# 方法

## 预处理

神经网络的训练过程，其实与人在学习事物的过程很相似，人在学习事物时倘若能够把问题简化，把握重心，那么不仅仅在于学习效率上，就连最终的学习结果都会变得更好，神经网络的训练亦是如此，预处理就是一个把握重心，简化问题的过程，这样操作之后便能够大大降低学习难度，缩短训练时间。

医学图像分割其实是计算机视觉中比较经典的语义分割任务，医学图像分割与自然图像分割有所不同，自然图像分割的分割对象分布比较随机，且没有一定的规律，而医学图像分割的分割对象比较聚焦，分割对象分布规律固定，自然图像分割大量应用于无人驾驶领域，所以自然图像分割更强调实时性，而医学图像分割则要求更精细，可以舍弃一定的速度要求来换取精度要求。

图像去噪，CT图像在传输保存的过程中，很容易引入各种噪声，而这些噪声会使得学习任务变难，为了追求更好的分割效果，应引入必要的去噪，使用了小波变换去噪。

先验规律裁剪，CT图像，拍摄角度固定，所以会有一些黑色边缘，可以通过裁剪将其消去，裁剪对于医学图像分割来说是十分有效的，一方面可以减小图像规模，另一方面可以突出重点，降低学习规模，然而简单的裁剪往往不能达到最好效果，而为了得到最好的裁剪效果，首先遍历训练集的标签，每一张病人的分割结果切片可以得到四个坐标，这四个坐标分别是图像中最上，最下，最左，最右的坐标，而后分别取最靠近边缘的值，而为了避免偶然性，上下左右四个方向外扩10个像素。

