计算机工程
Computer Engineering
ISSN 1000-3428,CN 31-1289/TP

《计算机工程》网络首发论文

题目：基于注意力机制和辅助任务的语义分割算法
作者：叶剑锋，徐轲，熊峻峰，王化明
DOI：10.19678/j.issn.1000-3428.0058447
网络首发日期：2020-10-12
引用格式：叶剑锋，徐轲，熊峻峰，王化明．基于注意力机制和辅助任务的语义分割算法．计算机工程. https://doi.org/10.19678/j.issn.1000-3428.0058447

基于注意力机制和辅助任务的语义分割算法

叶剑锋，徐轲，熊峻峰，王化明*

（南京航空航天大学 机电工程学院 南京 210008）

摘 要：为提升网络模型低层特征的离散度，提高语义分割算法的性能，以全卷积神经网络作为基础模型，引入图像分类中的辅助损失、机器学习中的多任务学习及自然语言处理中的注意力机制，提出一种基于辅助损失、边缘检测辅助任务和注意力机制的语义分割算法．首先重新设计网络模型的辅助损失分支，使网络低层特征编码更多语义信息；然后针对语义分割主任务选择边缘检测作为辅助任务，基于注意力机制设计针对边缘检测辅助任务的辅助任务分支，使网络低层特征更关注物体的形状和边缘信息；最后将基础模型、辅助损失分支、辅助任务分支集成构造为语义分割模型．在 VOC2012 数据集上进行实验的结果表明，该算法的平均交并比为 71.5%，比基础模型算法提高 6%，也优于其他基于该基础模型的语义分割算法．

关键词：注意力机制；辅助任务；辅助损失;多任务学习；语义分割

DOI:10.19678/j.issn.1000-3428.0058447

Research on semantic segmentation algorithm based on attention mechanism and auxiliary task

Ye Jianfeng, Xu Ke, Xiong Junfeng, and Wang Huaming*

(*College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing* 210008)

Abstract: *In order to improve the scatter of low-level features in neural network model and the performance of semantic segmentation algorithm with deep convolutional neural network, we choose fully convolutional network as basic model, introduce auxiliary loss in image classification, multi-task learning in machine learning and attention mechanism in natural language processing and propose a semantic segmentation algorithm based on auxiliary loss, edge detection auxiliary task and attention mechanism. First, we redesign the auxiliary branch to force the low-level feature to encode more semantic information. Second, an auxiliary task branch based on the attention mechanism is designed for edge detection selected as auxiliary task due to semantic segmentation, so that the low-level features of the network pay more attention to the shape and edge information of objects. Finally, the basic model, auxiliary loss branch and auxiliary task branch are integrated into the semantic segmentation model. We evaluate our algorithm on the VOC2012 dataset. The*

基金项目：国家自然科学基金(61363066)．作者简介：叶剑锋(1984—)，男，博士研究生,副教授，主要研究方向为图像识别、智能交通技术，ORCID:0000-0003-3100-5587;徐轲（1999 年--），女，机械工程专业，熊峻峰(1995—)，男，硕士研究生，主要研究方向为图像处理、深度学习；王化明(1973—)，男，博士，教授，博士生导师，论文通讯作者，主要研究方向为机器视觉、机器人控制,ORCID:0000-0002-4434-7482 E-mail：jfye0619@nuaa.edu.cn

experimental results show that mean intersection over union of our algorithm is 71.5%, outperforming the basic model algorithm by 6%. And the algorithm superior to other segmentation algorithms using the basic model.

Key words: Attention mechanism;Auxiliary tasks;Auxiliary loss;Multitasking;Semantic segmentation

0 概述

语义分割是计算机视觉的基础任务之一，其目的是将输入图像划分为不同的语义可解释的类别，即像素级别的多类别分类任务[1]，如图 1 所示. 目前，语义分割在自动驾驶、虚拟现实、城市交通规划等领域都有着广泛的应用.

传统的图像分割算法主要包括基于阈值的分割算法[2]、基于边缘的分割算法[3]和基于区域的分割算法[4]等，这些算法通常采用图像特征分类器来完成图像分割. 首先针对分割目标设计多个特征,分别针对每个特征设计一个结构复杂的特征提取器，最后构建一个分类器对所获取的特征进行分分类和识别.

近年来，深度学习方法，尤其是卷积神经网络[5-7]在图像分类任务上取得了显著的成果. 基于深度学习的图像处理算法与传统图像处理算法主要区别在于: 采用一种通用的学习过程从数据中主动学习得到特征，并不需要手工设计特征[1].

