



兰州大学

本科生毕业论文

论文题目（中文） 基于 U-net 的小样本图像分割

论文题目（英文） Few-Shot Learning for Semantic Segmentation
based on U-net

学生姓名 郁严贵

指导教师 刘振宇

学 院 信息科学与工程学院

专 业 计算机科学与技术

年 级 2018 级

兰州大学教务处

基于 U-net 的小样本图像分割

中文摘要

近年来，伴随着计算机硬件水平的发展，计算机的计算能力不断提高；同时互联网和物联网带来了大量的数据，使我们进入了大数据时代。计算能力的提升和海量的数据为深度学习的发展提供了客观条件，伴随着对深度学习大量的应用研究，深度学习开始在很多领域展示出巨大的潜力。当前，深度学习这一研究领域越来越受到研究者的关注。数据作为深度学习的重要基础，深度学习的成功往往离不开大量的数据。而在很多领域，经常会面临缺乏足够的数据用于训练的问题。小样本是一个普遍面临的问题，为了解决这一问题，小样本这一研究方向越来越受到关注。

本论文的主要研究内容为基于深度学习模型，在可获得的数据量不足的情况下，如何通过具有小样本特点的数据集，实现对深度学习模型更有效的训练，对图像实现更准确的分割效果，提高模型的泛化能力。U-net 是图像分割领域流行的网络模型架构，在医学图像分割方面表现处理优异的性能，同时所需的训练集较少便可以取得比较好的图像分割效果。本论文研究如何在传统的 U-net 网络结构进行改进，更好的解决小样本图像分割问题。本论文的主要研究内容如下：

第一，针对小样本数据量不足的问题，对训练数据集进行数据增强。通过对数据应用多种数据增强操作，分析比较有助于提高图像分割效果的数据增强操作类型。

第二，图像分割经常会面临细节分割效果不好，而在小样本问题下这一表现更加明显，对于边缘的分割效果不理想。常用与图像分割效果评估的损失函数为交叉熵损失函数（crossEntropy）与 Dice 相似系数。在此基础上，评估对边缘的分割效果，对边缘增加一定的权重，使网络对边缘分割更加敏感。

第三，介绍 U-net 网络模型及对其的改进，分析网络模型结构、参数，进行对比实验分析改进的 U-net 模型的训练效果。

关键词：深度学习、卷积神经网络、图像分割、U-net、小样本学习、数据增强

FEW-SHOT LEARNING FOR SEMANTIC SEGMENTATION BASED ON U-NET

Abstract

In recent years, the computing power of computers has been continuously improved with the development of computer hardware. At the same time, the Internet and the Internet of Things have brought a large amount of data, bring humanity into the era of big data. The improvement of computing power and massive data provide objective conditions for the development of deep learning, and with a large number of applied research on deep learning, deep learning has begun to show great potential in many fields. At present, the research field of deep learning is attracting more and more attention from researchers. Data is an important foundation for deep learning, and the success of deep learning is often inseparable from a large amount of data. In many fields, however, the problem of lacking sufficient data for training is often faced. Few shot learning is a commonly faced problem, and in order to solve this problem, the research field of few shot learning has attracted more and more attention.

The main focus of this thesis is on how to train deep learning models more effectively, achieve more accurate segmentation of images and improve the generalization ability of the models based on deep learning models with small sample data sets, given the insufficient amount of available data. U-net is a popular network model architecture in the field of image segmentation, and has shown excellent performance in medical image segmentation, while requiring a small training set to achieve better image segmentation results. This thesis investigates how the traditional U-net network architecture can be improved to better solve the small sample image segmentation problem. The main research contents of this thesis are as follows:

First, data augmentation is applied to the training dataset to address the problem of insufficient data. By applying multiple data augmentation operations to the dataset, the analysis compares the types of data augmentation operations that can help to improve the image segmentation results.

Secondly, image segmentation often faces poor detail segmentation, and this is even more evident in small sample problems, where the segmentation of edges is not ideal. Commonly used loss functions for evaluating image segmentation effects are crossentropy loss function and the Dice similarity coefficient. On this basis, the segmentation effect on edges is evaluated and a certain weight is added to the edges to make the network more sensitive to edge segmentation.

Thirdly, the U-net network model and its improvement are introduced, the network model structure and parameters are analyzed, and comparative experiments are conducted to analyze the

training effect of the improved U-net model.

Key Words: Deep learning, convolutional neural network, image segmentation, U-net, few-shot learning, data augmentation

目 录

中文摘要	I
英文摘要	II
第一章 绪论	1
1.1 研究背景与意义	1
1.2 研究现状综述	2
1.2.1 图像分割研究现状	2
1.2.2 小样本学习研究现状	2
1.3 面临的困难和挑战	4
1.4 论文结构和主要内容	4
第二章 相关原理和技术	6
2.1 下采样	6
2.1.1 卷积	6
2.1.2 池化	7
2.2 上采样	7
2.2.1 双线性插值	8
2.2.2 转置卷积	8
2.2.3 UnSampling	8
2.3 损失函数	9
2.3.1 MSE	9
2.3.2 Dice Loss	9
2.3.3 CrossEntropy	10
第三章 基于改进的 U-net 网络模型进行图像分割	11
3.1 数据集和数据处理	12
3.2 损失函数设计	14

3.3 U-net 模型改进	19
第四章 总结和展望.....	23
4.1 总结	23
4.2 展望	23
参考文献	25
致 谢	27

第一章 绪论

1.1 研究背景与意义

进入 21 世纪以来,随着计算机硬件的迅猛发展,计算机的计算能力得到了极大的提升。同时伴随着大数据时代的到来,在日常生活中与网络中产生了大量的数据,这为深度学习的发展起到了极大的促进作用。近年来,深度学习取得了极大的成功,并且人工智能与深度学习的研究方向也异常火热。

当前,深度学习在计算机视觉 (computer vision,CV),自然语言处理 (nature language process, NLP) 等领域都取得了极大的成功。作为计算机视觉领域的一个重要研究方向,图像分割将图像中感兴趣、关注的区域划分切割出来,是图像分析和处理的重要步骤之一。图像分割具有很大的研究价值和研究意义,在很多领域都有着广泛的应用,如在人脸识别中将人脸信息与背景区分开来,在行人检测中需要分割出每个行人实例,在医学领域中进行医学图像的分割。

随着深度学习的进展,应用深度学习理论在图像分割这一问题上表现出了卓越的性能,当前提出很多方法来解决图像分割这一问题。其中,基于卷积神经网络 (CNN) 的方法在图像分割问题上取得了极大的成功。但是, CNN 网络模型通常需要大量的训练数据。目前,深度学习的成功主要得益于大量的训练数据。然而大量可获得的训练数据在有些问题和情境下是很困难,或者不可能的。对图像分割的准确标注的代价是昂贵的,需要耗费大量的人力成本和精力。在图像分割这一问题下,对图像的标注需要精确到像素级的分类上一件耗时且困难的任务。这导致在很多情境下,如医学图像和工业领域中,缺乏大量的可以获得的训练数据。深层神经网络具有非常强的学习能力,在缺乏足够的训练数据对深度神经网络模型进行训练时,可能会出现过拟合的问题,使得模型的泛化性变差。

深度学习往往需要大量的训练数据进行训练,从大量的数据中提取信息,才能训练得到具有比较好的泛化性能的模型。与此相对,人类具有很强的学习能力,“举一反三”的推理能力,只需要少量的样本,便可以取得很好的学习效果。如何在可获得的训练数据不足的情况下,进行更有效的图像分割,这便是小样本图像分割要研究的问题。小样本学习的目标是在训练数据不足的情况下,在测试集上取得更好的表现效果。因此,研究在小样本这一情景下,如何实现基于深度学习网络模型对于图像实现更好的分割效果,从小样本中取得更好的学习效果,更符合实际要面临的实际情况,具有重要的现实和理论意义。

1.2 研究现状综述

本文的研究内容在图像分割与小样本学习两个方向开展，并将小样本学习的理论研究成果应用于图像分割这一问题。因此，在国内外的研究现状中，本文将从图像分割的研究现状和小样本学习的研究现状进行梳理。

1.2.1 图像分割研究现状

卷积神经网络 (convolutional neural network, CNN) 是一类强大的、表示性很强的网络，通常用于处理图像数据。当前在计算机视觉领域中，基于卷积神经网络的神经网络模型占有非常重要的地位。通常卷积神经网络在卷积层后会加上全连接层，将卷积层提取的特征图映射为一维特征向量输出。这适用于图像分类任务，而对于图像分割任务，可以看作像素级的二分类问题，神经网络的输出大小需要与输入图片的大小相同。如果采用全连接层，则会产生大量的参数，拟合如此多的参数需要大量的数据，同时训练的时间和空间开销将会异常庞大。

Jonathan Long 在 2015 年提出了全卷积神经网络 (FCN) [1]，从卷积图提取的特征图中恢复得到每个像素所属的类别，解决了图像语义分割的问题。利用卷积层对图像进行下采样提取特征，用反卷积层上采样，将特征图恢复到输入图像的尺寸，实现了从图像级别的分类到像素级别的分类，是使用深度学习模型解决图像分割问题的里程碑。

