文章编号:1671-4598(2024)01-0157-08 DOI:10.16526/j.cnki.11-4762/tp.2024.01.023 中图分类号:P391 文献标识码:A

融合残差连接的图像语义分割方法

王龙宝[1,2]，张珞弦[1]，张 帅[3]，徐 亮[4]，曾 昕[4]，徐淑芳[1,2]
(1. 河海大学 计算机与信息学院，南京 210000；
2. 河海大学 水利部水利大数据技术重点实验室，南京 210000；
3. 中国电建集团 昆明勘测设计研究院有限公司，昆明 650000；
4. 长江生态环保集团有限公司，武汉 430061)

摘要：由于传统 SegNet 模型在采样过程中产生了大量信息损失，导致图像语义分割精度较低，为此提出了一种融合残差连接的新型编一解码器网络结构：文中引入了多残差连接策略，更为全面地保留了多尺度图像中包含的大量细节信息，降低还原降采样所带来的信息损失；为进一步加速网络训练的收敛效率，改善样本的不平衡问题，设计了一种带平衡因子的交叉熵损失函数，对正负样本不平衡现象予以针对性的优化，使得模型的训练更加高效；实验表明该方法较好地解决了语义分割中信息损失以及分割不准确的问题，与 SegNet 相比，本网络在 Cityscapes 数据集上进行精细标注的 mIoU 值提高了约 13%。

关键词：语义分割；残差连接；交叉熵损失函数；SegNet 模型；深度学习

Image Semantic Segmentation Method for Fusion Residual Connection

WANG Longbao[1,2], ZHANG Luoxian[1], ZHANG Shuai[3], XU Liang[4], ZENG Xin[4], XU Shufang[1,2]
(1. College of Computer and Information, Hohai University, Nanjing 210000, China;
2. Key Laboratory of Water Big Data Technology of Ministry of Water Resources, Hohai University, Nanjing 210000, China;
3. Power China Kunming Engineering Corporation Limited, Kunming 650000, China;
4. Yangtze River Ecology and Environment Co., Ltd., Wuhan 430061, China)

Abstract: Due to the large amount of information loss generated by traditional SegNet model during the sampling process, it causes the accuracy of image semantic segmentation low. Therefore, a new encoder-decoder network structure with fusion residual connection is proposed. The multi-residual connection strategy is introduced to fully retain a large number of detailed information contained in multi-scale images, and reduce the information loss caused by sampling. In order to further accelerate the convergence efficiency of network training and improve the imbalance problem of samples, a cross-entropy loss function with balance factor is designed, and the imbalance phenomenon of positive and negative samples is emphatically optimized to train the model more efficient. Experimental results show that this method solves the problems of information loss and inaccurate segmentation in semantic segmentation, and compared with SegNet model, the fine labeling mean intersection over union (mIoU) index of the network on Cityscapes dataset is increased by about 13%.

Keywords: semantic segmentation; residual connection; cross entropy loss function; SegNet model; deep learning

0 引言

在过去的 30 年里，图像语义分割是计算机视觉中的关键任务之一，现实生活中也有越来越多的应用场景需要从影像中推理出相关的知识或语义。图像语义分割是在像素级别上的分类，属于同一类的像素都将被归为一类，即将图像中的所有像素划分为有意义的对象类，因此图像语义分割是从像素级别来理解图像的。图像语义分割与实例分割不同，语义分割不会将同一类的实例进行区分，只关注每个像素的类别，如果输入的对象中有两个具有相同类别的对象，那么语义分割不将其划分为单独的对象，而实例分割是需要对对象个体进行区分的，即实例分割对同一类的不同对象也会进行分割。

由于图像语义分割技术有助于理解图像中的具体内容，并且能够帮助人们确定物体之间的关系，因此图像分割的应用对于各种领域的发展都有所帮助[1-5]，比如自动驾驶、卫星图像分析、人脸识别、医学影像诊断等。具体而言，结合图像语义分割使得机器可以智能地对医疗影像进行分析，大大减少了运行诊断测试所需的时间的同时也很大程度的降低了医生的工作负担。此外，在自动驾驶过程中利用图像语义分割技术实时分割道路场景，使得自动驾驶汽车有环境感知的能力，以便自动驾驶车辆可以在道路上进

收稿日期:2023-02-16； **修回日期**:2023-04-03。
基金项目:云南省科技厅重大科技专项计划项目(202202AF080003);长江生态环保集团有限公司科研项目(HBZB2022005)。
作者简介:王龙宝(1977-),男,博士,高级工程师。
通讯作者:徐淑芳(1981-),女,博士,副教授。
引用格式:王龙宝,张珞弦,张 帅,等.融合残差连接的图像语义分割方法[J].计算机测量与控制,2024,32(1):157-164.

