AI2611 机器学习

大作业　基于 3DUNet 的肋骨骨折数据集检测分割

王怡闻 519030910367 曹嘉航 519030910347

June 15, 2021

1　完成情况

我们尝试性改进用于医学图像识别的 UNet，使用**深监督误差**与丰富的**数据增强操作**增强训练效果，添加**残差块结构**增强拟合效果，使用 **TverskyLoss** 综合 CE 和 DICE 指标作为损失函数。在肋骨骨折数据集上进行训练和验证，达到了验证指标的基本要求。训练及测试中比较有意义的指标见下表：

指标	FP=0.5	FP=1	FP=2	FP=4	FP=8	average	maximum
FROC	0.040396	0.080792	0.161584	0.260369	0.393646	0.1874	0.6207

2　问题描述

2.1　研究背景

随着人民健康意识的提升和数据化医疗设备的普及，CT 检查越来越普遍地应用于各种疾病的诊疗中，但这同时带来了医务工作者负担过重，难以区分需求急切的患者等问题。若基于较多的医学影像数据构建一智能分析系统，**快速筛选**出急需诊疗的骨折伤者，**标注出骨折位置**，可以有效提高医疗设施的运转效率，为医生与患者同时减负。

肋骨骨折位置的标注，将是未来人工智能技术在医学领域的突破性应用之一。

2.2　肋骨骨折数据集

本实验使用的数据集来源于`https://ribfrac.grand-challenge.org/`，包含一系列具有专家标注的肋骨骨折 CT **三维图像**。其中训练集 420 张，验证集 80 张，测试集（无标注）180 张。

数据与标注使用标准 NIFTI 图像的 gzip 压缩（.nii.gz）保存，NIFTI 图像在 ANALYZE 格式二进制图像资料的基础上增加了患者的方位信息，且更容易使用标准压缩软件压缩。

2.3 输出目标

模型在测试集上测试得出的预测标签将保存为.nii.gz 格式。医学影像的临床应用对精度与稳定性要求严格，模型的验证与测试效果将使用 FROC 进行评价。

3 模型设计及技巧

3.1 数据预处理：增强操作

尽管医疗图像数据点众多，本数据集中文件的数目尚且较少，实际应用中更难规避随机性的影响。因此要对图像进行带随机性的变换，使训练数据广泛性增强。我们参考 [4] 对训练集输入图像顺次进行如下变换，达到增强的效果：

1. 在高度尺度上，将图像无正类标签的区域去除
2. 将图像附加 0 为均值，0.1 为方差的高斯噪声
3. 将数据值到-200 至 1000 范围，即去除极端点
4. 将数据点归一化到 (-1,1) 范围
5. 将图像在各维度上随机缩放 0.8 至 1.2 比例
6. 以 0.5 的概率将图像左右翻转
7. 以 0.5 的概率将图像上下翻转

通过对数据的观察，骨折区域一般处于临近高度的不同位置，截取图像有利于特征提取，这与3.2中的随机采样有关，将在彼再次分析；增加噪声的操作有助于增强模型对模糊化噪声的适应能力；归一化操作增强模型对不同亮度条件的适应能力，使模型更关注形状特征；由于骨折处在不同方位，不同角度下特征相对相似，规模略大的骨折和略小的骨折相对相似，随机缩放与翻转相当于增加数据规模而不失一般性。

对图像进行缩放与翻转时，对标签也进行相同操作，而对于其它处理，标签不做变换，保证数据正确。

3.2 特征提取：感兴趣区域正负类

而在本问题中，二维图像的传统提取方式面对着如下三个主要困难：

1. 三维图像数据规模随着图像单向度尺寸的增加，指数爆炸效应极为明显
2. 骨折区域所占体素比例过小
3. 整张图像信息过于相似

可以预计若对整张图像直接进行特征提取效果较差，因此我们借鉴 RCNN 等算法，根据给定标签提取感兴趣区域（ROI）作为模型输入。ROI 为拥有 $64 \times 64 \times 64$ 体素大小的区域，相当于对原图像的小窗口采样。我们使用图形学算法提取标签中正类标签集中的区域，即每个被标注的骨折之

