**中图法分类号**:TP39  **文献标识码**:A **文章编号**:1006-8961(年)

结合U-Net和Lovasz-Crossentropy Softmax 损失函数的高分遥感图像语义分割

刘澄1，杨子豪2，万晓华2，邱德慧1

1.首都师范大学信息工程学院,北京100048；2.中国科学院计算技术研究所,北京 1001901

**摘 要：****目的** 针对高分遥感图像语义分割传统方法中存在运算速度慢效率低、依赖大数据集且难以处理遥感图像特有的道路的密集岔路、建筑物的细小缝隙等细节问题，本文提出了一种遥感图像语义分割的损失函数及其网络结构。**方法** 为了避免过深网络带来的梯度消失问题，以U-Net结构为基础，与ResNet结合形成Res-Unet网络，在保留清晰的U型编码解码结构的同时融入ResNet的残差模块。初始化设置为He初始化并引入全新的优化器随机权重平均（Stochastic Weight Averaging）帮助学习收敛。同时，为了更好地解决遥感图像语义分割任务中较小数据集学习效果一般、学习难以收敛等问题，本文设计将能够有效维持网络学习稳定收敛的交叉熵损失函数与能显著提升网络的mIoU得分的Lovasz Softmax损失函数相结合形成全新损失。**结果** 本文在“BDCI 2017佳格数据卫星影像的AI分类与识别 ”的竞赛数据集上进行对照实验，将本文损失与交叉熵（cross entropy）、tversky、Lovasz三种损失函数进行了图像分割性能对比。实验中Lovasz-Crossentropy Softmax损失mIoU分数高达0.84，明显超越了交叉熵的0.79与Lovasz的0.81；f1-score达到0.91，同样优于交叉熵的0.87与Lovasz的0.89。**结论** 实验结果表明，本文所提的方法在结合了交叉熵及Lovasz Softmax损失后，取得了更高的分割准确率。

**关键词**：语义分割；高分遥感图像；U-Net；损失函数；Lovasz损失；

U-Net and lovasz-crossentropy softmax loss function method for semantic segmentation of high-resolution remote sensing images

Liu Cheng1，Yang Zihao2，Wan Xiaohua2，Qiu Dehui1

*1.College of Information Technology, Capital Normal Uninversity，Beijing 100048，China;*

*2. Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190，China*

**Abstract:** **Objective** With the development of remote sensing technology in recent years, the number of remote sensing images is easier to obtain than the past, and the resolution is improved obviously, bringing huge challenges to manual annotation. At this time, people have adopted deep learning-based algorithms as the mainstream semantic segmentation scheme. In this context, it has become the main research direction to expand the segmentation scheme based on deep learning and try to extend it to the high resolution remote sensing image processing problem. At present, relevant deep learning technologies mainly involve full convolutional networks and their derivative models. At first, people mainly used convolutional neural network (CNN) as a semantic segmentation network, but its storage overhead was too large. Nevertheless, CNN only focus on the global information, which can’t achieve satisfactory result in segmentation task. Then, the Full Convolutional Network (FCN) replaced the last fully connected layer of CNN with a convolutional layer. forming a brand-new network solution. It uses up-sampling and iteration to successfully improve the accuracy of the network, and has become one of the mainstream semantic segmentation solutions models in recent years, including many famous fully convolutional network models such as SegNet and U-Net. SegNet utilized the most advanced VGG16 framework as the backbone at the time, and the remaining 13 pooled convolutional layers after removing the fully connected layer are reserved as the downsampling (ie encoder) part. Such a modification greatly reduces the number of training parameters required for down-sampling (from 134M to 14.7M), thereby reducing training overhead and model complexity. The upsampling part adopts the upsampling structure mirrored with downsampling, forming a system where each layer of decoder corresponds to a layer of encoder. The original design of U-Net is also similar to the structure of SegNet, which is also divided into two parts: upsampling and downsampling. The each obtained feature map after basic block (two convolutional layers and one maxpooling layer) has different scale, these feature maps will concatenate with the same scale feature maps in the upsampling part. U-Net has a clearer U-shaped structure and better adjustability, a large number of studies have shown that U-Net grafting excellent network models such as Inception, Mobile net, and ResNet can achieve outstanding results. In addition, the loss function has a kind of influence on network performance. The most classic loss function is the cross-entropy loss, which is one of the most commonly used losses in the deep learning field. It focuses on the distance relationship between two probability distributions. The smaller the cross entropy, the closer the two are. In recent years, researchers have proposed a new Lovasz Softmax loss in order to better complete tasks such as deep learning-based image semantic segmentation (improving the mIoU score). It was first proposed in the CVPR of 2018, and it had outstanding performance in the Kaggle competition (the models of the top five teams are different, but the loss function is almost all Lovasz). Unlike cross-entropy based on accuracy, Lovasz is a kind of The loss based on IoU design, which also means that it is more suitable for segmentation tasks. However, although Lovasz loss performed very well on Kaggle, it is different from the large number of categories on Kaggle and the remote sensing image information of up to hundreds of GB in that the data set used in this article is very small, with only five categories. In addition, the depth of our proposed network is far less complex than the network in the competition. In summary, the existing traditional methods for semantic segmentation of high-resolution remote sensing images have slow calculation speed, low efficiency and rely on larger data sets. It is difficult to deal with the detailed problems such as dense forks of roads and small gaps in buildings that are unique to remote sensing images. **Method** In order to solve the above problems, this paper proposes a loss function for remote sensing image semantic segmentation and its network structure. In order to avoid the problem of gradient disappearance caused by deep networks , U-Net with upsampling-downsampling structure is adopted on the main structure, and grafts the excellent classification network ResNet-34 to the encoder part, and combines with it to form a Res-Unet network, aiming to integrate the residual module of ResNet on the basis of keeping a clear U-shaped structure. On the basis of this network structure, the initialization is set to He initialization and the new optimizer Stochastic Weight Averaging (SWA) is introduced to help network converge more stably. The concrete realization of SWA innovatively adopts the form of SWA packaging Adam. At the same time, in order to better solve the problem of general learning effect of small datasets and difficulty innetwork convergence in remote sensing image semantic segmentation task, this paper designs a new loss. The loss design idea is to combine cross-entropy loss function (which can effectively maintain the stable convergence of network learning) and Lovasz Softmax loss function (which can significantly improve the network’s mIoU score). The experimental equipment in this article uses NVIDIA GeForce RTX 2060 as the GPU and Intel i7-9750H as the CPU, and it was carried out in the competition data set of "BDCI 2017 Jiage Data Satellite Image AI Classification and Recognition". **Result** The proposed loss function in this paper was compared with the three loss functions of cross entropy, tversky, and Lovasz. (The cross entropy and Lovasz were used as the weighted original loss. Contrast, and the introduction of tversky can eliminate the influence of the weighting algorithm itself on training and its results). In the experiment, the Lovasz-Crossentropy Softmax loss mIoU score is as high as 0.84, which obviously exceeds the cross entropy of 0.79, Lovasz's 0.81 and tversky's 0.76; f1-score reaches 0.91, which is also better than the cross entropy of 0.87, Lovasz's 0.89 and tversky's 0.86. **Conclusion** Experimental results show that the method proposed in this paper has achieved higher segmentation accuracy after combining cross entropy and Lovasz Softmax loss.