行安全行驶。

当前，图像语义分割方法分为传统图像语义分割方法和基于深度学习的图像语义分割方法。

传统的图像语义分割算法通常是基于聚类方法，并且往往还需要利用额外的轮廓、边缘等信息进行辅助分析[1,6]。假定同一区域内的像素点为同一类别，利用已有的聚类方法，将这些像素点聚类即可实现图像的分割。近年来，各研究者对基于聚类图像分割的技术进行了许多改进和发展，其中最著名、最重要的技术之一是使用马尔可夫过程进行建模。除此方法外，文献［7］将边缘提取、图像分割以及层次分析法结合起来。文献［8］对 SAR 影像无监督学习范围的扩大进行了研究。尽管传统的图像语义分割方法能够实现对图像的分割，并且对许多领域的发展都有一定的促进作用，但是，它仅仅是通过提取图片的低级特征信息来进行分割，并没有将图像的语义信息纳入到其中，所以，传统的图像语义分割方法的图像分割效果非常有限[9]。

与传统方法相比，基于深度学习的图像语义分割方法显著提高了分割效果，且从分割结果可以直接清楚的知道分割出来的具体是什么物体。基于深度学习的图像语义分割算法，可以有效地克服大部分传统的图像语义分割算法中所忽视的目标边缘问题，并且对椒盐噪声也具有鲁棒性[10-11]。深度学习[12-13]在计算机视觉中被广泛使用，通过增加模型的深度可以提高算法的性能和准确性，利用深度学习可以快速地从非常大的数据集中提取图像特征。

卷积神经网络（CNN，convolutional neural network）[14-15]是一种人工神经网络，其可以类似人一样具有简单的决定能力和简单的判断能力，在图像和语音识别方面可以给出更好的结果，在图像识别领域也被广泛应用。CNN 的结构可以分为三层，分别是卷积层、池化层和全连接层。卷积层的主要作用是进行特征提取以及特征映射；池化层进行下采样以降低空间分辨率和训练参数；全连接层就是一个完全连接的神经网络，通常在 CNN 尾部对卷积层以及池化层得出的特征进行重新拟合，通过调整权重和网络连接得到分类的结果，减少特征信息损失。CNN 本质上是多层感知器，成功的关键在于其网络连接和共享权重的方式。这种方法一方面降低了过度拟合的风险，另一方面减少了权重的数量，使得优化整个网络比其他方法更容易。然而，CNN 不能够训练不同大小的图像，由于全连接层的输入层中的神经元数量是固定的，因此卷积层的输入图像的尺寸大小是固定的。

全卷积神经网络（FCN，fully convolutional networks）将 CNN 最后一层全连接层替代为卷积层，消除了全连接层输入神经元个数的限制，解决了 CNN 卷积层必须是相同输入大小的问题，FCN 能够接受任意大小的输入图像。FCN 通过反卷积将上一层的特征图上采样，将其还原为与输入图像一致的尺寸大小，在保持原输入图像的空间信息的前提下，对每一幅图像都生成一个预测，并在此基础上对图像进行逐像素分类。此外，由于在卷积过程中避免了使用像素块带来的重复存储和计算卷积的问题，因此与 CNN 相比，FCN 减少了模型中的参数，提高了算法的运算效率。

然而，FCN 方法仍然存在一些问题，输出特征图通过卷积层和池化层的交替传播进行下采样，因此 FCN 直接预测通常是低分辨率的，目标边界也相对较为模糊。为了解决这个问题，最近提出了各种基于 FCN 的方法。例如，文献［16］中提出了一种多尺度卷积网络，包括多个具有不同分辨能力的子网络，以便逐步改进粗预测。文献［17］提出高低层特征融合，即在多层的输出后是一个反卷积层，用于对高密度的像素输出进行双线性的上采样，从而有效增强了图像语义信息特征以及空间信息特征。文献［18］为了精确地重构物体边界的高度非线性结构，用一个深度反卷积网络代替了文献［17］中的简单反卷积处理，以识别像素级的类别标记。此外，FCN 中全卷积的设计模式仍然保留使用了卷积神经网络中的池化层，忽略了高分辨率的特征图必然会导致边缘信息的丢失。同时，FCN 解码器中复用编码器特征图的方式使其在测试时显存消耗也很大，忽略了图像的位置信息以及减小了特征图的分辨率。

编码器和解码器结构是解决以上问题的关键，大多数基于深度学习的语义分割技术都使用编码器和解码器架构。编码器负责将输入转化为特征，解码器则负责将特征转化为目标。SegNet[19]和 U-Net[20]是两个典型的用于图像语义分割的编码－解码器结构。SegNet 是基于全卷积神经网络搭建的一种编码－解码器网络结构，通过编码器提取图像特征后，再通过解码器逐步还原到与原图相同分辨率的分割结果。U-Net 是为了帮助生物序列中的图像分割而创建的，它由两部分组成：收集上下文的收缩路径和用于识别精确位置的对称扩展路径相比于已有的深度卷积神经网络语义分割方法，该方法提出了一种更为稳定的网络结构。SegNet 的编码器部分使用了去除全连接层的 VGG-16 网络[21]，解码器部分使用了一系列上采样和卷积层，这样可以实现通过保留的最大池化层的最大值索引来恢复特征图分辨率，并利用可学习的后续卷积层来产生稠密特征。

尽管此方法提出了最大池化索引策略，尽可能保留了各特征图像中的关键信息，但是在编码器网络中仍旧不可避免的产生了大量信息损失，这些信息损失在解码器网络中往往是不可恢复的，导致语义分割结果精度的不理想。

因此，本文设计一种更加优化的网络模型，以降低 SegNet 在编码器网络中提取高维特征时产生的信息损失，同时，在解码时能够更加完整地勾勒分割边界，提高分割精度，并控制网络的参数总量和执行时的内存占比，从而能够在较低时间消耗和硬件需求的前提下，实现多目标的精确识别和多场景的全面理解。

