第 4 章 基于 L-SNet 的胰腺分割

4.1 引言

经典的 FCN 和 U-Net 模型在语义分割领域取得了巨大的成功，然而这些模型仍然有着难以克服的缺点，其中最重要的问题之一在于尺度的不均衡化。这个问题在医学图像处理中更为突出，真正需要分割的病灶/器官往往只占整个图片的很小一部分，并且由于病情不同，拍摄角度不同，需要分割的区域尺度变化较大。如图 4.1 所示，胰腺只占整图的很小一部分，同时不同切片的胰腺大小相差也较大。对于尺度不均衡的问题，已经有相当多的科研人员进行了相关研究。除了前面提到的 FPN、SPP、SNIP、TridentNet、Attention U-Net 等等外，另一个常见的方式是通过及联的方法直接连接多个模型。比如 Cascade R-CNN[91]通过及联 3 个网络逐步提升感兴趣区域的质量，在一步一步筛选和提升后最终在产生高质量的预测结果。论文[89]通过及联两个 Dense U-Net 进行 MRI 图像下的前列腺分割。论文[90]通过及联两个 2D FCN 和 3D FCN 利用多尺度信息分割肝脏和肿瘤的同时也节省模型显存的开销。

虽然很多科研人员在多尺度问题上做了诸多探索，如何利用不同尺度的特征进行精准分割在分割领域仍然是一个严峻的挑战。本文对于尺度不均衡的问题，提出一种全新的 L-SNet 模型架构，通过同时利用大尺度特征和小尺度特征从根本原因上缓解了尺度不均衡问题，并在公开数据集 TCIA Pancreas-CT 上达到 sota，验证了 L-SNet 的有效性。

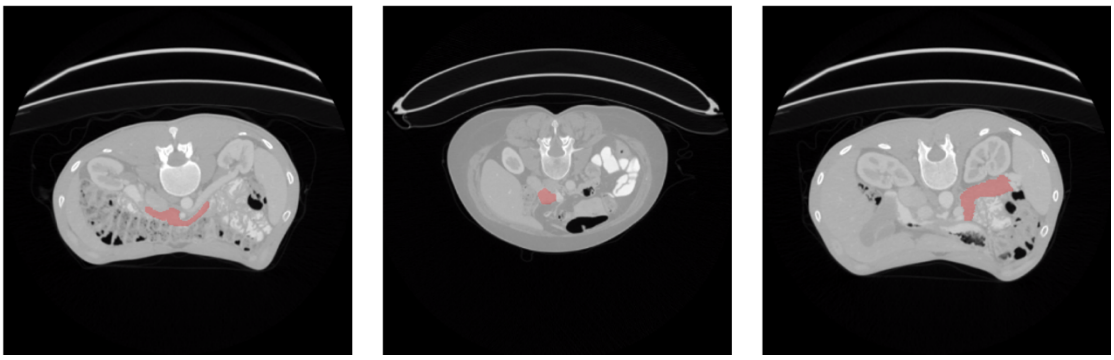

图 4.1 CT 切片中不同尺度大小的胰腺

4.1.1 数据集介绍

TCIA Pancreas-CT 数据集是由美国国立卫生研究院临床中心采集的公开数据集，该数据集中包含 53 例男性和 27 例女性的腹部增强 3D CT 扫描，共计 80

例样本（2020 年 9 月 10 号删除了两次重复数据）。其中 17 例受试者是进行肾脏切除术前的健康肾脏供体，另外 65 名受试者是由放射科专家挑选的即没有重大腹部疾病也没有胰腺癌病变的供体。受试者的年龄范围为 18 到 76 岁，平均年龄为46.8 ± 16.7岁。CT 扫描的切片厚度大小在 1.5 毫米至 2.5 毫米之间，分别旅大小为 512 * 512。扫描的机器型号为 Philips and Siemens MDCT scanners (120 kVp tube voltage)。以上数据由一名医学生逐帧标注分割信息，并通过一位专业专业放射科医生的验证。最后产生数据可以作为分割的 ground-truth 用于训练模型的验证模型。