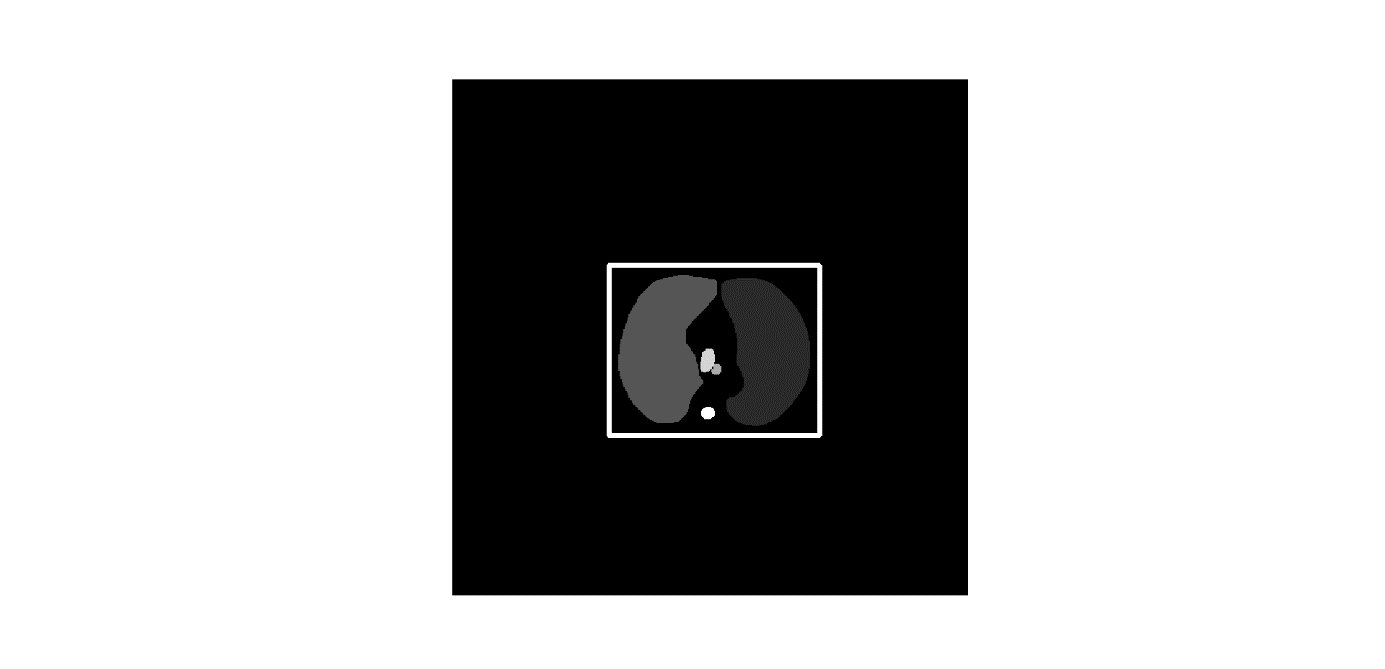


图1 单张切片搜索示意图

标准化，对CT图像进行z-score正则，将数据缩放到一个小区间，将问题规模缩小，降低学习的难度。

数据增强，在训练达到瓶颈的时候，或者训练正样本太少时，都可以引入数据增强，这里主要引入了空间域平移、翻转、旋转等数据增强手段，具体细节为：

表1 数据增强操作列表

|  |  |
| --- | --- |
| 平移 | 图像整体随机向四个方向平移0~10个像素，通过生成一个0~10的随机数来指示平移距离，并随机生成一个标志位来指示方向，而后构成平移矩阵对图像进行变换操作。 |
| 旋转 | 图像随机旋转0~5°，随机生成一个0~5的随机数来指示旋转角度。 |
| 翻转 | 图像随机水平、垂直翻转，随机生成一个标志位指示翻转类型，使得不翻转：水平翻转：垂直翻转：水平垂直翻转的比例为。 |

## 训练

### 训练策略

对于OAR的分割，往往是多个目标分割，而不同目标之间在分割效果上必定存在差异，例如，小目标的分割，肺部OAR中脊髓的分割效果往往很差，我们采用的是使用2D的病人Slice以及其对应的mask来进行训练，所以我们首先对训练集中的不同样本进行定义，已知图像掩膜通过不同灰度值来代表分割对象，而背景的灰度值为0，则首先对掩膜进行二值化，统计图像中非0的像素个数，非0像素个数大于阈值则为简单正样本，非0像素个数小于阈值大于0则定义为困难正样本，而完全为0则为负样本。

(1)

(2)

(3)

其中，为掩膜的长宽，是二值化后的掩膜，为掩膜中的器官有效面积，为一个长列表，指明某一片为何种样本。

正常人类的学习的过程就像一个上台阶的过程，显然最初时，一般是从相对简单的情况入手，而后再不断加大学习的难度，这样才能够有更高的学习效率，和学习效果。这里采用这个思路通过调整每个epoch中的困难样本与简单样本比例以求达到最好的效果。具体的操作方式是，首先将简单样本调出来，只有简单样本的情况，经过训练在验证集上的结果趋于稳定之后，然后再掺杂所有困难样本的进入到训练集，而后训练在验证集上的结果稳定后，再加入，直至所有的困难样本都被加入到训练集中，而为了提高鲁棒性，负样本也会有部分掺杂。

对于GTV的分割，则又有一个更麻烦的问题，由于GTV在病人的切片中，分布少，面积小，是分类问题中经典的数据不均衡现象（在这里之所以可以说分类问题，是因为本质上语义分割就只是像素级别的分类而已），根据公式(3)，我们同样定义三种样本，简单样本、困难样本、负样本，简单样本和困难样本都可以划分为正样本，在这里负样本不能再随意地掺杂，此时训练集中的负样本是远多于正样本地，为了解决这个问题，在GTV训练的过程中应该保证正负样本数量大致接近，而为了让训练效果更好，则依旧采用OAR中地策略，在每个epoch开始的时候，使用简单正样本与同等数量负样本构成训练集，而训练稳定之后，则加入的困难正样本进入数据集，当然同时也要增加一定的负样本来保持数据均衡，训练稳定之后，再继续加入困难正样本和负样本，直到所有困难正样本都被加入，而且训练达到稳定。

对于GTV的分割，还有一点就是，由于OAR与GTV使用的是同一套病人图像，而只是分割对象不同而已，那么就可以将OAR分割网络的权重部分读入GTV网络中做迁移学习。