1 相关理论基础

1.1 SegNet 模型

SegNet 模型核心是由一个编码器网络以及相应的解码器网络组成，整体架构如图 1 所示。

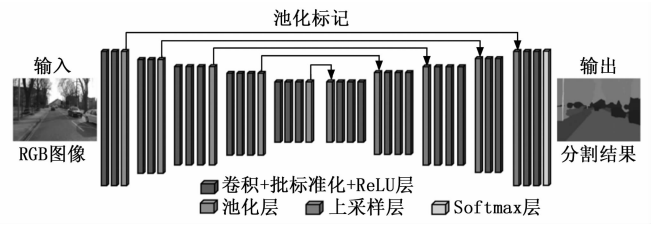

图 1　SegNet 模型结构

编码器网络主要由卷积层、批归一化层、ReLU 层和池化层组成。编码器网络中的卷积层对应于 VGG16 网络中的前 13 个卷积层。卷积层通过卷积提取特征，其使用的是 same padding 卷积，不会改变特征图的尺寸；批归一化层（Batch Normlisation）起到归一化的作用；ReLU 层应用逐元素非线性激活函数（ReLU）来加快此网络的收敛速度；池化层执行最大池化操作，记录最大值的索引位置并将结果输出。对于图像分类任务而言，多层最大池化和下采样由于平移不变性可以获得较好的鲁棒性，但同时也导致了特征图大小和空间信息的损失。为了解决这个问题，SegNet 只存储每个编码器特征映射的池化最大索引或每个池化窗口中最大特征值的位置。

解码器将编码器获取到的物体信息以及大致的位置信息与特定的像素点相对应，对缩小后的特征图像进行上采样，通过对上采样后的图像进行卷积处理，完善物体的几何形状，以补偿因编码器中的池化层将物体缩小造成的细节损失。解码器有与编码器相对应的上采样层、卷积层、批归一化层以及 ReLU 层。其中上采样层具体操作为对输入的特征图放大两倍，然后将输入的特征图数据根据池化层的最大索引位置放入，其他位置均为 0。解码器的最终输出被馈送到 soft-max 分类器，对每个像素进行独立分类，预测的分割结果对应于在每个像素处具有最大概率的类别。

SegNet 的创新之处在于解码器阶段的上采样层使用了编码器阶段池化层的最大池化索引来进行反池化。与 FCN 中利用双线性插值进行上采样的方式相比，反池化操作大大减少了模型的参数量。SegNet 相比其他架构更有效的原因正是由于其只存储特征图的最大池化索引，并在其解码器网络中使用它们来实现良好的性能。与 FCN 进行对比，SegNet 在达到较好的分割性能的同时，也具有较为均衡的内存占用率和准确率，反池化也提升了模型对边界的描述能力。与其他竞争架构相比，SegNet 结构在推理时间和有效的推理内存方面都体现出了较为良好的性能。

1.2　交叉熵损失函数

交叉熵损失函数是处理分类问题中常用的一种损失函数。交叉熵是用于描述两个概率分布之间的距离，交叉熵越小，两个概率的分布便越接近。交叉熵损失函数常常用在逻辑回归问题即求解离散的分类问题上，用来作为预测值和真实标签值的距离度量。模型在使用梯度下降更新参数时，模型训练的速度取决于学习率和偏导数值。偏导数的大小反映了模型的误差，值越大，模型效应越差，但同时模型训练则越快。因此，如果利用逻辑函数获得概率并且结合使用交叉熵损失函数，则模型效果不好时学习速度会更快，如果模型效果良好，学习速度会较慢。

标准的交叉熵损失函数（CE，cross-entropy loss）如下所示：

$$CE(p,y)=\begin{cases}-\log(p) & \text{if}(y=1)\\ -\log(1-p) & \text{otherwise}\end{cases} \tag{1}$$

其中：p 代表正样本的预测概率，y 代表样本标签，正类为 1，负类为 0。log 表示自然对数，底数为 e。可以看出，预测越准确，计算出的损失值就越小，如果预测完全正确，则计算的损失值就为 0，因此符合优化方向。为方便表示，简记如下：

$$p_t=\begin{cases}p & \text{if}(y=1)\\ 1-p & \text{otherwise}\end{cases} \tag{2}$$

则交叉熵可以表示为：

$$CE(p,y)=CE(p_t)=-\log(p_t) \tag{3}$$

交叉熵损失函数由于引入了类间竞争的特性，使得类间的互补性更强，但其仅仅覆盖了正确标记的正确率，并没有考虑其它非正确标记间的差别，导致所获得的特征有所偏离。

2　融合残差连接的语义分割网络结构

2.1　模型建立

本文所设计的网络结构是基于 SegNet 所提出的编－解码器结构，搭建一种残差连接的语义分割网络结构。对于一张普通拍摄照片而言，浅层 CNN 提取的特征往往包含更多的边界、纹理等直观视觉信息，深层 CNN 往往提取的是更高级的抽象特征，只有将二者有机结合，才能实现语义分割精度的提升。加深、加宽网络结构，虽然能够提高分割精度但是带来了大量的参数负担和冗余，因此需要引用残差连接和 concatenation（级联）操作，有效的将浅层视觉特征与深层语义特征进行结合。同时将已有的普通的层间连接调整为残差连接，总体来看，增加的参数量可以忽略不计。

如图 2 所示，其中输入 $H\times W$ 为大小的图片，输出为同样尺寸的分割结果。蓝色框表示卷积以及之后的线性激活函数层和批量归一化的操作，在框内的数字表示当前操作结果后图像大小及特征图数量。可以看出，在编码器网络阶段，通过三次降采样，将图像缩小到 $\frac{H}{8}\times\frac{W}{8}$，同时产生 256 个相同大小的特征图。在解码器网络阶段，利用特征图像上采样恢复至 $H\times W$ 大小并通过 softmax 函数输出像素所属语义类别的概率。