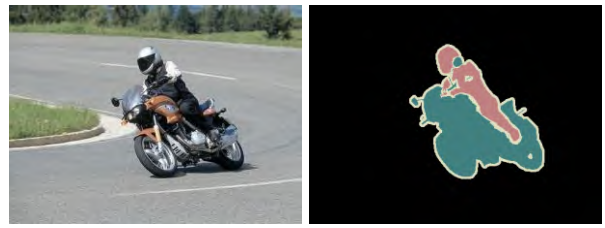
随着大量原本在图像分类、目标检测、自然语言处理等领域取得成功的深度学习方法被改进、迁移到语义分割领域，图像的语义分割技术也取得了很大的突破. 例如 Long 等[8]提出的全卷积神经网络(fully convolutional network, FCN)在图像分类网络视觉几何组网络(visual geometry group network, VGG)的基础上去除全连接层，加入多级上采样还原分辨率，实现端到端的语义分割,Faster Rcnn[9]与 Mask rcnn[10]由进一步进行了发展. Chen 等人[11]在网络模型中引入自然语言处理中的注意力机制以实现多尺寸特征图像的加权融合，提升算法的尺寸不变性.现有算法为了增大感受野、降低特征维度、减少计算量会对输入图像做多次下采样，但在此程中损失函数对特征的约束力越来越低，造成低层特征的离散度低，丢失大量空间细节信息，Yang 等人[12]提出一种区域级别的基于纹理基元块识别与合并的图像语义分割算法。该算法采用纹理基元等特征,考虑 到相邻像素点间的相互关系,保留物体间的棱角和边缘信息,分割出轮廓清晰的对象。但仍然存在对目标形状、边缘细节的分割欠缺等问题.

为提升网络模型低层特征的离散度[13]，使网络低层特征编码更多语义信息，更关注物体的形状和边缘信息，提高语义分割算法的性能，本文引入图像分类中的辅助损失、机器学习中的多任务学习及自然语言处理中的注意力机制[14]，提出一种基于辅助损失、边缘检测辅助任务和注意力机制的语义分割算法. 首先重新设计网络模型的辅助损失分支；然后针对语义分割主任务选择边缘检测作为辅助任务，基于自然语言处理中的注意力机制设计辅助任务分支；最后将基础模型、辅助损失分支、辅助任务分支集成构造为语义分割模型.

1 相关工作

1.1 语义分割

语义分割是计算机视觉应用如自动驾驶、场景理解等的关键技术. 得益于近年来卷积神经网络的高速发展，像素级语义分割也取得显著进展. 如图 1 所示. 最近的研究主要集中在以下两个方法:



a. 原图　　b. 语义分割

图 1 语义分割

Fig1. Semantic segmentation

编码器-解码器结构：将神经网络设计为编码器-解码器结构，通过在编码阶段保留更多的图像信息，同时在解码阶段更好地恢复图像损失的信息来生成更好的语义分割结果．例如，SegNet[15]利用在编码阶段保存的池化索引来恢复图像池化时丢失的空间信息；U-Net[16]设计跳跃连接结构，通过直接在解码阶段引入低层特征图来恢复图像所丢失的信息．

上下文信息：让神经网络聚合更多的图像上下文信息，连接不同采样率上的特征图像，解决尺度多样性问题，产生更精准的语义分割结果．例如，DeepLab[17]通过空洞空间金字塔池化结构在多尺寸图像上捕捉上下文信息；ParseNet[18]通过添加全局池化分支，在解码阶段引入全局上下文信息．

本文同时结合以上两种方法：一方面采用 FCN 作为基础模型，且可以更换为其他任意具有编码器-解码器结构的网络模型；另一方面采用注意力机制聚合更多上下文信息，采用跳跃连接结构连接不同采样率上的特征图像．

1.2 辅助损失

网络深度是神经网络的主要特征之一，但是神经网络过深会带来梯度消失、收敛困难等问题，使得神经网络训练失败或者达不到理想效果[7]．因此研究人员设计了多种训练方法和网络结构来解决这个难题，如 Dropout[19]、批归一化[20]、残差结构[7]等等．辅助损失，也叫做中间监督，通过直接在网络中间加入辅助损失分支，降低梯度消失、网络难以收敛的概率，使得深度网络训练更加容易．

本文将辅助损失引入语义分割网络中的主要目的并不是解决收敛困难等问题，而是迫使低层特征编码更多语义信息，提升低层特征的离散度．

1.3 多任务学习

如图 2 所示，多任务学习是指模型同时学习多个具有相关表征的任务，提升学习效率和预测准确率、改善泛化性能．多任务学习在机器学习、自然语言处理、计算机视觉等领域的许多应用中非常普遍[21-24]．MultiNet[21]设计了一种能够同时进行图像分割、目标检测和语义分割等视觉任务的网络结构．十字绣网络[22]针对性地研究多任务网络中神经元共享的方法，提出了可以通过端对端的学习来自动决定共享层的十字绣网络结构．

