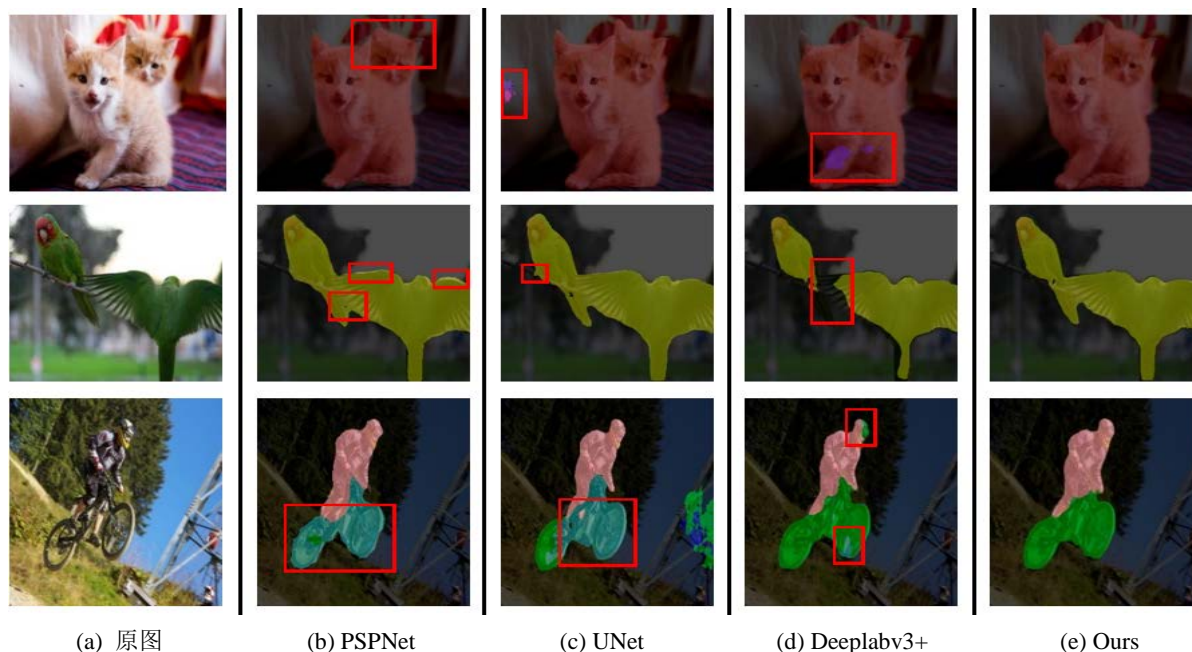
网络模型	MIoU%	MPA%	Accuracy%
未引入 Wavelet 模块	76.44	87.54	93.47
引入 Wavelet 模块	77.23	87.85	93.58

#### 4.7. 主观评价

原图如图 4(a)所示。使用 PSPNet 处理后的结果如图 4(b)所示, 猫的耳朵没有完成分割; 鸚鵡身体出现过分割; 运动员分割完全但是自行车后轮仅仅只分割 1/3。使用 UNet 模型处理后的结果如图 4(c)所示, 图像不包含猫的部分却仍然进行小部分分割; 鸟的身体分割过度; 自行车轮胎没有分割完全, 仅仅只是将后部分车轮分割 2/3。使用 DeepLabv3+处理后的结果如图 4(d)所示, 很明显发现 DeepLabv3+存在分割不清楚问题, 猫的轮廓出现阴影; 鸚鵡翅膀没有分割完全; 运动员头部和前轮还有部分没有完成分割。



对比我们改进后的模型得到的分割图像如图 4(e)所示, 分割细节以及效果从主观角度看均有明显的提升。



**Figure 4.** Comparison of Segmentation effects of different models

**图 4.** 不同模型分割效果对比

## 5. 结论

本文提出了一种基于四通道加性小波与 DeepLabv3+网络结合的语义分割模型, 能够较好地提取图像的特征, 使得分割对象具有更清楚的细节信息。实验结果表明, 对比语义分割领域经典模型 PSPNet, UNet 以及 DeepLabv3+模型, 本文模型效果更好, 明显提升了模型性能。将主干网络替换为 MobileNetv2, 减少了网络的参数, 使得模型更轻量级; 加入 CBAM 模块提高物体分割精度; 加入小波变换后, 明显提升网络的边缘学习效果, 图像的高频是图像轮廓信息, 我们将图像经过小波处理放入模型中提升了模型对于细节信息的捕获。无论是主观还是客观结果均显示出我们的模型优于原模型。