特别地，在 SegNet 中的图像恢复阶段，在高级语义特征提取过程中产生的降采样特征图像 $H\times W$ 、$\frac{H}{2}\times\frac{W}{2}$ 和 $\frac{H}{4}\times\frac{W}{4}$ 同样具备大量丰富的图像特征信息，然而由于网络结构的局限性导致图像上采样时无法通过大量稀疏的特征映射产生密集的特征映射图。在此过程中，不可避免的损失

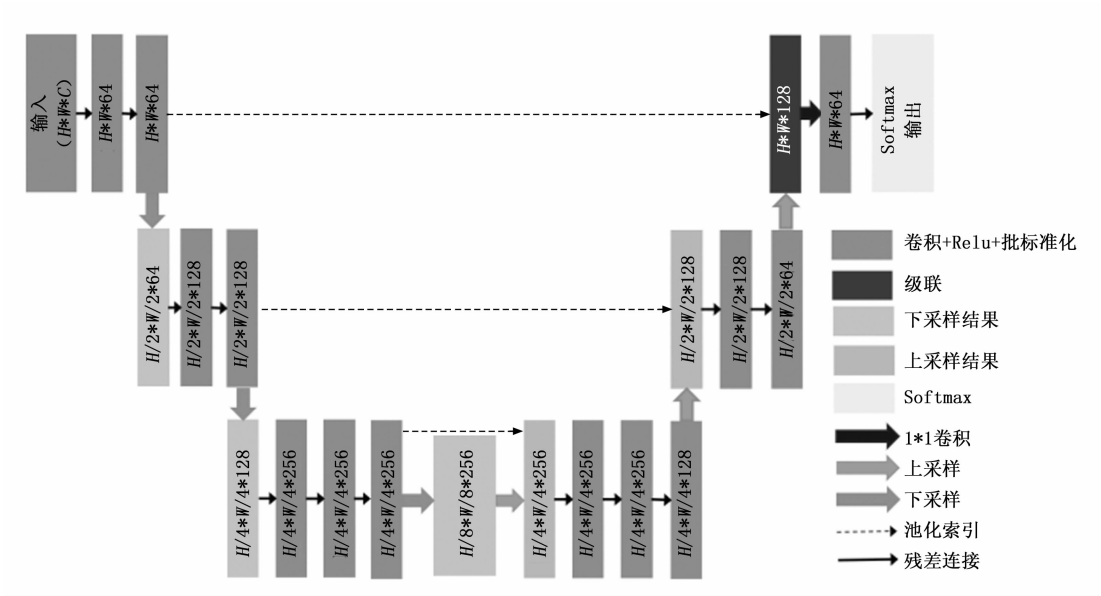

图 2 融合残差连接的语义分割网络结构图

了大量关键信息，使得最终的图像分割精度无法满足需求。本网络结构结合最大池化索引和残差连接，将编码器阶段提取的浅层特征映射输入解码器的非线性上采样阶段，通过反卷积产生的密集的特征映射图像，最大程度的保留原始图像的色彩、纹理和边界等特征。

将一张带训练图像输入此改进的 SegNet 网络结构，其在网络中共经过一下几步过程：

1）将图像进行卷积操作，得到 $H*W*64$ 个通道的特征图像，记为 F_1。

2）下采样得到 $H/2*W/2*64$，然后再进行卷积操作得到 $H/2*W/2*128$，记为 F_2。

3）下采样得到 $H/4*W/4*128$，然后进行卷积操作得到 $H/4*W/4*256$，记为 F_3。

4）下采样得到 $H/8*W/8*256$，记为 F_4。

5）上采样得到 $H/4*W/4*256$，记为 F'_3，其计算公式如下：

$$F'_3 = Fuse(PI(F_3), F_4) \tag{4}$$

其中：PI 表示池化索引 $Fuse$ 表示特征映射图的融合。然后将 F'_3 通过反卷积得到 $H/4*W/4*128$ 的特征图像，记为 $De(F'_3)$。

6）上采样得到 $H/2*W/2*128$，记为 F'_2，其计算公式为：

$$F'_2 = Fuse(PI(F_2), De(F'_3)) \tag{5}$$

然后将 F'_2 通过反卷积得到 $H/2*W/2*64$，记为 $De(F'_2)$。

7）将 $De(F'_2)$ 上采样后恢复分辨率至 $H*W$，此时结合 F_1，通过级联操作得到 $H*W*128$，记为 F_c，其计算公式为：

$$F_c = Conc(Fuse(PI(F_1), De(F'_2)), F_1) \tag{6}$$

8）最终通过 softmax 函数对每一像素所属类别予以赋值并输出相应的语义分割结果。

2.2 模型训练

图 3 本方法网络模型训练流程图。首先对数据集进行预处理以及训练集和验证集划分。其次将处理好的数据输入初始化参数的语义分割网络模型。根据分割结果的交叉熵损失最小原则，不断迭代网络更新模型参数，直至收敛并达到最小损失。最后输出最优网络模型和参数。