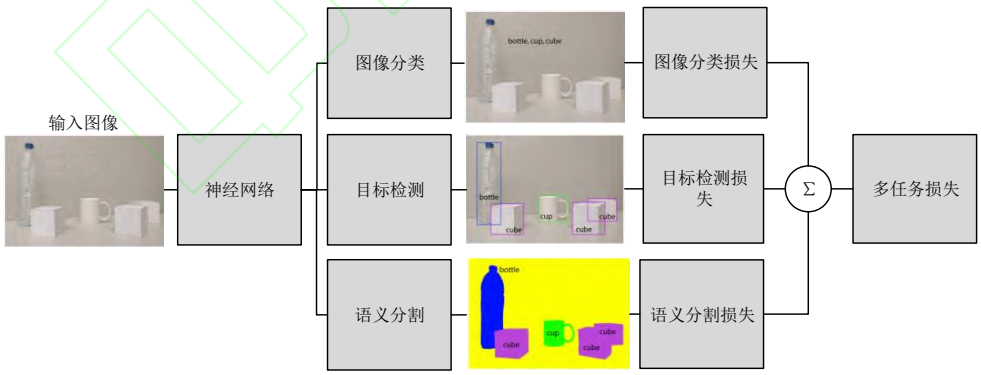


图 2 多任务学习

Fig2.Multi-task learning

多任务学习神经网络专注于计算机视觉任务的并行学习，需要在多个任务的结果之间权衡，不能保证单个任务取得最优结果．本文提出的辅助任务则是在多任务学习中区分主任务与辅助任务，只专注于主任务的训练效果，使用辅助任务的训练信号中所拥有的特定领域信息来提升主任务的泛化效果，使主任务取得最优结果．

1.4 注意力机制

注意力机制(attention mechanism)在自然语言处理领域中应用广泛[23,24]．近来，如何将注意力机制引入到计算机视觉中也成为研究

热点. Hu 等人[25]提出目标关系模组来建模一系列目标间的关系提升目标检测效果. Chen 等人[11]提出多尺寸注意力机制来自适应融合多尺寸图像提升语义分割效果.

本文将自注意力机制(self-attetion mechanism)和残差模块结合，设计针对边缘检测任务的辅助任务分支. 自注意力机制可以根据通道间的依赖关系自适应地增强相关语义的通道图，提升残差模块相关语义的表达能力.

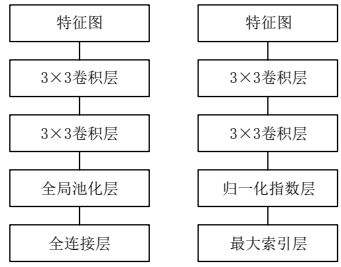
2 网络结构

2.1 辅助损失分支

深度神经网络中加入辅助损失的目的是降低梯度消失、网络难以收敛的概率，便于深度网络的训练. 但最近的研究[4,26-28]发现，在精心设计的网络结构及采用其他训练方法的情况下，超过 100 层的深度神经网络不采用辅助损失也不会出现无法收敛的问题. 甚至在部分浅层的图像分类网络上使用辅助损失会降低分类准确率.

但是语义分割网络中引入辅助损失的主要目的并不是解决收敛困难等问题，而是提升低层特征的质量. 为了在辅助损失分支中输出语义信息，低层特征需要编码更多语义信息，提升低层特征的离散度，这有利于其后的特征融合. 因此对于浅层网络，加入辅助损失后，虽然其分类准确率变化不大甚至降低，但是依旧可以提升其分割的准确度，即平均交并比.

图像分类只有一维输出如图 3a，而语义分割的输出是二维图像，因此辅助损失分支的结构设计也不一样，针对语义分割的辅助损失分支的结构如图 3b. 最上方为通过基础模型中间某层所得到的特征图，经过多层卷积处理降维后，再通过双线性插值进行拉伸，得到与原图像尺寸一致的特征图，最后计算特征图与目标图像的二维交叉熵损失. 而算法的总损失函数为基础模型损失和所有辅助损失的加权和. 训练完成后，将移除所有辅助损失分支，仅使用基础模型进行推断，避免辅助损失分支带来的额外的内存及时间消耗.



a. 图像分类辅助分支　　b. 语义分割辅助分支

图 3 辅助损失分支结构

Fig3.Auxiliary loss branching structure

2.2 辅助任务分支

2.2.1 辅助任务

虽然同样是多个任务并行训练，但和多任务学习不一样的是，本文的算法专注于提升主任务的训练效果，其余任务均为辅助任务. 对于辅助任务，其本身的训练效果并不重要，重要的是其对主任务的训练效果带来的提升.