## 基金项目

国家自然科学基金面上项目(No. 61471160)。

## 参考文献

- [1] Ess, A., Müller, T., Grabner, H., *et al.* (2009) Segmentation-Based Urban Traffic Scene Understanding. *British Machine Vision Conference, BMVC 2009*, London, 7-10 September 2009, 2. <https://doi.org/10.5244/C.23.84>
- [2] Geiger, A., Lenz, P. and Urtasun, R. (2012) Are We Ready for Autonomous Driving? The KITTI Vision Benchmark Suite. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, Providence, 16-21 June 2012, 3354-3361. <https://doi.org/10.1109/CVPR.2012.6248074>
- [3] Cordts, M., Omran, M., Ramos, S., *et al.* (2016) The Cityscapes Dataset for Semantic Urban Scene Understanding. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 3213-3223. <https://doi.org/10.1109/CVPR.2016.350>
- [4] Oberweger, M., Wohlhart, P. and Lepetit, V. (2015) Hands Deep in Deep Learning for Hand Pose Estimation.

- [5] Yoon, Y., Jeon, H.G., Yoo, D., *et al.* (2015) Learning a Deep Convolutional Network for Light-Field Image Super-Resolution. *Proceedings of the IEEE International Conference on Computer Vision Workshops*, Santiago, 7-13 December 2015, 24-32. <https://doi.org/10.1109/ICCVW.2015.17>
- [6] Wan, J., Wang, D., Hoi, S.C.H., *et al.* (2014) Deep Learning for Content-Based Image Retrieval: A Comprehensive Study. *Proceedings of the 22nd ACM International Conference on Multimedia*, Orlando, 3-7 November 2014, 157-166. <https://doi.org/10.1145/2647868.2654948>
- [7] 邱艺东. 基于图像分割的增强现实导航方法研究[D]: [硕士学位论文]. 福州: 福州大学, 2020.
- [8] Rosenfeld, A. (1981) The Max Roberts Operator Is a Hueckel-Type Edge Detector. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-3**, 101-103. <https://doi.org/10.1109/TPAMI.1981.4767056>
- [9] Lang, Y. and Zheng, D. (2016) An Improved Sobel Edge Detection Operator. 2016 *6th International Conference on Mechatronics, Computer and Education Informationization (MCEI 2016)*, Shenyang, 11-13 November 2016, 590-593. <https://doi.org/10.2991/mcei-16.2016.123>
- [10] Yang, L., Wu, X., Zhao, D., *et al.* (2011) An Improved Prewitt Algorithm for Edge Detection Based on Noised Image. 2011 *4th International Congress on Image and Signal Processing*, Vol. 3, 1197-1200. <https://doi.org/10.1109/CISP.2011.6100495>
- [11] Zhang, Y.J. (2006) An Overview of Image and Video Segmentation in the Last 40 Years. In: Zhang, Y.-J., Ed., *Advances in Image and Video Segmentation*, IGI Global, Hershey, 1-16. <https://doi.org/10.4018/978-1-59140-753-9.ch001>
- [12] Pham, D.L., Xu, C. and Prince, J.L. (2000) A Survey of Current Methods in Medical Image Segmentation. *Annual Review of Biomedical Engineering*, **2**, 315-337. <https://doi.org/10.1146/annurev.bioeng.2.1.315>
- [13] Tremeau, A. and Borel, N. (1997) A Region Growing and Merging Algorithm to Color Segmentation. *Pattern Recognition*, **30**, 1191-1203. [https://doi.org/10.1016/S0031-3203\(96\)00147-1](https://doi.org/10.1016/S0031-3203(96)00147-1)
- [14] Cheng, Y. (1995) Mean Shift, Mode Seeking, and Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**, 790-799. <https://doi.org/10.1109/34.400568>
- [15] Fukunaga, K. and Hostetler, L. (1975) The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition. *IEEE Transactions on Information Theory*, **21**, 32-40. <https://doi.org/10.1109/TIT.1975.1055330>
- [16] Sheikh, Y.A., Khan, E.A. and Kanade, T. (2007) Mode-Seeking by Medoidshifts. 2007 *IEEE 11th International Conference on Computer Vision*, Rio de Janeiro, 14-21 October 2007, 1-8. <https://doi.org/10.1109/ICCV.2007.4408978>
- [17] Boykov, Y.Y. and Jolly, M.P. (2001) Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in ND Images. *Proceedings 8th IEEE International Conference on Computer Vision, ICCV 2001*, Vol. 1, 105-112.
- [18] Rother, C., Kolmogorov, V. and Blake, A. (2004) "GrabCut" Interactive Foreground Extraction Using Iterated Graph Cuts. *ACM Transactions on Graphics (TOG)*, **23**, 309-314. <https://doi.org/10.1145/1015706.1015720>
- [19] Tang, M., Gorelick, L., Veksler, O., *et al.* (2013) Grabcut in One Cut. *Proceedings of the IEEE International Conference on Computer Vision*, Sydney, 1-8 December 2013, 1769-1776. <https://doi.org/10.1109/ICCV.2013.222>
- [20] Long, J., Shelhamer, E. and Darrell, T. (2015) Fully Convolutional Networks for Semantic Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 7-12 June 2015, 3431-3440. <https://doi.org/10.1109/CVPR.2015.7298965>
- [21] Badrinarayanan, V., Kendall, A. and Cipolla, R. (2017) Segnet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 2481-2495. <https://doi.org/10.1109/TPAMI.2016.2644615>
- [22] Chen, L.C., Papandreou, G., Schroff, F., *et al.* (2017) Rethinking Atrous Convolution for Semantic Image Segmentation.
- [23] Zhao, H., Shi, J., Qi, X., *et al.* (2017) Pyramid Scene Parsing Network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 2881-2890. <https://doi.org/10.1109/CVPR.2017.660>
- [24] Chen, L.C., Zhu, Y., Papandreou, G., *et al.* (2018) Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, 8-14 September 2018, 801-818. [https://doi.org/10.1007/978-3-030-01234-2\\_49](https://doi.org/10.1007/978-3-030-01234-2_49)
- [25] 徐长友, 樊绍胜, 朱航. 采用通道域注意力机制DeepLabv3+算法的遥感影像语义分割[J]. 控制工程, 2023, 30(2): 368-375.
- [26] 郑斌军, 孔玲君. 基于DeepLabv3+的图像语义分割优化方法[J]. 包装工程, 2022, 43(1): 187-194.
- [27] 马朝永, 马兴杰, 胥永刚. 基于DeepLabv3+网络的滚动轴承故障特征识别[J]. 轴承, 2023(2): 74-81.
- [28] 张鑫禄, 张崇涛, 戴晨光, 季虹良, 王映雪. 基于DeepLabv3架构的高分辨率遥感图像分类[J]. 海洋测绘, 2019, 39(2): 40-44.