在此之后，不断有用于图像分割任务的模型被提出。

1.2.2 小样本学习研究现状

当前，对于小样本学习的研究方向主要分为数据处理方向和模型算法方向。

基于数据增强的小样本学习

数据处理主要是通过数据增强的方式来弥补训练数据不足的问题，数据增强作为一种有效的扩充训练样本容量的方式，不仅可以用于小样本学习的情况下；在训练数据充足的情况下，也可以进行数据增强，进一步提升模型性能。一个模型的性能除了和网络模型结构本身有关，数据增强在很大程度上也有助于提高一个模型的性能。当前，在图像分类这一问题下，提出了很多数据增强策略。

AutoAugment [2]，从数据中学习数据增强策略，是一种自动数据增强策略。AutoAugment 定义了一个搜索空间，包括可以选择的数据增强操作，进行该操作的概率以及图像增强的幅度。通过在搜索空间上进行搜索，得到对应数据增强策略。

RandAugment [3]，每次随机的从候选的数据增强操作中选择一些数据增强操作，对数据进行数据增强。有两个超参数，进行图像增强操作的数量 N 和图像增强幅度 M 。

TrivialAugment [4]，每次随机的选择图像增强操作和图像增强幅度，对图像进行数据增强。

RandomErasing^[5]，从图像中随机的去除一部分区域，保留图像的标签信息。

MixUp^[6]，在图像分类问题中，对每个图像有其对应的标签，mixup 随机的将两个图像进行线性组合，同时对它们的标签信息也进行相应的线性组合。

基于模型的小样本学习

目前，用于解决小样本学习问题的模型主要可以分为度量学习和元学习^[7,8]。度量学习通过评估样本与支撑集中的距离，选择度量距离最接近的类别，借助最近邻的思想完成分类。元学习的方法通过学习到的先验知识实现在少量样本上参数的快速更新。

• 度量学习

在数学概念下，度量（又称为距离函数）是衡量一个定义集合中不同元素之间距离的函数^[9]。度量学习，即相似度学习，是为了评估两个样本之间的距离，从而衡量样本之间的相似程度。^[10] 度量学习的目标是使得相同类别的样本具有比较大的相似度，而不同类别的样本具有比较小的相似度。目前，度量学习的方法主要应用于图像分类、人脸识别、音乐识别等领域。

2015 年，Koch 等人提出了孪生神经网络 (siamese neural network)^[11]。孪生神经网络是一个双路的神经网络，进行单样本图像类别识别，是一种相似度度量模型，用于判断两个样本是否属于同一个类别。训练时，输入是一对样本，通过组合构造不同的样本对进行训练。在预测时，将测试样本与支撑集中进行比较，与支撑集相似度最高的所属类别作为测试样本的类别输出。

2016 年，Vinyals 等提出了匹配网络 (matching network)^[12]。将样本映射到低维向量空间上，计算要分类的样本与带标签样本的相似度，相似度高的作为预测标签输出。

2017 年，Snell 等提出了原型网络 (prototype network)^[13]。认为每个类别存在一个原型，通过将图片映射为向量，对于图片的分类问题变为求与测试样本最接近距离的对应类别原型。

2018 年，Sung 等提出了关系网络 (Relation network)^[14]。关系网络不采用距离函数来计算样本之间的相似度，而是使用神经网络的方式对距离的度量方式进行学习。

• 元学习

元学习，即学会去学习 (learning to learn)^[15]。元学习利用以往的学习知识来帮助新任务的学习，让模型获得学习能力，这种能力可以帮助模型学习到一些元知识。当前深度学习任务针对特定的任务一般需要在数据集上从头开始训练模型，神经网络的参数等都要先进行随机初始化。而元学习可以在当前的模型训练过程之外，通过对大量的先验任务中学习到的元知识，来帮助模型在新任务上更好更快的学习。元知识包括超参数、神经网络的初始参数等。这符合人类的认知模式，在基于记忆带来的

大量先验知识的基础上，对于新的物品只需要少量样本便可以学习到它的特征。

2016 年，Santoro 等人提出基于记忆增强的神经网络（memory-augmented neural networks）^[16]。使用记忆增强的方法来解决单样本学习问题。Santoro 提出的神经网络基于神经图灵机（neural Turing machine）。神经图灵机既可以进行长时记忆，又可以实现短期记忆。通过神经图灵机学习将样本类型存入记忆的策略，并学习如何用这些类型来进行预测。

2017 年，Munkhdalai 等人提出元网络（meta network）^[17]。元网络包括两个部分，base-learner 和 meta-learner。在不同的任务中学习元知识，学习不同任务之间的泛化信息对权重缓慢更新；在新的要处理的任务上实现权重的快速更新。

1.3 面临的困难和挑战

目前的数据增强策略和小样本学习模型主要针对于图像分类问题，对于每个图像有一个相对应的标签，而图像分割需要的是对每个像素的分类，这有别于图像分类问题。对于图像分割这种对于每个像素的分类问题，输出大小需要与输入大小相同，以得到对于图像的分割图。

受限于搜索有效的数据增强策略所需的计算资源，同时本文要用于训练的数据集规模比较小。数据增强作为一种通用的扩充训练数据规模的方法，有别于 Autoaugment 在搜索空间中搜索得到有效的数据增强操作和 Randaugment 进行随机的数据增强，本文通过对数据集分别进行常用的数据增强操作，并通过训练与基准数据集进行比较得到有助于解决图像分割这一问题的数据增强操作。

在目前的相关研究中，对于小样本学习的问题一般更具体的分为三个方面：小样本学习（few-shot learning）、单样本学习（one-shot learning）和零样本学习（zero-shot learning）。本文关注于小样本学习的研究，每个类别的可以有多个样本，这有别于单样本学习和零样本学习。对小样本的研究主要集中在图像分类领域，孪生神经网络和匹配网络等都是解决单样本学习的图像分类问题。而元学习需要在大量的先验任务上学习得到先验知识，在将学习到的元知识应用于当前的任务。

本文要做的是只基于当前的数据集，没有通过别的数据集进行模型预训练和通过别的任务学习得到先验知识情形下的小样本学习问题。基于小样本面临的训练数据不足的问题，对数据进行数据增强操作，分析比较有助于提升图像分割的数据增强操作；本文借鉴元学习从先验知识快速更新权重的方法，通过对模型的改进，减少模型的参数量，使得在反向求导的过程中也可以实现对参数的快速更新，减轻因为深度网络层次加深导致梯度消失而学习过慢的问题。

1.4 论文结构和主要内容

本论文选择图像分割作为小样本这一领域的研究方向，着重于语义分割领域的小样本学习：

本文的内容分为四个章节，章节组织结构及其主要内容如下。

第一章，主要介绍了课题研究的背景及研究意义。介绍深度学习的发展，图像分割以及小样本学习的研究背景。进而明晰研究在数据量不足，即小样本的情况，进行图像分割研究的意义。

第二章，主要介绍了本文所采用的 U-net 卷积神经网络模型所涉及的基础理论知识，包括卷积、池化操作、转置卷积、损失函数。

第三章，主要介绍了在数据处理方向对小样本学习的研究，所采用的数据集，常用的数据增强方法及其在图像分割处理中的效果。对 U-net 网络模型的改进，使其更适用于小样本问题。采用改进的网络模型进行的实验过程和结果分析比较。

第四章，总结和展望，完成对基于深度学习的小样本图像分割这一课题的研究。提出当前所进行的研究工作中的一些不足之处，及可能的改进方向。

第二章 相关原理和技术

目前图像分割的方法主要分为：阈值分割方法、区域分割方法和基于特定理论的分割方法。当前，基于卷积神经网络的方法在图像分割领域取得了极大的成功。卷积神经网络是一类强大的，为处理图像数据而设计的网络模型。对比全连接神经网络，卷积神经网络具有结构信息，更深层的卷积操作可以提取更高维的特征信息；同时卷积神经网络因为共用一个卷积核对图像进行处理，从图像中提取出特征图。参数量相比全连接神经网络大大减少。在卷积层中，需要训练的参数为卷积核权重和偏置项。

对于图像分割这一领域更具体的分为：

- 语义分割（semantic segmentation）：对一个图像中的所有像素点进行分类。
- 实例分割（instance segmentation）：对图像中的目标实例进行区域划分，需要区分同一物体的不同实例。

本文要研究的是图像语义分割，下面介绍一下在使用卷积神经进行语义分割涉及到的相关原理和技术。

2.1 下采样

下采样，即对图像进行缩小。在卷积神经网络中，下采样的方式主要有卷积和池化。在卷积神经网络中进行下采样的主要目的降低输入大小，提取输入特征；增大感受野，更深层的卷积核可以学习到范围更大的特征信息。

2.1.1 卷积

在数学中，两个函数的卷积定义为：

$$f(x) * g(x) = \int f(t)g(x-t)dt \quad (2.1)$$

其中， t 为积分变量。在卷积操作中，将函数 $g(x)$ 进行翻转并平移一定量时，对函数 $f(x)$ 进行加权，测量两个函数重叠部分的相应积分；当函数为离散对象时，积分变成求和运算。

