智能终端软件开发技术课程大作业

地表建筑物识别

一、项目说明

本组项目的具体实践案例选自Datawhale与天池联合发起的入门系列赛事 —— 语义分割之地表建筑物识别。

<https://tianchi.aliyun.com/competition/entrance/531872>

小组成员：陈瑞，梁雅琪，朱庭磊，邓卓，吴嘉麒，章程，杨鑫，何东旭，陈海彬，李嘉怡，单培民，郑语

二、项目背景

遥感技术已成为获取地表覆盖信息最为行之有效的手段，遥感技术已经成功应用于地表覆盖检测、植被面积检测和建筑物检测任务。以计算机视觉为背景，使用给定的航拍图像训练模型并完成地表建筑物识别任务，将地表航拍图像素划分为有建筑物和无建筑物两类。

如下图，左边为原始航拍图，右边为对应的建筑物标注。



三、实验设备

kaggle平台

四、实验原理

从题目可以很明显的得出，这是一个语义划分的二分类任务，因此采用常见的语义分割模型FCN模型、U-net模型，U-net++模型对图像进行语义分割，然后选取最佳模型。

五、实验过程

1、数据分析与数据预处理

1.1数据分析

在一些图像处理任务中，对像素点数值进行预处理可以提高后续处理步骤的效果。通过统计占比，可以确定一些特定数值的阈值或区间，并对图像进行二值化、阈值化或其他处理，以突出感兴趣的区域或去除噪声。这里我们随机选取了一个image中的图片进行其像素点的统计，并进行其归一化。由此我们能够统计其均值和标准差。



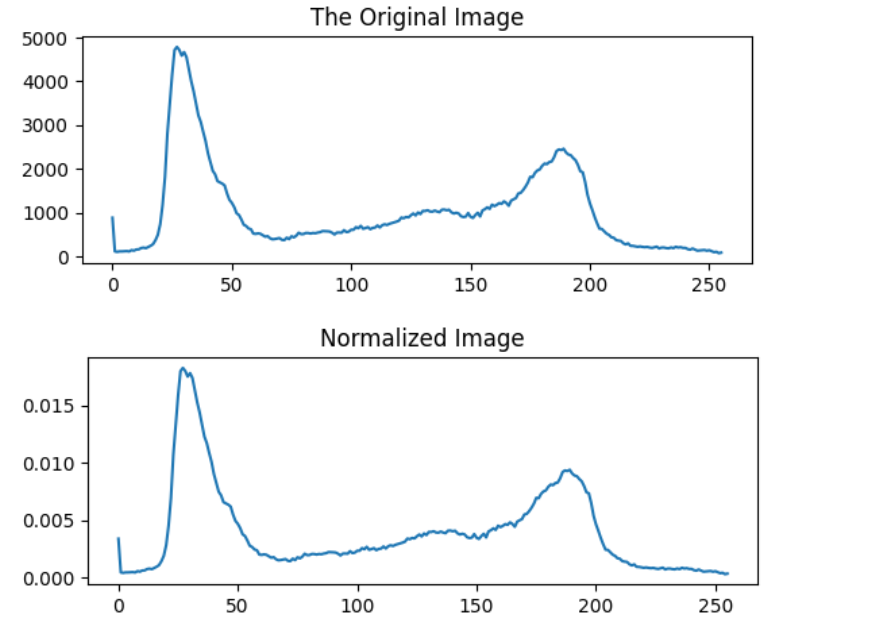


图1 图像数值统计和归一化后的图像数值统计

可见数据集中在0-50与150-200之间。

1.2数据增强

在数据预处理方向，我们选择了用进行数据增强进行的数据预处理。数据增强旨在通过图像增强技术提高图像数据集的多样性和鲁棒性，以增强训练模型的泛化能力。

这里我们选择的数据增强是对训练集的随机选取1000张image和mask一一对应进行随机水平翻转和垂直翻转操作，这个操作可以改变图像的方向和视角，从而增加了数据集的变化程度。这对于训练模型来说是有益的，因为模型可以学习到更多不同的图像特征和模式。