**Key Words:** Semantic segmentation; high-resolution remote sensing image; U-Net; loss function; Lovasz loss;

**0 引 言**

遥感技术是人们于20世纪60年代开始广泛发展的新技术，而其在促进生活中各行各业经济发展的同时也对军事等众多特殊领域提供了更加广泛的应用可能性，而这些也成为了遥感相关技术前进的新方向。目前我国已经发射了数十发长光卫星“造星”，遥感数据所获取到的空间分辨率和时间分辨率已经大幅提高，在各行各业中都有着广阔的应用前景。其中，遥感图像的语义分割处理是最重要的处理任务之一；其能高效地将图像中的地物类型进行分类标注，是后续深层次处理运算的基础。

然而与普通图像分割任务不同的是高分遥感图像中海量信息带来的指数性增长的工作复杂度。最初，图像分析相关任务均由人工作业完成标注；随着技术的飞速进步，深度学习也为图像语义分割提出的新的解决方案。近几年，深度学习已经成功地应用在普通图像的语义分割任务上，无论是Long等人（2009）提出的全卷积网络（Fully Convolutional Networks，FCN），还是Fisher Yu等人（2016）发表的空洞卷积（dilated/atrous），都取得了相对出色的效果。然而高分遥感图像则存在自身的工作特殊性。首先遥感图像不同于常规图像，本身图像具有尺寸超大（目前业内常用尺寸至少在5000x5000以上）、包含地物种类极其丰富复杂（往往超过二十种不同的种类），使得传统方案计算效率极其低下；除此之外，高分遥感图像又受到光照、云层、拍摄角度等多方面因素影响，使得图像噪声较大，语义分割的准确率又成为了问题；同时，深度学习网络训练依赖大量的标记数据集，但人工标注数据费时费力效率低下，再加上大尺寸及多分类等影响，准确率也难以保证。这些弊端也使得高分辨遥感图像语义分割目前面临重大挑战。如最传统的全卷积模型SegNet网络，本身基于经典的VGG16设计，无论是网络深度还是结构都过于简单，面临高分遥感图像的处理工作时则显现出准确率不达标等问题，难以完成对于高分遥感图像的分析工作。与之相比，基础U-Net网络则具有更清晰的U型上下采样结构以及长连接（skip connection）来尽可能防止上采样丢失的空间信息。但其许多参数设置都并不能作为最优解（例如采样为四层设置，周纵苇等人（2020）就发现其实四层采样并不是通解），而这些设置在本文实验中的最终准确率等指标的成绩（尤其是本文为小数据集的情况下）也不能令人满意。

对于基于深度学习网络进行的遥感图像语义分割任务，近年来相关研究成果也有许多，但仍存在一些问题及弊端。如前文提到经典语义分割网络U-Net，苏健民等人（2019）在其基本的U型结构的基础上，将网络内的ReLU替换为ELU激活，改进后的U-Net模型在其论文中取得了相对优秀的效果。但在本文实验数据集中，其对比U-Net的优化效果非常有限，并没有显著提升mIoU等重要指标的成绩；同样在U-Net的框架下，Rakhlin等人（2018）也尝试用新的Lovasz Softmax损失（由Berman等人（2018）提出）提升模型学习效果，但其论文内实验成绩相对一般，且虽然使用了加权的思想但也没有更深入地讨论计算超参数设置等细节；Selim等人（2019）提出使用同样类似于FCN的FPN（Feature Pyramid Network）来解决相关问题，但其金字塔结构所需资源较本文实验资源相对庞大，且在小数据下表现一般；Chao等人（2019）也基于经典的DenseNet提出了自己的DFCNet作为新的网络框架解决遥感图像土地分类问题，但其网络规模过深且运算负荷较重，难以在本文环境下完成。

至此，高分遥感图像语义分割传统方法中存在运算速度慢效率低等问题已经初现端倪。综合来看问题主要由以下几点引起：

1、过于依赖大数据集。为了取得更好的实验成绩，庞大的数据集必将成为最有力的数据支撑，但其最大的弊端便在于使得整个网络的训练周期显著延长。

2、网络框架较为繁复。与第一点相照应的，为了获得更好的成绩，复杂化整体网络来提升拟合能力是最简单直观的方案。但其弊端便在于同时也增大了资源及计算时间的开销。

为了解决上文提出的问题，实验需要在较小训练数据集条件下有效利用深度学习网络解决高分遥感图像语义分割的问题并取得更优的实验结果，即需要设计出结构相对简练运算效率高的同时保证优秀成绩的网络。由此，本文基于经典的U-Net结构设计了一个全新的网络，并同时搭载独立设计的加权综合损失。首先网络架构上采用U-Net的U型结构，并将其下采样部分嫁接为ResNet-34网络，以此来避免梯度消失等问题的发生；接着在其中引入He初始化设置，同时搭载随机权重平均（Stochastic Weight Averaging）优化模块来优化整体网络学习收敛。然后为了更好地提升整体网络学习效果及最终实验成绩，本文设计将能够有效维持网络学习稳定收敛的交叉熵损失函数与能显著提升网络的mIoU得分的Lovasz Softmax损失函数相结合形成全新损失。最终的实验中将本文损失在准确率及mIoU等指标上的得分对比了交叉熵损失、Trajdos等人（2017）论文中的Tversky损失及Berman等人（2018）提出的Lovasz Softmax损失，最终也证明本文损失取得了更优得成绩，也就体现出了更出色的语义分割效果。

**1、网络设计**

本文网络结构主要采用Res-Unet网络，并在其基础上使用He初始化函数，且搭载SWA优化器。其中Res-Unet网络可以在保证深度的同时尽可能规避梯度消失等问题，而He初始化更契合本文激活函数，SWA封装则可以协助网络进行更稳定的学习收敛。其具体研究内容如下介绍。

**1.1网络模型Res-Unet**

空洞卷积与全卷积网络是目前比较主流的深度学习算法。由于空洞卷积对于细小物体的分类准确度有限，不适用于本文高分遥感影像的实验目标，故选择全卷积方案。在全卷积网络中，出于对实验资源及数据集等因素的考量，本文选择具有清晰U型结构以及高度可调（嫁接）性的U-Net作为框架。

接着将U-Net的下采样部分替换嫁接成34层ResNet网络结构。其残差块的设计可以抑制网络深度上升后带来的梯度消失问题，同时也能更效率地利用有限的实验资源。而这样形成的新网络结构称作Res-Unet。具体结构如图1所示。

图中每一个方块都标志着一组运算（卷积池化、upsampling或concatenate），为了更好地表述其特征，图中将一些重复的运算打包成了运算块（conv block、conv block（short cut）及sampling block）。块的上方表示重复运算次数，而下方则是本次运算块内卷积的过滤器数。

本文的Res-Unet结构中存在两种残差块模式（conv block及conv block（short cut）），其中普通的卷积运算块中则是正常的addition，而另一个则加入了1×1的卷积用来调整特征图的维度（因为该运算块内的卷积会使得尺寸变为1/2）。

值得一提的是，每一次池化操作虽然能去除一些冗余信息，但同时也会不可避免地丢失掉一些关键的空间信息。本文网络中为了提升每个特征图的提取功能和结果，在能保证整体运算顺利进行的提下移去了原ResNet上第一层卷积上的Max Pooling，力求能尽量多地保留特征信息。而其后所有特征图尺寸减半的操作均由步长为2的卷积运算完成，使得整个网络尽可能避免了最大池化。此举过后会使得初始阶段只有第一层7×7卷积核的卷积操作会使得整体尺寸减半，在进入接下来的下采样进程时保留了128×128的大尺寸，且上采样也同样可以减少一次反卷积操作。而在本文的实验进行中也证明，虽然会小幅度地增加计算开销，但确实取得了有效的优化效果。