- 
- [29] Sandler, M., Howard, A., Zhu, M., *et al.* (2018) Mobilenetv2: Inverted Residuals and Linear Bottlenecks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 4510-4520. <https://doi.org/10.1109/CVPR.2018.00474>
- [30] 刘斌, 彭嘉雄. 基于四通道不可分加性小波的多光谱图像融合[J]. 计算机学报, 2009, 32(2): 350-356.
- [31] Liu, B. and Peng, J.X. (2008) Multi-Spectral Image Fusion Method Based on Two Channels Non-Separable Wavelets. *Science in China Series F: Information Sciences*, **51**, 2022-2032. <https://doi.org/10.1007/s11432-008-0162-6>
- [32] Chen, Q., Micchelli, C.A., Peng, S., *et al.* (2003) Multivariate Filter Banks Having Matrix Factorizations. *SIAM Journal on Matrix Analysis and Applications*, **25**, 517-531. <https://doi.org/10.1137/S0895479802412735>
- [33] Woo, S., Park, J., Lee, J.Y., *et al.* (2018) Cbam: Convolutional Block Attention Module. *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, 8-14 September 2018, 3-19. [https://doi.org/10.1007/978-3-030-01234-2\\_1](https://doi.org/10.1007/978-3-030-01234-2_1)
- [34] Hu, J., Shen, L. and Sun, G. (2018) Squeeze-and-Excitation Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7132-7141. <https://doi.org/10.1109/CVPR.2018.00745>
- [35] Lin, M., Chen, Q. and Yan, S. (2013) Network in Network.
- [36] 刘斌, 熊静雯. 不可分小波构造及其在图像水印中的应用[J]. 现代电子技术, 2020, 43(11): 87-91+96.