①image



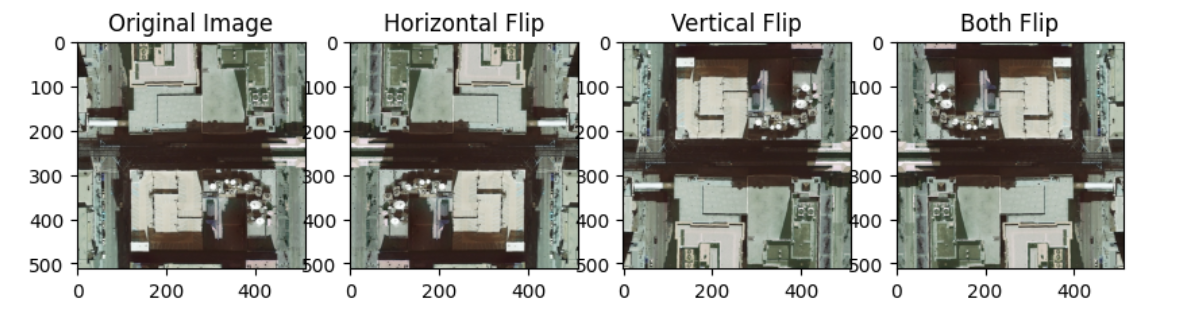


图2 image数据增强

从左往右分别为原始图，水平翻转，竖直翻转，水平竖直翻转。

②mask



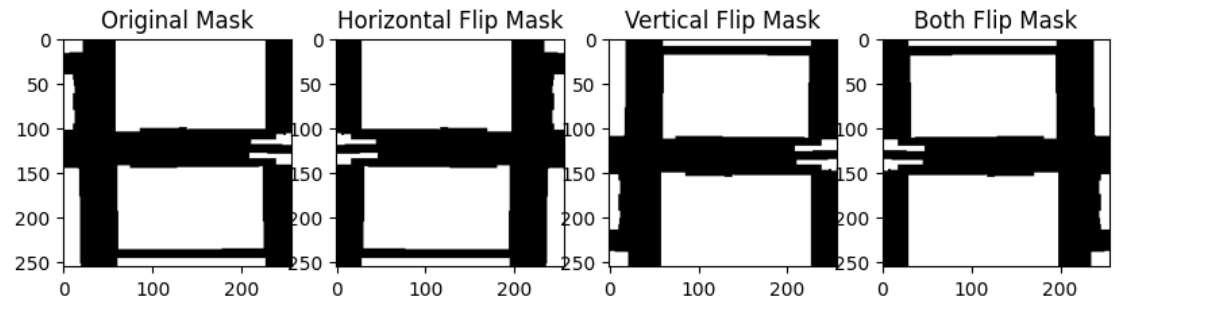


图3 mask数据增强

1.3图像处理

锐化和高斯模糊是图像处理中常用的滤波操作。其中锐化操作可以增强图像的边缘和细节，使其更加清晰和锐利。在图像分割中，锐化可以帮助凸显图像中的边界和重要特征，有助于更好地分离目标区域和背景。通过增强图像的边缘和细节信息，锐化操作可以提高分割算法对细微结构和形态特征的感知能力，从而改善分割结果的准确性。高斯模糊是一种平滑图像的操作，它可以模糊图像中的细节和噪声。在图像分割中，高斯模糊常用于去除图像中的噪声和平滑不规则的纹理。通过降低图像的高频细节信息，高斯模糊有助于消除图像中的噪声干扰，并使分割算法更加稳定和鲁棒。此外，高斯模糊还可以帮助平滑不连续的边界，减少分割结果中的伪影和碎片。