辅助任务能提升模型分割效果的原因主要有以下几点：1) 辅助任务能为模型提供归纳偏置，提高模型的泛化能力. 2) 辅助任务可以提供额外的数据信息，可以视作一种数据增广算法. 3) 辅助任务所提供的信息也有可能成为噪声，Holmstrom 等人的研究[29]表明偶尔在训练过程中加入噪声能够增强网络模型的泛化能力. 综上所述，辅助任务的选择应该满足以下要求：1) 主任务的概念层次应高于辅助任务，且主任务的目标域应与辅助任务的目标域存在交集. 2) 主任务和辅助任务的训练图像应一致或辅助任务的标注图像应便于从主任务的标注图像中获得.

本文研究的主任务为语义分割，根据上述原则，本文选择的辅助任务为边缘检测. 边缘检测是传统图像处理中的基本问题之一，目的是提取图像中对象与背景间的交界线. 它可以使低层共享网络更关注于物体的形状和边缘信息，获取更多关于物体类内差异的特征. 而且边缘检测所需的标注图可以非常方便的从语义分割的标注图中获取，边缘检测标注图如图 4 所示.

a. 原图　　b. 语义分割标注图　c. 边缘检测标注图

图 4 语义分割和边缘检测的标注图

Fig4.Semantic segmentation and edge detection

2.2.2 注意力残差模块

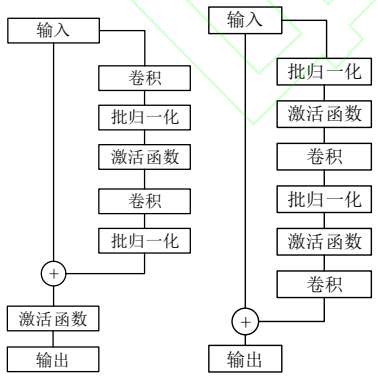
本文结合自注意力机制与残差模块设计注意力残差模块(attention residual module, ARM)，结构，将注意力残差模块堆叠即可得到辅助任务分支，以下阐述注意力残差模块的构建过程.

原始残差模块如图 5a 所示，即:

$$y_l = F(x_l, W_l) \tag{1}$$

$$x_{l+1} = h(x_l) + f(y_l) \tag{2}$$

x_l 和 x_{l+1} 分别为第 l 层的输入和输出，F 为残差函数，h 为恒等映射函数，f 为整流线性激活函数. 虽然残差模块内恒等映射函数可以保证信息流无损流动，但由于激活函数的存在，整个网络的信息流并不能无损流动. 因此为保证信息流无损地在各层间流动，于是将 f 也变为恒等映射函数，得到改进后的残差模块，即恒等残差模块[30]，如图 5b.



a. 原始残差模块　　b. 恒等残差模块

图 5 残差模块

Fig5.Residual module

其数学表达式为:

$$x_{l+1} = x_l + F(x_l, W_l) \tag{3}$$

$$x_L = x_1 + \sum_{i=1}^{L-1} F(x_i, W_i) \tag{4}$$

根据反向传播链式法则有:

$$\frac{\partial Loss}{\partial x_1} = \frac{\partial \varepsilon}{\partial x_L}\frac{\partial x_L}{\partial x_1} = \frac{\partial \varepsilon}{\partial x_L}\left(1 + \frac{\partial \sum_{i=1}^{L-1} F(x_i, W_i)}{\partial x_L}\right) \tag{5}$$

由式 5 可以看出损失梯度可以无损地传递到任意残差模块，甚至任意残差模块的损失梯度都可以无损地传递其余任意残差模块，因此减小了梯度消失的概率.

但是特征图的每个通道可以被视为特定于某种语义特征的响应图，并且不同的语义特征彼此相关联. 显然残差模块中 x_l 与 y_l 语义特征并不一致，不能直接相加. 因此在恒等残差模块 x_l 与 y_l 的融合中引入自注意力机制，用于显式建模 x_l 与 y_l 各语义特征之间的相互依赖关系. 通过利用通道之间的相互依赖性，可以增强相互依赖的特征并改进特定语义的特征表示.

$$y_l = F(x_l, W_l) \tag{6}$$

$$x_{l+1} = x_l + y_l P(x_l, y_l) \tag{7}$$

注意力残差模块结构如图 6 所示，输入特征图为 $X \in \mathbb{R}^{C\times H\times W}$，经过两轮批归一化、激活函数和卷积后可得到新特征图 $Y \in \mathbb{R}^{C\times H\times W}$，然后将 X 和 Y 分别重排为 $X' \in \mathbb{R}^{C\times N}$ 和 $Y' \in \mathbb{R}^{C\times N}$，对 X' 和 Y' 的转置作矩阵乘法，再经过归一化指数函数后即可得到通道注意力图 $A \in \mathbb{R}^{C\times C}$.