Pancreas-CT 的 80 例 CT 数据中，除去不包含有用信息的噪音帧后，一共含有 17749 帧扫描图片，平均每例包含 222 帧图片。总计的 17749 帧图片中，有 10936 帧图片不包含胰腺。包含胰腺的图片中，平均每帧的胰腺大小为 3668 像素，仅占整图大小的 1.4%。胰腺的大小最小仅为 6 像素，最大为 16578 像素，尺度相差 2763 倍。中位数大小为 3474，尺度也相差 4.77 倍。从数据分析中可以看 Pancreas-CT 数据集的尺度不均衡问题较明显，主要体现在前景大小差异较大，前景和背景差异较悬殊。出本文使用 2D 的方式对 CT 扫描图片进行处理，首先讲数据按病例随机分成 4 份。对每一份数据中的 CT 图像，分别使用-1024 和 600 对 HU 值进行截断，再通过以下公式归一化到 0-1 之间：

$$X = \frac{X - \min(X)}{max(X) - min(X)} \tag{3.1}$$

式中X代表图片矩阵，对归一化后的图片*255，就可以得到可以可视化的灰度图像，图 3.1 列出了部分 CT 图像的可视化效果。从图中可以看出需要分割的部位只占整图的很小一部分，并且胰腺边缘比较模糊，分割难度较大。

4.2 L-SNet 模型设计

4.2.1 模型整体结构

L-SNet 的核心思想在于通过本文提出的 Location Network（LNet）检测出需要分割的目标，再通过仿射变换得出需要检测的图片，最后使用（Segmentation Network）SNet 对检出图片进行分割。由此将分割任务完全集中在同一尺度下进行，从根本上缓解尺度不均衡的问题。模型整体设计架构如图 4.2。