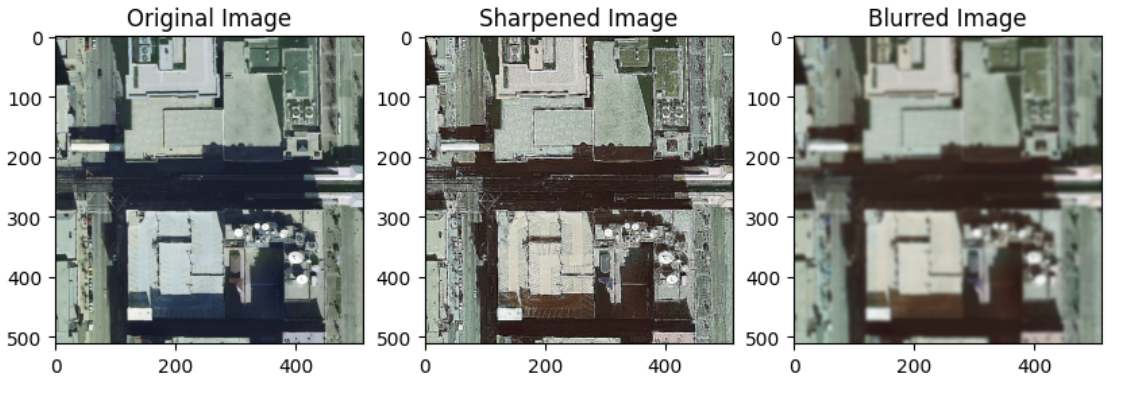


图4 原图锐化图高斯图对比

我们使用锐化和高斯模糊后，图片细节的对比。



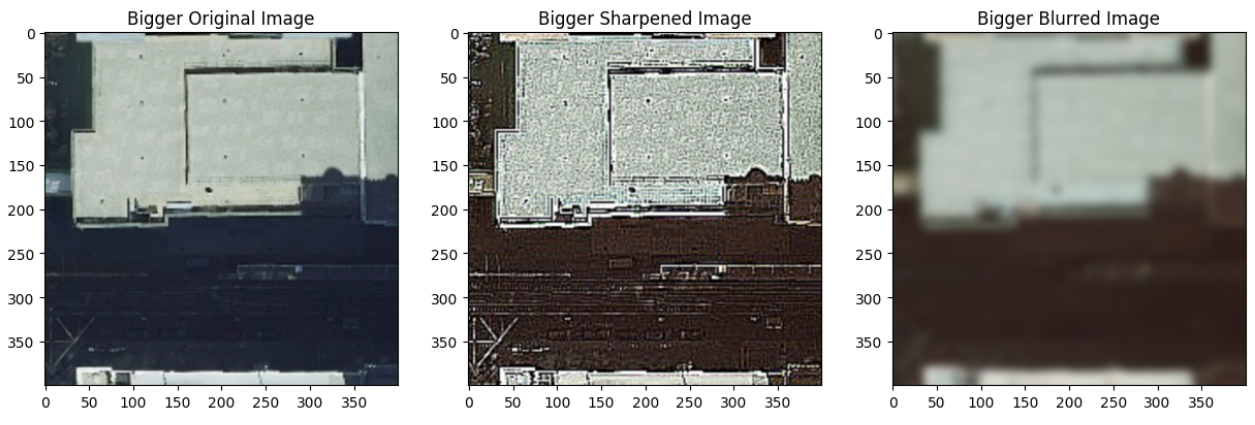


图5 原图锐化高斯图边缘对比

1.4数据集构建与加载

采用的torch深度学习框架，所以通过设置类继承Dataset来构建数据集，传入图片的路径，和缩放大小scale，同时传入transform对图片进行处理，这里我只对图片进行维度转换，将HWC转换为CHW，同时对图片做归一化处理，最后处理完以字典的形式返回。对于测试集我们不传入mask的路径，将其替换为图片的路径，方便预测时保存图片。