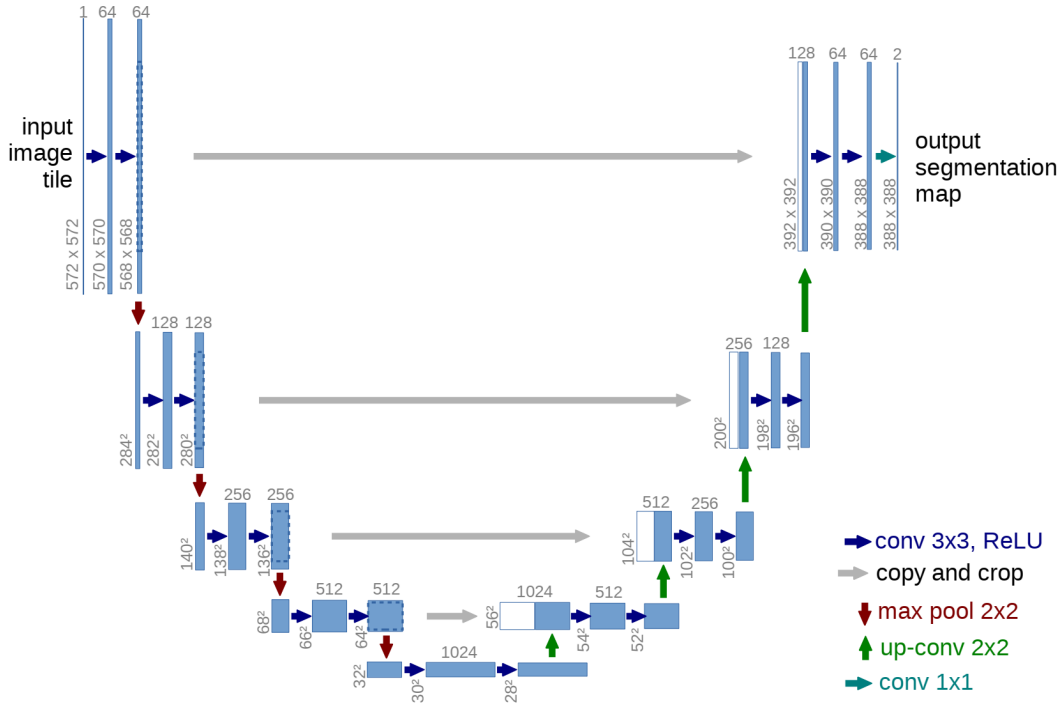


Figure 1: UNet 的输入输出结构

处，以其为中心得出 ROI 区域为正类 ROI，为输入网络的图像，标签为该区域在标签上的对应位置，数据中由专家标注的多处骨折处统一记为标签 1。

本问题是骨折-非骨折的二分类问题，而如上提取的 ROI 标签几乎必定绝大多数为正类体素，输入信息存在很大特异性，网络无法很好区分正负类。因此我们同时使用如下思路提取负类 ROI 区域：

一般来说，骨骼结构左右对称，而骨折位置不会左右对称，因此提取与正类 ROI 左右对称的区域为负类 ROI，这样的输入数据受骨骼位置的不同影响小，可以使网络精准提取骨折特征。为了进一步增大模型泛化性，我们同时在距图像边界各 32 单位的体素网格中随机取样与正类 ROI 相同数目的中心点，构建类似的 ROI，希望使图像边界信息、标签交叉区域信息被充分提取。

这样取得的 ROI 数量随着原始标注个数的不同而不同，直接输入网络将会导致计算资源占用不稳定，因此最终网络的输入为从所有 ROI 中提取的固定量样本。在考察原始标注密度后，采样数目定为每张图像 16 个 ROI。