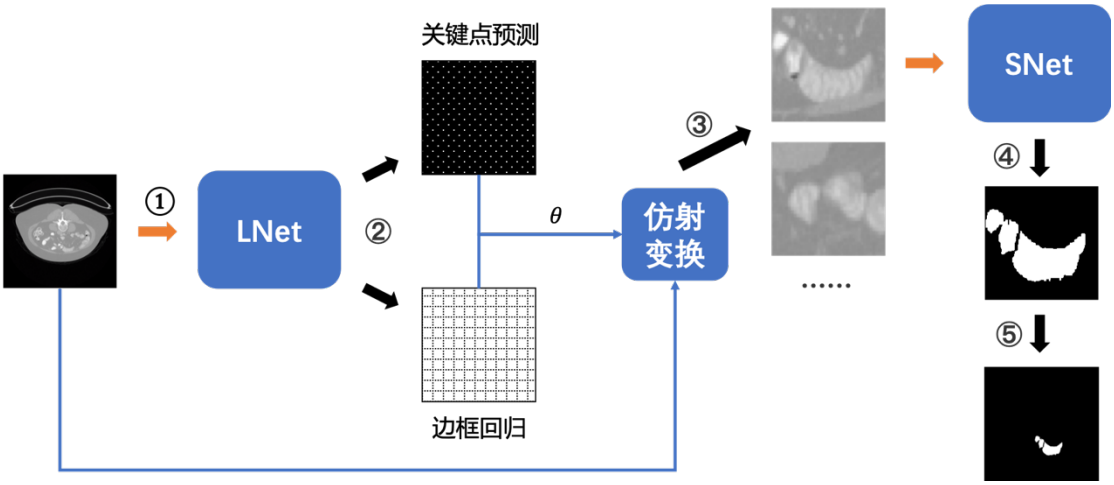

图 4.2 L-SNet 整体结构

4.2.2 Location Network

LNet 为该模型的核心部分，LNet 通过对目标物体的中心点生成热图进行关键点回归，同时对物体内的点回归 left，right，top，bottom 四个值计算目标物体的大小和位置。LNet 与 FCOS[92]的区别在于 1.LNet 直接使用关键点预测目标的中心点，而 FCOS 通过 mask 和 centernees 分支计算目标点在目标中心的概率。2.LNet 不需要非常精细的检测结果，所以不包含复杂的 FPN 模块，同时输出的结果是原图输入的 1/4。3.LNet 使用 DIOU loss 进行边框回归，而 FCOS 使用 IOU loss/GIOU loss 进行边框回归。LNet 与 CenterNet[93]的主要区别在于 1.LNet 直接回归点到边框的距离，而 CenterNet 回归边框的大小以及中心点的偏移。2.LNet 只检出前景，而 CenterNet 在检出前景的同时也进行了分类。3.LNet 直接采用连续的坐标位置，不存在偏移。而 CenterNet 使用离散的坐标，同时预测偏移量。4.LNet 回归关键点附近的所有点，而 CenterNet 只对关键点进行回归。

LNet 可以有效检测出目标物体的位置，为接下来的分割做准备，模型采用编解码结构，设计如下：

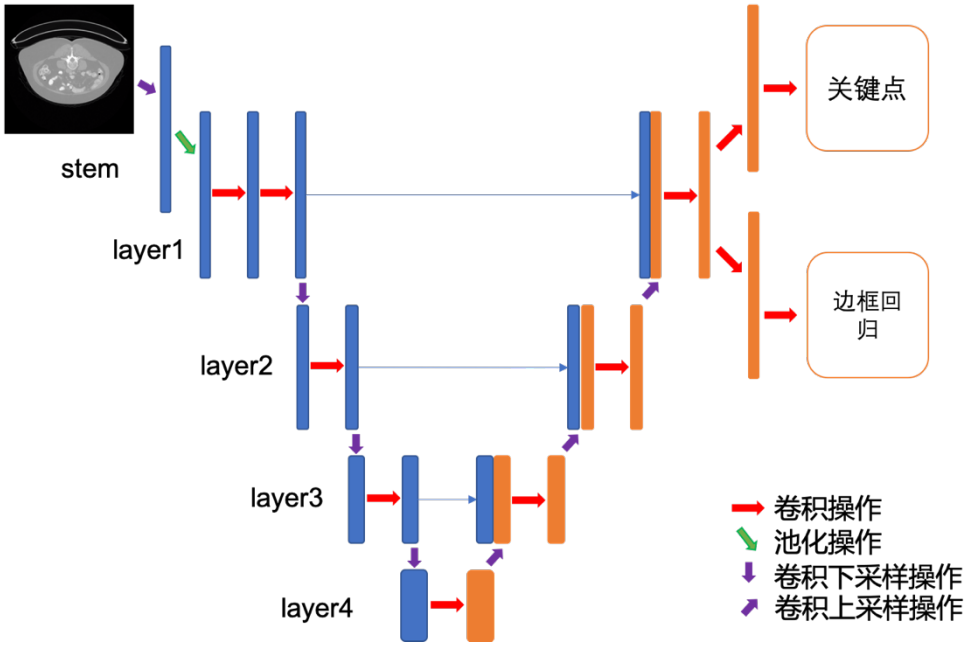

图 4.3 LNet 结构

LNet 模型基于 ResNet 改进并设计，输入大小为 320*320 的图片，经过卷积和 5 次下采样后，生成大小为 10*10 的特征图。再通过卷积和 3 次上采样，恢复特征大小到 80*80，在上采样的过程中，同时如图示连接潜层的特征，浅层高分辨率的信息与深层低分辨率语义信息融合后，可以有效防止模型在降采样后导致分辨率过低，通过 concat 浅层和深层特征的方式丰富模型的语义信息的同时保持较高的分辨率。最后，通过两个分支分别进行关键点预测和边框回归。值得一提的是，本文采样点的方式不同于以往论文中使用步长计算特征点对应原图位置的方式，而是直接取等分点的浮点坐标，同时在生成关键点目标信息和边框回归目标信息时，均保持精确的浮点操作，因此不会因为取整而出现位置偏移，更有利与模型学习。