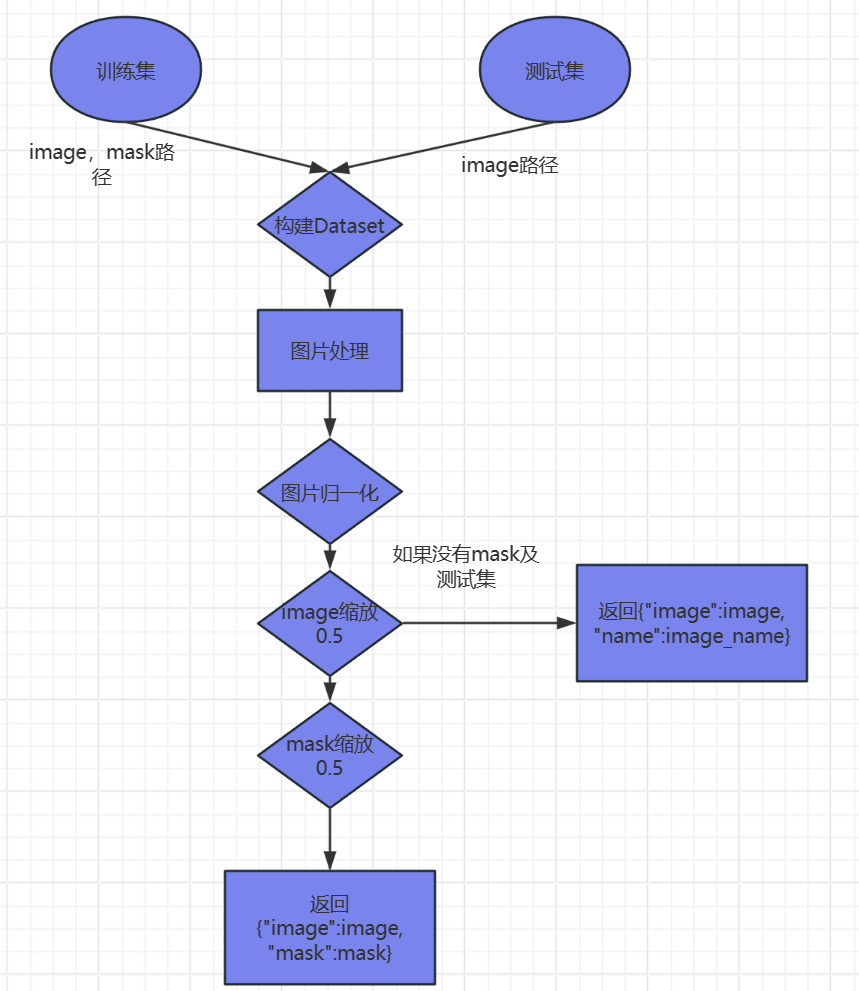


图6 数据集的构建

在完成数据集的构建后我们对其进行划分选择70%的数据作为训练集，30%的作为验证集，最后通过DataLoad加载数据。

**2、模型构建与实现**

2.1U-net模型的构建

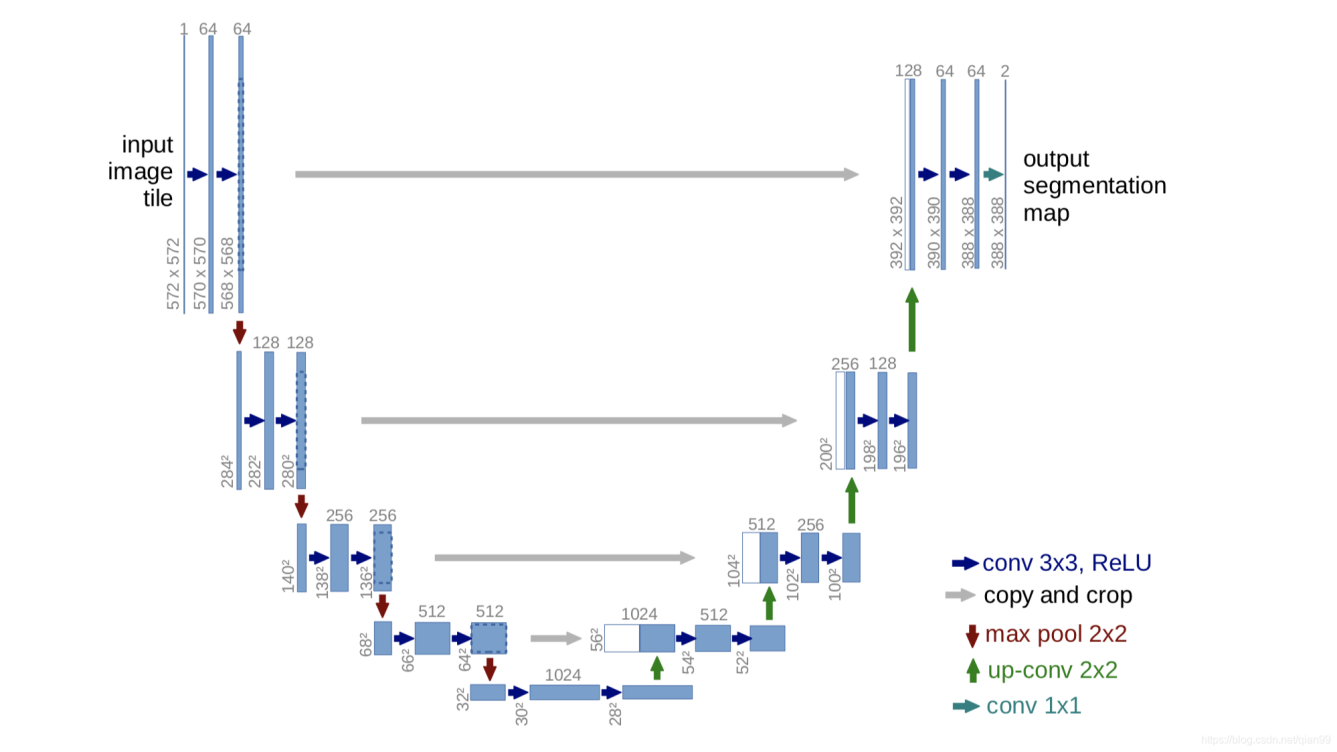


图7 U-net模型结构结构图

如上图，U-net 网络结构是对称的，形似英文字母 U 所以被称为 U-net。整张图都是由蓝/白色框与各种颜色的箭头组成，其中，蓝/白色框表示 feature map；蓝色箭头表示 3x3 卷积，用于特征提取；灰色箭头表示 skip-connection，用于特征融合；红色箭头表示池化 pooling，用于降低维度；绿色箭头表示上采样 upsample，用于恢复维度；青色箭头表示 1x1 卷积，用于输出结果。

