

Machine Learning Course Project

MR Image Segmentation

夏卓凡 自 66
2016011496

xiazf16@mails.tsinghua.edu.cn

景道冉 自 66
2016011506

jdr16@mails.tsinghua.edu.cn

摘要

在本项目中，我们实现了基于 *U-Net* 模型的 3D、2D 以及 2D-3D 混合模型，针对给出的 40 组肝脏肿瘤 MRI 图像进行分割。我们从压缩模型大小，利用全局信息以及空间连续性等方面入手，先将 3D 模型转化成 2D 模型，改善了模型预测结果的准确率；将 2D 模型转化为 2D-3D 混合模型，进而解决了 2D 模型在预测时的 z 轴不连续现象。其中 2D-3D 混合模型的表现在这两个时期的图像上表现较为出色，能够给出令人基本满意的分割预测结果。

1. 问题背景介绍

原发性肝癌是全球第二大常见癌症死亡原因，也是第六大常见癌症。根据实体瘤反应评估标准 (RECIST) 可知，肿瘤大小往往与肿瘤良性恶性有较大关系。在临床分析过程中，医生通常会使用计算机断层扫描 (CT)、核磁共振图像 (MRI) 等影像学方法实现对人体结构的三维建模，并依此实现对肝脏肿瘤是否存在、肿瘤大小及良性恶性进行判断。

然而，因为人工诊断图像费用昂贵、耗费时间长、可再现性差且其分割结果与诊断医师的专业程度十分相关。因此，我们希望根据 MRI 构建的三维图像训练模型，并实现对肝脏肿瘤的自动分割。

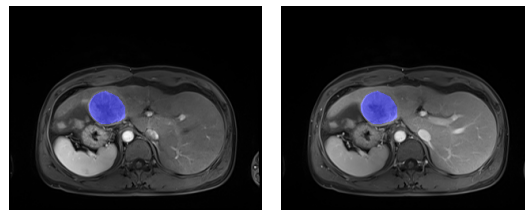
本项目的主要任务是：根据预标注的 40 组 MRI 图像，实现对肝脏肿瘤细胞位置的预测。在本次图像分割任务中，我们将图像中的肿瘤部分标记为 1，非肿瘤部分标记为 0，需要实现对所有像素点的分类。

本项目中主要问题有：肝脏肿瘤的位置与形状未知，且不同器官与组织在 MRI 图像中，也存在一定强度差异，因此，我们需要全局信息以实现肿瘤的更好预测。同时，项目所使用的 MRI 图像中，正样本在总样本中的占比低于 1%，这也给模型训练带来了难度。

2. 数据集处理

2.1. 核磁共振 (MRI) 数据集简介

在本项目中，主要使用的数据集为助教提供的内部数据，采集了来自 40 名肝脏肿瘤病人的核磁共振检查的图像。一组核磁共振图像数据，包括平扫期 (pre phase)、动脉期 (artery phase)、门脉期 (portal phase) 和延迟期 (delay phase) 4 个时期的检查图像以及肿瘤的分割标注。



(a) 动脉期 MRI 图像

(b) 门脉期 MRI 图像

图 2: MRI 数据样例切片 (FZMC001, $z = 45$)

如图 2 所示蓝色区域，可以肉眼可见左图动脉期和门脉期中对应位置有明显的类似肿瘤的区域，这表示这张图片的标注是基本有效的。事实上，数据和标注都是 3D 图像，肿瘤的分布类似球体而非平面上的圆形区域。在不同的 z 轴位置对应的切片上，肿瘤区域的占比