3.3 网络架构：3D-ResUNet

UNet 是一类为医学图像处理设计的网络结构，有着兼顾网络深层信息与浅层信息的特点。[5] 3D-UNet 由二维版本的 UNet 迁移而来，而 3D-ResUNet 为 3D-UNet 添加残差块结构的模型。

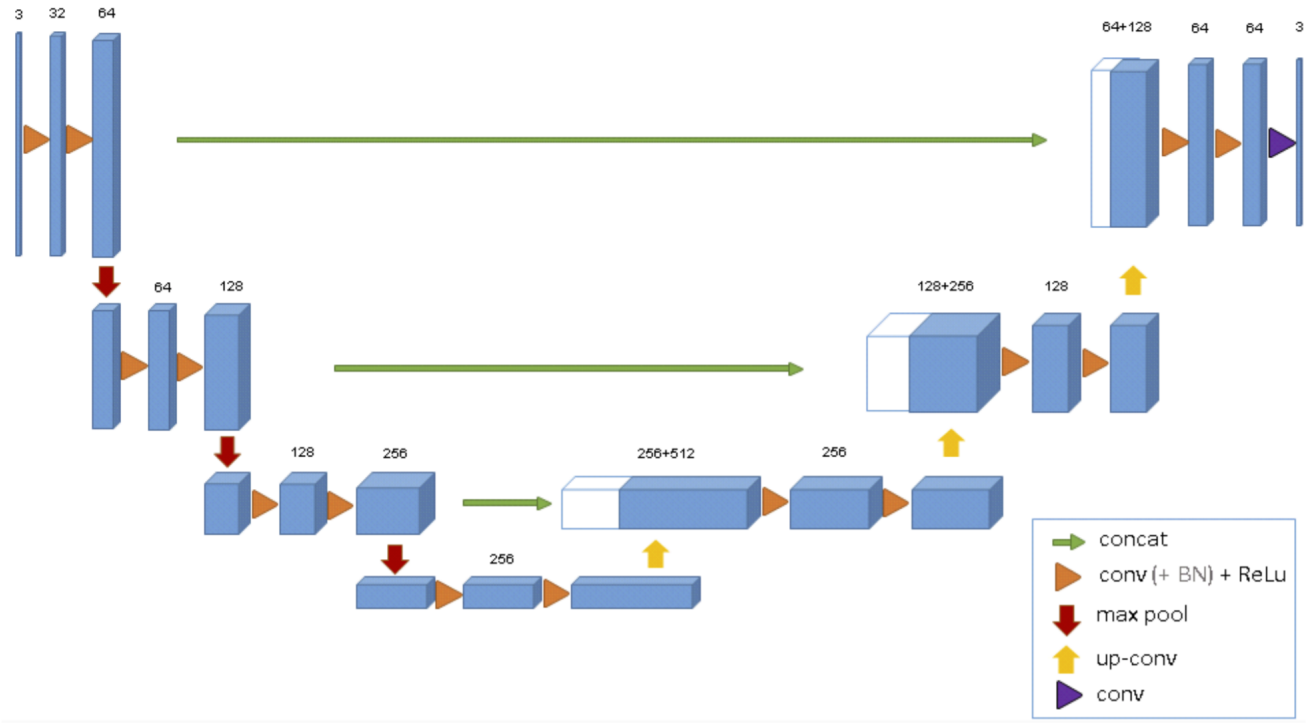
3.3.1 UNet

编码部分 Encoder：左半部分，由两个 3x3 的卷积层连接 2x2 的 max polling（stride=2）层不断级联组成，每经过一次下采样，通道数翻倍；
解码部分 Decoder：右半部分，由一个 2x2 的上采样卷积层 +Concatenation（crop[3] 对应的 Encoder 层的输出 feature map 然后与 Decoder 层的上采样结果相加）+2 个 3x3 的卷积层（ReLU）反复构成；最后一层通过一个 1x1 卷积将通道数变成期望的类别数。
卷积层的激活函数都使用 PReLU，考虑到网络较深，信息较多，使用该激活可以缓解过拟合与梯度消失。
数据维度变化：UNet 的 encoder 下采样 4 次，共下采样 16 倍，对称地，其 decoder 也相应上采样 4 次，将 encoder 得到的高级语义特征图恢复到原图片的分辨率。相比于其拥有类似编码-解码结构的前身：FCN 和 Deeplab，UNet 共进行了 4 次上采样，并在同一个 stage 使用了 skip connection，而不是直接在高级语义特征上进行监督和 loss 反传，这样就保证了最后恢复出来的特征图融合了更多的 low-level 的 feature，也使得不同 scale 的 feature 得到了的融合，从而可以进行多尺度预测和 DeepSupervision。4 次上采样也使得分割图恢复边缘等信息更加精细。
为什么适合于医学图像？医学图像有如下几个特点：

1. 图像语义较为简单、结构较为固定。
2. 数据量少。医学影像的数据获取相对难，很多比赛只提供不到 100 例数据。所以设计的模型不宜多大，参数过多，很容易导致过拟合。