Unet网络的结构比较简单，左侧分支每一层包含两个重复的卷积，我们命名为DoubleConv，从第二层开始，都是max\_pool+DoubleConv；右侧分支每一层都是up\_conv+copy\_crop+DoubleConv，在最后输出层，有一个1x1 conv。所以，我们可以将以上网络简化为4个模块，分别是：

①输入层的DoubleConv模块；

②左侧分支从第二层开始的max\_pool+DoubleConv，称为Down模块；

③右侧分支的up\_conv+copy\_crop+DoubleConv，称为Up模块；

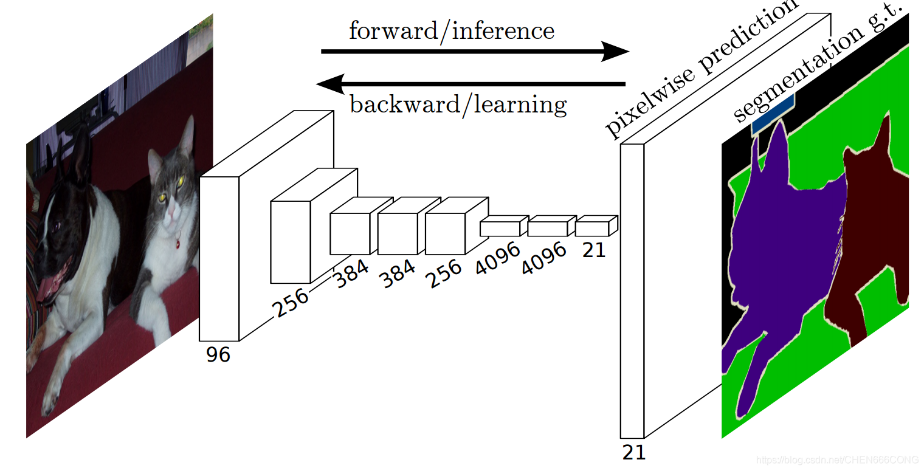
首先经过一个上采样或转置卷积，然后从左侧路径的同一层feature map中截取相同的size（从图中很容易可以看出，左侧同一层中的feature map比右侧的size要大一些），与右侧feature map合并，最后再进行DoubleConv。