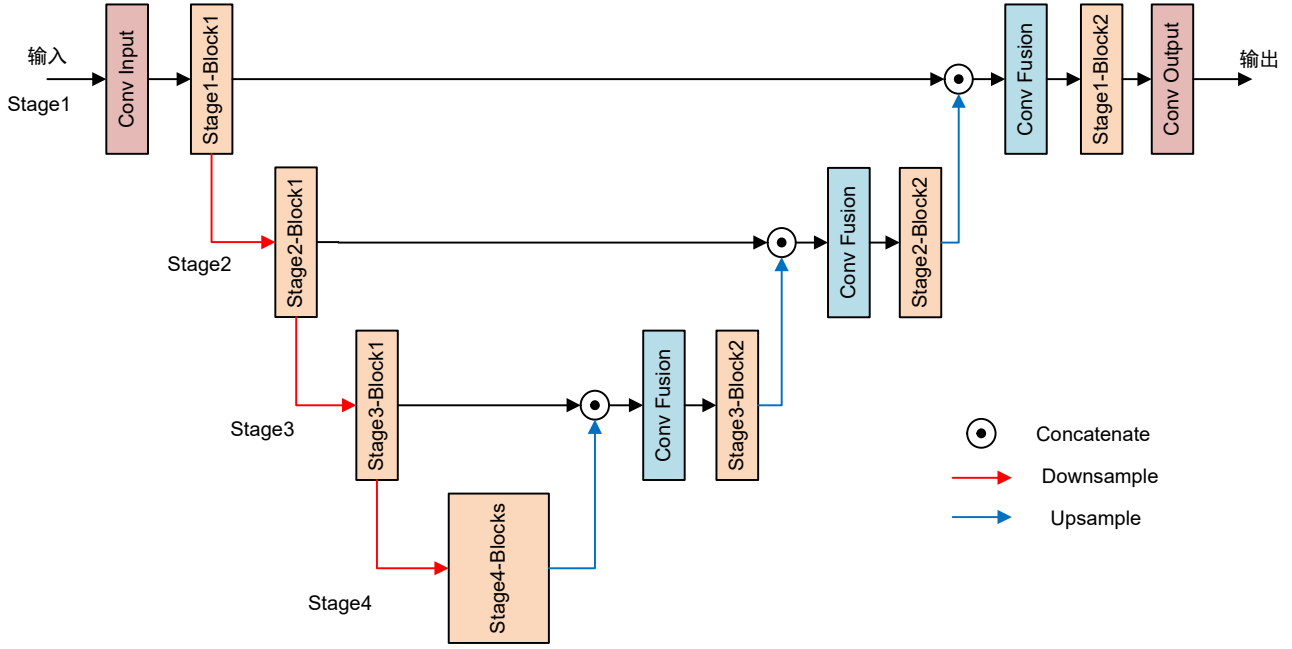


图 1: 整体模型结构

并不相同，这是与 2D 医疗图像分割的显著不同。

2.2. 数据的处理与生成

原始的数据为 40 个压缩包，每个包内包含 4 个时期的扫描图像和 1 个对应的标注，大小基本为 $320 \times 256 \times 72$ ，但有一些数据的 z 轴厚度不一致。考虑到模型处理的一致性以及取整方便，我们对原始数据进行了如下的处理。

首先，将 40 组数据进行简单整理，对 3D 形体数据进行插值，缩放到标准的 $320 \times 256 \times 72$ 大小。3 个模型所需的训练数据处理方式各不相同，但测试数据的处理完全相同。在所有的实验中，将测试集对应的图像和标注封装在 1 个 `.tfrecord` 文件中保存。

对于训练集数据，原始的 3D 模型需要大量的小型切块，在这里需要取多组 $64 \times 64 \times 64$ 大小的训练数据块。为了正负样本的均衡，在有肿瘤标注的区域取出一定数量的切块后，再在没有肿瘤标注的区域取相等数量的切块，分别存入 `.tfrecord` 文件中保存。2D 模型则需要对 z 轴切片，将每一片切片作为数据的单位保存。2D-3D 混合模型在 z 轴上切一定厚度的切块，按块保存。

其中，为了保证重采样和切块时数据数值上的一致性，对各个时期的图像采用三线性插值，对对应的标注采用最近邻插值，并将图像数据保存为 `int16` 类型，标注数据保存为 `uint8` 类型。

3. 提出的方法

针对核磁共振图像分割问题，我们渐进地提出了 3 个模型，获得了不断提升的效果。

3.1. 3D-UNet 模型与简化

首先，我们想到的是直接将文献 [3] 和 [2] 中的 U-Net 框架直接应用在 MRI 图像分割上。为此，我们设计了一个类似 V-Net 的网络结构，如图 1 所示。