### 模型1 Unet

Unet[1]是非常经典的语义分割网络，虽然在自然场景分割方面，由于其实时性较差以及对于自然场景的适应性不强，其早已不是最热门的算法之一，但是在医学图像分割领域，却是十分炙手可热的存在，这是因为医学图像本身的特点决定的，医学图像结构固定，语义信息不那么强，所以Unet的跳跃连接结构比较有效，而且医学图像样本少，数据获取难，参数过多的网络比较容易过拟合，所以这里选择将Unet（图2）作为baseline以形成比较：

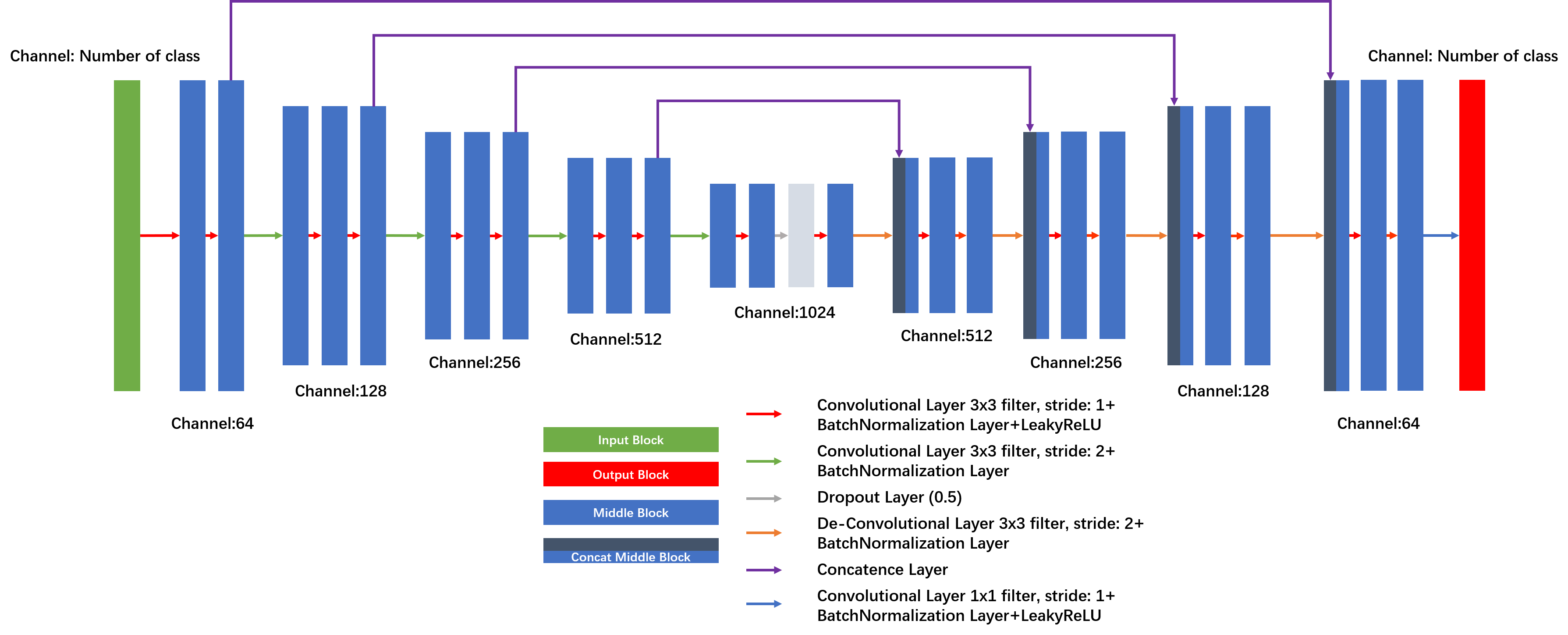


图2 Unet(baseline)

### 模型2 Unet+Channel Attention Block

目前许多高效的医学图像分割的神经网络模型都是以Unet作为backbone进行调改的，例如，Vnet[2]、Attention U-net[3]、Unet++[4]、MultiResUNet[5]、CE net[6]等，这里我们依然选用Unet的网络结构作为backbone，不同尺寸下的多次卷积中插入Senet[7]中的典型结构SE-Block（图4），这样做的理由是，实际上Unet本身的编码解码结构不需要做相当大的调整，做太多花哨的操作反而会让训练的时间变得过长，而效果提升甚微，而考虑到对于多个器官进行分割，则器官与器官之间的重要程度不同，所以应该引入注意力机制，而不断下采样的不断上采样的过程都是不同的编码解码阶段，而每一次编码、每一次解码都应该让网络对于学习对象有一定的偏重，这样才能使得各个分割对象都有较好的分割结果，于是我们提出以下新结构：