$$a_{i,j} = \frac{\exp(x_i, y_j)}{\sum_{i=1}^{C} \exp(x_i, y_j)} \tag{8}$$

$a_{i,j}$ 即为 X 的第 i 个通道对 Y 的第 j 个通道的影响因子. 对 A 和 Y' 作矩阵乘法，再重排为 $E \in \mathbb{R}^{C\times H\times W}$ 即为增强后的特征图. 将 E 与 X 作元素加操作即可得到最终输出特征图

$O \in \mathbb{R}^{C\times H\times W}$.结构如图 6 所示。

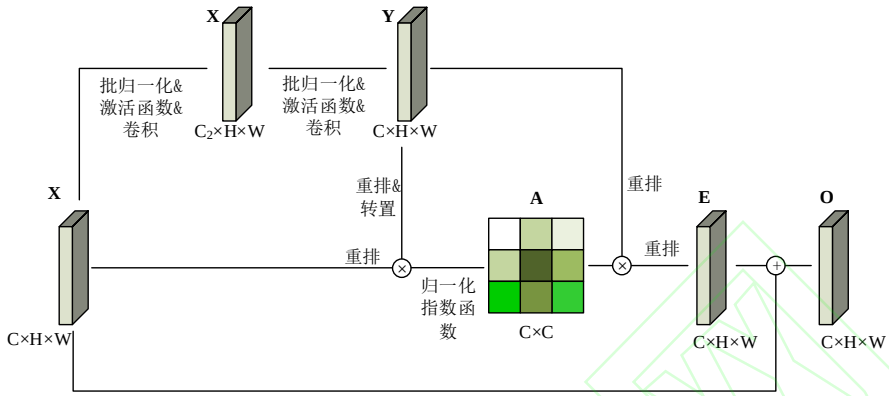


图 6 注意力残差模块

Fig6.Attention residual module

与辅助损失分支一样，训练完成后，将移除所有辅助任务分支，仅使用基础模型进行推断，避免辅助任务分支带来的额外的内存及时间消耗.

2.3 整体结构

FCN 在深度图像分类神经网络 VGG 的基础上去除全连接层，大幅降低网络的参数量，提高计算速度. 由于 FCN 只剩卷积层和激活函数，因此可以看作一个大型卷积核，能接受任意尺寸图像作为输入图像. 最后，加入多级上采样还原分辨率，实现端到端的语义分割. 整体结构如图 7 所示。

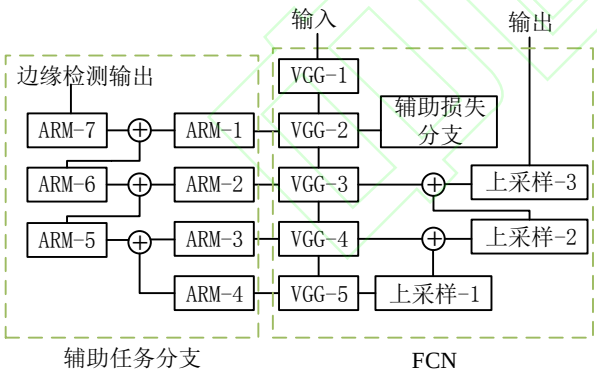


图 7 网络模型整体结构

Fig7.Overall structure of network model

由于 FCN 具有轻量化、高精度、结构简单，且能接受任意尺寸图像作为输入图像的特点，便于实现复杂算法并快速验证的同时依旧保持高精度.

故本文采用 FCN 作为基础模型.

辅助任务分支中所有注意力残差模块后都加入双线性上采样层还原分辨率. 网络模型损失函数如下:

$$L_{总} = L_{主任务} + \alpha L_{辅助任务分支} + \beta L_{辅助损失分支} \quad (9)$$

总损失能量为主任务、辅助任务分支和辅助损失分支的损失能量加权和，主任务、辅助任务分支和辅助损失分支的损失能量加权和均为交叉熵损失函数，并取$\alpha=0.1$、$\beta=1$.

3 实验

为验证所提出的算法，本文在 PASCAL VOC2012 大型数据集上进行测试，采用像素准确率和平均交并比来衡量分割真值(groud truth, GT)与实际分割结果的差异. PASCAL VOC2012 是目前最受欢迎的语义分割数据集，其拥有 1464 张训练集图像，14449 张验证集图像和 1456 张测试集图像，本文实验环境如表 1 所示.