1. 关键点预测

LNet 定义每个目标物体的中心为关键点，并且在一个 channel 上同时预测一类物体的所有关键点。关键点使用热图预测，热图生成的公式如下：

$$p_{ij} = \sum_c e^{-\frac{max(0,\ |x_i - x_c| - d)^2 + max(0,\ |y_j - y_c| - d)^2}{2 * var}} \tag{4.1}$$

式中p_{ij}为坐标$x_i y_j$的属于类别c的概率。$x_c y_c$为物体的中心点。var为方差系数，用于控制生产的关键点热力图辐射范围，本文默认设置为 20。d为范围系数，用于扩大关键点选取的范围，这里我们设置d为 3.96，因为最终输出的大小只有原图长宽只有原图的 1/4，为了保证至少距离原图最近的点能以概率 1 设为关键点，对距离要求条件放宽，同时用 max 函数保证距离大于 0。同一类别的关键点热图概率直接叠加，但保证概率最大为 1。生成的热图如下：

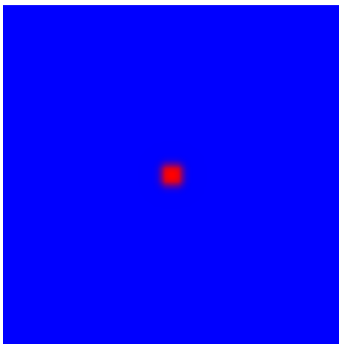

图 4.4 d取 3.96 时生成的热力图，颜色越红概率越接近 1，颜色越蓝概率越接近 0。

从图 4.4 中可以看出每一个目标都有足够的点用于检测出目标的中心位置，同时由于本文采用的是高斯核生成热力图，给模型学习提供了比简单 0/1 二值化方式更加精细的目标，更有利于模型学习到合理的信息。

2. 边框回归

边框回归是单独用来回归目标大小的分支，本文采用中心点到边框距离的四个值作为回归目标，分别记为l,r,t,b，四个值分别对应中心点到左端，右端，顶端，底端四个方向的距离，如下图所示：

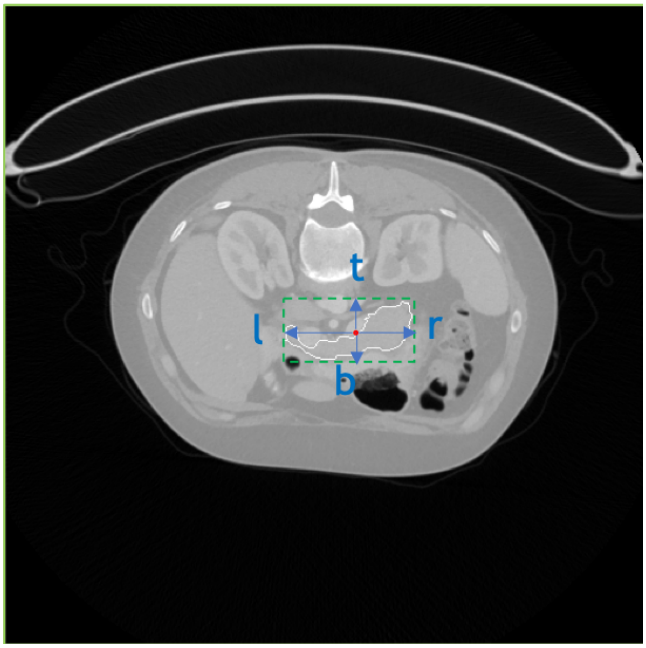

图 4.5 回归分支预测目标

4.2.3 SNet 部分

SNet 负责检出整图中的目标区域位置，通过仿射变换提取出目标区域后，使用 SNet 进行精细的分割，此时 SNet 可以专注于分割感兴趣区域，能减少背景和尺度不一致对模型分割效果的影响。本文中 SNet 部分采用了了传统 U-Net 的设计，使用 ResNet34 为 backbone 构建了 U-Net 网络，具体网络结构如下图所示：