3. 多模态。相比自然影像，医疗影像是具有多种模态的。以 ISLES 脑梗竞赛为例，其官方提供了 CBF,MTT,CBV,TMAX,CTP 等多种模态的数据。
4. 可解释性重要。医疗图像有大量的医学理论作为支持，对特征信息的充分了解有助于模型调试与实际诊疗。

UNet 利用了底层的特征（同分辨率级联）改善上采样的信息不足，融合了不同尺度的特征，同时 skip-connection 保证上采样恢复出来的特征不会很粗糙。
下面来介绍 UNet-3D[1]。

3.3.2 3D-UNet



输入图像先通过编码路径，再通过解码路径，每一条都有 4 个分辨率级别，编码路径的到的信息将作为新通道拼接（concat）到解码路径信息中。

编码路径每一层包含两个 3×3×3 卷积，每一个都后接一个 ReLU 层，然后是一个 2×2×2 的每个方向上步长都为 2 的最大池化层。

在解码路径，每一层包含一个步长为 2 的 2×2×2 的反卷积层，紧跟两个 3×3×3 的卷积层，每一个都后接一个 RuLU 层。

通过 shortcut，将编码路径中相同分辨率的层传递到解码路径，为其提供原始的高分辨率特征。

最后一层为 1×1×1 的卷积层，可以减少输出的通道数，最后的输出通道数为标签的类别数量。

网络的输入为 3 通道的 132×132×116 的像素集合。输出的大小为 44×44×28。在 ReLU 之前使用了 batch normalization。

一个重要的部分是加权 softmax 损失函数，使得网络可以使用稀疏注释的数据进行训练。

将未标记的像素的权重设置为零，使得网络可以仅从有标记的像素中学习，并推广到整个立体数据。

UNet 的改进版 ResUNet[2] 结构启发了我们对网络进行改进。

3.3.3 ResUNet

我们从 [小作业] 中获得启示，将最初的 UNet 中编码块和解码块中的卷积层替换成了残差块。残差网络很好地解决了深度神经网络的退化问题，并在 ImageNet 和 CIFAR-10 等图像任务上取得了非常好的结果，同等层数的前提下残差网络也收敛得更快。这使得前馈神经网络可以采用更深的设计。除此之外，去除个别神经网络层，残差网络的表现不会受到显著影响，这与传统的前馈神经网

