基于深度学习的小样本图像分割

郁严贵

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

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

# 绪论

## 背景

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。图像分割作为计算机视觉领域的一个重要问题，具有很大的研究价值与研究意义，一直是深度研究的一个热点领域，受到大量研究者的广泛关注。

同时，当前深度学习取得的成功离不开大量的数据。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。有些问题本身缺乏大量的可以获得的训练数据，或者对于大量数据的标注需要耗费极大的人力成本和精力。例如，在医学图像领域，便存在着用于图像分割的标注数据不足的问题。

## 研究目的与意义

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。而深度学习往往需要大量的标注样本进行训练，才能训练得到具有比较好的泛化性能的模型。而在深度学习的过程中，不可避免的会面临数据量不足的问题。深层神经网络具有非常强的学习能力，在缺乏足够的标注样本对深度学习模型进行训练时，可能会出现过拟合的问题，使得模型的泛化性变差。因此对于小样本这一问题的研究，在可获得的标注数据不足的情况下，实现深度学习网络模型更好的泛化性，具有重要的现实和理论意义。深度学习带来更深层的神经网络。

为了让深度学习更接近人类的认知能力和推理能力，小样本学习方向的研究至关重要，其研究的进展将极大的推进人工智能的前进和发展。

## 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，

​ 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深度学习网络模型，包括网络结构图示及其特点。

​ 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。

​ 第四章，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。

​ 第五章，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。

# U-net网络模型

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很重要的地位。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年发表了《Fully Convolutional Networks for Semantic Segmentation》，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。

U-net由Olaf Ronneberger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。与别的神经网络相比，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[[fig1]](#fig1)所示：

image

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。

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。

# 数据集和数据处理

## 数据集

数据集来源于Kaggle的车辆边界识别与车辆实例分割,数据集包括不同种类的车辆在不同背景下的照片以及车辆对应的分割图。以及同一车辆在不同角度的图片。

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。

## 数据增强

小样本学习的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是，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，在缺乏足够的数据支撑时，很难训练得到比较优的参数。同时，如论文Why do deep convolutional networks generalize so poorly to small image transformations指出的那样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。增强模型的泛化性。

* 数据增强时机  
  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），即在训练的过程中不断的在小批量上数据上进行转换。
* 数据增强策略  
  对于图片的数据增强可以有很多种不同的操作，不同的操作也可以有不同的增强幅度。那么如何找到适合于当前数据集及当前深度学习任务的数据增强方式，对于图像分类任务，目前已经有AutoAugment，RandAugment等数据增强策略。

因为本论文研究的问题为小样本图像分割，面对的问题是数据流不足的情况，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。同时由于数据增强策略需要在定义的搜索空间中来找到适用于当前任务的数据增强操作。这在训练的过程中是一项非常耗费时间的工作，这里不采用这种数据增强策略。而是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：

p5cm<p8cm<p2.3cm< 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)

对原始数据集进行数据增强，并验证效果，[[fig2]](#fig2)给出所采用的增强操作（True表示采用当前增强操作，- 和False表示不采用当前增强操作），[[fig3]](#fig3)在验证集上通过Dice相似系数评估的结果：

image

image

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，[[fig4]](#fig4)和[[fig5]](#fig5)分别给出原始图像与反转图像的对比：

image [fig4]

image [fig5]

虽然从上图对比来看，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。同时高斯模糊这一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6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。

image

[[fig6]](#fig6)中ruby-shape-158为进行6中数据增强操作后将原始数据集扩充6倍后进行5次迭代的运行情况，atomic-moon-156为将迭代次数增加6倍，进行35次迭代的情况，step的差异是由于后者进行了更多次的验证操作。对后者进行35次迭代的原因是控制两次运行所需要的总训练图像数量相同，在此基础上评估运行结果，数据增强操作对最终训练结果带来了明显的提升效果。

# 基于改进的U-net网络进行图像分割

## 损失函数设计

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，我们的目标是最小化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。

### 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 Loss  
  Dice Loss定义如下，其值越大，代表X与Y的差异程度越高。

### 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：

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。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。