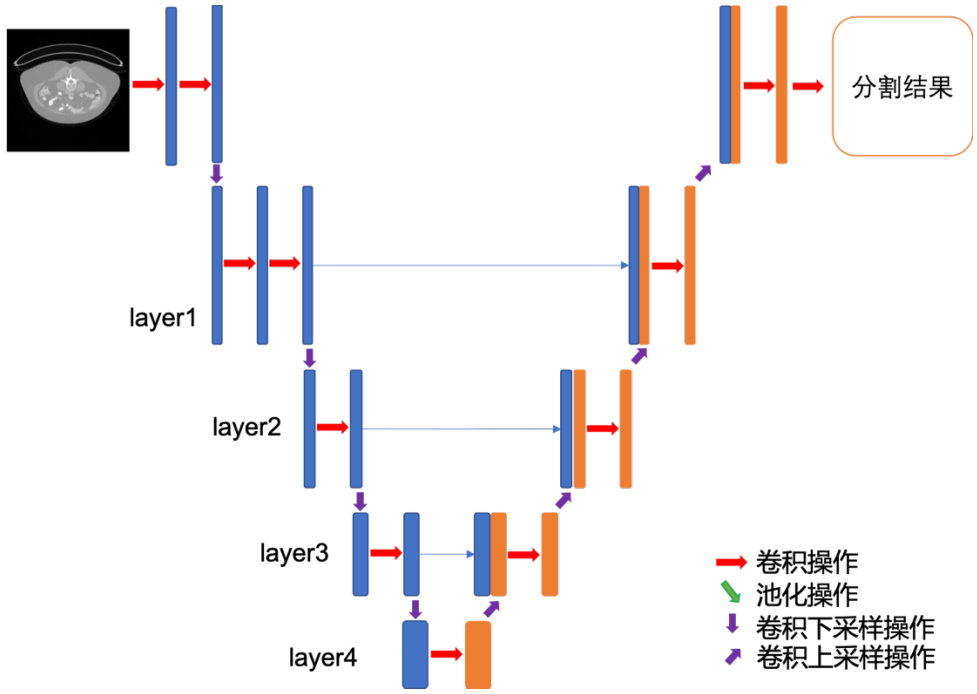

图 4.6 SNet 结构

4.2.4 loss 损失函数设计

损失函数包含两部分，一部分是定位模型 LNet 的损失函数，另一部分是分割模型 SNet 的损失函数。LNet 的 loss 函数用于引导模型训练关键点检测和边框回归，SNet 的 loss 函数用于引导模型学习正确的分割结果，同时由于整个网络结构是完全打通并且可导的，SNet 的 loss 也会影响 LNet 模型的学习。

Loss 定义如下：

$$loss = loss_{LNet} + loss_{SNet} \tag{4.2}$$

$$loss_{LNet} = \beta_1 * Dice\ Loss(pr_k, gt_k) + Focal\ Loss(pr_k, gt_k) + \lambda * DIoU\ Loss(pr_{bbox}, gt_{bbox}) \tag{4.3}$$

$$loss_{SNet} = \beta_2 * Dice\ Loss(pr_m, gt_m) + Focal\ Loss(pr_m, gt_m) \tag{4.4}$$

式中$loss_{LNet}$部分pr_k为关键点分支预测的结果，gt_k为对应的关键点 ground truth。此处使用了$Dice\ Loss$和$Focal\ Loss$共同监督学习关键点的信息，这里使用$Focal\ Loss$可以更好的学习关键点的信息，因为整图上的关键点数量非常小，样本严重不均衡，而$Focal\ Loss$可以很好的应对这种情况。结合$Dice\ Loss$可以更好的学习关键点的分布，通过调整$Dice\ Loss$和$Focal\ Loss$权重可以得到一个更佳的损失函数，本文设置β_1和β_2均为 0.2，$DIoU\ Loss$是针对$IoU\ Loss$的改进，可以更佳快速的收敛边框位置，并且达到更好的收敛效果。$loss_{SNet}$和$loss_{LNet}$类似，直接对分割预测优化$Dice\ Loss$和$Focal\ Loss$。