卷积神经网络是一类强大的网络，在图像处理领域表现出了非常好的效果。卷积神经网络中的卷积操作，通过一组卷积核对图像进行加权处理，从原始图像中提取得到特征图。卷积层通过将卷积核作用于输入，并添加偏置后产生相应的输出结果。将卷积核作用于输入上，从左上角开始，从左到右，从上到下进行滑动。对应于卷积核，在输入上有一个对应的窗口，将卷积核与对应窗口的元素进行按元素相乘后求和得到对应的标量值。通常卷积核大小为 3×3 ，步幅为 1，假设输入大小 $w \times h$ ，则输出大小为 $(w-1) \times (h-1)$ 。可以

通过输入四周进行填充 (padding) 使得输出尺寸与输入尺寸相同。如图 2.1 所示, 在不添加填充的情况下, 将 3×3 的卷积核作用于 5×5 的矩阵, 在步幅为 1 的情况下, 得到输出为 4×4 的矩阵。

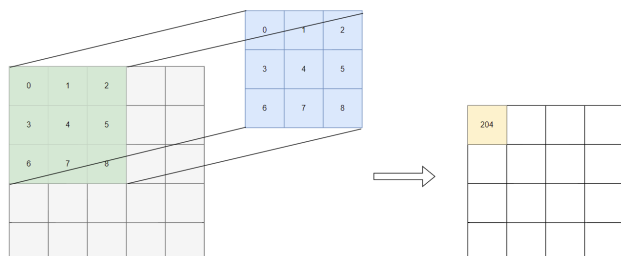


图 2.1 无填充、步幅为 1 的卷积操作

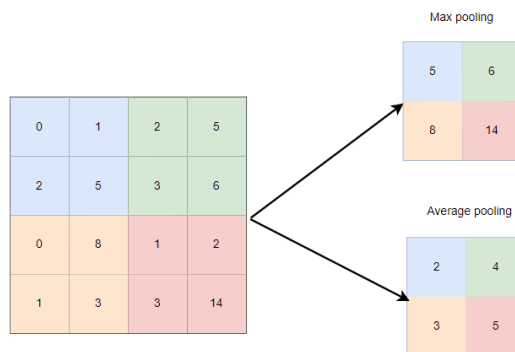


图 2.2 最大池化和平均池化操作

2.1.2 池化

池化操作对输入进行下采样, 降低输入的尺寸大小, 进行特征的过滤和提取。池化层可以降低卷积层对于位置的敏感性, 常用的池化操作有最大池化和平均池化。最大池化操作计算窗口区域的最大值, 平均池化操作计算窗口区域内的平均值。最大池化和平均池化都降低图片分辨率, 但最大池化保留池化窗口中的最大值, 即图片的最突出特征, 通常效果更好一些。

在图像处理中, 池化层一般作用于卷积层之后, 通过对图片输入进行填充, 使得卷积层不改变输入大小。然后通过池化层进行下采样操作, 提取数据特征。池化层不需要学习, 通常池化层采用 2×2 大小, 步幅为 2, 对图像进行 2 倍下采样操作, 使得输入的高度和宽度减半。如图 2.2 所示, 分别为最大池化和平均池化操作。

2.2 上采样

上采样, 即对图像进行放大, 提升图像的分辨率。上采样的方法主要有双线性插值 (bilinear)、转置卷积 (transposed convolution) 和 UnSampling 等。在对图片进行下采样提取特征的过程中, 会使图片变小。而对于图像分割问题, 我们希望输出图像与输入图像大小相同, 以得到对于输入图像的分割图。这一目标需要通过上采样操作来实现, 上采样可以放大图像, 使图像逐步恢复到原图大小。

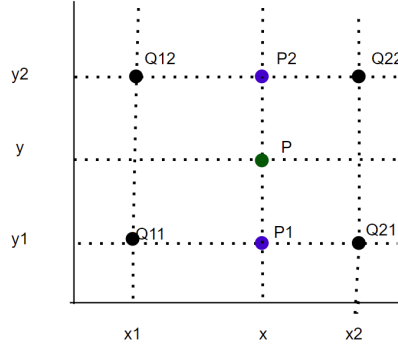


图 2.3 双线性插值

2.2.1 双线性插值

双线性插值作用于具有两个变量的函数，分别在两个方向上各进行一次线性差值。假设已知 $f(Q_{11}(x_1, y_1), Q_{21}(x_2, y_1), Q_{12}(x_1, y_2), Q_{22}(x_2, y_2)$ 处的值，要求得 $f(P(x, y))$ 处的值，如图 2.3。首先在 x 方向上进行一次线性差值：

$$f(P_1) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{11}) + \frac{x - x_1}{x_2 - x_1} f(Q_{21}) \quad f(P_2) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{12}) + \frac{x - x_1}{x_2 - x_1} f(Q_{22}) \quad (2.2)$$

在 x 方向的得到两个点 $P_1(x, y_1), P_2(x, y_2)$ 处的值，对这两个点在 y 方向上在进行一次差值，得到点 $P(x, y)$ 处的值。

$$f(P) = \frac{y_2 - y}{y_2 - y_1} f(P_2) + \frac{y - y_1}{y_2 - y_1} f(P_1) \quad (2.3)$$

在计算机视觉和图像处理中，双线性差值利用周围四个特征点的值来求得一个新的像素点信息，在进行图片放大时可以接近原图片的视觉效果，减少图片失真。

2.2.2 转置卷积

卷积和池化操作通常会减少输入大小，而对于图像分割，需要输出大小与输入大小相同，可以看作像素级的分类问题。转置卷积进行上采样增大输入大小，把提取的高维特征进行解码。通过卷积核对来自输入的元素进行广播，逆转下采样导致的空间尺寸变小。图 2.4 为将 2×2 转置卷积核作用于 2×2 的输入矩阵，步幅为 2 时的输出：

2.2.3 UnSampling

UnSampling 通过对相应位置的值填充复制到一定区域内，实现对图片的放大，类似于逆平均池化层的操作，图 2.5 描述了 UnSampling 操作。

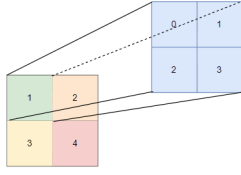
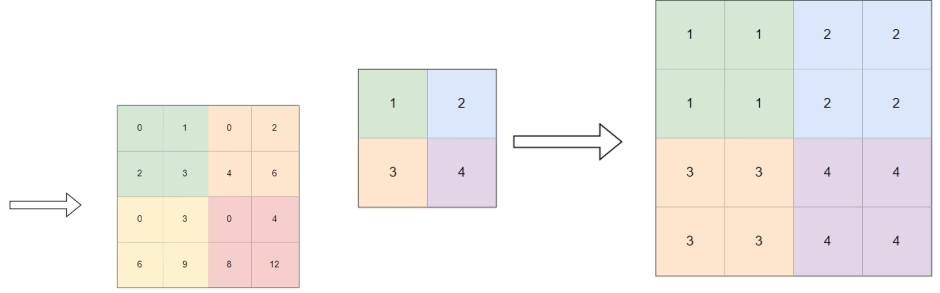
图 2.4 2×2 , 步幅为 2 的转置卷积操作

图 2.5 UnSampling 操作

2.3 损失函数

在深度学习模型中，我们需要有一个函数来评估模型的优劣程度，这被称为损失函数。损失函数是评估模型网络的性能指标，我们的目标是 최소화损失函数来学习模型参数的最优解。在深度学习中，常用的损失函数有平方损失函数（MSE），交叉熵损失函数（CrossEntropy）损失函数，合页损失函数（Hinge Loss）等。在图像分割领域中，常用 CrossEntropy loss 与 Dice loss 来设计损失函数。

2.3.1 MSE

平方损失函数（Mean Squared Error, MSE）计算预测值 \hat{y}_i 与真实值 y_i 的接近程度，常用于回归模型。其定义如下：

$$MSE = \frac{1}{2n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.4)$$

2.3.2 Dice Loss

在介绍 Dice loss 损失函数前，先介绍一下 Dice 相似系数。

- Dice 相似系数

Dice 相似系数，是一种集合相似度度量指标，用于计算两个样本的相似度，Dice 相似系数的取值范围为 [0,1]。Dice 相似系数的取值越高，代表两个样本的相似度越高。X 与 Y 的 Dice 相似系数定义如下：

$$Dice(X, Y) = \frac{2|X \cap Y|}{|X| + |Y|} \quad (2.5)$$

- Dice Loss

Dice Loss 定义如下，其值越大，代表 X 与 Y 的差异程度越高。

$$DiceLoss(X, Y) = 1 - \frac{2|X \cap Y|}{|X| + |Y|} \quad (2.6)$$

2.3.3 CrossEntropy

CrossEntropy 用于量化两个概率分布之间的差异，假设真实概率分布为 P ，预测概率分布为 Q ，损失函数定义如下：

$$H(P, Q) = - \sum_x P(x) \log Q(x) \quad (2.7)$$

$P(x)$ 是真实概率分布，在分类问题中通常用长度为 x 的独热编码向量来表示 x 个类别，对应类别的值为 1，其余为 0，而图像分割问题可以看作每个像素的分类问题。 $-\log Q(X)$ 取得在相应的真实类别预测概率，对预测概率越大时，这一项取值越小，表明预测的结果越准确。