模型接受大小为 $64 \times 64 \times 64$ 的切块输入，经过一个输入卷积层提取特征之后，分别使用 4 级卷积模块提取各个尺度上的信息，之后再行上采样，将不同尺度的信息逐级整合解码，最后使用一个 1×1 卷积层输出预测的 `logit`。

在这个模型设计中，各层的特征处理块设计为 2 组堆叠在一起的 Conv-BN-LeakyReLU 级联，这一设计与原始的 U-Net 结构 [3] 一致，如图 3(a) 所示。下采样层

使用 Maxpooling 操作，上采样层使用 ConvTranspose。

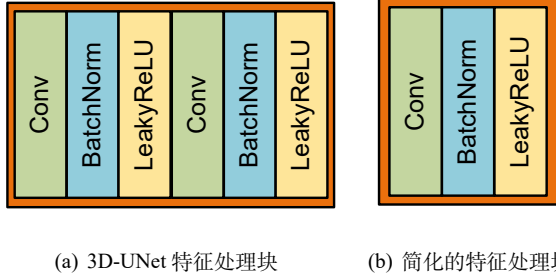


图 3: 模型特征处理模块的设计

由于采用了 3D 卷积，这个模型在基础通道数为 32 时的参数量已经达到了 29M，相较于数万量级的训练数据，模型的复杂程度已经非常高。根据统计学习理论，这样的复杂程度不利于模型的泛化能力，很容易陷入过拟合问题之中。此外，训练的速度变得缓慢，我们无法在合理的时间之内得到一组有效训练的模型。

因此，我们决定简化模型，将特征处理模块的 2 组卷积替换为 1 组，并将上采样的 ConvTranspose 替换为 3D 的最近邻插值。这种情况下，参数量约为 2M，在合理的范围之内，该模型命名为 SimplifiedUNet。

3.2. 损失函数的设计

图像的分割输出为图像中每个点的二分类 logit 值，一种自然的想法是使用分类损失 Binary Cross Entropy 监督这些分类，如下式所示

$$L_{bce} = \sum_{i=1}^N y_i \ln \hat{y}_i + (1 - y_i) \ln(1 - \hat{y}_i). \quad (1)$$

其中 $y_i \in \{0, 1\}$ 为标注值， $\hat{y}_i \in (0, 1)$ 为 sigmoid 函数激活后的预测值。但是由于正负样本不均衡的问题十分严重，直接使用这个损失会导致模型非常倾向预测负样本，甚至无法给出正样本的预测。

事实上我们观察到，即使是在生成的正样本切块中，标记为 1 的点（实际正样本）的个数仍然低于 5%，正负样本比例严重失衡。因此，准确率（accuracy）作为衡量分割性能的指标就失去了意义，因为在大量负样本的情况下，对正样本不做预测仍然可以获得极高的准确率，这显然不是我们希望得到的结果。

为了正确测量这种情况下的模型性能，我们关注

正样本的 Dice 系数，表达式为

$$\text{Dice} = \frac{2 \sum_{i=1}^N y_i \hat{y}_i}{\sum_{i=1}^N \hat{y}_i + \sum_{i=1}^N y_i}. \quad (2)$$

特别地，在二分类问题中，这一指标可以认为是精度（precision）与召回率（recall）的一种联合考虑，因为

$$\text{Dice} = \frac{2TP}{2TP + FP + FN}, \quad (3)$$

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (4)$$

$$\text{Recall} = \frac{TP}{TP + FN}. \quad (5)$$

其中 TP , FP , FN 为真正率、假正率和假负率，分别对应两类统计学错误。因为 Dice 系数在 (0, 1) 之内，且越接近 1 表示性能越好，因此可以设计一种 Dice 的损失函数如下

$$L_{\text{dice1}} = 1 - \text{Dice} = 1 - \frac{2 \sum_{i=1}^N y_i \hat{y}_i + \epsilon}{\sum_{i=1}^N \hat{y}_i + \sum_{i=1}^N y_i + \epsilon}. \quad (6)$$

但实际上我们使用的是 Dice 损失的变体 [2] 作为损失函数，数学表达形式如下

$$L_{\text{dice2}} = 1 - \frac{2 \sum_{i=1}^N \hat{y}_i y_i + \epsilon}{\sum_{i=1}^N \hat{y}_i^2 + \sum_{i=1}^N y_i^2 + \epsilon}. \quad (7)$$