4.3 模型训练和推理

4.3.1 训练过程

本文采用交替的方式训练 LNet 和 L-SNet，详细参数信息如下：LNet 和 L-SNet 均使用 Adam 优化器，设置初始学习率 learning rate 均为 0.0001，每 25 个 epoch 降低学习率为原本的 1/10，设置 batch size 大小为 16。Backbone 中的 ResNet 结构部分加载 ImageNet 的预训练参数，模型共训练 60 个 epoch，前 15 个 epoch 交替训练 LNet 和 L-SNet，由于训练 L-SNet 的监督无法显式规定 LNet 的优化方向,而是以微调的形式优化 LNet 并减少两个步骤之前输入的偏差，我们在更新 L-SNet 时,对 LNet 模型部分的参数学习率缩小为 1/100。15 个 epoch 之后着重优化模型的分割效果，交替训练 LNet 和 L-SNet 时，每训练一次 LNet，训练三次 L-SNet，并且只有第一次更新 LNet 部分的权重。L-SNet 中间的的输入为 LNet 预测的结果，每次从 LNet 预测的关键点中，从每张输入图片中概率前 3 大的关键点中随机选取一个关键点,使用对应位置处的边框回归结果通过仿射变换取出要分割的区域。

在训练时，统一将输入图片 resize 到 320 * 320 的大小。为了能让模型有更好的泛化性，在训练时采用了在线增强的方式对数据进行了扩充。增强方式包括以 0.5 的概率进行 0.1 尺度范围内的随机缩放，15 度范围内随机旋转，0.1 尺度范围内的随机偏移，随机 90 度旋转，随机水平翻转，随机垂直翻转。以 0.8 的概率进行亮度和对比度变换。

4.3.2 推理过程

推理过程和训练过程类似，主要的区别在于感兴趣区域的选取方式不同，在推理时，检测 LNet 输出的关键点轮廓，对于每一个轮廓产生一个单独的感兴趣区域，此处不再使用最大概率检出关键点，而是取所有概率大于 0.5 的关键点形成的轮廓中心，作为检出的关键点。对每一个检出的关键点，都分别通过网络进行分割,最后把分割结果映射回原图的大小并且进行合并。整体流程如下图所示：

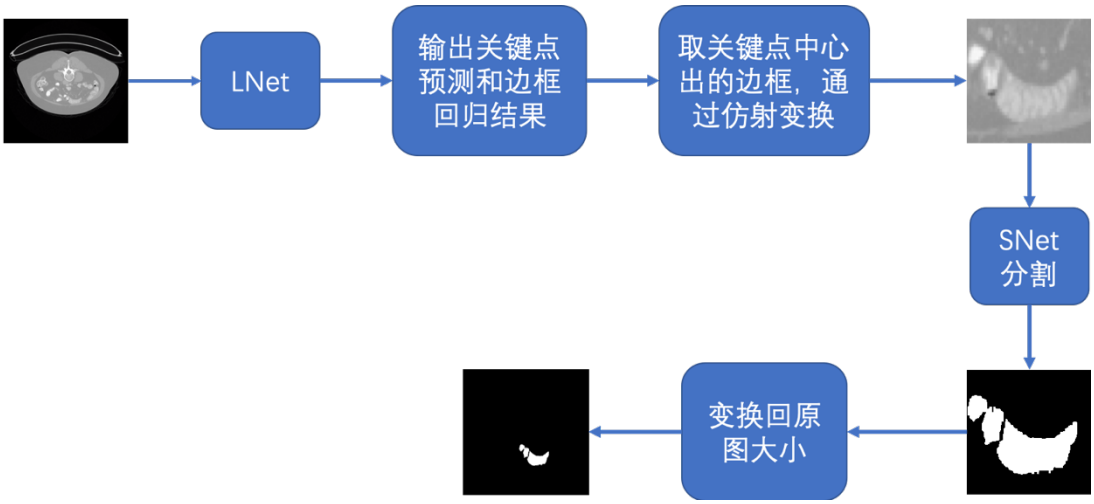

图 4.7 SN-U-Net 整体结构