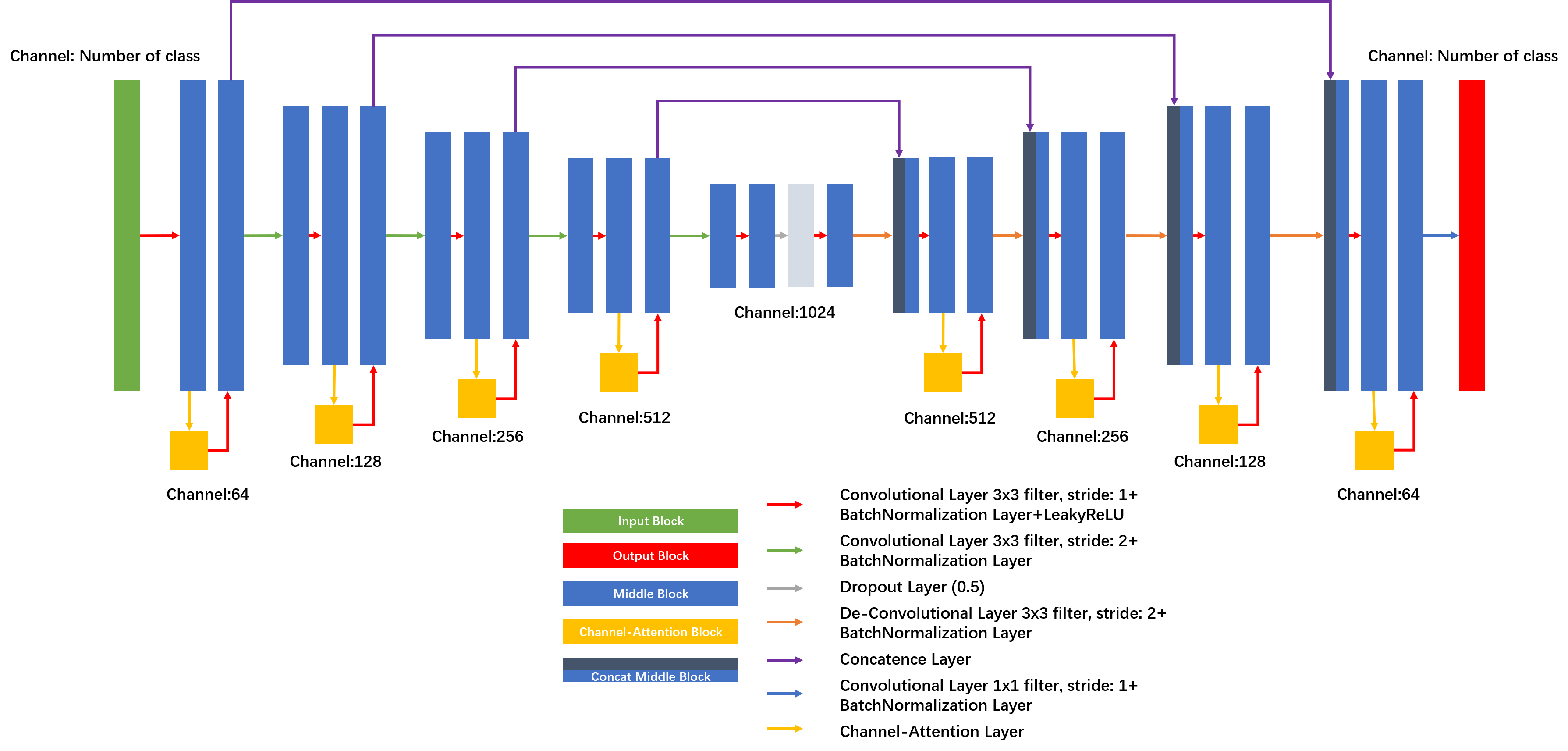


图3 Unet-Channel Attention Block

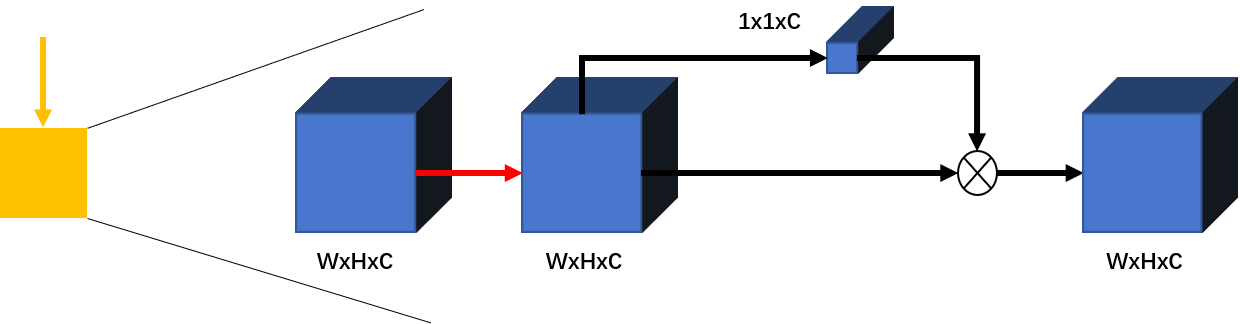


图4 Channel-Attention Block

### 损失函数与评价指标

交叉熵(CE,Cross Entropy)是一种十分优秀的损失函数，在分类问题中得到广泛的应用，对于数据较为均衡，学习对象较少的情况下，交叉熵的性能较好，而面对学习对象多而数据不均衡问题较严重时则往往效果较差。

(4)

而这里，我们主要使用加权交叉熵(WCE,Weighted Cross Entropy)[8]作为损失函数，选择理由主要是因为，其能够对于不同的分割对象给出不同的重视程度，对于多目标的分割有着较好的效果，并且对数据不均衡现象也能够有一定的缓解。

(5)

评价指标为戴斯相似性系数，这在图像分割中应用十分广泛，且计算起来方便简单，于是我们选择其作为我们模型的评价指标：

(6)

## 后处理

后处理的主要是一些形态学上的操作，例如腐蚀与膨胀，这样操作的原因是因为不同OAR之间虽然整体特征有很大区别，然而在边缘的特征上却比较相似，所以在预测中经常会有不同OAR的边缘相互混淆的情况，腐蚀膨胀则能够去掉那些不太合理的边缘，其具体的操作方法是：

1、获取神经网络推理结果，并将结果按通道拆分。

2、对逐个通道作腐蚀操作。

3、对逐个通道作膨胀操作。

4、通道合并。

而后还有一些针对性的特殊处理，对验证集上的输出结果进行观察发现，对于对称结构可能会出现互相嵌套的现象，而后通过判断是否嵌套类型，然后去除并补充，后处理的效果如下：