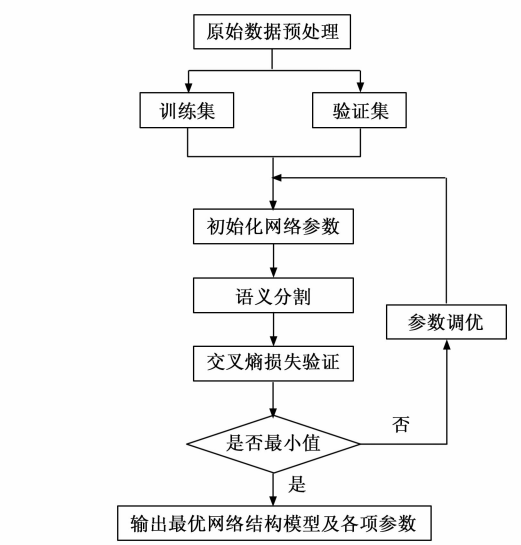

图 3 网络训练流程图

2.3 改进的交叉熵损失函数

标准的交叉熵损失函数计算公式中所有样本的权重都是相同的，因此如果正、负样本不均衡，大量简单的负样本会占据主导地位，少量的难样本与正样本会起不到作用，导致精度变差。

因此，我们引入平衡因子，取值在［0，1］区间内。

$$\beta=\begin{cases}\beta & \text{if}(y=1)\\ 1-\beta & \text{otherwise}\end{cases} \tag{7}$$

设计的改进的交叉熵损失公式（B-CE，balanced cross-entropy loss）如下：

$$CE(p,y)=-\beta\log(p_t) \tag{8}$$

引入平衡因子的交叉熵损失函数，在收敛效率上比原函数更快，主要是在不均衡分布的类别像素上，其迭代优化的效率更具备针对性，因此整体收敛效率得到了提升。

3　实验结果与分析

3.1　实验环境

本文实验系统为 Windows 10 professional，24 GB RAM，处理器为 Intel（R）Core i7- 8750H，2.20 GHZ，GPU 为 NVIDIA GeForce GTX 1 060 6 GB。实验平台为 Matlab 2018 b，基于 MatconvNet 和 visual C++ 2015 搭建深度学习网络模型，模型训练和测试是基于 cuda9.0 搭建的 GPU 环境。

3.2　评价指标

IoU（Intersection over Union）的全称为交并比，具体是指预测候选边界集和真实边界集的交集和并集的比值，是当前目标识别和语义分割研究最通用的评价指标。IoU 是一个较为简单的测量标准，只要是在输出中得出一个预测范围的任务都可以用 IoU 来测量。交并比的数学含义如图 4 所示。最理想情况是候选边界集与真实边界集完全重叠，即比值为 1，即预测精确度越高。交并比的计算公式如下：

$$IoU=\frac{area(C)\bigcap area(G)}{area(C)\bigcup area(G)} \tag{9}$$

一般约定，0.5 是阈值，用来判断预测的边界框是否正确，IoU 越高，边界框越精确。

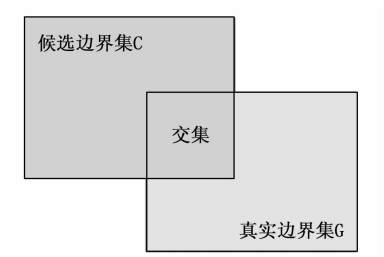

图 4　交并比的数学含义

3.3　实验结果与分析 PASCAL VOC 2012 数据集

以下内容将展示在 PASCAL VOC 2012 数据集上进行实验的结果，包括各类别 IoU 的数值统计与分析和随机样本的视觉解析，并从评价指标和视觉效果两方面全方位评估所设计网络结构的有效性和先进性。

3.3.1　数据集简介及参数设置

PASCAL VOC 2012 作为基准数据之一，数据集包含原图片总共 17 125 张及其对应的标注图。在对象检测、图像分割网络对比实验与模型效果评估中被频频使用。Pascal VOC 2012 数据集针对视觉任务中监督学习提供了标签数据，它主要有 4 个大类别，分别是人、常见动物、交通车辆、室内家具用品，并可细分为二十个类别：

1）person：person；

2）animal：bird，cat，cow，dog，horse，sheep；

3）vehicle：aeroplane，bicycle，boat，bus，car，motorbike，train；

4）indoor：bottle，chair，dining table，potted plant，sofa，tv/monitor。

此外，针对该数据集的实验中各项参数设置如表 1 所示，测试数据为完全随机抽选，在网络训练时采用带动量的随机梯度下降法作为优化器，学习率和动量参数设置为 0.1 和 0.9。

表 1　PASCAL VOC 2012 数据集超参数设置

名称	数值
训练图像数量/幅	5 717
验证图像数量/幅	5 823
测试数据数量/幅	1 000
学习率(learning rate)	0.1
动量	0.9
迭代上限(epoch)	200

3.3.2　实验结果对比与分析

如表 2 所示，为本方法、SegNet 在 PASCAL VOC 2012 数据集上的分割表现。可以看出，相比于其他两种方法，本方法整体分割精度表现优异，其中 Bird 等 7 类物体交并比超过 90%，13 类物体超过 80%，mIoU 达到 80.81%，相比于 SegNet 提高了约 8 个百分点。

表 2　PASCAL VOC 2012 测试集各类别 IoU

类别	IoU	
	本文	SegNet
Aeroplane	92.71	89.9
Bicycle	60.33	39.3
Bird	94.05	79.7
Boat	74.94	63.9
Bottle	82.86	68.2
Bus	95.10	87.4
Car	88.54	81.2
Cat	94.71	86.1
Chair	45.52	28.5
Cow	91.38	77.0
Dining table	76.32	62.0
Dog	90.58	79.0
Horse	91.76	80.3
Motorbike	88.13	83.6
Person	87.98	80.2
Potted plant	69.87	58.8
Sheep	82.83	83.4
Sofa	60.91	54.3
Train	80.75	80.7
Tv/monitor	66.83	65.0
Mean	80.81	72.5