### EdgeLoss

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[[fig7]](#fig7)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。

image

将得到的四个图分别与原始分割图进行按位异或操作，得到[[fig8]](#fig8)第一个箭头右侧的4个图，然后将这4个图进行按位或操作，便可以得到右侧的图像。这样可以得到最右侧的图，将蓝色背景的称为边缘(edge)，绿色背景的称为轮廓（outline）。更进一步，将这个图与原始分割图进行按位与操作可以提取出边缘图，将这个图与边缘图进行按位异或操作可以提取出轮廓图。提取出这两个部分后，我们便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:

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。表示两个集合对应元素点乘，将结果相加求和。

image

如下图所示，为通过上述方式提取出的边缘图：

mask图和edge图

mask图和edge图

采用Dice相似系数作为模型的评估指标，如[4.2](#edgeloss)所示为在损失函数中添加EdgeLoss项与不添加EdgeLoss项的对比情况。 其中左侧图像为不进行数据增强操作，训练5 epochs;右侧atomic-moon-156表示进行6种数据增强操作，训练5 epoch，ruby-shape-158不进行数据增强，训练35 epochs以弥补数据增强带来的训练样本数量差异。 devited-sound-152、atomic-moon-156添加EdgeLoss损失项。 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。

添加EdgeLoss项与不添加EdgeLoss项对比

添加EdgeLoss项与不添加EdgeLoss项对比

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，[[mask\_compare]](#mask_compare)显示了他们之间的对比情况，添加了EdgeLoss项后对图像边缘的分割较不添加更精确一些：

## U-net模型改进

对U-net网络模型结构的改进如[[model\_improve]](#model_improve)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，具体如下：

image

* 卷积操作  
  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。U-net网络模型包括四次下采样，和四次上采样，每一步进行两次卷积核操作。对其调整，采用进行每一通道的特征提取，在通过的卷积核来进行不同通道的特征融合。 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，输出通道个数为：
* 以输入3通道的RGB图像通过卷积变为64通道为例：
* 通过上述操作很大程度上减少了网络模型的参数个数，同时在ReLU激活函数前添加批量规范化层（batch normalization）。批量规范化是一种很有效的技术，通过对输入进行正则化，可以减弱每一层的变化剧烈程度，加速深层网络的收敛速度。
* 转置卷积操作  
  四次转置卷积操作逐步恢复原始图片的尺寸，U-net网络模型在进行转置卷积操作时同时将通道数减半，将来自对应下采样层的特征输出裁剪拼接到转置卷积操作的输出中。这一操作将来自下采样的低频特征与上采样的高频特征进行组合，在改进的U-net模型中，去掉copy and crop这一操作，同时在转置卷积操作时不改变通道数量。

改进的U-net模型和U-net在训练集上的表现情况如所示，legendary-oath-189表示改进U-net模型在验证集上的表现，通过对网络模型的修改， 减少了模型的参数量，减少了训练样本偏少时可能出现的过拟合问题，在验证集上的表现得以提升。

image [model\_compare]

image [model\_loss]

# 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。

## 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。

## 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，同时对计算得需求比较大。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，这一方法将不在适用。
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。

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。

99 Olaf Ronneberger, Philipp Fischer, and Thomas Brox., U-Net: Convolutional Networks for Biomedical Image Segmentation. Carvana Image Masking Challenge.,https://www.kaggle.com/c/carvana-image-masking-challenge. Jonathan Long, Evan Shelhamer, Trevor Darrell., Fully Convolutional Networks for Semantic Segmentation. Aharon Azulay, Yair Weiss., Why do deep convolutional networks generalize so poorly to small image transformations?. Kaiming He, Georgia Gkioxari, Piotr Dollar, Ross Girshick., Mask R-CNN. Gen Li, Varun Jampani, Laura Sevilla-Lara, Deqing Sun, Jonghyun Kim, Joongkyu Kim., Adaptive Prototype Learning and Allocation for Few-Shot Segmentation. Bingfeng Zhang, Jimin Xiao1, Terry Qin., Self-Guided and Cross-Guided Learning for Few-Shot Segmentation.