第三章 基于改进的 U-net 网络模型进行图像分割

U-net 由 Olaf Ronneberger 等^[18]在 2015 年提出,是一种全卷积神经网络,基于 FCN 改进得到。其网络模型结构包括一条用于提取图像特征图的收缩路径,和一条恢复图片尺寸的扩张路径。其结构看起来像是字母 U,因此称为 U-net。U-net 最初用于医学领域的图像分割,并且取得了当时非常好的效果。在医学图像分割任务下,可获得的标注样本数量通常面临着数量不足的问题,U-net 在比较少的训练集下对医学图像分割取得了非常好的分割效果。与别的网络模型相比较,U-net 所需训练样本少,对图像的分割效果好。U-net 网络模型结构如图 3.1^[18]所示:

它由一个左侧的收缩路径和一个右侧的扩张路径组成。左侧的收缩路径是典型的卷积神经网络结构,在每一层包括两次卷积操作,每次卷积操作后用 ReLU 作为激活函数。采用步幅为 2 的最大池化层将图片尺寸减半。在每次下采样操作中将特征通道数加倍。共进行 4 次下采样操作,对输入图像下采样 16 倍。

右侧扩张路径在每一步中对特征图进行上采样,采用 2×2 的转置卷积核逐步加倍恢复原始图片尺寸,同时特征通道数减半。并与来自对应收缩路径层裁剪的特征图进行拼接,构成特征通道数加倍的特征图。然后分别进行两次卷积操作,每次卷积操作采用 ReLU 作为激活函数。对应于 4 次下采样操作,进行 4 次上采样操作,恢复到原图片大小。

当前,针对 U-net 这一经典的图像分割模型已经提出了很多改进策略。在卷积神经网络中,浅层的卷积层的感受野比较小,对于图片细节的学习感知能力比较强。深层的卷积层,因为下采样操作降低了图片的分辨率,所以感受野相对增大,可以学习到图片的整体特征。U-net 使用了长连接,将来自浅层卷积层提取的特征图与深层卷积层通过上采样恢复的特征图在通道上进行跳跃拼接 (skip connection)。Zongwei Zhou 等认为直接将来自于下采样

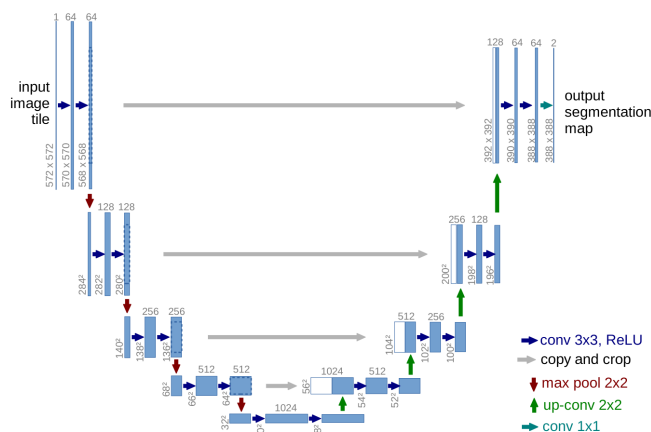


图 3.1 U-net

的浅层特征与上采样的深层特征进行连接会带来 semantic gap, 在此基础上提出了 Unet++ 网络^[19]。U-net++ 网络在 U-net 网络的基础上添加了更多的长连接操作。[19]

在何凯明等提出残差网络 (ResNet)^[20] 后, Zhang Zhengxin 等结合 ResNet 和 U-net 提出了 ResUnet^[20]。ResNet 很好的解决了深度神经网络的训练问题, 本文参照 ResUnet 对于 U-net 网络的改进思路, 将 U-net 网络每一层的两次卷积操作添加残差块。同时由于深度神经网络强大的学习能力, 在进行小样本训练时, 很容易出现过拟合的问题, 对其中的卷积操作进行改造, 减少模型需要训练的参数量。

3.1 数据集和数据处理

数据集来源于 Kaggle 的车辆边界识别与车辆实例分割, 这一数据集包括不同种类的车辆在不同背景下的照片, 同一车辆在不同角度的图片以及车辆对应的分割图。选择这个数据集的原因首先是因为它的图片构成要素比较简单, 没有复杂的背景信息, 便于分析和比较; 其次, 每个图片只有一个要分割的实例。同时, 不同车辆又有比较大的差异化特征。从中选择 10 种汽车, 每种汽车 5 张图片, 共 50 张图片作为训练集; 同时选择 200 张图片作为验证集, 验证对于图片的分割效果。因为对于不同的任务, 不同的数据集, 其有效的数据增强策略一般是不同的。同时受限于时间和本人水平的因素, 只采用了这一个数据集来进行小样本图像分割的研究。

小样本学习要面对主要问题便是可用于进行训练的数据不足, 在训练神经网络模型时, 我们的目标是使模型损失不断降低, 尽可能得到最优解。而深度神经网络有着大量的参数, 在缺乏足够的支撑数据时, 很难训练网络模型求解得到比较优的参数。同时, 论文^[21]指出, 神经网络在最初的时候并没有那么聪明, 对小的图片转换操作具有比较大的敏感性。通过对现有的数据增强, 对图片应用翻转, 移位, 旋转, 水平变换等操作, 我们可以扩充数据规模, 同时也有助于模型学习在进行转换前后图片中的相关性与不变形。关于数据增强时机的选择: 一种被称为线下增强 (offline augmentation), 即首先对所有数据进行数据增强操作后在进行模型的训练。另一种是线上增强 (online augmentation)^[22], 即在训练的过程中不断的在小批量数据上进行转换。因为本论文研究的问题为小样本图像分割, 面对的问题是数据量不足的情况, 因此采用线下增强的方式, 对数据集完成所有的数据增强转换操作后, 在进行模型的训练。在常见的对图片的数据增强技术中, 分析考虑可能对当前的图像分割任务可能有益的数据增强操作, 并通过实验分析对比和验证他们的结果。

下表给出了本论文中采用的数据增强操作: 通过对原始数据集进行不同的数据增强操作, 并与不进行数据增强进行对比, 分析可以用于当前任务的数据增强操作。下表给出所采用的增强操作以及在验证集上通过 Dice 相似系数评估的结果: 不进行数据增强时, 分别训练了 5 次和 10 次, 进行数据增强时都训练了 5 次。进行一种数据增强操作时, 会将现有数据集规模扩充一倍。因此将不进行数据增强, 进行 10 次训练的作为对比, 与进行数据增

表 3.1 对图片进行的数据增强操作

数据增强操作	说明
invert	反转图像颜色
hflip	对图像水平翻转
rotate	对图像旋转一定角度
affineScale	对图像缩放
translateX	水平移位
GaussianBlur	使用高斯模糊核对图像进行高斯模糊
ColorJitter_hue0.5	随机改变图片色调, 变化幅度为 (-0.5, 0.5)
ColorJitter_contrast0.5	随机改变图片的对比度, 变化幅度为 (0,0.5)

表 3.2 数据增强操作与训练结果

数据增强操作	训练次数 (epoch)	Dice 相似系数
不进行数据增强操作	5	0.8902
不进行数据增强操作	10	0.9019
invert	5	0.5905
hflip	5	0.9311
rotate	5	0.9288
affineScale	5	0.9301
translateX	5	0.8976
GaussianBlur	5	0.9036
ColorJitter_hue0.5	5	0.9231
ColorJitter_contrast0.5	5	0.9047

强的训练结果进行比较。

通过对上述结果的分析比较,对图像进行颜色反转 (invert) 并没有带来结果的提升,反而给验证精度带来了波动状态。图 3.4 为进行 invert 数据增强操作后,进行 5 次迭代训练,每次迭代在验证集上评估 10 rounds 时得到 Dice 相似系数情况,3.1 节和 3.1 节分别给出原始图像与反转图像的对比。



图 3.2 原始图像



图 3.3 invert 操作图像

虽然从上图对比来看, invert 操作并没有给图片形状带来太大的改变。但分析认为这与图片的计算机存储方式有关,彩色图片采用 RGB 三通道进行存储,对于每一个像素点,在 RGB 通道中分别有一个 0-255 的数值表示,假设某一像素点的数值为 $[x, y, z]$, invert 操作为 $[255-x, 255-y, 255-z]$,从数值层面分析,对图像进行反转增强操作给原始图像带来的改变比较剧烈,因此对图像特征的影响比较大。同时高斯模糊和对图像进行水平移位的增强操作也没有对图像分割的结果有所提升。基于以上分析,采用剩下的 5 种数据增强操作作为当前图像分割数据集的数据增强策略。其中,对图片进行水平翻转和旋转的数据增强操作都提高了 Dice 相似系数。因为数据集中对一种汽车包括其在不同角度的 16 张图片,而训练数据集则只有 5 张图像,即五个角度的图片。对于图片的水平翻转和旋转操作补足了同一汽车在不同角度的图像,因此对于结果起到了提升作用。