如图 5 所示，数据样本的视觉展示进一步对分割效果进行了评估。为更充分和全面验证方法的分割能力，从数据集中随机选取的 5 个样本几乎包含数据集具有的所有类别的物体。尽管结果相似度较高，但仍能直观的从分割结果看出本方法分割精度更高。首先从整体上看，SegNet 与本方法均能够较好的实现图像语义分割任务，基本上能够将图像中的目标物体识别并标注出。然而，在部分关键细节处，本方法表现更佳。如图 5（a）中所示，自行车轮廓的准确勾勒需要准确的高频边界信息，相比于 SegNet，本方法对低级别的高频边界信息进行了更大程度的保留并应用于解码器网络，使得最终分割结果中自行车轮廓更为清晰。同样地，如图 5（b）中鸟类双脚的分叉处，图 5（c）中椅子的轮廓边界，图 5（e）中自行车轮廓和远端人物边界等高频细节信息处，本方法更具针对性的低级别特征与高级别语义特征融合方法使得分割结果更接近真实标注图。

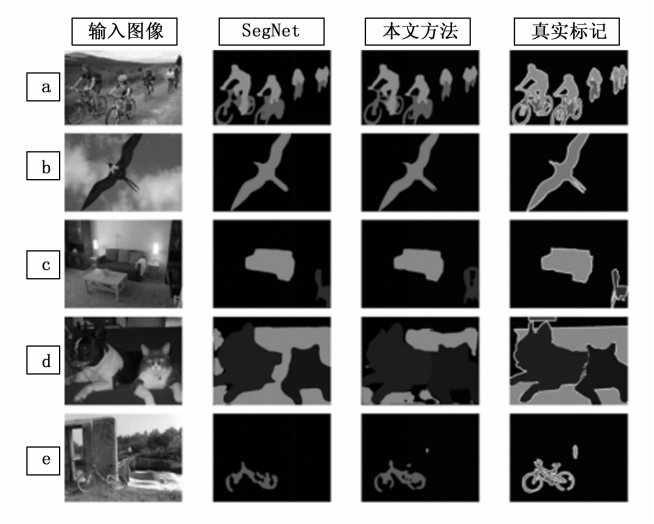

图 5 PASCAL VOC 2012 随机样本测试结果展示

综上可得，在添加了有效的多残差连接之后，该语义分割网络所提取的特征保真度更高，能够与原图保持更高的相关性，使得图像的像素级分类结果和边界定位效果更优于 SegNet。从视觉感受的定性分析情况以及各类别交并比的定量分析结果来看，本方法综合利用了最大池化索引的有效性和多残差连接的灵活性，使得图像语义分割结果达到更高的精度，更加满足实际应用需求。

3.4 实验结果与分析－Cityscapes 数据集

以下内容为在 Cityscapes 数据集上进行实验对比结果与分析，包括各类别 IoU 的数值统计与分析和随机样本的视觉解析，分别从评价指标和视觉效果两方面全方位评估所设计网络结构的在不同类型数据集上的鲁棒性和有效性。

3.4.1 数据集简介

Cityscapes 数据集是由包含戴姆勒在内的三家德国单位联合提供的，是一个新的大规模数据集，主要关注城市环境中驾驶场景的图像。Cityscapes 数据集涵盖了 50 个城市的不同季节、不同时段的街道场景，包括 5 000 张精标注图片和 20 000 粗标注图片，其中精标注图片数据集被划分为训练集（2 975 张）、验证集（500 张）和测试集（1 525 张）。Cityscapes 数据共有两种数据标注格式，分别是实例分割和语义分割所采用的分割图格式以及多边形边框的 json 格式。精标注数据集中的每张图片都同时拥有 3 个标注文件，即实例分割标注、语义分割标注、多边形标注。标注类别共分为 8 组，每组的具体类别如下所示：

1）flat：road，sidewalk，parking＋，rail track＋；

2）human：person＊，rider＊；

3）vehicle：car＊，truck＊，bus＊，on rails＊，motorcycle＊，bicycle＊，caravan＊＋，trailer＊＋；

4）construction：building，wall，fence，guard rail＋，bridge＋，tunnel＋；

5）object：pole，pole group＋，traffic sign，traffic light；

6）nature：vegetation，terrain；

7）sky：sky；

8）void：ground＋，dynamic＋，static＋。

其中＊表示部分区域连在一起的实例，会作为一个整体来标注；＋表示该类别不包含在验证集中，并被视为无效标注。

本文在精标注数据集上进行分割实验并与 SegNet 方法在此数据集上的分割表现进行对比分析。如表 3 所示，为本文在部署实验时的各项参数数值。同样地，采用带动量的随机梯度下降法作为优化器，其学习率和动量数值分别预设为 0.1 和 0.9。

表 3 Cityscapes 数据集超参数设置

名称	参数数值
训练图像数量/幅	2 975
验证图像数量/幅	500
测试数据数量/幅	1 525
学习率(learning rate)	0.1
动量	0.9
迭代上限(epoch)	200