表 1 训练与测试环境

Table 1 Training and test environment

项目	参数
操作系统	Ubuntu 16.04 LTS
算法框架	Pytorch
CPU	Intel i7-4710MQ

GPU	NVIDIA GTX950m
显存	2GB
内存	8GB
CUDA 版本	10.0

3.1 辅助损失分支实验及分析

首先在浅层网络模型 ResNet50 上进行实验以验证 2.1.1 节中的理论. 在 CIFAR-10 验证集上测试分类准确率, 在 VOC2012 验证集上测试平均交并比. 如表 2 所示, 加入辅助损失后, 虽然其分类准确率变化不大, 但是其平均交并比提升 1.78%. 这证明浅层网络模型加入辅助损失后, 分割的准确度即平均交并比确实得到提升. 低层特征需要编码更多的语义信息, 提升低层特征的质量, 最终提高分割的准确度.

表 2 ResNet18 加入辅助损失前后的性能对比

Table 2 ResNet18+ Auxiliary loss performance comparison

网络模型	分类准确率	平均交并比
ResNet18	92.6%	62.4%
ResNet18+辅助损失	92.6%	63.18%

在 FCN 中加入辅助损失分支来提升其低层特征的质量. 如表 3 所示, FCN 加入辅助损失分支后取得 66.2%的平均交并比, 相比于基准模型提升了 0.7%, 验证了辅助损失算法的有效性. 从表 3 中可以看出, 随着辅助损失在网络模型中位置变深, 网络模型性能反而降低, 这可能是随着辅助损失的层数在网络模型中位置越来越深, 其对低层特征的约束力越来越弱, 因此提升效果越来越差.

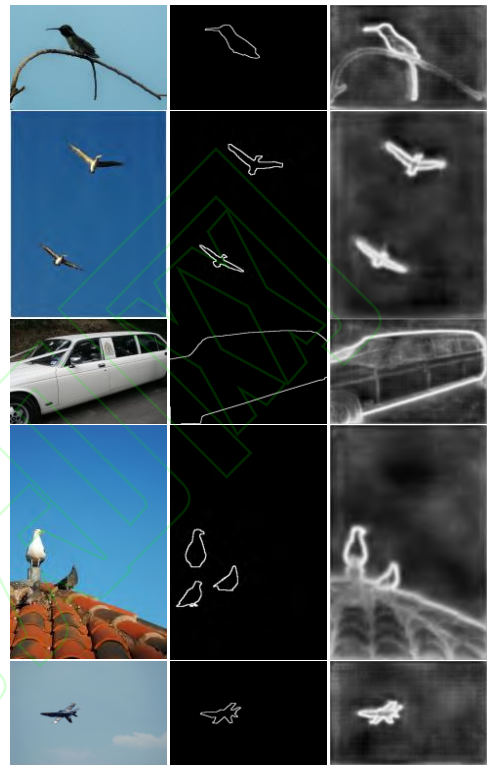
表 3 FCN 不同层加入辅助损失的验证集性能对比

Table 3 Performance comparison of validation sets with different FCN layers + auxiliary losses

网络模型	像素准确率	平均交并比
FCN	91.2%	65.5%
辅助损失(VGG-1)	91.3%	66.2%
辅助损失(VGG-2)	91.3%	66.1%
辅助损失(VGG-3)	91.2%	65.6%
辅助损失(VGG-4)	91.2%	64.5%

3.2 辅助任务分支实验及分析

辅助任务进一步加强网络低层特征的离散度. 本文采用图像分类与语义分割多任务作为对比. 如表 4 所示, 加入辅助任务后网络模型的平均交并比从 65.5%提升到 70.7%, 而多任务反而产生了负面的影响, 造成网络性能降低了 5.7%, 验证了辅助任务算法的有效性.



a. 原图 b. 边缘 GT 图 c. 边缘检测结果图

图 8 边缘检测结果对比

Fig8.Comparison of edge detection results

图 8 为边缘检测辅助网络的分割结果. 可以看出, 边缘检测辅助任务分支正常学习到语义边界特征.