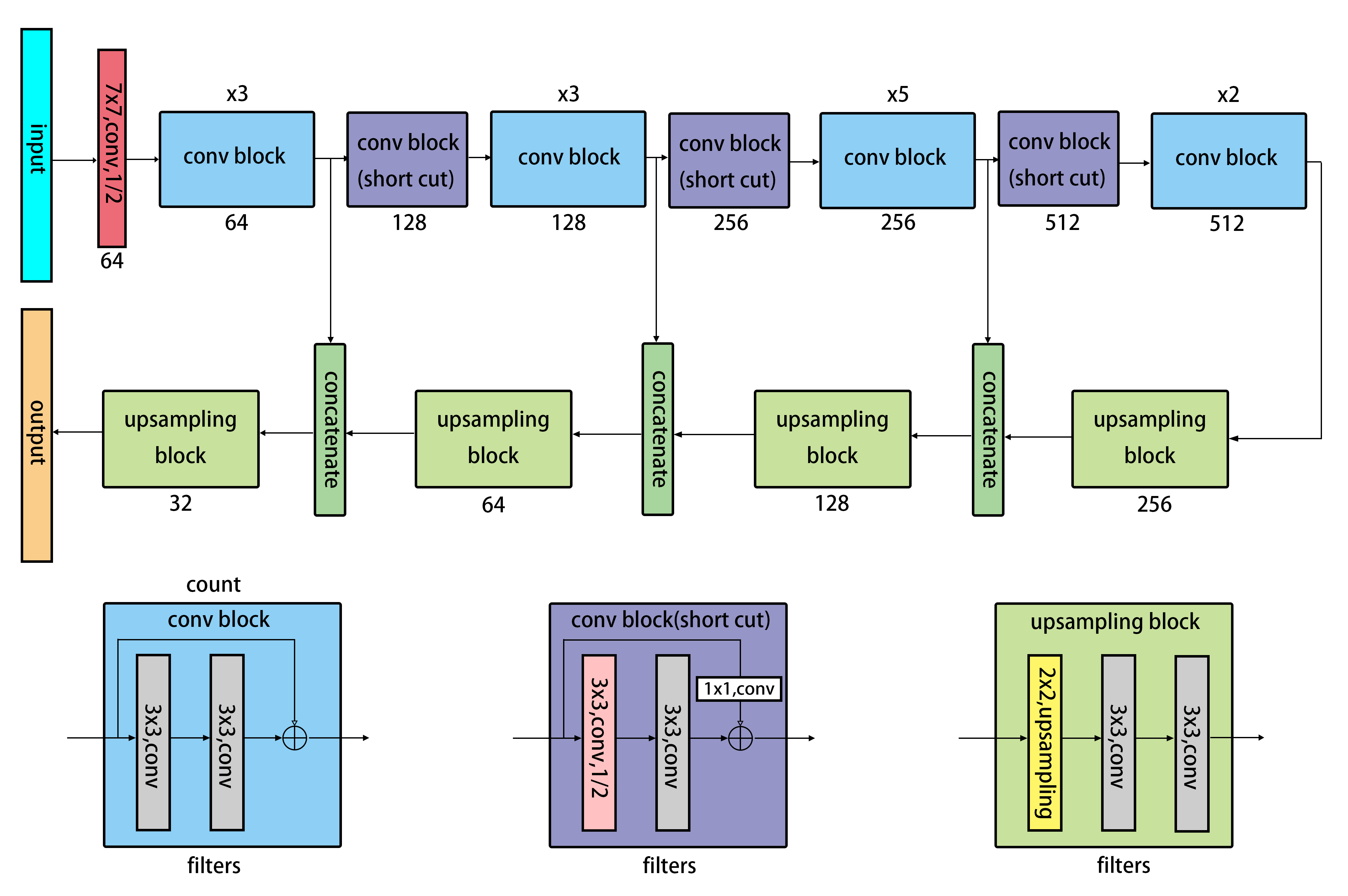


图1 Res-Unet结构图

Fig.1 Res-Unet structure diagram

**1.2初始化函数**

本文实验网络设计中均采用He初始化方案。

最初，随机初始化曾是最常用的初始化方案。核心思想在于通过高斯分布随机变量为所有参数赋予一个随机值，然后再将这个值大幅缩小（一般是与0.01做乘法），因为一个比较大的权重矩阵会让tanh和Sigmoid等激活函数非常难收敛，模型学习效果就会大幅下降。

由于随机初始化存在着随着层数的加深，输出会迅速趋向于0使得参数难以更新。而Xavier通过保持输入输出方差一致，来避免输出趋向于0的问题。这个方案对于tanh很有效，但是对于ReLU则还是会存在趋零问题。

鉴于Xavier在常用的激活函数ReLU上的并不优秀的表现，何凯明提出了针对ReLU的初始化方式。He的核心思想在于如果使用ReLU做激活，那么假定每一层有一半的神经元成功被激活，而另外一半则为0。于是为了保持差额不变，在Xavier的基础上再乘0.5即可以得到解决。效果如图2所示：