其中， $\epsilon = 1.0$ 为平滑因子，防止在极端情形下出现数值错误，其中使用 L_{dice2} 可以获得更加平滑的损失，训练过程更稳定。

在训练接近收敛的时候，我们引入了另一种损失函数来对结果进行微调，称为 Label Related Dropout Element Hinge Loss (LRDEHL)。首先根据标注的正样本，从全体负样本中取预测值最大的那些，使得与正样本数目相同，这样就控制了每次梯度传播的有效数据点，认为平衡了正负样本，并做了难样本挖掘。与之配套的损失函数采用的是 MSE 或者 MAE 损失，即

$$L_{\text{lrdehl1}} = \frac{1}{N_{\text{positive}}} \sum_{i=1}^N |\text{mask}(y_i) - \text{mask}(\hat{y}_i)|, \quad (8)$$

以及

$$L_{\text{lrdehl2}} = \frac{1}{2N_{\text{positive}}} \sum_{i=1}^N (\text{mask}(y_i) - \text{mask}(\hat{y}_i))^2. \quad (9)$$

数据集	模型	指标	样本 1	样本 2	样本 3	样本 4	样本 5	样本 6	样本 7	样本 8	平均值
动脉期	3D	Precision	0.5043	0.9886	0.0000	0.9105	0.3065	0.6823	0.3255	0.5731	0.5364
		Recall	0.4926	0.1387	0.0000	0.4032	0.0196	0.1773	0.0740	0.1374	0.1803
		Dice	0.4984	0.2433	0.0000	0.5589	0.0369	0.2814	0.1206	0.2216	0.2451
	2D	Precision	0.9696	0.9302	0.3377	0.8186	0.5004	0.8517	0.3991	0.5906	0.6760
		Recall	0.3210	0.1427	0.0568	0.2315	0.1353	0.1208	0.4952	0.0994	0.2003
		Dice	0.4823	0.2474	0.0972	0.3610	0.2104	0.2115	0.4420	0.1701	0.2782
	混合	Precision	0.9952	0.8339	0.6728	0.7745	0.2373	0.8435	0.7364	0.9500	0.7554
		Recall	0.2602	0.4648	0.0522	0.1436	0.1184	0.0638	0.8521	0.3413	0.2870
		Dice	0.4125	0.5969	0.0969	0.2422	0.1580	0.1187	0.7900	0.5021	0.3647
门脉期	3D	Precision	0.5043	0.9886	0.0000	0.9105	0.3065	0.6823	0.3255	0.5731	0.5364
		Recall	0.4926	0.1387	0.0000	0.4032	0.0196	0.1773	0.0740	0.1374	0.1803
		Dice	0.4984	0.2433	0.0000	0.5589	0.0369	0.2814	0.1206	0.2216	0.2451
	2D	Precision	0.9350	0.9422	0.3442	0.5489	0.8011	0.9122	0.0098	0.9007	0.6743
		Recall	0.1992	0.0977	0.0146	0.0445	0.3776	0.2974	0.0034	0.2219	0.1570
		Dice	0.3285	0.1770	0.0280	0.0823	0.5133	0.4485	0.0050	0.3561	0.2423
	混合	Precision	0.8408	0.5860	0.4940	0.4730	0.6463	0.8220	0.0726	0.5719	0.5980
		Recall	0.4471	0.3397	0.2025	0.2668	0.5270	0.4386	0.2393	0.6221	0.3854
		Dice	0.5838	0.4301	0.2873	0.3926	0.5806	0.5720	0.1114	0.5998	0.4447

表 1: 测试数据集上的性能对比

3.3. 2D-UNet 模型

由于 3D-UNet 模型只处理了局部信息,容易将 MRI 图像中不同器官之间的强度差异也误判为肿瘤结构。因此,我们基于原有 3D-UNet 的模型结构设计了 2D-UNet 的网络结构。由于 2D-UNet 计算负载较小,我们可以一次输入一整个的 xy 平面进行训练与测试,这样就使得网络可以有效地考虑全局信息,做出更准确的预测。