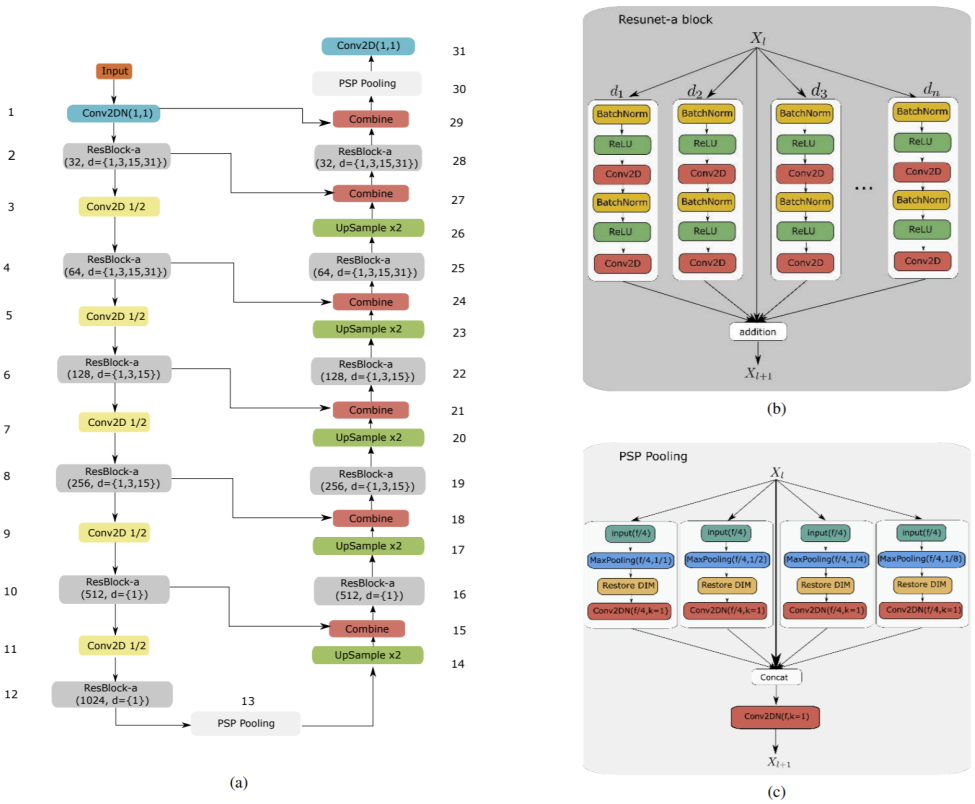


Figure 2: ResUNet 在 [2] 中的结构，我们去除了相对本问题冗余的部分

络大相径庭。

3.4 损失函数

由于网络各层输出维度变化过大，这里选择基于 DICELoss 的 TverskyLoss 进行反向传播。

3.4.1 Dice 系数与 Dice Loss

Dice 系数是一种集合相似度度量函数，通常用于计算两个样本的相似度，取值范围在 [0,1]：

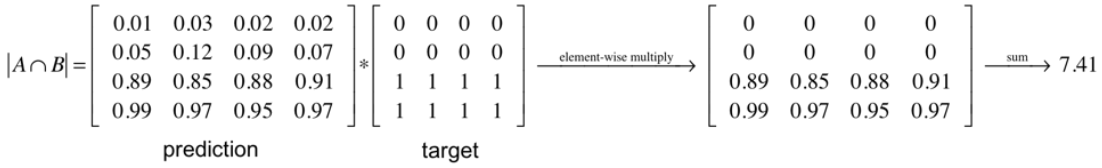
$$s = \frac{2|X \bigcap Y|}{|X| + |Y|} \tag{1}$$

其中 $|X \bigcap Y|$ 是 X 和 Y 之间的交集，|X| 和 |Y| 分表表示 X 和 Y 的元素的个数，其中，分子的系数为 2，是因为分母存在重复计算 X 和 Y 之间的共同元素的原因。