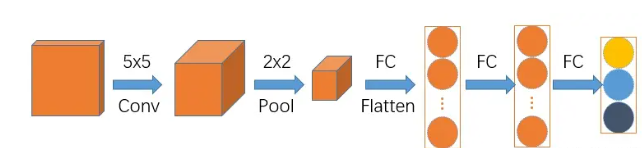
④输出层的1x1卷积，称为OutConv模块。

U-Net网络能从极少的训练图像中，依靠数据增强将有效的标注数据更为有效地使用。U-Net与其他常见的分割网络（如FCN）有一点非常不同的地方：U-net采用了完全不同的特征融合方式：拼接（Concat），U-Net采用将特征在channel维度拼接在一起，形成更厚的特征。

2.2 FCN50模型构建

FCN对图像进行像素级的分类，从而解决了语义级别的图像分割（semantic segmentation）问题。与经典的CNN在卷积层之后使用全连接层得到固定长度的特征向量进行分类（全联接层＋softmax输出）不同，FCN将CNN中的全连接层换成了卷积操作。换成全卷积操作后，由于没有了全连接层的输入层神经元个数的限制，所以卷积层的输入可以接受不同尺寸的图像。





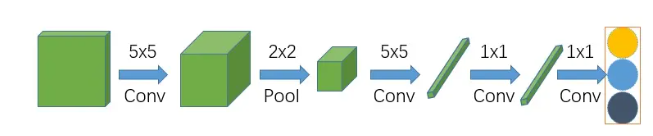


图8 FCN模型

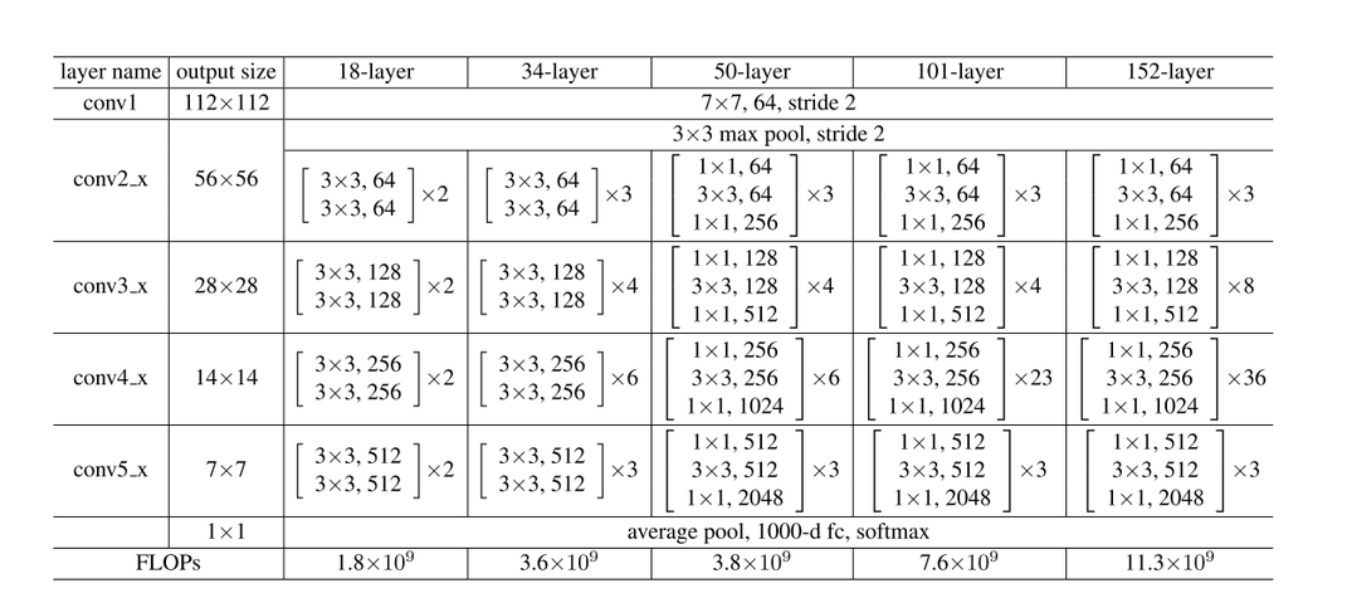


图9 FCN50模型结构图

fcn-50是一种深度残差网络，它的主要特点是通过引入残差块来解决深度网络训练中的梯度消失和梯度爆炸问题。

fcn-50的模型结构组成：

①卷积层和池化层（前处理）： 最初的卷积层和池化层用于对输入图像进行初步处理。

②残差块（Residual Block）： ResNet-50使用了多个残差块，每个残差块由两个分支组成，一个是恒等映射，另一个是包含卷积层的主分支。这种结构使得网络可以学习残差，从而更容易训练非常深的网络。每个残差块可以有不同的配置，但通常包括两个3x3的卷积层。

③网络深度： ResNet-50的网络深度较大，包括多个残差块。具体地，包含3个残差块的stage1，4个残差块的stage2，6个残差块的stage3，和3个残差块的stage4。

④全局平均池化层： 在网络的最后，全局平均池化层用于将最后一个残差块的输出特征图转换成固定大小的向量。

⑤全连接层： 最后是一个全连接层，用于将全局平均池化的输出映射到类别分数。

2.3 U-net++模型构建