2D-UNet 模型接受大小为 320×256 的切片输入,也是经过一个入卷积层提取特征值,之后分别使用 4 级卷积模块提取各个尺度信息,并进行上采样,逐级混合解码。最终使用一个 1×1 卷积层,得到预测的逻辑输出。

在 2D-UNet 模型中,考虑到训练数据只有几千,为了防止过拟合,我们依然对原始 UNet 模型进行了简化,采用了如图 3(b)所示的设计,在基础通道数为 32 时,参数量在 1M 左右。下采样为 2D 的 Maxpooling,上采样为 2D 双线性插值。

3.4. 2D-3D 混合模型

在后面的实验部分,我们观察到 2D 模型由于只能在一个平面上预测,容易出现 z 轴输出不连续的情况。也就是说,2D 模型没有考虑到另一个方向上数据分布的连续性。因此,在不放弃使用 xy 平面的全局信息的基础上,我们将 z 轴上的连续 3 层输入模型,增加对 z 轴方向上的信息融合,设计出了 2D-3D 混合模型 (HybridUNetX)。

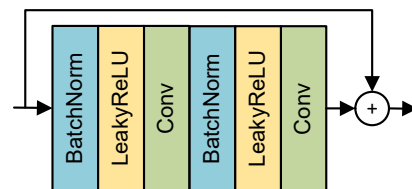


图 4: ResNetv2 的构建块

2D-3D 混合模型与 2D 模型的上下采样机制是相同的,使用 2D 的双线性插值能够更好地保证上采样数据

的光滑性。除此之外，特征处理模块替换为如图 4所示 ResNetv2[1] 的构建块，这样做可以共用多组激活函数，并增加网络的训练效率。我们在核心区域（Stage4）使用 3 组级联的构建块提升性能，但考虑到参数量，使用 16 基础通道保证一个在 1M 左右参数的模型来平衡。

4. 实验与结果分析

4.1. 3D-UNet 模型

3D-UNet 模型中，我们将初始学习率设置为 1×10^{-3} （动脉期，门脉期），每迭代 5000 次则减小为原来的 0.5 倍，选择 Adam 作为优化器，参数分别设置为 $\beta_1 = 0.9, \beta_2 = 0.99$ ，使用 3.0×10^{-5} 的 L2 正则化抑制过拟合。训练使用 1 块 NVIDIA GeForce RTX 2080 Ti 显卡 15 小时左右，完成 40k 迭代；测试结果如表1所示。

其中 3D-UNet 的训练数据为在肿瘤附近进行采样的， $64 \times 64 \times 64$ 的正方体。测试数据为在 MRI 三维图像上平移采样得到的，与训练数据形状相同的立方体，最终的测试结果时由这些立方体拼接得到。

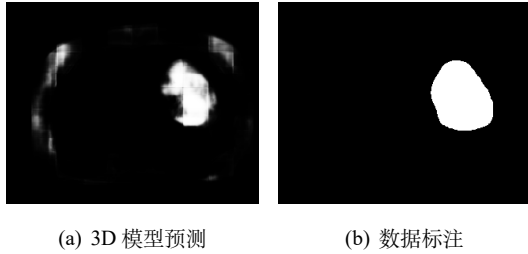


图 5: 3D 模型预测结果样例 (FZMC034, $z = 16$)

其实验结果样例如上图5(a)和5(b)所示，观察可知：由于最终的 MRI 图像中的肿瘤预测是根据多个小立方体直接拼接而成，因此可以观察到明显的拼接痕迹。而图像中存在较多非肿瘤区域被标注为肿瘤，而这些假阳类错误多出现在不同器官的分隔处。经过分析后，我们认为这种情况出现的原因是：模型仅仅学习到了肿瘤与周围环境之间的局部差异，而未考虑全局信息导致的。

为引入全局信息并节约计算量，我们引入了 2D-UNet 模型进行实验。



(a) 2D 模型预测



(b) 2D 模型预测 z 轴方向

图 6: 2D 模型预测结果样例 (FZMC034, $z = 16, y = 136$)