Dice 系数差异函数 Dice Loss：

$$s = 1 - \frac{2|X \bigcap Y|}{|X| + |Y|} \tag{2}$$

以下是一个 DICE 损失函数计算的例子：



3.4.2 UNet

3.4.3 TverskyLoss

在 [6] 中提出了一种基于 Tversky 指数的广义损失函数，以解决数据不平衡的问题，在精度和召回率之间找到更好的平衡。在磁共振图像上进行多发性硬化病灶分割的实验结果显示，测试数据中的 F2 评分，Dice 系数和精确召回曲线下的面积均得到改善。指出的测试和训练速度很快，并使用整个样本学习局部和全局图像特征。

$$T(A,B) = \frac{|A \cap B|}{|A \cap B| + \alpha|A - B| + \beta|B - A|} \tag{3}$$

|A-B| 意味着是假阳性（FP），而 |B-A| 意味着是假阴性（FN）：α 和 β 分别控制 FP 和 FN，调整两者之间的权衡。

3.4.4 深监督学习

受模型兼顾深层信息与浅层信息的启发，我们尝试在训练上也二者兼顾，即使用深度监督技巧：在训练时对解码路径上每个模块的输出上采样到输出数据大小，分别与标签计算损失函数，若第一、第二和第三解码模块和输出对应的损失函数分别为 $\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3, \mathcal{L}_4$，则最终参与反向传播的损失函数为

$$\mathcal{L} = \mathcal{L}_4 + \alpha(\mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3)$$

其中 α 为深度监督系数，随着训练轮次的增大，为保证模型稳定性，深层信息的影响应逐渐衰减。本实验中，α 初始为 0.4，每 30 轮变为原来的 80%。

3.5 模型输出

3D-ResUNet 的模型输入为3.2步骤提取的正负类 ROI，标签为这些 ROI 对应位置的标签，输出为与 ROI 同样大小的预测值。

3.6 评价指标：FROC

由 [3] 提出，经典的 ROC 方法不能解决对一幅图像上多个异常进行评价的实际问题，70 年代提出了无限制 ROC 的概念 (free-response ROC;FROC)。FROC 允许对每幅图像上的任意异常进行评价。

本实验中验证集上 FROC 的分数为 FP=0.5,1,2,4,8 的 FROC 的均值。

3.7 验证函数

首先，我们先创建一个新的模型，模型参数和结构与之前在训练集上训练的模型保持一致，然后将在训练过程中验证集上表现最佳的的模型参数（保存在 best_model.pt）载入这个模型。
以 ROI 大小 $64 \times 64 \times 64$ 为滑动窗口大小，以 32 为三个维度的步长，扫描测试集的图像，得到符合模型输入大小的数据作为网络输入。将模型输出经过 Sigmoid 函数，防止原输出结果再 $[0, 1]$ 范围外造成的混乱，得到该区域各体素骨折可能性的打分值。

这样处理的区域之间存在重叠，对于重叠区域将与该区域原有的打分值取均值，最终表现为区域打分为包含此区域的滑动窗口对此区域打分的均值。滑动窗口遍历整组数据得出与该图像同样大小的预测。默认对预测值进行如下后处理（由 post-processing 参数控制）：

1. 去除掉一些预测概率低于 0.64 的区域；
2. 标记高亮度部分为骨骼，去除在垂直尺度上亮度和最大的连通区域，即脊柱；
3. 将预测中体积比较小的物体移除，这些部分一般为其它组织或噪声；
4. 去除区域打分值均值较低的结果，进一步去除噪声与误识。