表 4 多任务和辅助任务验证集性能对比

Table 4 Compares the performance of multi-task and secondary task validation sets

网络模型	像素准确率	平均交并比
FCN	91.2%	65.5%
FCN+辅助任务	92.3%	70.7%
FCN+多任务	90.5%	59.8%

3.3 整体结构实验及分析

为测试本文算法性能, 在 VOC2012 测试集上对加入辅助任务分支和辅助损失分支后的完整算法进行实验, 同时和基于同样基础模型 FCN 的主流语义分割算法进行对比. 表 5 列出

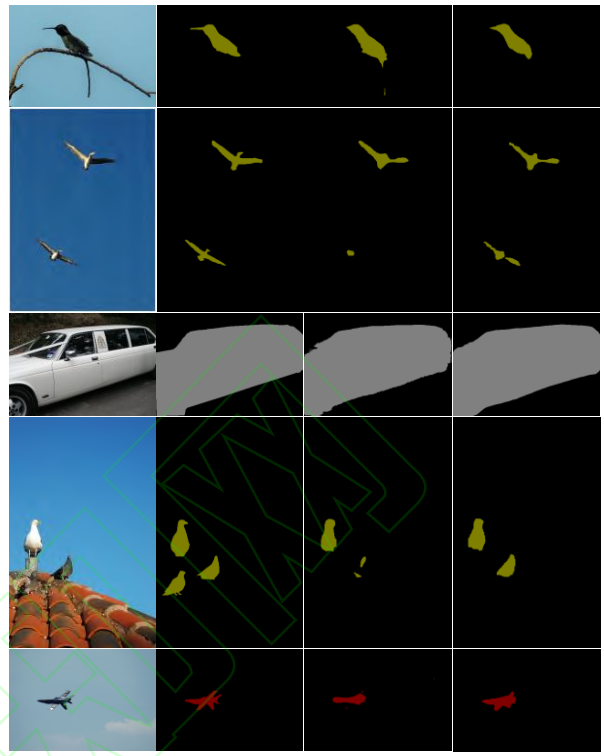
了测试集上的分割结果，FCN-A 为本文算法模型，基础模型为 FCN. 本文也将辅助任务分支与辅助损失分支应用到 SegNet 上，即 SegNet-A. 最终 FCN-A 取得了平均交并比 71.5%的结果，比基础模型算法高 6%,推断速度仅增加 30ms，验证了本文算法的有效性. 同时，SegNet-A 取得了 72.2%的结果，比之前最好的 ParseNet 高 2.4%，推断速度仅增加 5ms，验证了本文算法的可扩展性.

表 5　不同算法 VOC2012 测试集性能对比

Table 5 Comparison of VOC2012 test sets

网络模型	辅助任务	辅助损失	平均交并比	推断时间 ms
FCN[8]			65.5%	500
DeepLab-LargeFOV[17]			65.9%	1200
SegNet[15]			67.4%	80
ParseNet[18]			69.8%	320
FCN-A	√	√	71.5%	530
SegNet-A	√	√	72.2%	85

从图 9 可以看出，加入边缘检测辅助任务分支的网络模型对于物体的形状、语义边界的分割效果更好. 这证明边缘检测辅助任务分支确实使网络模型更关注物体的形状和边缘信息，获取更多关于物体类内差异的特征，提升了网络模型低层特征的离散度，优化了基础模型分割结果的语义边缘. 但是，可以从图 9 第四行的分割结果中看出，本文模型对物体与背景纹理、颜色近似度高的情况分割结果并不理想. 这可能是因为网络模型特征的类间差异度较低，需要学习更多关于纹理、颜色的特征. 后续可以尝试加入最大化类间差异度的损失函数或结构等.



a. 原图　b. GT　c. FCN　d. FCN-A

图 9　网络模型的预测结果对比

Fig9.Comparison of prediction results of network model

4　结　语

本文提出一种基于辅助损失、边缘检测辅助任务和注意力机制的语义分割算法. 首先引入图像分类中的辅助损失并为其重新设计网络模型的辅助损失分支，使网络低层特征编码更多语义信息；其次引入机器学习领域中的多任务学习，选择边缘检测作为辅助任务，基于自然语言处理中的注意力机制为其设计辅助任务分支，使网络模型低层特征更关注物体的形状和边缘信息；最后将基础模型、辅助损失分支、辅助任务分支集成构造为语义分割模型. 在 VOC2012 数据集上进行验证实验，给出了辅助损失分支、辅助任务分支和整体结构的验证实验结果及分析，并与其他主流语义分割算法进行了对比. 实验结果表明，本文算法在 VOC2012 测试集上的平均交并比达到 71.4%，比 FCN 算法提高

6%，也优于其他基于 FCN 的语义分割算法，验证了本文算法的有效性．将基础模型更换为 SegNet 后，平均交并比达到 72.2%，比基础模型提高 4.8%，验证了本文算法的可扩展性.在后续研究中，一方面对辅助任务机制的内在数学机理进行研究，另一方面结合新的特征提取网络研究成果进行注意力机制和辅助任务的泛化性进行研究。

参考文献(References):

[1] Lecun Y, Bengio Y, Hinton GE, *et al*. Deep learning[J]. Nature, 2015, 521(7553): 436-444