图 3.5 是对数据集进行颜色增强操作与使用原始数据集时对图像的分割情况,其中第一行左边的为原始图像,后面的图像为改变其颜色的图像,第二行为用颜色相关数据增强后得到的模型对上面图像的分割情况,第三行为不进行数据增强训练得到的模型对第一行图像的分割情况。通过对比,可以发现进行数据增强后,对四幅图像得到了差不多的分割效果;而不进行数据增强,对于四副图像的分割差异性比较大。因此,对于颜色的变换降低了网络模型对于颜色的敏感性。同时,颜色是汽车一个重要差异性指标,对于颜色的数据增强操作在当前数据集可以表现出比较好的效果。

3.2 损失函数设计

在图像分割领域中,常用 CrossEntropy loss 与 Dice loss 来设计损失函数。

而在图像分割领域,经常面临的问题时是对分割的细节不准确,对边缘的分割效果不

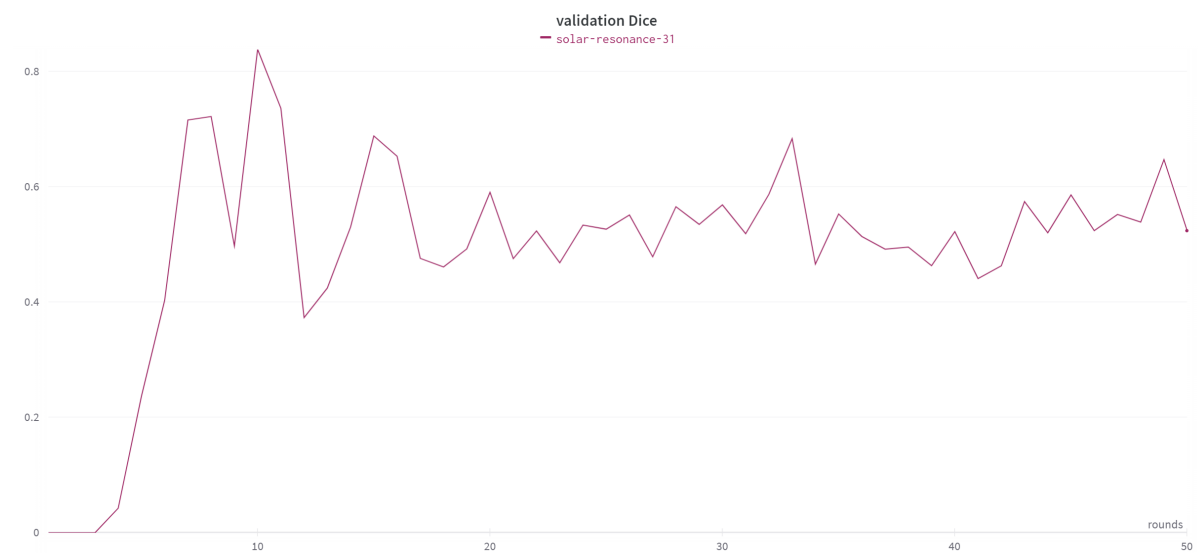


图 3.4 进行 invert 数据增强在验证集上的 Dice 相似系数

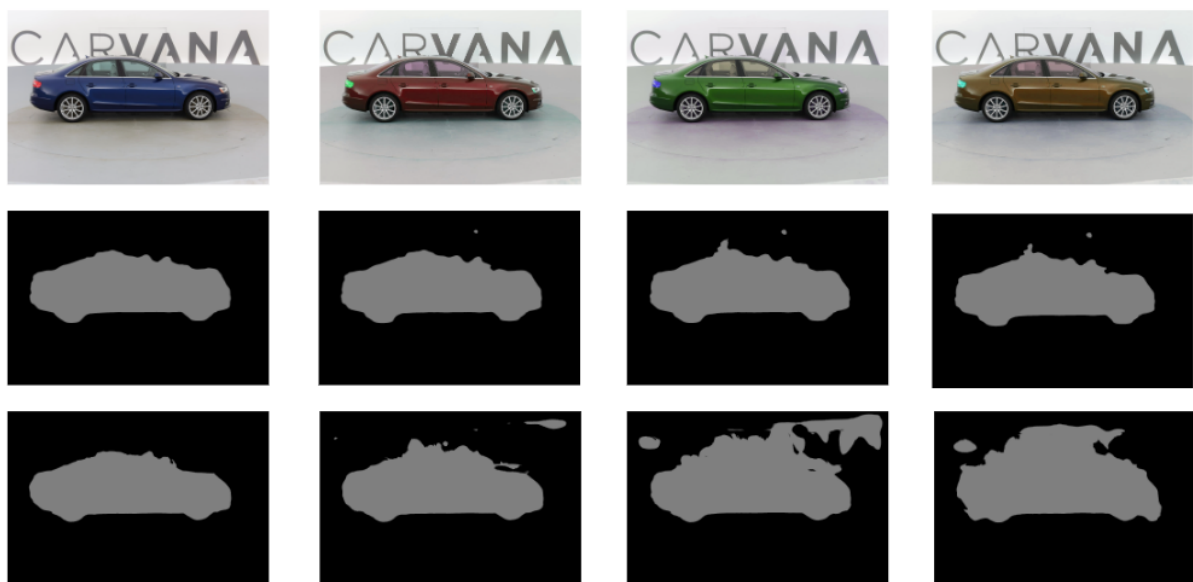


图 3.5 对数据集进行颜色增强操作与进行颜色操作时对图像的分割情况

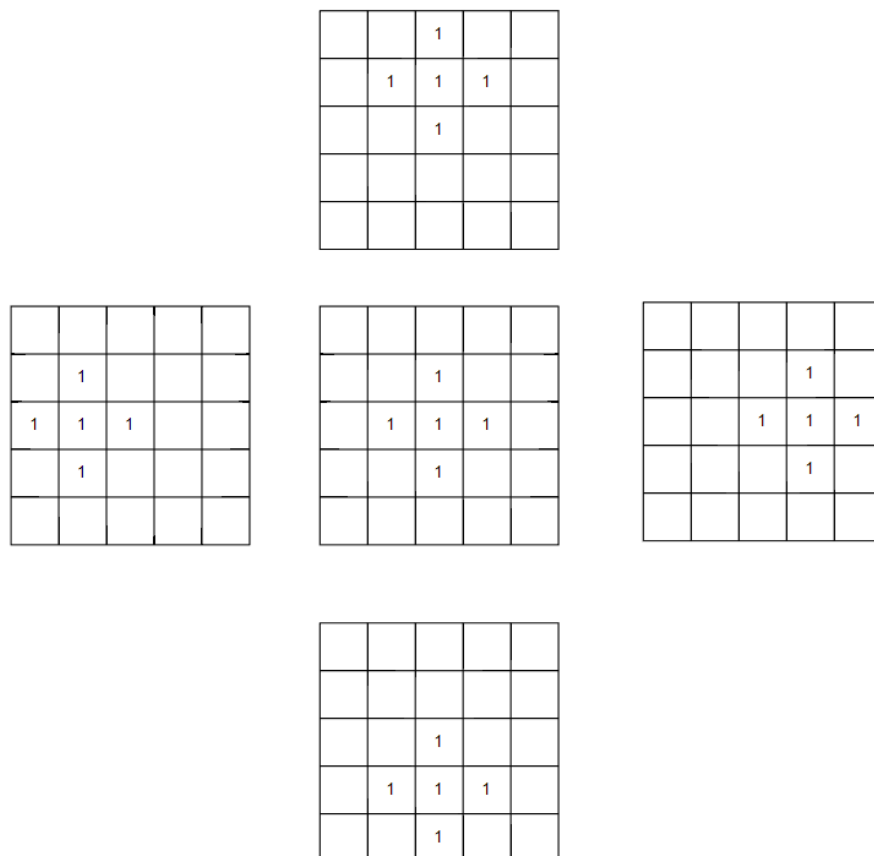


图 3.6 分割图 (mask) 与向四个方向平移后的图像

够理想，而这一问题在只有小样本用于训练时，会表现的更为明显。因此考虑在损失函数中添加一项对边缘分割效果的评估，并对边缘添加一定的权重，让神经网络对边缘的分割更加敏感。

要评估边缘分割效果，首先需要获取图像的边缘，而在数据集中，通常有的都是对实例的分割图，而没有单独的图像边缘图。基于当前采用的数据集特点，对每一个像素都是二分类问题，即这个像素要么属于背景，要么属于要分割的实例图。结合学过的位运算知识，提出一种从分割图中提取出图像边缘的方法：

如图 3.6 所示，中间的图为原始分割图，1 表示是实例的一部分，0 表示背景。将原始图像分别向上下左右平移 1 个像素位置，空缺的位置填补 0，得到对应四个方向的图像。

图 3.7 描述了从分割图 (mask) 中提取出边缘的流程，流程图中 `edge_outline` 对应图 3.8 中最右侧的图，将蓝色背景的称为边缘 (edge)，绿色背景的称为轮廓 (outline)。`edge_outline` 既包括实例的最外面的属于实例的一个像素，也包括这一像素外的一个像素。进一步，流程图描述了如何从这个图进一步提取出边缘和轮廓的过程。提取出这两个部分后，我们可以对其进行分别加权来设计损失函数，同时这两个部分也是对于分割图边缘分割效果影