最后在测试集上得到的预测标签打包成.gz 文件存储到 experiment/3DUNet/pred 下。

4 性能分析

4.1 分类结果

本实验使用独立的验证集进行验证，最终模型参数为在验证集上 DICELoss 最小的轮次时模型参数。若上述模型参数在 30 轮次内未改变，训练将提前终止。

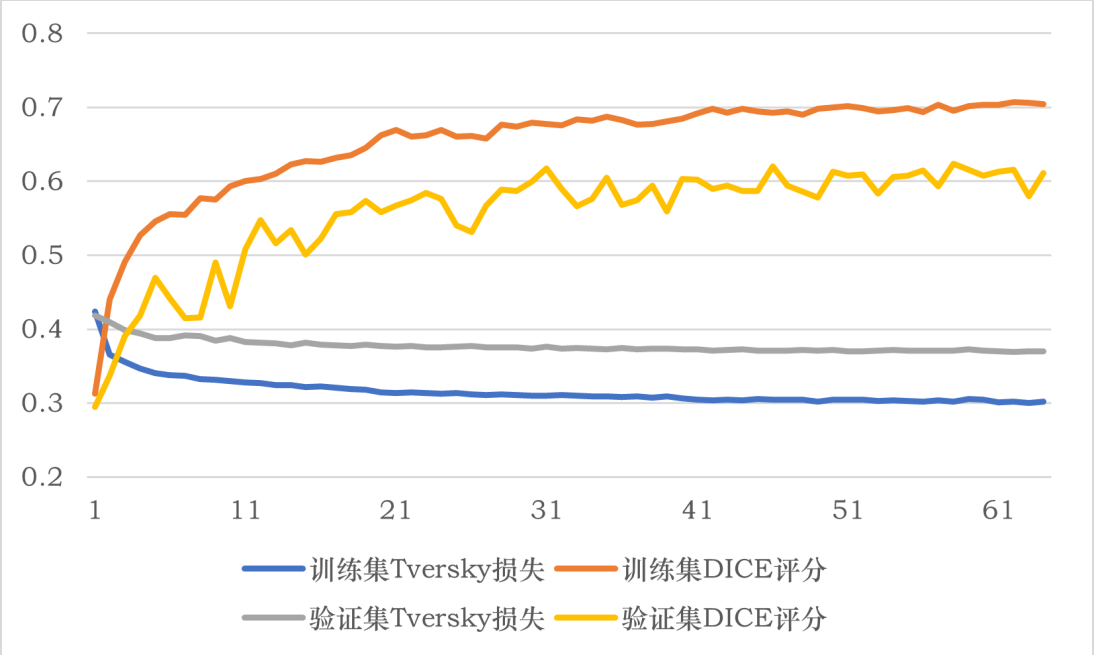
在验证集上运行时，直接输出各区域的似然值进行3.7中操作作为验证结果，与验证集标签计算 FROC，结果如下表：

Table 1: 不同 FP 下的召回率

指标	FP=0.5	FP=1	FP=2	FP=4	FP=8	average	maximum
FROC	0.040396	0.080792	0.161584	0.260369	0.393646	0.1874	0.6207