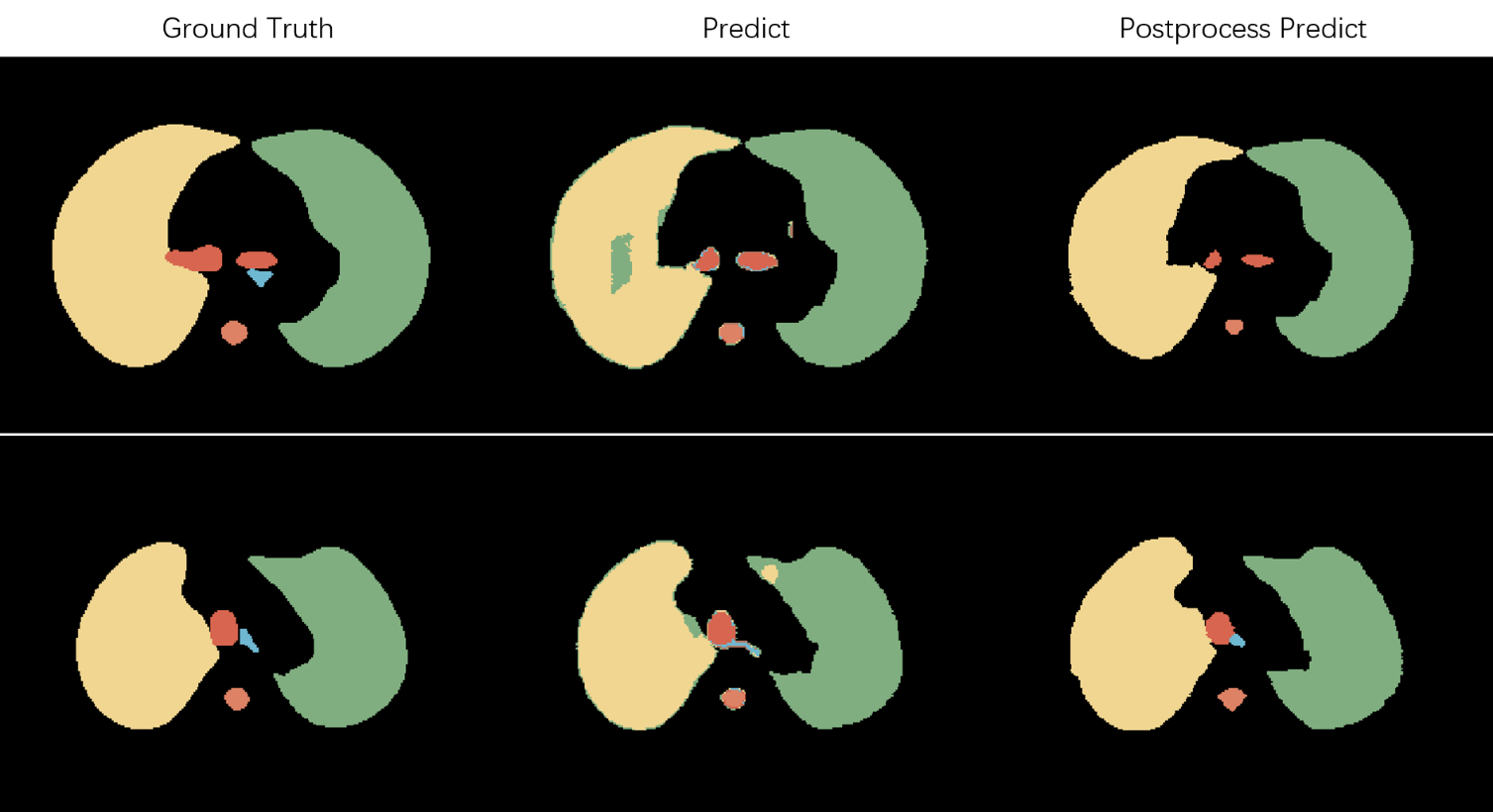


图5 后处理效果示意图

# 实验

这套方法训练集为Cstro Challenge的训练集，测试使用Cstro Challenge的测试集，共有四个任务：鼻咽癌GTV分割、鼻咽癌OAR分割、肺癌GTV分割、肺癌OAR分割，对于四个任务我们分别训练两个网络，实验设备为GTX1060 6G显存+GTX1060 4G显存。

## Unet training set

设置为200的最大epoch数，初始学习率为0.0001，学习率衰减策略为累计3个epoch评价指标不上升则学习率折半衰减，使用Adam优化器，Batch size为6，将数据集以8:2的比例划分为训练集和验证集，最终鼻咽癌GTV在98 epoch时收敛得到最好的验证集结果、鼻咽癌OAR在87 epoch时收敛得到最好的验证集结果、肺癌GTV在91 epoch时收敛得到最好的验证结果、肺癌OAR在47 epoch时收敛得到最好的验证结果。

## Unet+Channel-Attention Block training set

设置为200的最大epoch数，初始学习率为0.0001，学习率衰减策略为累计3个epoch评价指标不上升则学习率折半衰减，使用Adam优化器，Batch size为4（这是因为该网络的参数较原始Unet要多，而显卡显存限制其无法调到足够大的Batch size），将数据集以8:2的比例划分为训练集和验证集，最终鼻咽癌GTV在113 epoch时收敛得到最好的验证集结果、鼻咽癌OAR在77 epoch时收敛得到最好的验证集结果、肺癌GTV在92 epoch时收敛得到最好的验证结果、肺癌OAR在53 epoch时收敛得到最好的验证结果。