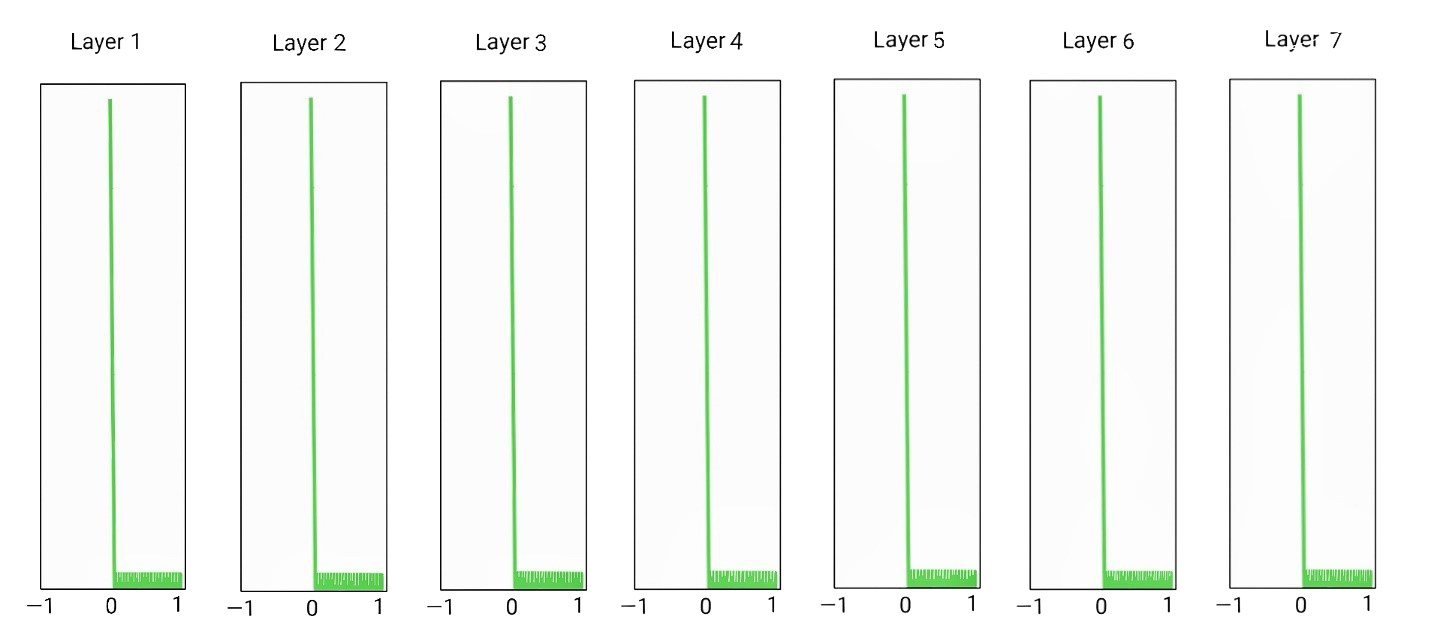


图2 He初始化输出情况

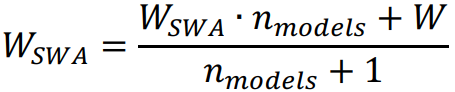
Fig.2 He initialization output

图2中描述的是ReLU激活函数的输出值（0~1）分布情况，从左往右代表着层数的逐渐加深（第1层~第7层），可以看出直至第七层也没有明显的差异。由此可见，He initialization并没有因为层数的加深导致出现输出值趋零现象（分布情况向0聚拢导致其他值极少甚至不取），相反甚至基本没有因为变深而产生任何变化，保持了很好的稳定性和平均度，非常适合给ReLU做初始化。

综上可以看出，He初始器可以与ReLU有更优的适配度，而本文中为了网络效果也采用了ReLU/ELU的激活函数，故也选择He\_normal做初始化函数。

**1.3优化器**

本文采用新兴的随机权重平均（Stochastic Weight Averaging，SWA）作为优化器配置。它可以借助经验观察等思想对权重进行动态更新，从而帮助网络进行自学习。其中权重更新公式为：



（1）

式中*W*表示当前模型的权重，*WSWA*代表平均权重（也是更新的目标），*nmodels*则表示目前迭代更新周期。

SWA通过平均化这些不同的损失值获取到更低的损失，最终帮助网络实现自学习，找到当前学习状况下的最优解。

它的方案与FGE（Fast Geometric Ensembling，快速几何集成）以及Snapshot Ensembling（快照集成）类似，但集成模型的性能更优于上述两种。作者在其论文中展示的与SGD模型的对比如图3所示。

图片左图中的W1，W2 和W3分别代表三个保持差异、独立进行学习训练的网络，而Wswa则是三种网络结果的平均值。虽然从中间的图中可以看到SWA在125次迭代时损失要高于SGD，但右图可以发现SWA在测试集上的表现显著优于SGD，可以证明其更优的训练效果。

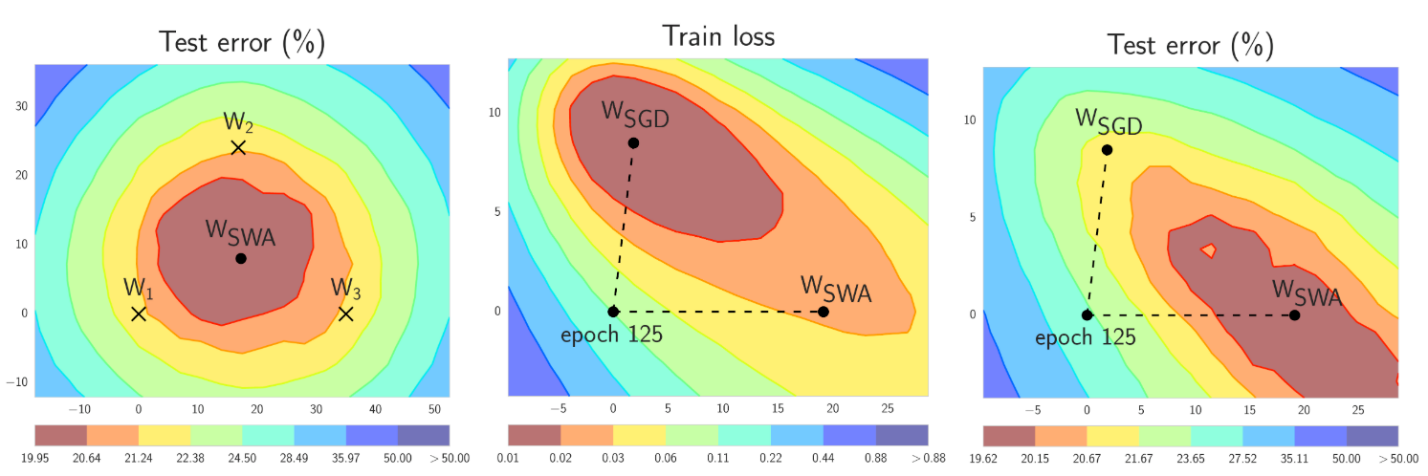
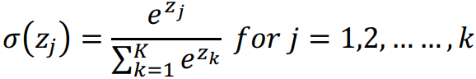


图3 在CIFAR-100上带有预激活ResNet-164的SWA和SGD对比图

Fig.3 Illustrations of SWA and SGD with a Preactivation ResNet-164 on CIFAR-100

**1.4其他细节设计**

除了前文介绍的网络结构外，网络中最后一层的激活为Softmax，其公式为：



（2）

在该公式中，*σ*表示Softmax函数，*σ（Zj）*表示将Softmax函数应用于数字*Zj*。*Zj*表示单个原始输出值，*j*表示当前运算的输出值。

优化器使用SWA（随机权重平均）方案优化的Adam优化器，学习率采用动态学习率，在0.001至0.0005之间浮动。

分批方面Batch size为8，防止过拟合而设置的Dropout层位于上采样的conv4与conv5之间，系数为0.2。初始化方案采用He\_normal初始器，其为从均值为0，标准差公式为：



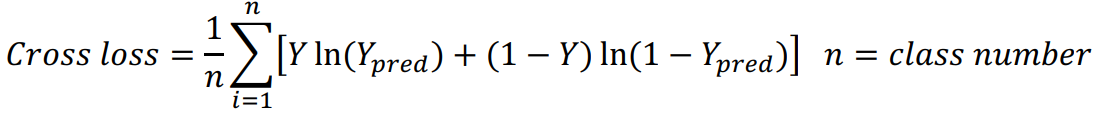
（3）

其中*a*表示后一层的激活函数中负斜率（默认为ReLU激活，即*a*=0），为权值张量中的输入单位的数量。

**2、损失函数算法**

**2.1交叉熵损失**

交叉熵（Cross Entropy）损失是多分类网络中最常用也是最稳定的损失函数之一。其着重刻画了两个概率分布之间的距离关系，交叉熵越小也就说明两者越接近。因为交叉熵更多地表达了预测输入样本属于判断类别中的某一类的概率，所以在分类网络中表现突出。交叉熵的公式如下：



（4）

式中*n*表示类别的数量，*Y*表示样本的label即真实值，*Ypred*表示样本预测为真的概率，即预测值。

交叉熵作为多分类下最常用的经典损失，其具有能够帮助网络稳定收敛及有效迭代提升准确率等特点，在任何分类任务中都能取得较稳定出色的效果。值得一提的是，交叉熵损失与其他常用损失类似，均基于正确率（accuracy）设计。