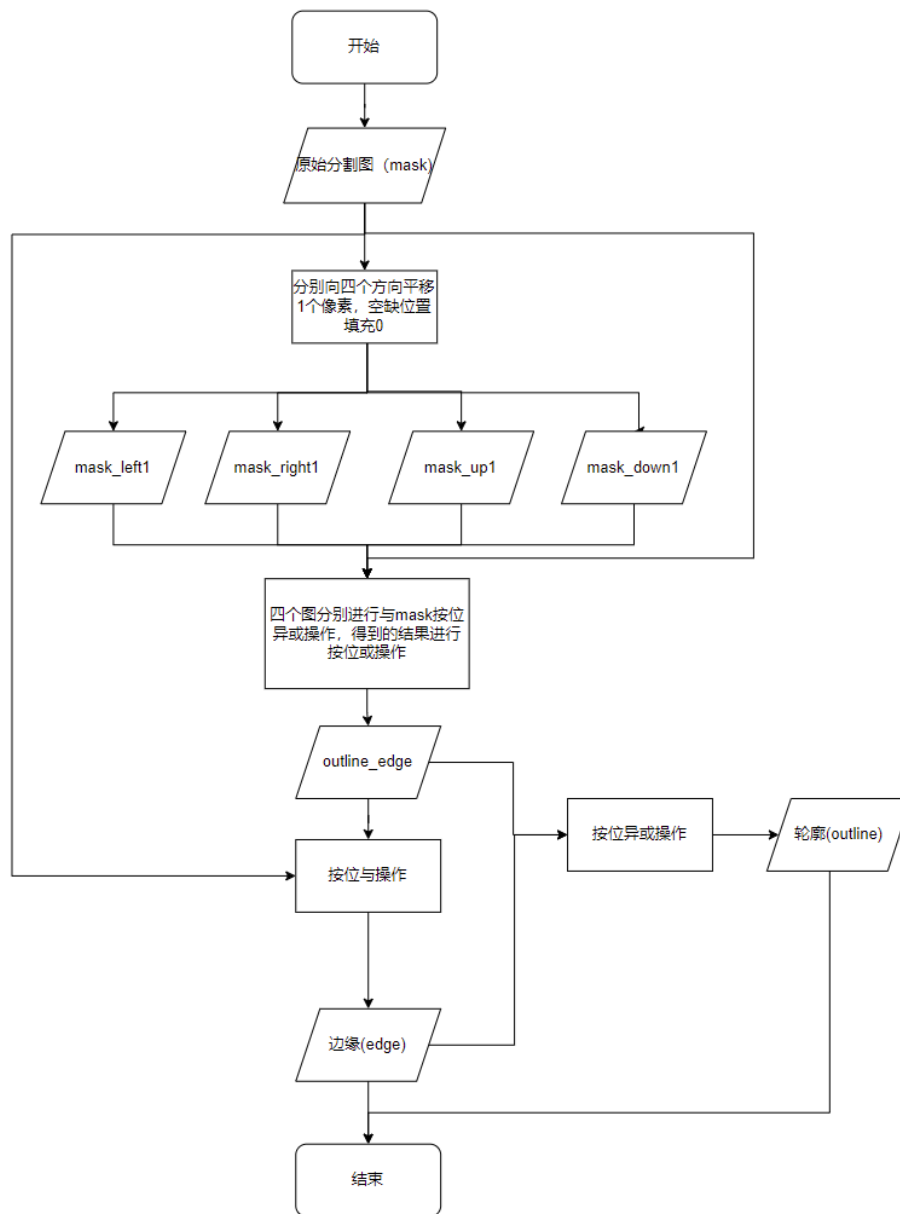


图 3.7 从分割图 (mask) 提取出边缘和轮廓的流程

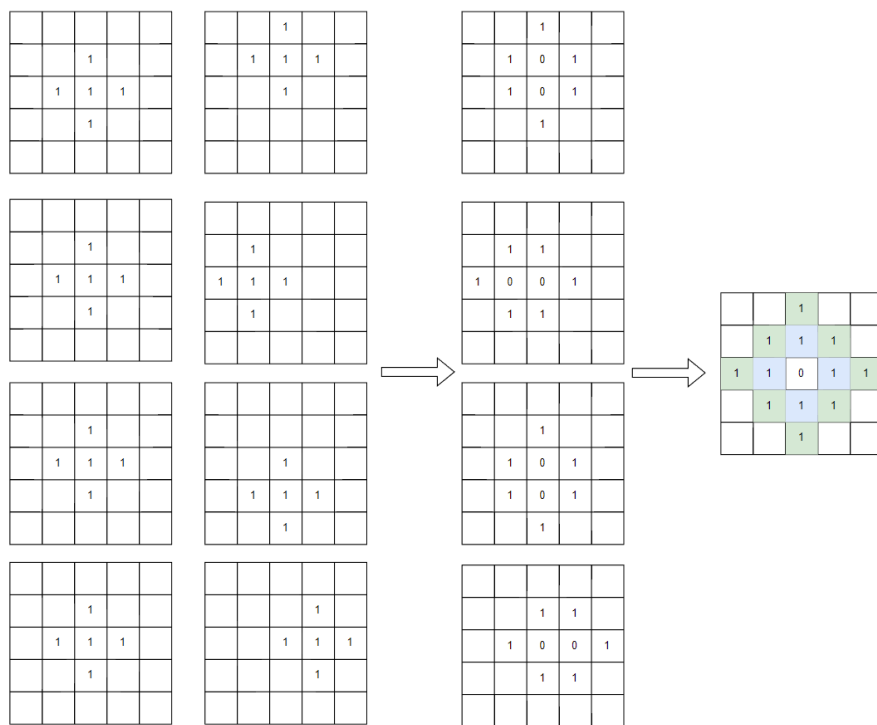


图 3.8 从 mask 图得到 edge_outline 图

响最大的部分。然后对于其的损失函数计算方式采用如下的计算方式:

$$EdgeLoss(X,Y) = \frac{|X \cap Y|}{|Y|} \quad (3.1)$$

X 表示预测的概率, Y 表示提取出的边缘图, 这里采用与 Dice Loss 有些差异的计算方式, 因为这一损失项只关注对边缘的分割情况。 $|X \cap Y|$ 表示两个集合对应元素点乘, 将结果相加求和。

如下图所示, 为通过上述方式提取出的边缘图:



图 3.9 mask 图和 edge 图

采用的模型的损失函数如下:

$$loss = CrossEntropy + DiceLoss + EdgeLoss \quad (3.2)$$

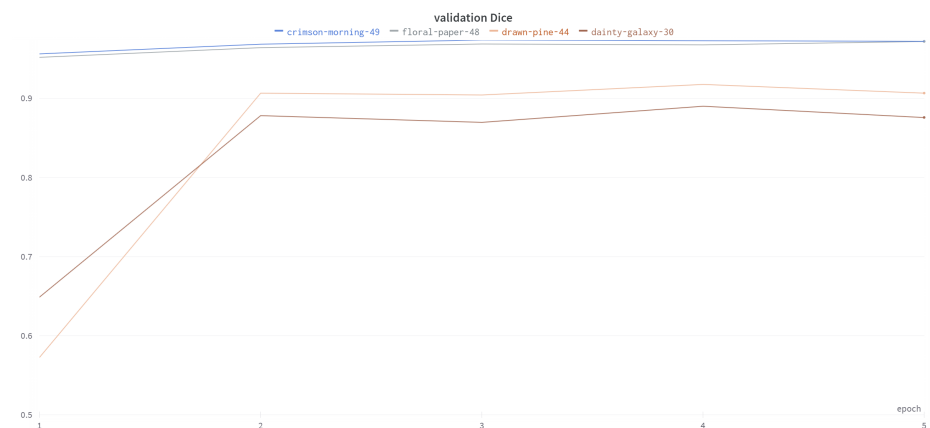


图 3.10 在损失函数中添加 EdgeLoss 损失项与不添加这一项时的结果

通过对结果的比较, 添加了这一项后深度学习网络模型的性能得以提高。

采用 Dice 相似系数作为模型的评估指标, 图 3.10所示为在损失函数中添加 EdgeLoss 项与不添加 EdgeLoss 项的对比情况。其中 drawn-pine-44、dainty-galaxy-30 为不进行数据增强操作, crimson-morning-49、floral-paper-48 进行上节提到的 5 种数据增强操作。dainty-galaxy-30 和 floral-paper-48 都未添加 EdgeLoss 损失函数项。通过对实验结果的比较可以发现, 当在损失函数中添加 EdgeLoss 这一项后, 模型的性能得到提高。

更近一步, 可以比较添加这一项与不添加这一项时对图像的边缘分割情况, 图 3.11显示了他们之间的对比情况, 添加了 EdgeLoss 损失项后对边缘的分割效果更好。通过在损失函数中添加了 EdgeLoss 项, 使得模型对边缘的分割效果更加敏感。当模型对边缘的分割效果不好时, EdgeLoss 项使得损失函数增大, 从而使得模型更多的关注于边缘的分割。