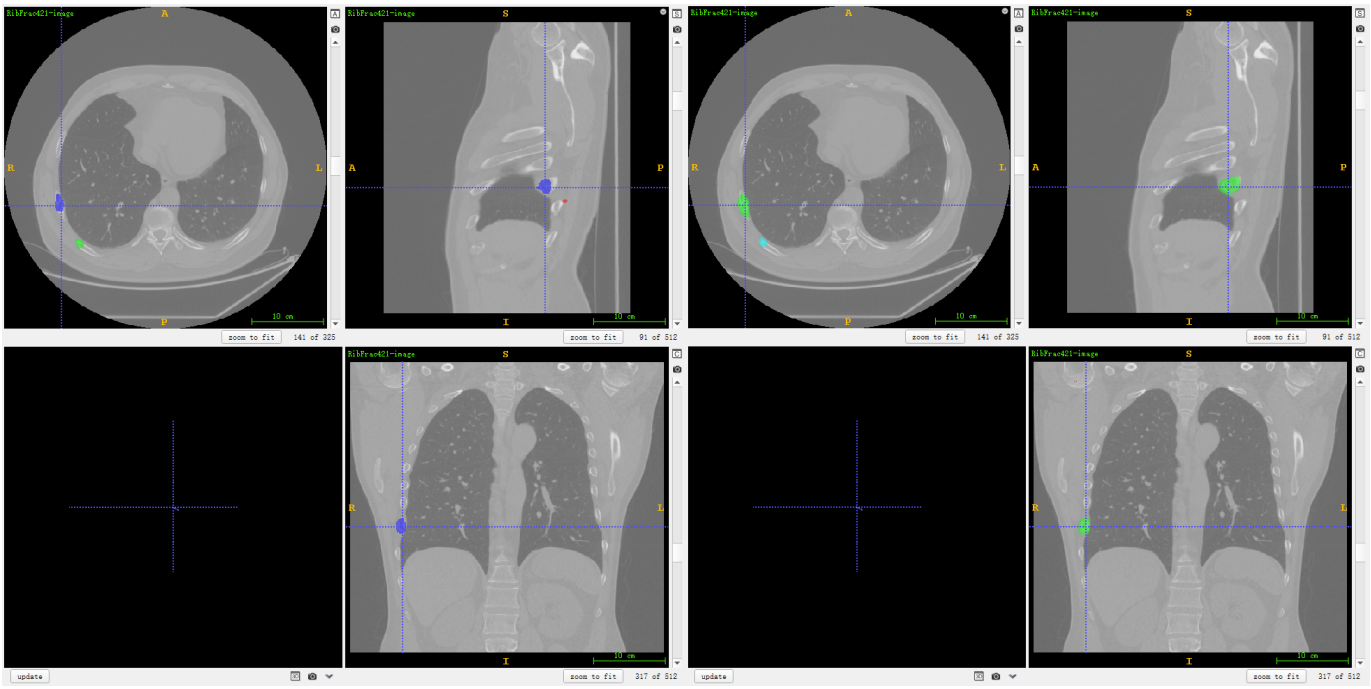
4.2 训练效果

训练集与验证集上的损失值如下图，同时也体现了 DICE 损失函数稳定性差，不适合作为训练指标。



4.3 输出示例

模型对绝大多数正类标签都可以实现分割，在边界上略微存在差异，在实际应用中，并不需要得到严格精确的边界。输出示例如 Figure3。



(a) (b)

Figure 3: 在验证集上进行测试的输出结果：(a) 真实标签 (b) 预测结果

4.4 时间性能

本实验数据量巨大，算法复杂，但我们在预处理、数据载入数据集以及输出测试结果时均做了并行优化，将在 CPU 上的操作进行提速。优化后的模型在训练时运行速度平均每轮 13 分钟，预测结果的速度为平均每张图片 1.68 分钟。在约 24 小时的训练后可以较快速得到基本满足诊疗需求的结果。

以上结果使用的计算资源为 2 块 NVIDIA GeForce RTX 2080 Ti GPU。

4.5 空间性能

本实验对计算资源的要求比较合理，注意显存的及时释放，训练-验证同时进行约占用显存 $16GB$，由于服务器系统漏洞，内存占用计算未能计算。

5 实验感悟

通过本次使用神经网络对三维图像实现检测分割的实验，我们感受到数据维度增加使特征提取、模型预测等环节的困难都成倍增加，意识到降维与合理采样的重要性。我们同样感受到医学图像精度高、图像间信息相似等特点，深刻体会到分类问题不同评价指标的评价角度差异。这些收获使我们对机器学习适合解决的问题有了进一步理解。

References

[1] Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention* (2016), Springer, pp. 424–432.

[2] Diakogiannis, F. I., Waldner, F., Caccetta, P., and Wu, C. Resunet-a: a deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing 162* (2020), 94–114.

[3] Egan, J. P., Greenberg, G. Z., and Schulman, A. I. Operating characteristics, signal detectability, and the method of free response. *The Journal of the Acoustical Society of America 33*, 8 (1961), 993–1007.

[4] He, T., Zhang, Z., Zhang, H., Zhang, Z., Xie, J., and Li, M. Bag of tricks for image classification with convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2019).

[5] Ronneberger, O., Fischer, P., and Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (2015), Springer, pp. 234–241.

[6] Salehi, S. S. M., Erdogmus, D., and Gholipour, A. Tversky loss function for image segmentation using 3d fully convolutional deep networks. In *International workshop on machine learning in medical imaging* (2017), Springer, pp. 379–387.