**2.2 Lovasz Softmax 损失**

2018年Berman等人（2018）首次提出Lovasz-Softmax 损失。作为一种基于IoU计算的损失，论文表述其实验中的收敛效果好于经典的交叉熵（cross\_entropy），并能在分割任务中使用。值得一提的是，在近年的Kaggle图像语义分割比赛上，前几名均使用了该损失。

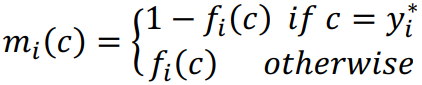
论文中除了对于二分类问题的优化，对于多分类也有多个突出贡献：

1、提出了Lovasz-Softmax 损失对于多类别的参数设定。

2、设计了基于mini-batch的优化IoU。

3、原文提出的损失对经典分割方案的效果做出巨大提升。

其具体计算过程如下：



（5）



式中*fi(c)*表示第c类经过激活（softmax）后的预测值，(Lovasz extension)等价于将原模函数的输出作为基值进行插值，而这些基的运算在ground truth中取一部分作为预测结果，其余则不做预测。

然而，虽然Lovasz 损失在Kaggle上表现非常优秀，但不同于Kaggle上的大量类别、多达几百GB的遥感图像信息，本文所用数据集较小且仅有五个类别；除此之外，网络深度也远没有比赛用网络庞大复杂。综合这两点考虑，为了在已有数据集上取得更好的结果，本文在实验中设计了一种新的加权损失作为解决方案。

**2.3 Lovasz-Crossentropy加权综合损失(LCE Loss)**

随着深度学习的不断发展，损失函数设置上将加权算法作为损失的设计思路已经较为常见。其中tversky 损失就是一种将Dice及Jaccard作加权的综合损失，有大量的实验证明其最终的收敛效果也比较出色。

为了更好地解决上述实验问题，本文借鉴了其加权的设计思路，将Lovasz 与交叉熵损失作了综合，形成了一个全新的加权综合损失。其公式如下：



（6）

公式中*α*代表Lovasz损失在综合损失中所占权重，而剩余权重便分在交叉熵损失上。其中α的具体取值问题将在实验中进行讨论。

由于该损失函数由Lovasz与CE损失加权计算获得，故命名为Lovasz-Crossentropy（LCE）loss。

LCE损失的设计上主要目标是借助加权算法更好的统筹融合Lovasz及交叉熵损失两者的优势。具体思路如下：

1、如前文所说，Lovasz是基于IoU设计的损失函数，而IoU也是近年来图像语义分割领域最权威的评价指标，其可以帮助网络模型取得一个较优的评测成绩。而Lovasz 损失对于网络及数据集等要求也较高，因为在高分遥感图像语义分割中表现不佳。

2、这样的前提下，交叉熵损失作为一个经典而可行性极强的损失函数，虽然是基于传统的准确率为基准设计，但是其稳定性也能够在一定程度上帮助网络顺利收敛学习。

3、将两者加权综合后，理论上既能够追求较高IoU得分，又能减缓数据集等因素的限制影响，帮助网络顺利学习收敛。因此，加权综合损失函数更适用于较小数据集下的遥感图像语义分割任务。

综上，本文借助加权的思想设计一种全新的加权综合损失函数，旨在能结合交叉熵及Lovasz损失两者的学习优势，帮助网络在较小数据集下更好地完成相对复杂的语义分割任务。

**3、实验与分析**

**3.1 数据集及实验环境介绍**

本实验中所选用数据集出自2017年《BDCI 2017佳格数据卫星影像的AI分类与识别》竞赛数据集，总数据集为五张大尺寸高分遥感图像及其人工标注图，尺寸大小分别为5664×5142、7969×7939、3357×6116、7969×7939、4011×2470（单位均为像素）。

本文实验中选取了其中四张图像（尺寸分别为5664×5142、7969×7939、7969×7939、4011×2470）作为训练集和验证集，剩下的图像（3357×6116）作为测试集检验模型。

实验中为了降低计算负荷及受显存限制，所有高清遥感大图和其label均切割成了多个256×256像素小图再送入通道，后按8个作为一个批次(batch)分批输入网络，即网络每次运算的单位是一个小批次，这也有效降低了资源的负荷。最终经过图像增强及切割后，共有50000张小图用于网络训练。

本实验软硬件如表1所示：

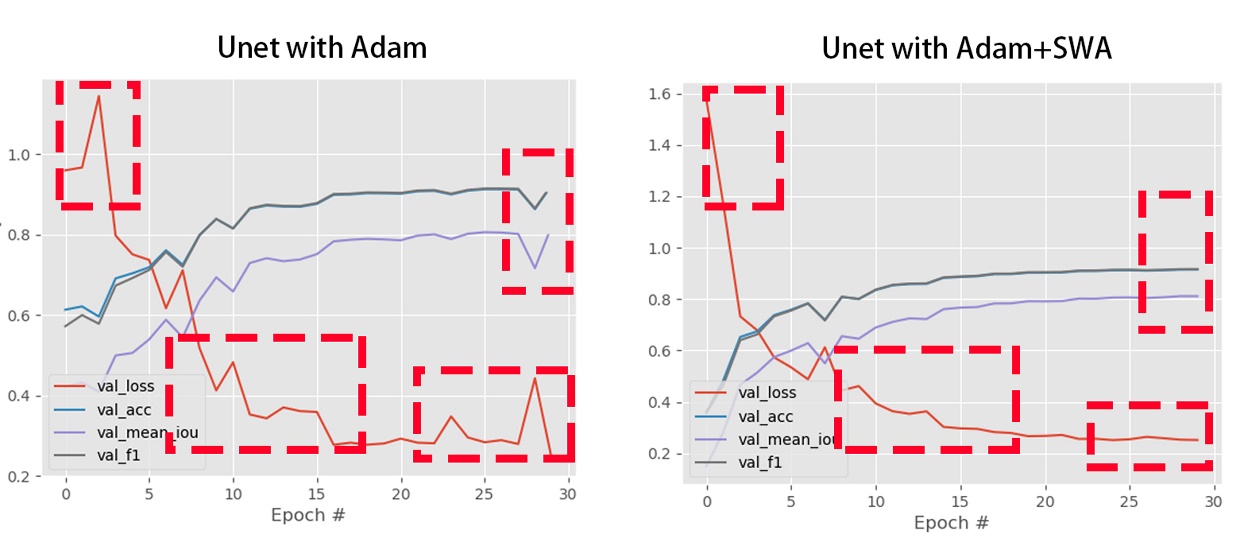
**表1 实验软硬件情况**

**Table 1 Experimental hardware and software**

|  |  |
| --- | --- |
| 软硬件名称 | 版本/型号 |
| 深度学习框架 | Keras 2.1.0 |
| Opencv | 3.4.2 |
| CPU型号 | Intel® Core(TM)i7-9750H CPU @ 2.60GHz 2.59GHz |
| GPU型号 | NVIDIA GeForce RTX 2060(5980MB) |

**3.2 SWA性能比较实验**

首先从训练曲线中分析优化器SWA的优势。实验采取对比实验，将两个其他参数均相同的网络进行结果比较。网络结构为Res-Unet，损失函数为lovasz与交叉熵加权损失。其中一个网络采用Adam优化器，另一个为用SWA优化过的Adam优化器。其训练相关参数曲线如图4所示。



（a）未经SWA优化网络学习曲线 （b）经SWA优化后网络学习曲线

图4 SWA优化效果

Fig.4 SWA optimization effect

(a) Network learning song without SWA optimization; (b) Network learning curve after SWA optimization