3.3 U-net 模型改进

对 U-net 网络模型结构的改进如图 3.12所示, 整体网络模型与 U-net 一样, 在此基础上进行了一些改进。

图图 3.13和图 3.14给出了在网络模型在第一层从输入、到第一次下采样池化操作的具体模型比较。在卷积神经网络中, 卷积核是主要要通过训练学习得到的参数。卷积核的大小很大程度上影响着卷积神经网络的参数数量, 因此卷积层一般使用 3×3 大小的卷积核。

当网络模型有很多参数时, 需要给模型足够数量的样本, 在训练的过程中才能不断的拟合这些参数。而小样本要面对样本数量不足的问题, 因此要尽量减少模型的参数, 同时不影响对图像的分割效果。对于用于卷积层的 3×3 卷积核, 其形状是一个四维张量 $[hw]$ 。对应于每一个输出通道, 需要输入通道个数的卷积核分别作用于对应的输入通道进行不同通道的融合, 并将结果叠加得到一个输出通道的结果。U-net 网络模型包括四次下采样, 和四次上采样, 每一步进行两次 3×3 卷积核操作。因此 3×3 的卷积核所包括的参数个数是当前模型主要需要训练的参数。基于此, 对 3×3 卷积和进行调整。使得对于每一个输出通

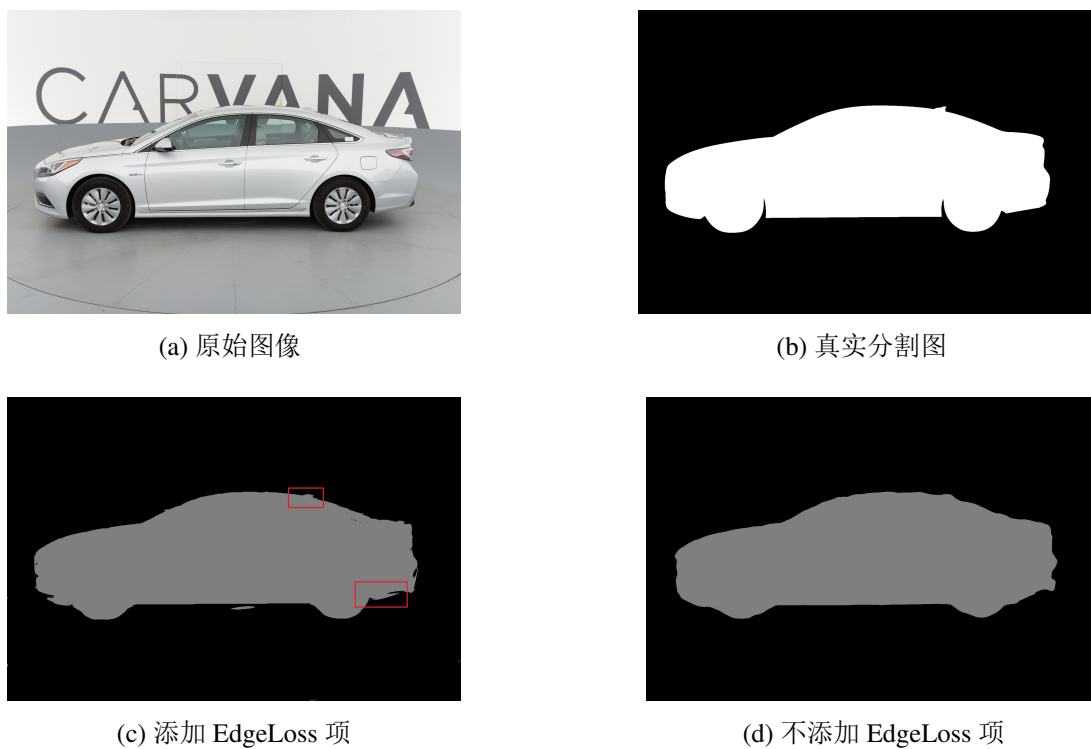


图 3.11 分割图对比

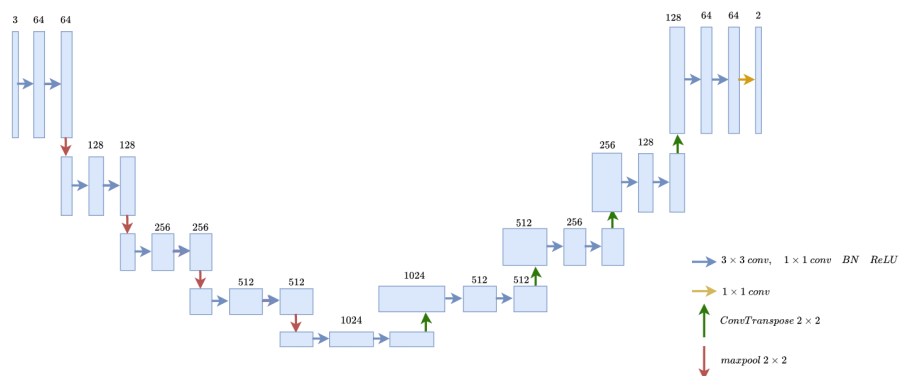


图 3.12 改进的 U-net 模型结构，矩形上方是对应的通道数

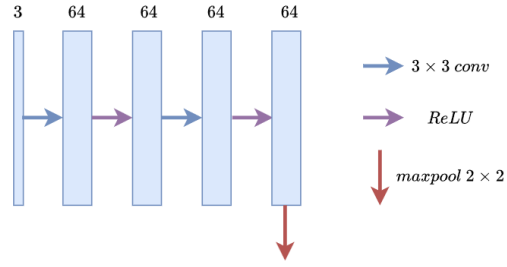


图 3.13 U-net 模型第一层结构

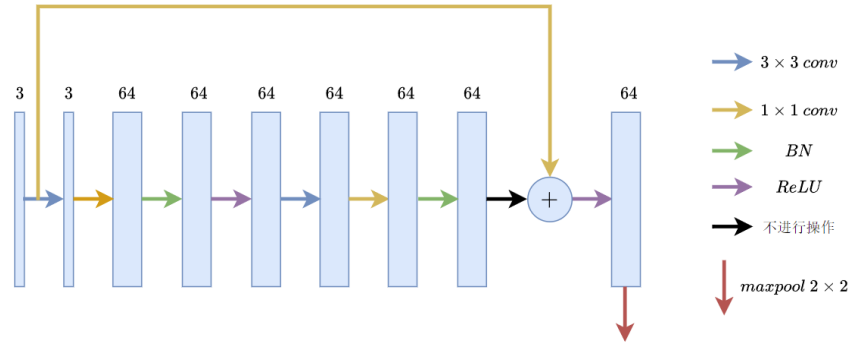


图 3.14 改进的 U-net 模型第一层结构

道，只用一个卷积核作用于输入通道进行通道特征提取，此时这一操作保证输入通道与输出通道个数相同。对于通道的调整，不同通道特征的融合则通过 1×1 的卷积核来实现。下面计算一下他们所需要的参数个数，假设输入通道个数为 C_{in} ，输出通道个数为 C_{out} ：

$$3 \times 3 \text{卷积核进行特征提取和通道融合: } C_{out} \times C_{in} \times 3 \times 3 \quad (3.3)$$

$$3 \times 3 \text{卷积核特征提取} + 1 \times 1 \text{通道融合: } C_{in} \times 1 \times 3 \times 3 + C_{out} \times C_{in} \times 1 \times 1$$

以输入 3 通道的 RGB 图像通过卷积变为 64 通道为例，计算它们分别的参数个数，同时网络使用了残差块结构作为卷积的模块，需要将 3 通道输入用 1×1 卷积变为 64 通道输出，使得其可以与卷积作用的输出进行相加：

$$3 \times 3 \text{卷积核进行特征提取和通道融合: } 64 \times 3 \times 3 \times 3 = 1728$$

$$3 \times 3 \text{卷积核特征提取} + 1 \times 1 \text{通道融合: } 3 \times 1 \times 3 \times 3 + 64 \times 3 \times 1 \times 1 = 219$$

$$3 \times 3 \text{卷积核进行特征提取和通道融合: } 64 \times 3 \times 1 \times 1 = 192$$

$$\frac{1728 - (219 + 192)}{219} \approx 6.01$$

通过上面的计算结果，可以发现将 3×3 卷积核操作变为 $3 \times 3 + 1 \times 1$ 可以很大程度上减少网络模型的参数个数。较少的参数个数更有助于在小样本数据集上进行学习。