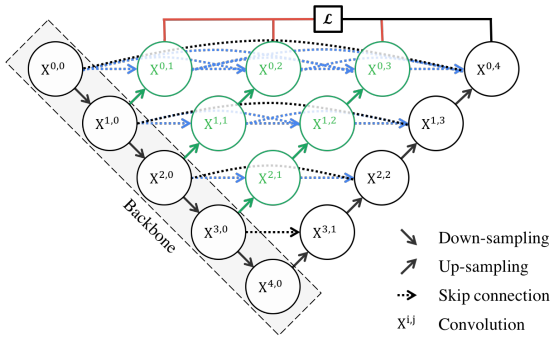


图10 U-net++结构图

U-net++是一个用于图像分割的深度神经网络，它在U-net的基础上增加了多尺度的特征融合和深度监督的机制，以提高分割的精度和鲁棒性。

U-net++的主要改进有以下几点：

①它使用了密集的跳跃连接，将编码器和解码器中不同深度的特征图进行拼接，从而利用多尺度的信息。

②它使用了全尺度的跳跃连接，将编码器中每一层的特征图直接连接到最终的输出层，从而减少信息损失。

③它使用了深度监督，在解码器中每一层都添加了一个辅助输出层，从而增强梯度反向传播和网络训练。

U-net++相比于U-net，在多个数据集上都取得了更好的分割性能，同时使用了更少的参数。它也可以灵活地使用不同的编码器结构作为特征提取器。

2.4模型优化

（1）调参

采用自适应学习率调整和余弦退火策略，初始学习率的设置范围为[1e-2,1e-6]，权重衰退设置为1e-3，较大的batch\_size可使模型更快的收敛以及稳定的参数更新，最终batch\_size设置为[40-50]， epoch设置为[20-30]。

（2）优化器

在优化器上，在我们的尝试下，最终选择使用adam效果最佳

Adam优化器：Adam是一种结合了动量和自适应学习率的优化算法。它基于梯度的一阶矩估计和二阶矩估计来调整每个参数的学习率，适用于大多数深度学习任务，包括语义分割。  
 RMSProp优化器：RMSProp也是一种自适应学习率的优化算法，它通过平均梯度的平方根来调整参数的学习率。RMSProp在处理非平稳目标函数或具有大幅度不同尺度的参数时表现良好。  
 SGD优化器：随机梯度下降是最基本的优化算法之一，在语义分割中仍然被广泛使用。它通过不断迭代更新参数以最小化损失函数。  
 AdaGrad优化器：AdaGrad是一种自适应学习率的优化算法，它根据每个参数的历史梯度信息来调整学习率，适用于处理稀疏数据。  
 AdaDelta优化器：AdaDelta是一种自适应学习率的优化算法，它类似于AdaGrad但添加了对学习率的自适应调整，可以减少AdaGrad中学习率不断减小导致的问题。