从图4中可以看出，虽然SWA会使得初始Loss相对高一点，但是并不影响其收敛速度及效果，而且最终结果也稍优于没有使用SWA的模型。初始Loss较高在作者论文中也出现，原因推测是因为最开始的权重由于初始化的缘故并不是最佳的，导致Loss略高，但并不影响整体网络效果。除此之外SWA主要作用在于防止相关系数震荡。从图中可见SWA网络在收敛中抖动非常少，曲线平稳。而相同时期的几处曲线对比已经由红色框图框出，可以明显看出在没有SWA的情况下即便是Adam优化器也会使得曲线有大幅震荡，而相同时期相同网络有SWA的情况则大不相同，几乎是平稳过渡。

至此可以看出SWA清晰的学习优化效果，后续实验优化器均采用SWA进行包装。

**3.3损失函数性能比较实验**

在进行损失函数的性能对比实验前，首先需要通过测试Lovasz-Crossentropy加权损失中α系数的不同赋值来寻找最佳的权重超参数。



（6）

实验中分别对α系数进行了7个不同值进行测试，赋值分别为0、0.2、0.4、0.5、0.6、0.8、1。赋值测试主要通过mIoU及accuracy两个指标进行评估。结果如表2所示。

**表2 不同α值的得分对比**

**Table 2 Comparison of scores for different α values**

|  |  |  |
| --- | --- | --- |
| α | accuracy | mIoU |
| 0 | 0.9248 | 0.7937 |
| 0.2 | 0.9230 | 0.7948 |
| 0.4 | 0.9249 | 0.8289 |
| 0.5 | **0.9257** | 0.8379 |
| 0.6 | 0.9140 | **0.8381** |
| 0.8 | 0.9134 | 0.8356 |
| 1 | 0.9046 | 0.8148 |

注：加粗加下划线字体为每列最优值。

从表2中可以看到，评估结果中取得最佳accuracy值0.9257时α值为0.5，而最佳mIoU值0.8381时α为0.6。综合来看，在α值小于0.5时，mIoU值的得分会随着α的下降明显衰减；反之其大于0.5时，accuracy值得分情况也逐渐下落。总的来说当α取值在0.4~0.6之间时，网络能够取得一个较优的综合成绩。而在该区间内0.5在实验中表现稍优于0.4及0.6（取得了最佳accuracy得分，且mIoU得分只与0.6的相应得分相差0.0002），故后文实验中Lovasz-Crossentropy加权损失的α超参数均设置为0.5。

接着进行对比损失函数的主要实验。实验模式采取对比实验，主要由本文方法的Lovasz-Crossentropy加权损失与已有的tversky损失、交叉熵（cross entropy）损失、Lovasz损失作对比。其中交叉熵与Lovasz作为加权的原损失进行比较分析，tversky则是作为同是加权综合损失与本文损失进行比对，以此排除加权算法本身对于训练及其结果的影响。

从表3中可以看出，本文所用损失的accuracy值为0.9257，mIoU值为0.8379，而f1-score值为0.9109，kappa值为0.8811，precision为0.8989，recall值为0.9232。相较于其他网络的表现，本文损失函数基本在全方位都达到了最佳。

更值得注意的是本文损失与交叉熵及Lovasz 损失的对比情况。如前文所说本文模型所用损失是用加权交叉熵及Lovasz 损失的方式设计的。如前文所说损失函数的计算基准不同，数据中也确实体现出交叉熵损失的accuracy得分更好，而Lovasz则是在mIoU上取得了较出彩的成绩；然而在实验数据中可以看出本文损失甚至在其优势得分项目上都比原损失表现更好（准确率优于交叉熵，mIoU得分优于Lovasz）。可以说加权综合算法起到了“1+1>2”的出众效果，比较成功地达成了设计预期。同时也印证了加权算法确实有效地结合了两者的优势，形成了一个更加全面优秀的损失函数。

**表3 各损失函数最终得分情况**

**Table 3 The final score of each loss function**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | accuracy | mIoU | f1-score | kappa | precision | recall |
| crossentropy | 0.9248 | 0.7937 | 0.8786 | **0.8824** | 0.8418 | 0.9188 |
| tversky | 0.9046 | 0.7648 | 0.8610 | 0.8481 | 0.8279 | 0.8969 |
| Lovasz | 0.9046 | 0.8148 | 0.8976 | 0.8468 | 0.8949 | 0.9003 |
| Lovasz- crossentropy | **0.9257** | **0.8379** | **0.9109** | 0.8811 | **0.8989** | **0.9232** |

注：加粗加下划线字体为每列最优值。

不过，同时我们可以看出交叉熵loss的kappa系数得分微高于本文损失函数，但包括mIouU、f1-score、precision等成绩均远不如后者。从指标意义来看，kappa系数是一种用于一致性检验的指标，核心意义在于衡量分类精度。与常用的accuracy不同，它更多的表现在关注目标分类的偏向性。直观上分析，交叉熵loss虽然正确率及交并比不及本文loss，但其对于分类的偏向性极低。初步判断认为这正是一些简单经典loss的优势所在，即对于不同分类的偏向性较低。同时也要考虑数据集中本身分类占比就不平衡的前提。不过其他指标上本文网络均要表现更佳，而即便是kappa系数也只低了0.01，不影响判断出本文loss对于目标图像的预测精度更高于交叉熵loss。

除了各个评判指标的成绩，其预测效果图也需要以视觉维度来观察分析。经过染色后的预测图如图5所示。

直观地去看本文网络预测图与真实标注图（最后两列），从整体分析的角度来看，实验网络优势已经比较明显。临近建筑物的间隔勾勒以及道路的完整性应是该测试集最大难点，在这两点上本文网络表现比较出色（尤其是第四张图右上角的建筑物非常难区分，只有该网络将大致轮廓描述正确）。除此之外，对于一些较窄的路分辨得也比较清楚，对于小建筑物的形状把握也相对一致。

与之相对的，交叉熵损失预测的图像虽然在结构、细节等方面表现也很不错，但是出现了较多的预测错误，让整体预测图像出了比较多明显的差异（例如第三张图像中央的几块黄色像素、第四张图右侧的绿色像素），导致预测效果不如实验网络。

而Lovasz损失的预测图虽然基本没出现过错误预测像素，但是出现本应连续的像素却有明显断裂的情况（例如第二张图的道路），而且对于密集建筑物的间隔刻画也比较糟糕（第三张图右下、第四张图右上，都基本没能分离建筑物），可以看出对细节处理还是与实验网络有明显差距。

综上，可以看出综合损失有效地规避了交叉熵损失的错点过多、Lovasz的细节分割糟糕等问题，并结合两者的优势，在细节分类和整体预测上都取得了不错的成效，这也可以印证本文设计