## Result

表2 实验结果(dice值比较)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Dice score | | | |
| **HaN\_OAR** | **Naso\_GTV** | **Thoracic\_OAR** | **Lung\_GTV** |
| Unet-CAB | **0.8373** | 0.5985 | **0.9223** | 0.4271 |
| Unet | 0.7773 | 0.5844 | 0.9051 | 0.3931 |
| Cstro sota | 0.8156 | **0.7936** | 0.8926 | **0.6363** |

可以看出，我们的方法在OAR的勾画上有一定的先进意义，而在GTV方面也相对于传统Unet上有一定的提升，这可能是由于加入了一定量的Channel-Attention Block然后使得每次编码解码的过程中，对于多个目标有不同的关注度，所以引入了过多的Channel-Attention Block，所以在大量的学习中能够对分割结果有所提升。

于是我们再更细地探索OAR分割结果，由于Cstro Challenge没有公开单个器官的分割结果，所以我们只好让Unet-CAB(Unet+Channel-Attention Block)与Unet比较。

### 肺部OAR

表3 Unet与Unet-CAB在肺部OAR上各个器官分割结果比较

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Dice Score | | | | | | |
|  | **Lung(R)** | **Lung(L)** | **OAR3** | **OAR4** | **OAR5** | **OAR6** | **Overall** |
| Unet | 0.945 | 0.958 | 0.824 | 0.466 | 0.657 | 0.692 | 0.905 |
| Unet-CAB | 0.922 | 0.949 | 0.823 | 0.612 | 0.829 | 0.834 | 0.922 |

### 鼻咽部OAR

表4 Unet与Unet-CAB在鼻咽部OAR上各个器官分割结果比较

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Dice score | | | | | |
|  | **OAR1** | **OAR2** | **OAR3** | **OAR4** | **OAR5** | **OAR6** |
| **OAR7** | **OAR8** | **OAR9** | **OAR10** | **OAR11** | **OAR12** |
| **OAR13** | **OAR14** | **OAR15** | **OAR16** | **OAR17** | **OAR18** |
| **OAR19** | **OAR20** | **OAR21** | **OAR22** | **Overall** |  |
| Unet | 0.766 | 0.801 | 0.795 | 0.000 | 0.000 | 0.000 |
| 0.000 | 0.000 | 0.775 | 0.798 | 0.000 | 0.770 |
| 0.759 | 0.732 | 0.674 | 0.673 | 0.687 | 0.349 |
| 0.352 | 0.636 | 0.852 | 0.850 | 0.777 |  |
| Unet-CAB | 0.836 | 0.878 | 0.867 | 0.357 | 0.208 | 0.382 |
| 0.232 | 0.010 | 0.828 | 0.840 | 0.575 | 0.831 |
| 0.829 | 0.801 | 0.806 | 0.779 | 0.785 | 0.672 |
| 0.656 | 0.805 | 0.882 | 0.891 | 0.837 |  |

### 结论

可以见得，Unet-CAB引入的Channel-Attention Block相当有效，对于一些小器官效果相比于Unet来说有了显著提升，每次编解码的注意力机制的引入，使得不同层次的特征提取有了偏重，获得信息是加权之后的信息，于是小分割目标也得到了响应的重视，所以小分割目标也能取得较好的结果。