（3）选取最佳模型

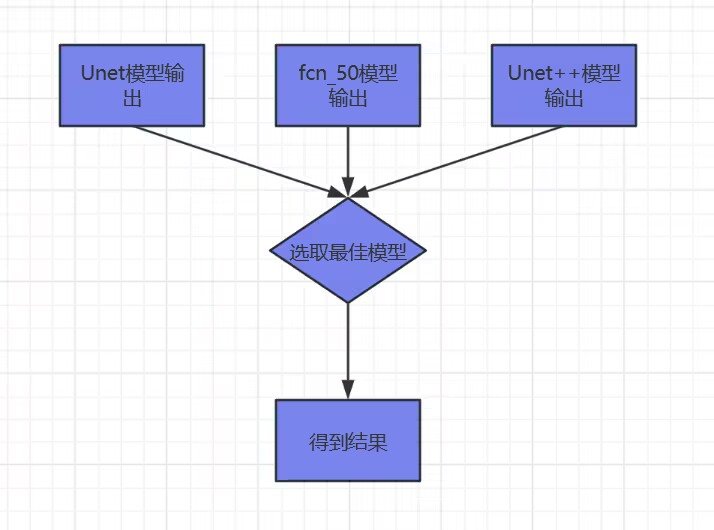


图11 最佳模型

六、心得体会

近年来，深度学习在图像分析领域取得了显著的成就。本组在使用PyTorch框架中的UNet、FCN8和Unet++模型对牙齿图像进行语义分割的实践中，积累了一些有关模型性能和使用经验的心得体会。

UNet模型在图像分割任务中表现出色。UNet架构具有对称的编码器-解码器结构，可以有效地提取图像特征并重建细节。通过跳跃连接，在解码器中的每个阶段都与对应编码器层级的特征进行连接，使得模型可以利用不同层级的信息，从而提高分割结果的准确性和细节保留能力。我们发现，UNet模型对于图像中的边缘信息和细微结构的分割效果较好。

FCN50模型在图像分割任务中也表现出了很好的性能。FCN将全连接层替换为全卷积网络，使其可以接受任意大小的输入图像，并输出相应大小的分割图。FCN50模型使用了VGG16作为基础网络，并在不同分辨率上生成了三个分割结果，将这些结果进行上采样和叠加，最终得到最终的分割结果。我们发现，FCN50模型在牙齿图像的整体分割上表现较好，能够捕捉到牙齿的整体形状和分布情况。

Unet++模型是对UNet模型的改进和扩展，它通过增加更多的跳跃连接来进一步提升分割精度。Unet++模型在编码器中使用了多尺度的特征融合方法，将不同层级的特征进行融合，提高了模型对不同尺度和细节的感知能力。在对图像进行语义分割时，Unet++模型展现了更好的性能和分割结果，尤其是对于一些细小的牙齿结构和边缘。

在实践过程中，我们还发现了一些重要的技巧和注意事项。首先，对于深度学习模型的训练，需要充分准备数据集，包括标注准确的牙齿分割图和对应的原始图像。同时，数据增强技术如随机翻转、随机旋转和弹性变形等可以帮助提高模型的鲁棒性和泛化能力。其次，使用适当的损失函数对模型进行训练非常重要。交叉熵损失函数通常可以满足大多数分割任务的要求，但对于类别不平衡的情况，可以考虑使用加权交叉熵损失函数或Dice系数损失函数。最后，模型的调参和优化也是关键步骤，学习率的选择、优化器的选用以及学习率调度的设置都需要仔细考虑和调整。

使用UNet、FCN50和Unet++模型对图像进行语义分割是一项有挑战性但有意义的任务。这些模型在不同方面都展现了很好的性能，可以根据实际需求和数据集特点选择适合的模型。同时，通过合理的数据预处理、损失函数选择以及模型的训练和调优，我们可以进一步提高分割结果的准确性和稳定性。深度学习在图像分析中的应用潜力巨大。

对UNet、FCN50和Unet++这三个模型进行模型融合是一个有效的策略，可以进一步提高语义分割的准确性和鲁棒性。通过将不同模型的预测结果进行融合，可以弥补单个模型的不足，并取得更好的整体性能。