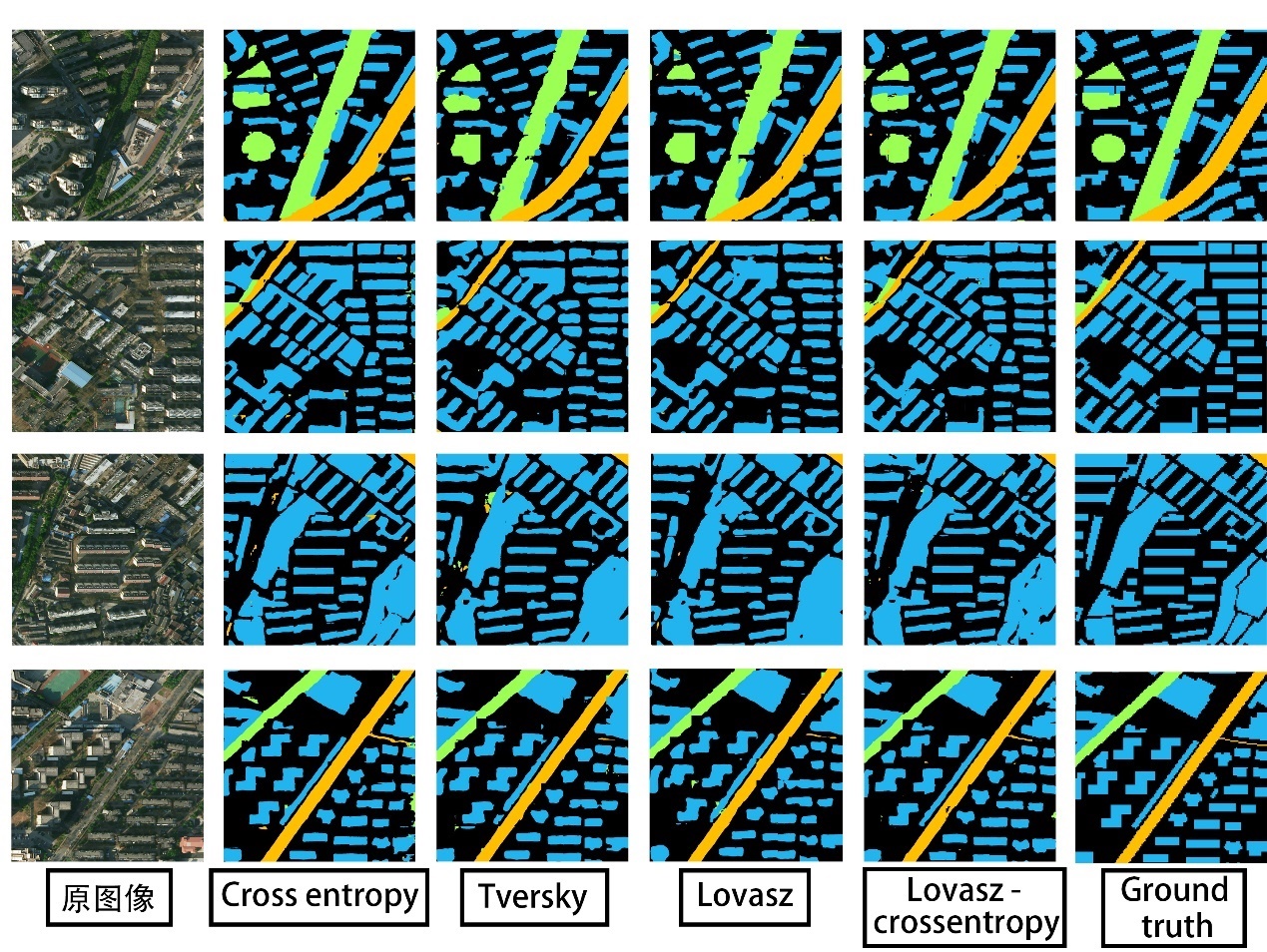


图5 各损失函数预测效果

Fig.5 The prediction effect of each loss function

**表4 各模块优化得分情况**

**Table 4 Optimization score of each module**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | accuracy | mIoU | f1-score | kappa | precision | recall |
| SegNet  U-Net | 0.8736  0.8736 | 0.7599  0.7701 | 0.8712  0.8651 | 0.8298  0.8451 | 0.8516  0.8415 | 0.8820  0.8902 |
| Res-Unet | 0.8931 | 0.7851 | 0.8788 | 0.8282 | 0.8670 | 0.8909 |
| Res-Unet+SWA | 0.9089 | 0.8008 | 0.8947 | 0.8537 | 0.8875 | 0.9022 |
| Res-Unet+SWA+LCE loss | **0.9257** | **0.8379** | **0.9109** | **0.8811** | **0.8989** | **0.9232** |

注：加粗加下划线字体为每列最优值。

损失（LCE）借助加权的思想高效地结合了两个优秀损失函数的优势及特点，在视觉效果图上也展现出来明显的提升效果。

**3.4网络核心模块消融实验**

经过了上述实验与结论分析，可以证实本文网络整体设计在实验中取得了相对出色的成绩。为了进一步的保证实验的严谨性，本文根据网络设计中的三个关键核心算法（即网络模型Res-Unet、优化器SWA及损失函数Lovasz-crossentropy（LCE）加权综合loss）进行了相关消融实验。

实验设计中首先引入传统SegNet网络与U-Net进行框架上的对比；接着将原始U-Net网络作为基准（base line），即下采样部分不做任何嫁接，优化器采用Adam，损失函数为交叉熵loss；其后依次完成将网络结构替换为Res-Unet、优化器加入SWA进行包装、损失函数优化成Lovasz-crossentropy加权综合loss三个操作，从而形成五种网络模型进行对比实验。实验中的评价指标依然采取前文实验中的六个指标进行评估。评估结果如表4所示；

从表4中进行分析后可以看出，引言中提到的经典全卷积网络SegNet与作为基准的初始U-Net成绩相近（整体略低于U-Net），即它们虽然也具有顺利完成训练的能力，但在三种成绩上表现均非常一般；而在将下采样网络嫁接为网络

ResNet后，各个指标均出现了比较明显的上涨；类似地，SWA的优化效果也比较显著，其使得整体网络成绩进一步得到了提升，准确率已经超过了90%；最终在完成交叉熵到本文设计的Lovasz-crossentropy加权综合loss的替换后，整体指标均得到再次优化的同时，mIoU分数更是直接提高了3.8个百分点，得到了飞跃式地快速增长。总的来说，Res-Unet网络、SWA优化器及LCE loss均体现出了比较明显的优化效果，其中以本文loss对于mIoU指标的提升最为显著。

综上所述，实验设计的各个模块均体现出了不同程度的优化效果，且每个模块对于网络得分的提升幅度都比较明显。其中最为突出的为损失函数模块的Lovasz-crossentropy加权综合loss。其作为全新设计的损失函数，不仅在accuracy及f1-score成绩的优化效果上不逊于其余模块，且在mIoU分数上帮助网络取得大幅度提升。

**3.5网络架构运算效率对比实验**

除消融实验之外，实验还对各个网络架构的运算效率进行了对比。与优化效果对比实验类似地，引入架构包括SegNet网络、U-Net网络、Res-Unet网络及本文设计的Res-Unet+SWA+LCE架构。需要注意的是，实验中将是否去除第一个最大池化层（位于7\*7的卷积层后）的两种情况作为不同的两个架构加入了对比实验。同时实验还引入了更复杂的网络ResNext来协助进行运算效率的分析。指标

**表5 各架构运算效率情况**

**Table 5 Computing efficiency of each architecture**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Time/epoch | accuracy | mIoU |
| SegNet  U-Net | 780s  **711s** | 0.8736  0.8736 | 0.7599  0.7701 |
| Res-Unet | 797s | 0.8931 | 0.7851 |
| Res-Unet+SWA+ LCE loss | 880s | 0.9134 | 0.8192 |
| Res-Unet(without 1st maxpooling)+SWA+ LCE loss | 910s | **0.9257** | **0.8379** |
| ResNext-Unet | 1987s | 0.9112 | 0.7989 |