3.4.2 实验结果对比与分析

表 4 展示了本文所提方法和 SegNet 在 Cityscapes 数据集进行精细标注的各类别交并比以及平均交并比。总的来看，本文所设计网络结构实现了更高的交并比表现，其预测的分割结果更相近于真实标记数据集。相比于 SegNet，本网络在 mIoU 值上提高了约十三个百分点。

如图 6 所示，随机选择 6 张测试样本做视觉展示，分别通过 SegNet 和本文方法进行分割预测并产生精细标注分割图，最右侧为真实标记结果。整体分割结果上来看，SegNet 与本文方法均能实现较好的分割结果。由于是车载摄像装置拍摄的图像，其中前方马路等主体大范围目标均可以实现较为准确的分割，这一结果和 IoU 值形成对应。

表 4　Cityscapes 测试集各类别 IoU

类别	SegNet	本文
Road	96.4	96.6
Sidewalk	73.2	83.3
Building	84	90.3
Wall	28.5	36.8
Fence	29	36.4
Pole	35.7	56.1
Traffic light	39.8	66.9
Traffic sign	45.2	71.3
Vegetation	87	91.8
Terrain	63.8	71.6
Sky	91.8	92.6
Person	62.8	77.3
Rider	42.8	51.9
Car	89.3	94.5
Truck	38.1	48.2
Bus	43.1	76.1
Train	44.2	64.3
Motorcycle	35.8	62.2
Bicycle	51.9	68.7
Mean	57	70.4

然而由于 SegNet 的细节处理不尽精细，导致其部分分割结果不能达到满意。具体来说，如图 6（a）中左侧人群部分的分割结果比较模糊，然而本方法的分割结果能够将人群中的不同个人进行一个较为优化的分割，其结果也更趋近于最右侧的真是标注结果。如图 6（b）中的最右侧交通指示牌、左侧绿色植物右上侧的交通指示牌，SegNet 的分割结果无法达到较为精确的分割，指示牌识别上出现明显的少分、漏分情况，而本文方法的分割结果则更为精细和准确。同样的情况包括如图 6（c）中的路灯、图 6（d）中的立柱、行人、图 6（f）中的自行车及车手等目标分割状况。综上所述，本文所设计的网络结构在细节标注及边界勾勒时的表现全面优于 SegNet 的分割性能。

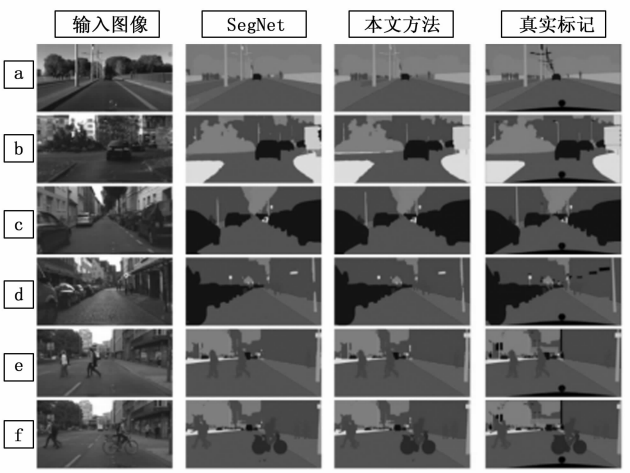

图 6　Cityscapes 随机测试样本结果展示

4　消融实验——损失函数

本文将通过实验验证所设计的带有平衡因子的交叉熵损失函数的影响，尤其在模型训练的收敛效率方面的表现。以 PASCAL VOC 2012 数据集作为训练数据，在训练过程中的损失曲线如图 7 所示。

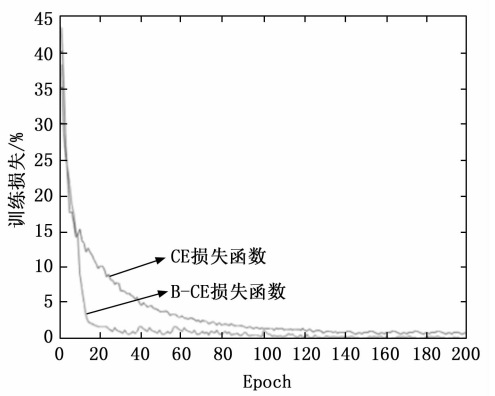

图 7　不同损失函数的训练损失

从图 7 中可以得出采用 B-CE 损失函数能够有效提高迭代效率，在采用 B-CE 损失函数之后在大约第 80 个 epoch 开始达到收敛状态。相对地，使用原始的 CE 损失函数尽管收敛过程较为稳定，但是收敛速率较之更慢，在大约100 epoch 时才能达到收敛状态；从损失层面来看，B-CE 可以帮助网络结构实现更少的损失，分析其主要原因可能来自负样本带来的损失的进一步减少，因为提高了对负样本、难分类样本的关注度，训练更具针对性，并且本来易分的样本也不会因为损失函数的微调而导致错误分类。从实际的训练和损失曲线中动态关系中，我们进一步验证了所改进损失函数的有效性。

5　结束语

由于 SegNet 模型在多次降采样和上采样过程中产生了大量信息损失，语义分割精度表现因此而受到较大限制。为解决此问题，本文设计了一种融合多残差连接的新型编一解码器网络结构，在不新增大量参数负担的前提下，通过引入若干残差连接，使得不同分辨率的低级别空间信息特征和高级别语义特征得以充分利用，进而显著减小上采样和下采样过程产生的信息损失。此外，为缓解类别非均衡分布带来的消极影响，本文基于交叉熵损失函数设计了一种带平衡因子的交叉熵损失函数，不仅促进了模型收敛效率，同时在达到收敛状态时降低了大量损失，使得模型具备更强的性能，实现更高的分割精度。通过在 PASCAL VOC 2012 和 Cityscapes 数据集上进行实验对比和分析，通过量化评价指标和视觉分析效果证实本方法的分割表现明显优于 SegNet。

参考文献：

[1] WEINLAND D, RONFARD R, BOYER E. A survey of vision-

based methods for action representation, segmentation and recognition [J]. Computer Vision and Image Understanding, 2011, 115 (2): 224 - 241.