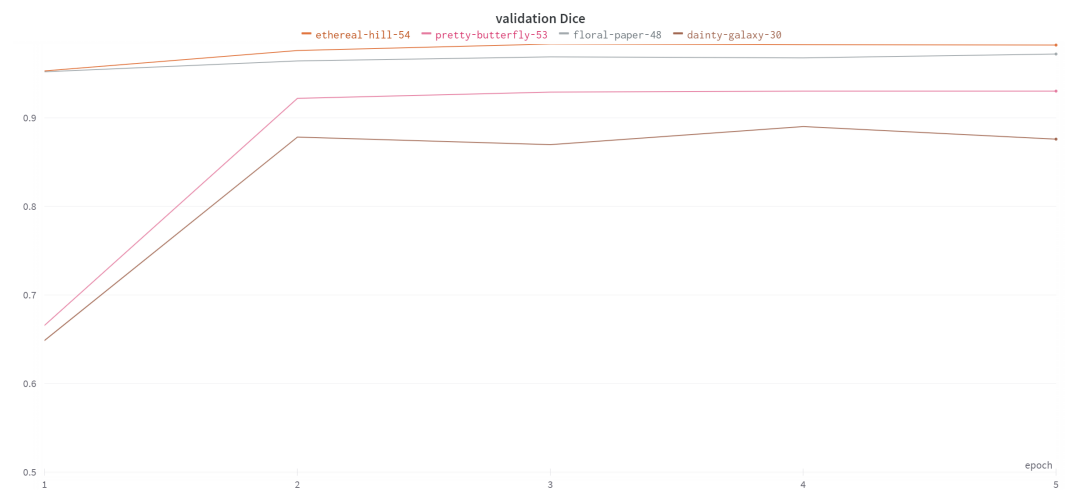


图 3.15 U-net 和改进 U-net 的表现效果

Sergey Ioffe^[2]等提出了 Batch Normalization (BN)，对每一层的输入进行归一化。BN 是一种很有效的技术，通过对输入进行正则化，可以减轻每一层的变化剧烈程度，加速深度神经网络的收敛速度。在训练深度神经网络时，BN 是一种流行有效的技术，可以持续加速深度神经网络的收敛。现在的卷积神经网络模型一般都会在卷积操作之后加入批量规范化层 (BN)。在改进的 U-net 模型中，在 1×1 卷积操作之后，ReLU 激活函数前添加批量规范化层 (batch normalization)。

Resunet^[23] 将残差连接加在批量规范化层 (BN) 之前，同时两次卷积操作与 U-net 网络相同。对 U-net 模型的改进参考这一思路将残差连接加在 ReLU 激活函数之前。

改进的 U-net 模型和 U-net 在训练集上的表现情况如??所示，floral-paper-48 和 ethereal-hill-54 进行数据增强操作。ethereal-hill-54 和 pretty-butterfly-53 是在改进 U-net 模型上训练的结果，floral-paper-48 和 dainty-galaxy-30 是在 U-net 模型上训练的结果。通过对结果的比较，改进的 U-net 模型在测试集得到了较高的 Dice 相似系数，取得了比较好的效果。

第四章 总结和展望

深度学习作为人工智能领域的一项突破，近些年来在很多领域不断展现出其巨大的潜力。而要训练得到有效的深度学习网络模型，通常都需要大量的数据作为基础。论文首先介绍了进行小样本学习研究的背景和意义，其次介绍了常用于图像分割的网络模型结构 U-net，针对小样本样本数据不足的问题，进行一些解决方案的研究，对训练数据集进行数据增强，以及对网络模型进行改造以适应小样本学习问题。通过训练结果的分析比较，验证了上述方法的有效性。

4.1 总结

本论文的主要研究工作总结如下：

1. 基于小样本样本不足的问题，对训练数据集进行数据增强。采用多种数据增强方式，分析比较数据增强后训练得到的网络模型在验证集上的表现。
2. 设计损失函数，针对图像分割对于细节边缘的分割效果不理想的问题，提高模型对于边缘分割的敏感性。
3. 对 U-net 网络模型进行改进，针对深度学习网络模型学习能力强，在数据量不足的情况下可能出现的过拟合问题，减少 U-net 网络模型的参数。

4.2 展望

由于进行毕业论文的时间和实验条件的限制，所进行的工作还有很多不足之处。本文所在的工作是在当前常用于解决小样本问题的思路，给出适用于所进行的研究方向的解决方案，进行实验和验证，对解决方案的有效性进行分析和论证，还有很多地方需要进一步完善：

1. 数据增强是常用于解决数据量不足的方法和手段，但不同的任务和目标下，所需要进行的数据增强操作会有比较大的差异性。如何针对目标任务找到有效的数据增强策略，更有效的利用训练数据，同时平衡与效率。
2. 本文从分割图中提取边缘的操作具有很大的局限性，同时对计算需求比较大。利用了位运算的特性，这只适用于分割图的数值表示为 0,1 的情况下，即对每一个像素是一个二分类问题。如果图像中有多个需要分割的实例，同时需要对这些实例进行区分的情况下，这一方法将不再适用。
3. 当前数据集中的图片背景比较简单，只有一个需要分割出的实例。当前所进行研究的

工作对于数据集有效，但对于具有复杂背景的图片，具有多个实例需要分割的图片，其有效性还有待进一步验证。

参考文献

- [1] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]. 2015. 3431–3440.
- [2] Cubuk E D, Zoph B, Mane D, et al. Autoaugment: Learning augmentation strategies from data[C]. 2019. 113–123.
- [3] Cubuk E D, Zoph B, Shlens J, et al. Randaugment: Practical automated data augmentation with a reduced search space[C]. 2020. 702–703.
- [4] Müller S G, Hutter F. Trivialaugment: Tuning-free yet state-of-the-art data augmentation[C]. 2021. 774–782.
- [5] Zhong Z, Zheng L, Kang G, et al. Random erasing data augmentation[C]. 2020. 13001–13008.
- [6] Zhang H, Cisse M, Dauphin Y N, et al. mixup: Beyond empirical risk minimization[J]. *arXiv preprint arXiv:1710.09412*, 2017.
- [7] 赵凯琳, 靳小龙, 王元卓. 小样本学习研究综述 [J]. *Journal of Software*, 2021, 32(2).
- [8] 李新叶, 龙慎鹏, 朱婧. 基于深度神经网络的少样本学习综述 [J]. *计算机应用研究*, 2020, 37(8):2241–2247.
- [9] Shen Y, Yan Y, Wang H. Recent advances on supervised distance metric learning algorithms[J]. *Acta Automatica Sinica*, 2014, 40(12):2673–2686.
- [10] Bellet A, Habrard A, Sebban M. A survey on metric learning for feature vectors and structured data[J]. *arXiv preprint arXiv:1306.6709*, 2013.
- [11] Koch G, Zemel R, Salakhutdinov R, et al. Siamese neural networks for one-shot image recognition[C]. 2015. 0.
- [12] Vinyals O, Blundell C, Lillicrap T, et al. Matching networks for one shot learning[J]. *Advances in neural information processing systems*, 2016, 29.
- [13] Snell J, Swersky K, Zemel R. Prototypical networks for few-shot learning[J]. *Advances in neural information processing systems*, 2017, 30.
- [14] Sung F, Yang Y, Zhang L, et al. Learning to compare: Relation network for few-shot learning[C]. 2018. 1199–1208.
- [15] Thrun S, Pratt L. Learning to learn: Introduction and overview[C]. *Proceedings of Learning to learn*. Springer, 1998: 3–17.
- [16] Santoro A, Bartunov S, Botvinick M, et al. Meta-learning with memory-augmented neural networks[C]. 2016. 1842–1850.
- [17] Munkhdalai T, Yu H. Meta networks[C]. 2017. 2554–2563.
- [18] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]. 2015. 234–241.

- [19] Zhou Z, Rahman Siddiquee M M, Tajbakhsh N, et al. Unet++: A nested u-net architecture for medical image segmentation[C]. Proceedings of Deep learning in medical image analysis and multimodal learning for clinical decision support. Springer, 2018: 3–11.
- [20] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. 2016. 770–778.
- [21] Azulay A, Weiss Y. Why do deep convolutional networks generalize so poorly to small image transformations?[J]. *J. Mach. Learn. Res.*, 2019, 20:184:1–184:25.
- [22] Tang Z, Gao Y, Karlinsky L, et al. Onlineaugmt: Online data augmentation with less domain knowledge[C]. 2020. 313–329.
- [23] Zhang Z, Liu Q, Wang Y. Road extraction by deep residual u-net[J]. *IEEE Geoscience and Remote Sensing Letters*, 2018, 15(5):749–753.

致 谢

大学四年的时光转瞬即逝，当年入学时的情景仿佛就在昨天。而随着毕业论文的完成，我也即将毕业，离开这个生活了许久的校园。

首先要感谢的是我的导师刘振宇，在进行毕业论文期间，一直关注这我毕业论文的进展状况，并及时的给予指导。在我因拖延一段时间都没有进行毕业论文时，及时的提醒我，督促我，使我最终得以及时的完成毕业论文。

其次，要感谢我的家人。我的父母和姐姐，在我进行毕业论文期间，理解和支持我。并在我陷入失落和没有信心的转态时，给予我信心和鼓励，给予我精神上的支持。

同时，要感谢信息科学与工程学院，为我带来了丰富充实的大学四年生活。这四年的大学生活，丰富了我的知识，提升了我的专业素养。在信息科学与工程学院这个舞台，培养了我的人生观和价值观，为我今后的道路指明了方向。

最后，感谢我的母校——兰州大学。在我人生最重要的阶段遇到了你，“自强不息，厚德载物”的校训在我的脑海里留下深深的烙印。在之后的人生道路中，来自母校的影响将不断的影响和塑造着我，让我更加坚定自己的方向。