[2] Ostu N.A Threshold Selection Method from Gray-Level Histograms[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1979, 9(1): 62-66

[3] Canny JF. A Computational Approach to Edge Detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986, 8(6): 679-698

[4] XU Xiaowei , Martin Ester, Hans-Peter Kriegel, J,et al. A distribution-based clustering algorithm for mining in large spatial databases[C].Proceedings of International Conference on Data Engineering. Orlando: IEEE Computer Society, 1998: 324-331

[5] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C] .Proceedings of the Neural Information Processing Systems. New York: Curran Associates, 2012: 1106－1114

[6] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[OL]. [2019-07-10].arxiv.org/abs/1409.1556

[7] HE Kaiming ,ZHANG Xiangyu, *et al*. Deep residual learning for image recognition[C].Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016:770-778

[8] Shelhamer E, Long J, Darrell T, *et al*. Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651

[9] HE Kaiming , Gkioxari G, Dollar P, *et al*. Mask R-CNN[C].Proceedings of the International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2017: 2980-2988

[10] REN Shaoqing,HE Kaiming, Girshick R B, *et al*. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149

[11] CHEN Liang-chie,YANG Yi, *et al*. Attention to Scale: Scale-Aware Semantic Image Segmentation[C].Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016:3640-3649

[12] YANG Xue,FAN Yong,GAO Lin,QIU Yunchun. Image Semantic Segmentation Based on Texture Element Block Recognition and Merging[J]. Computer Engineering, doi: 10.3969/j.issn.1000-3428.2015.03.047.
杨雪,范勇,高琳,邱运春. 基于纹理基元块识别与合并的图像语义分割[J]. 计算机工程, 2015, 41(3): 253 - 257.

[13] Szegedy C, LIU Wei , *et al*. Going deeper with convolutions[C].Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2015:1-9

[14] Vaswani A, Shazeer N, Parmar N, *et al*. Attention is All you Need[J]. Neural Information Processing Systems, Long Beach CA, USA, 2017: 5998-6008

[15] Vijay B, Alex K, *et al*. SegNet:A Deep Convolutional

Encoder-Decoder Architecture for Image Segmentation[J]. Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12):2481-2495

[16] Ronneberger O, Fischer P, *et al*. U-net: Convolutional networks for biomedical image segmentation[C] .Proceedings of the Medical Image Computing and Computer-Assisted Intervention. Heidelberg: Springer, 2015:234-241

[17] CHEN,Liang-Chieh,George Papandreou , *et al*. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs[J]. Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4):834-848

[18] LIU Wei,Rabinovich A, *et al*. ParseNet: Looking Wider to See Better[OL]. [2019-07-10]. arxiv.org/abs/1506.04579v2

[19] Srivastava N , Hinton G , Krizhevsky A , *et al*. Dropout: A Simple Way to Prevent Neural Networks from Overfitting[J]. Journal of Machine Learning Research, 2014, 15(1):1929-1958.

[20] ZHAO Hengshuang ,SHI Jinpeng , *et al*. Pyramid Scene Parsing Network[C].Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017:6230-6239

[21] Marvin T, Michael W, *et al*. MultiNet: Real-time Joint Semantic Reasoning for Autonomous Driving[C].Proceedings of the Intelligent Vehicles Symposium, Piscataway: IEEE 2018:1013-1020

[22] Ishan M, Abhinav S, *et al*. Cross-Stitch Networks for Multi-task Learning[C].Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016:3994-4003

[23] Mccann B, Keskar N S, Xiong C, *et al*. The Natural Language Decathlon: Multitask Learning as Question Answering[OL].[2019-07-10]. arxiv.org/abs/1806.08730

[24] ZHANG Zhanpeng ,LUO Peng,*et al*. Facial Landmark Detection by Deep Multi-task Learning[C].Proceedings of the European Conference on Computer Vision. Heidelberg: Springer, 2014:94-108

[25] HU Han, GU Jiayuan , et al. Relation Networks for Object Detection[C].Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 3588-3597

[26] HU Jie, LI Shen, et al. Squeeze-and-Excitation Networks[C].Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018:7132-7141

[27] XIE Saining , Ross Girshick,et al. Aggregated residual transformations for deep neural networks[C] .Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 5987-5995

[28] Oyedotun O K, Shabayek, A R, et al. Highway Network Block with Gates Constraints for Training Very Deep Networks[C].Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018:1739-1745

[29] Holmstrom L, Koistinen P. Using additive noise in back-propagation training[J]. Transactions on Neural Networks, 1992, 3(1):24-38

[30] HE Kaiming, ZHANG Xiangyu, et al. Identity

Mappings in Deep Residual Networks[C]. Proceedings of the European Conference on Computer Vision, Heidelberg: Springer, 2016:630-645