4.2. 2D-UNet 模型

2D-UNet 模型中，我们将初始学习率设置为 1×10^{-3} （动脉期） 1×10^{-4} （门脉期），每迭代 5000 次则减小为原来的 0.5 倍，选择 Adam 作为优化器，参数分别设置为 $\beta_1 = 0.9, \beta_2 = 0.99$ ，L2 正则化强度为 5.0×10^{-4} ；选用 LRDEHL 作为 30k 迭代后的损失函数。我们使用 1 块 NVIDIA GeForce RTX 2080 Ti 显卡训练 40k 迭代步数，用时 2.44 小时。

与 3D-UNet 模型不同的是，2D-UNet 模型在三维图像的每个横切面上训练，数据大小为 256×320 。

测试实验结果如下图 6(a)所示。观察可知，在 xy 平面来看，相比较于 3D 模型，2D-UNet 的假阳率明显下降，预测肿瘤的位置与真实肿瘤位置差别更小，对肿瘤大小的预测也偏小。

然而，观察模型在样本图像中 z 轴上对肿瘤的预测，发现其存在肉眼可分辨的断层现象即不连续性，如图 6(b)所示。为了改善这一结果，我们提出了 2D-3D 混合的 Hybrid-UNet 模型。

4.3. 2D-3D 混合模型

HybridUNetX 模型中，我们将初始学习率设为 1×10^{-4} （动脉期）和 1×10^{-5} （门脉期），每迭代 5000 次则减小为原来的 0.5 倍，选择 Adam 作为优化器，参数分别设置为 $\beta_1 = 0.9, \beta_2 = 0.99$ ，L2 正则化强度为 5.0×10^{-4} 。

我们使用 1 块 NVIDIA GeForce RTX 2080 Ti 显卡训练 40k 迭代步数，用时 8.1 小时。

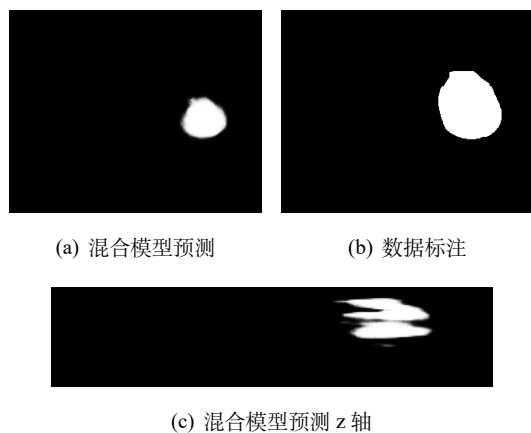


图 7: 2D 模型预测结果样例 (FZMC034, $z = 32, y = 120$)

如图 7(a)所示，2D-3D 混合模型保持了 2D 模型对全局信息的处理，能显著减少假阳性错误；同时，如图 7(c)所示，2D-3D 混合模型的边缘更加平滑， z 轴的连续性更好，这也提升了 recall。如表 1 所示，2D-3D 模型在多个设置下都取得了最佳的效果。

5. 结论

针对肝部肿瘤 MRI 图像的小样本图像分割问题，在使用深度神经网络进行自动诊断时，UNet 及其衍生网络能够得到可行的结果。为了优化模型对肿瘤的预测结果，我们需要考虑这些因素：小样本均衡问题、图像的全局信息和预测结果的连续性。在这个思路下，我们实现了 UNet 原始模型在 3D、2D、2D-3D 混合情况下的应用，并为了减少计算量和参数量、防止过拟合对 UNet 模型进行改进与实验验证，并获得了可以接受的结果。然而，我们的模型仍然存在着较大的改进空间，特别是 Recall 仍然比较低，对比图 7(b)可以看出，模型仍然倾向于预测较少的正样本。接下来我们可以考虑将目标检测的手段结合进来，或者采用其他的样本均衡策略。

6. 任务分工

夏卓凡：训练数据处理与框架搭建、HybridUNet 模型搭建、模型训练与调试及报告内相关部分书写。

景道冉：数据处理及训练框架修改、3D-UNet 与 2D-UNet 模型搭建、模型训练与调试及报告相关部分书写。

参考文献

- [1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *European conference on computer vision*, pages 630–645. Springer, 2016.
- [2] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 565–571. IEEE, 2016.
- [3] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.