[2] PATIL D, DEORE S. Medical image segmentation: a review [J]. International Journal of Computer Science and Mobile Computing, 2013, 2 (1): 22 - 27.

[3] KEN C, RELJA A, OMKAR P, et al. On-the-fly learning for visual search of large-scale image and video datasets [J]. International Journal of Multimedia Information Retrieval, 2015 (4): 75 - 93.

[4] MALLIK A, CHAUDHURY S. Acquisition of multimedia ontology: an application in preservation of cultural heritage [J]. International Journal of Multimedia Information Retrieval, 2012, 1 (4): 249 - 262.

[5] ATMOSUKARTO I, SHAPIRO L. 3D object retrieval using salient views [J]. International Journal of Multimedia Information Retrieval, 2013, 2 (2): 103 - 115.

[6] IIEA D, WHELAN P. Image segmentation based on the integration of colour-texture descriptors-a review [J]. Pattern Recognition, 2011, 44 (10 - 11): 2479 - 2501.

[7] ARBELAEZ P, MAIRE M, FOWLKES C, et al. Contour detection and hierarchical image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33 (5): 898 - 916.

[8] YU P, QIN A, CLAUSI D. Unsupervised polarimetric SAR image segmentation and classification using region growing with edge penalty [J]. IEEE Transactions on Geoscience and Remote Sensing, 2012, 50 (4): 1302 - 1317.

[9] 赵小强, 徐慧萍. 分级特征融合的图像语义分割 [J]. 计算机学与探索, 2021, 15 (5): 949 - 957.

[10] XIN P, JIAN Z, JUN X. An end-to-end and localized postprocessing method for correcting high-resolution remote sensing classification result images [J]. Remote Sensing, 2020, 12 (5): 1 - 21.

[11] ZHUO K P, JIA S X, YU B G. Deep learning segmentation and classification for urban village using a worldview satellite image based on U-net [J]. Remote Sensing, 2020, 12 (10): 1 - 17.

[12] LECUN Y, BENGIO Y, HINTON G. Deep learning [J]. Nature, 2015, 521 (7553): 436 - 444.

[13] GAO J C, WANG H Y, SHEN H Y. Task failure prediction in cloud data centers using deep learning [J]. IEEE Transactions on Services Computing, 2022, 15 (3): 1411 - 1422.

[14] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770 - 778.

[15] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C] // Proceedings of the Advances in neural information processing systems, 2012: 1097 - 1105.

[16] EIGEN D, FERGUS R. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture [C] // Proceedings of the IEEE International Conference on Computer Vision, 2015: 2650 - 2658.

[17] 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.

[18] NOH H, HONG S, HAN B. Learning Deconvolution Network for Semantic Segmentation [C] // Proceedings of the IEEE International Conference on Computer Vision, 2015: 1520 - 1528.

[19] VIJAY B, ALEX K, ROBERTO C. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Scene Segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39 (12): 2481 - 2495.

[20] OLAF R, PHILIPP F, THOMAS B. U-net: Convolutional networks for biomedical image segmentation [C] // Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015: 234 - 241.

[21] 徐昭洪, 刘 宇, 全吉成, 等. 基于 VGG16 预编码的遥感图像建筑物语义分割 [J]. 科学技术与工程, 2019, 19 (17): 250 - 255.

(上接第 156 页)

[12] LONG H, SANG L, WU Z, et al. Image-Based abnormal data detection and cleaning algorithm via wind power curve [J]. IEEE Transactions on Sustainable Energy, 2019, 11 (2): 938 - 946.

[13] 朱倩雯, 叶 林, 赵永宁, 等. 风电场输出功率异常数据识别与重构方法研究 [J]. 电力系统保护与控制, 2015, 43 (3): 38 - 45.

[14] 胡 阳, 乔依林. 基于置信等效边界模型的风功率数据清洗方法 [J]. 电力系统自动化, 2018, 42 (15): 18 - 23.

[15] 雷启龙. 基于阶比分析的风电机组故障预警与诊断研究 [D]. 保定: 华北电力大学, 2015.

[16] KUSIAK A, ZHENG H, ZHE S. On-line monitoring of power curves [J]. Renewable Energy, 2009, 34 (6): 1487 - 1493.

[17] PARK J Y, LEE J K, OH K Y, et al. Development of a novel power curve monitoring method for wind turbines and its field tests [J]. IEEE Transactions on Energy Conversion, 2014, 29 (1): 119 - 128.

[18] 刘 帅, 刘长良, 曾华清. 基于核极限学习机的风电机组齿轮箱故障预警研究 [J]. 中国测试, 2019, 045 (2): 121 - 127.

[19] LIANG G, SU Y, CHEN F, et al. Wind power curve data cleaning by image thresholding based on class uncertainty and shape dissimilarity [J]. IEEE Transactions on Sustainable Energy, 2020, 12: 1383 - 1393.