注：加粗加下划线字体为每列最优值。

上使用每个epoch的平均运算时间（总训练均为30个epoch）、accuracy得分及mIou得分进行比较。实验结果如表5所示。

分析表5的数据后可以看出，随着网络复杂程度及计算性能的提升，整体运算时长不可避免得出现了上升的趋势。虽然本文提出网络架构较简单架构时间消耗有一定上升（比最快的U-Net平均慢了200s/epoch），但较更复杂的架构ResNext-Unet依然体现出较快的运行速度（其平均较U-Net慢接近1300s/epoch，较本文网络架构也有接近1100s/epoch的差距）。除此之外，也可以看到删除了第一个池化层平均只增加了30s/epoch的消耗。总得来说，本文设计架构还是成功地将运算时间的上升控制在了一个较小的区间内。

虽然运算时间上本文架构并不是最快网络，但其mIou及accuracy得分较其他网络体现出了绝对优势（包括理论上运算能力最强的ResNext）。同时也可以看出，删除第一个池化层使得整体网络在accuracy得分上提高了1.23个百分点，mIoU更是提高了1.87；可以说相较于仅有30s/epoch的时间消耗上升，成绩的提高更为显著而出色。由此也可以分析出，删除第一个池化层的设计优势远远高于弊端。

综上，运算效率上本文架构也体现出了优秀的实验成绩。在运算能力大幅提高的同时将运算时间控制在较小的浮动里（较ResNext），且依然取得了最优的成绩；此外，数据对比也证明了删除第一个池化层的设计确实在少量增加时间消耗的基础上显著提高了整体实验成绩，可以进一步证明出本文整体架构设计的合理性。

**4、总结**

本文以U-Net作为基础架构，提出了一种引入残差块模块的Res-Unet模型，并在其上搭载SWA优化包装；除此之外又独立设计全新的Lovasz-crossentropy加权综合loss作为损失函数。该网络模型为前文提出的需要在有限资源下有效利用深度学习网络解决高分遥感图像语义分割的问题提出了一个较优的新解决方案。本文实验中不仅通过取值的比对实验找到了更优的综合权重α，且证明了SWA的优化效果；在本文实验数据集下，本文网络在包括accuracy及mIoU在内的六种指标的评估下取得了比较出色的成绩，其中mIoU分数高达0.84，明显超越了交叉熵的0.79与Lovasz的0.81；f1-score达到0.91，同样优于交叉熵的0.87与Lovasz的0.89；而最后的消融实验证明了模型各个模块均起到了有效的优化效果，效率对比实验也体现出了整体架构的高效及设计合理性。综上，最终实验也确实证明了本文网络模型及损失超越了一些已有的经典网络设计，取得了相对优越的综合成绩，为高分遥感图像的语义分割问题提出了新的解决方案。

本文设计的网络及损失函数虽然使得最终图像分割效果有较好的实验结果，在细节方面也表现出了相对优秀的处理识别能力，但是还是存在一些问题及需要解决的难点。首先在预测图中依然出现少量错点，而这种错误预测的像素在连续图形中显得比较突兀，即便错误像素很少也会影响整体预测效果；其次针对于类似建筑物缝隙的细小物体识别问题，虽然已经是本文实验表现最佳的模型，但依然出现分割不清等问题。故本文网络仍存在很大进步空间，无论是在结构上还是损失函数上都可以进一步优化。在未来的研究中也会侧重于这方面学习，努力为图像语义分割领域提出更优秀的解决方案。

**参考文献(References)**

Berman M , Triki A R and Blaschko M B . 2018. The Lov'asz-Softmax Loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks// 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE: 2291-2230 [ DOI: 10.1109/CVPR.2018.00464]

He K , Zhang X , Ren S and et al. 2016. Deep Residual Learning for Image Recognition// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE: 770-778 [ DOI: 10.1109/CVPR.2016.90]

He K , Zhang X , Ren S and et al. 2015. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification// 2015 IEEE International Conference on Computer Vision (ICCV). Chile: IEEE: 1026-1034 [ DOI: 10.1109/ICCV.2015.123]

Izmailov P, Podoprikhin D, Garipov T and et al. 2018. Averaging Weights Leads to Wider Optima and Better Generalization// Uncertainty in Artificial Intelligence (UAI). Montery: IEEE: 876-885 [ DOI: 10.1.1.370.2913]

Ma H L, Wang Z L and Li H J. 2019. 2010-2016 Urban Spatial Expansion Research in Gaoling District Based on Multi-source Remote Sensing Images and Geographical National Conditions Monitoring Technology. Surveying and mapping technical equipment, 2019(2):20-24 (马红利,王祖亮,李宏建.2019.基于多源遥感影像和地理国情监测技术的高陵区2010-2016年城镇空间扩展研究. 测绘技术装备, 2019(2):20-24) [ DOI：10.3969/j.issn.1674-4950.2019.02.006]

Mishkin D, Sergievskiy N, Matas J. 2017. Systematic evaluation of convolution neural network advances on the Imagenet. Computer vision and image understanding, 161(aug.):11-19. [ DOI: 10.1016/j.cviu.2017.05.007]

Ronneberger, Olaf, Fischer, Philipp, Brox and Thomas. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. Lecture Notes in Computer Science, 11(18): 234-241. [ DOI: 10.1007/978-3-319-24574-4\_28]

Rakhlin A, Davydow A and Nikolenko S. 2018. Land Cover Classification from Satellite Imagery with U-Net and Lovász-Softmax Loss// 2018 CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE: 262-266 [ DOI: 10.1109/CVPRW.2018.00048]

Roy A G, Navab N and Wachinger C. 2018. Concurrent Spatial and Channel Squeeze & Excitation in Fully Convolutional Networks// 2018 International Conference on Medical Image Computing and Computer-Assisted Intervention. Granada: IEEE: 421-429 [ DOI: 10.1007/978-3-030-00928-1\_48]

Su J M, Yang L X, Jing W P. 2019. U-Net-based semantic segmentation method for high-resolution remote sensing images. Computer Engineering and Application, 55(07):207-213(苏健民,杨岚心,景维鹏. 2019.基于U-Net的高分辨率遥感图像语义分割方法计算机工程与应用, 55(07):207-213). [ DOI: 10.3778/j.issn.1002-8331.1806-0024]

Trajdos P and Kurzynski M. 2019.An approximated decision-theoretic algorithm for minimization of the Tversky Loss under the multi-label framework. Pattern Analysis and Applications, 22(2): 389-416) [ DOI: 10.1007/S10044-017-0651-6]

Zhou, Zongwei and et al. 2020. UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation//IEEE Transactions on Medical Imaging. US: IEEE: 1856–1867 [ DOI: 10.1109/TMI.2019.2959609]

**作者简介**



刘澄，1998年生，男，本科生，研究方向为深度学习、图像处理与计算机视觉。

E-mail:blue\_fairy\_star@126.com



万晓华，1980年生，女，副研究员，主要研究方向为图像处理、高性能计算。

E-mail: wanxiaohua@ict.ac.cn

杨子豪，男，硕士生，研究方向为遥感图像处理、深度学习。E-mail: yangzihao19S@ict.ac.cn

邱德慧，女，副教授，主要研究方向：深度学习加速器设计、FPGA嵌入式系统设计。

Email: qiudehui@cnu.edu.cn