### 结果展示

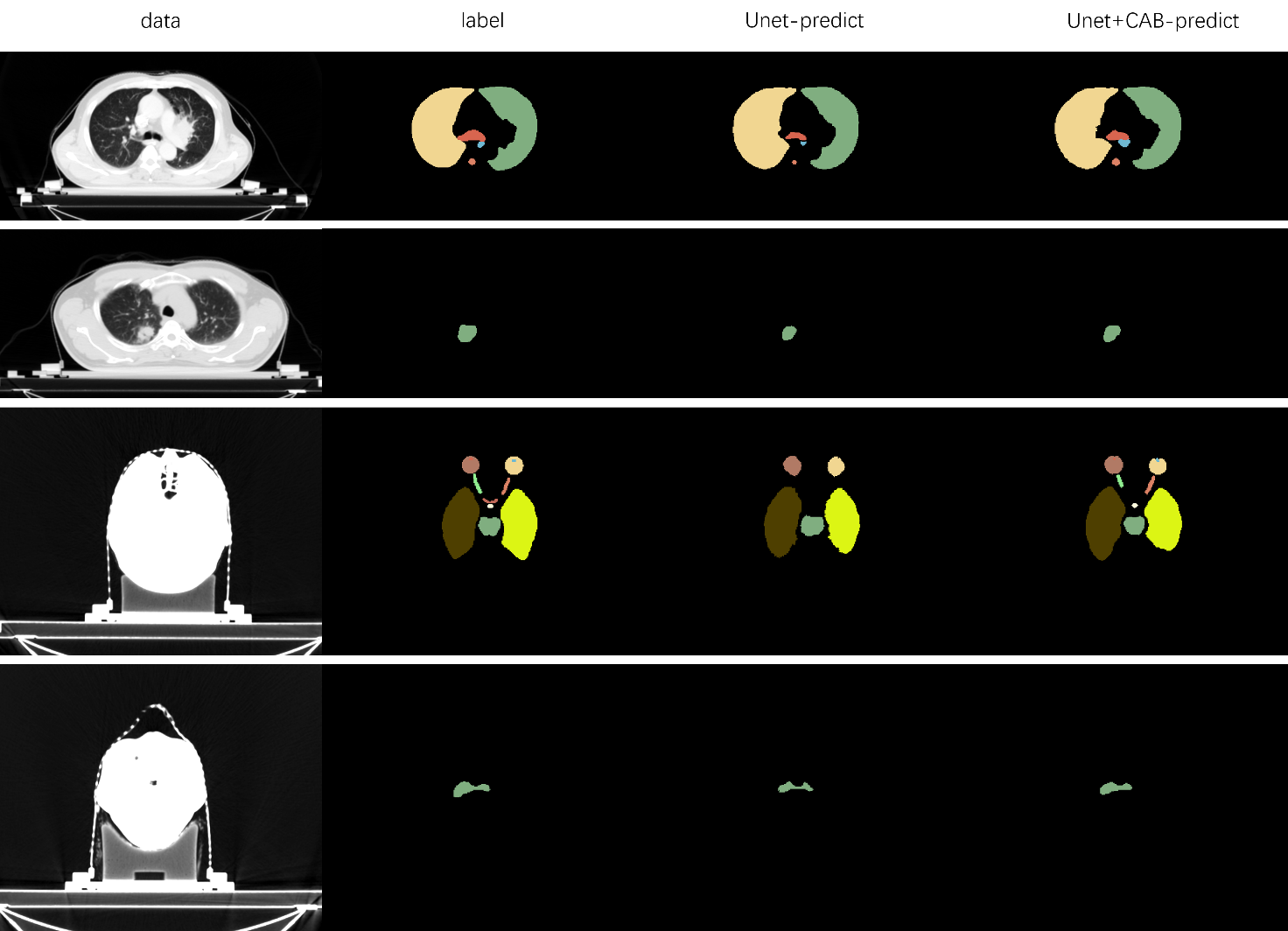


图6 肺癌、鼻咽癌的OAR与GTV在两种网络下的分割结果比较

[1] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.

[2] Milletari F, Navab N, Ahmadi S A. V-net: Fully convolutional neural networks for volumetric medical image segmentation[C]//2016 Fourth International Conference on 3D Vision (3DV). IEEE, 2016: 565-571.

[3] Oktay O, Schlemper J, Folgoc L L, et al. Attention u-net: Learning where to look for the pancreas[J]. arXiv preprint arXiv:1804.03999, 2018.

[4] Zhou Z, Siddiquee M M R, Tajbakhsh N, et al. Unet++: A nested u-net architecture for medical image segmentation[M]//Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer, Cham, 2018: 3-11.

[5] Ibtehaz N, Rahman M S. MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation[J]. Neural Networks, 2020, 121: 74-87.

[6] Gu Z, Cheng J, Fu H, et al. CE-Net: context encoder network for 2D medical image segmentation[J]. IEEE transactions on medical imaging, 2019, 38(10): 2281-2292.

[7] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.

[8] Iqbal S, Ghani M U, Saba T, et al. Brain tumor segmentation in multi‐spectral MRI using convolutional neural networks (CNN)[J]. Microscopy research and technique, 2018, 81(4): 419-427.