高效自动分割　　多级肺动脉分割：PARSE 挑战赛

Gongning Luo[a], KuanquanWang[a], Jun Liu[*a], Shuo Li[b], XinjieLiang[a],XiangyuLi[a],ShaoweiGan[c],Wei Wang[*a],Suyu Dong[c], Wenyi Wang[a], Pengxin Yu[d], Enyou Liu[d], Hongrong Wei[e],NaWang[e],Jia Guo[f],Huiqi Li[f],Zhao Zhang[g],Ziwei Zhao[g],Na Gao[h], NanAn[i], Ashkan Pakzad[j], Bojidar Rangelov[j], Jiaqi Dou[k],SongTian[l],Zeyu Liu[m],YiWang[m],Ampatishan Sivalingam[n], KumaradevanPunithakumar[n], Zhaowen Qiu[*c], Xin Gao[*o]

[a]哈尔滨工业大学计算机科学与技术学院，哈尔滨 *150001*，中国

[b]凯斯西储大学计算机与数据科学系和生物医学工程系，美国克利夫兰 *44106*。[c]东北林业大学信息与计算机工程研究所，中国哈尔滨*150040*[d]推想医疗科技股份有限公司，中国北京[e]商汤科技，中国上海[f]北京理工大学，中国北京[g]北京大学，中国北京[h]东北大学，中国沈阳[i]上海交通大学，中国上海[j]伦敦大学学院（*UCL*），英国伦敦[k]清华大学生物医学影像研究中心，中国北京；[l]飞利浦医疗，中国[m]重庆大学，中国重庆[n]阿尔伯塔大学，加拿大埃德蒙顿[o]计算机、电气与数学科学& 工程部，阿卜杜拉国王科技大学，*4700 KAUST*，沙特阿拉伯图瓦勒 *23955*

arXiv:2304.03708v2 [eess.IV] 9 Aug 2024

文 章 信 息

文章历史：收稿日期：202x年xx月
xx日 终稿接收日期：202x年xx月
xx日 接受日期：202x年xx月xx日
在线发表日期：202x年xx月xx日

关键词：分割，肺动脉，多级，效率

摘 要

对 CTPA 图像中的多级（即主肺动脉和分支肺动脉）进行高效自动分割，对临床应用具有重要意义。然而，现有方法大多只分别关注主肺动脉或分支肺动脉分割，并忽视分割效率。此外，目前没有专门针对肺动脉分割的公开大规模数据集，这使得不同方法之间的比较极具挑战性。为评测多级肺动脉分割算法，我们组织了首届**肺部动脉分割**（PARSE）挑战赛。一方面，我们同时关注主肺动脉和分支肺动脉分割；另一方面，为更好地服务临床应用，在保证肺动脉分割准确性的同时，我们为分割效率（主要指推理期间的运行时间和 GPU 内存消耗）赋予同等的评分权重。我们总结了顶尖算法，并就如何实现高效、准确的多级肺动脉自动分割提出了一些建议。我们将 PARSE 挑战赛开放访问，供社区评测未来算法的发展：https://parse2022.grand-challenge.org/Parse2022/。

© 2024 爱思唯尔公司。保留所有权利。

1. 引言

作为多级血管系统，肺动脉（PA）由肺动脉干、左肺主动脉、右肺主动脉（统称为主肺动脉）以及肺内大量肺叶段支（统称为分支肺动脉）组成，负责将脱氧血从心脏输送出去。肺动脉与多种疾病相关，例如肺动脉高压（PH）（Gali`e 等人，2016年；Shahin 等人，2022年）和肺栓塞（Konstantinides 等人，2020年）。计算机断层肺血管造影（CTPA）是一种广泛用于多种肺动脉疾病诊断、可视化和治疗规划的医学成像技术（Gali`e 等人，2016年；Konstantinides 等人，2020年）。一项前提步骤是从CTPA中准确识别并分割多级肺动脉结构，

*通讯作者：

准确性和效率（Est´epar 等人，2013；Shahin 等人，2022年）。

近年来，越来越多的研究表明，肺部分支血管的提取对于肺部疾病的诊断和治疗具有至关重要的意义。Shahin等人（Shahin 等人，2022年）重点研究了肺部血管和小肺血管的大小及体积，并发现它们与肺动脉高压（PH）的亚型相关。Poletti等人（Poletti 等人，2022年）将肺部区域划分为三个等体积部分，包括外周区、中间区和中央区，分别对应不同大小的血管。他们分析了正常肺和患病肺（流感和严重急性呼吸综合征冠状病毒2型）三个区域中的血管体积，并发现血管体积发生了显著变化，包括中间区和中央区的分支血管，以及中央区的主血管。

然而，如图 1 所示，准确且高效的肺动脉分割仍然具有挑战性。**(1)复杂的拓扑结构：** 从主肺动脉到分支肺动脉，存在大量分叉（图1 (b)），尤其是在分支的远端。沿动脉的分叉过程伴随着血管横截面积的减小和血管密度的增加，这表明肺动脉分割的难度不断提高。肺动脉还具有与肺部气道和肺静脉相似的拓扑结构，尤其是肺静脉，如图 1 (a)所示，肺静脉通常与肺动脉相互交织。**(2)大而精细的感兴趣区域（感兴趣区域）：** 需要分割的肺动脉区域范围很大，主肺动脉和精细的分支肺动脉都需要进行准确分割。因此，很难同时保证分割准确率和效率。**(3)类间不平衡：** 前景（肺动脉区域）与背景之间存在严重的不平衡，尤其是在分支血管末端。**(4)类内不平衡：** 主肺动脉在 CT图像中包含更多体素，因此相对容易分割。然而，分支肺动脉包含的体素较少，在分割任务中容易出现断裂或漏分。**(5)较大的个体差异：** 不同个体自然肺动脉的外观和几何形状各不相同，加上疾病因素的影响，进一步导致了变异性。造影剂的使用以及成像设备参数的设置可能会引入伪影和固有图像噪声，从而增加成像变异性。**(6) 标注难度：** 由于拓扑结构极其复杂，难以对所有肺动脉进行完整且准确的标注。标注可能存在噪声。

在过去几十年里，据我们所知，鲜有公开数据集专门针对肺动脉分割。VESSEL12（Est´epar 等，2013）挑战旨在实现计算机断层扫描（CT）影像中的自动肺血管分割，仅发布了三个标注样本。CARVE14（Charbonnier 等，2015）旨在基于 55 例非增强 CT 扫描进行肺动静脉分离和分类，这些扫描来自 ANODE 挑战（Van Ginneken 等，2010）。然而，VESSEL12 和 CARVE14 数据集的人工标注都仅针对感兴趣点，而非整个血管树，因此难以用于深度学习训练。Tan 等人（Tan 等，2021）回顾了 2020 年国际图像计算与数字医学研讨会挑战中的自动肺血管分割算法，该挑战包含 16 组 CT 平扫图像和 16 组计算机断层扫描血管造影（CTA）增强图像，但未提供数据集下载链接。因此，目前仍缺少专门用于肺动脉分割的大规模数据集。

为解决这些局限性，我们在新加坡举办的2022年国际医学图像计算与计算机辅助手术大会（MICCAI）期间，组织了肺动脉分割（PARSE）挑战赛。参赛者需开发自动主肺动脉和分支肺动脉分割算法，并根据分割准确率和效率进行评估，以确定最终排名。

本文介绍了 PARSE 挑战赛的概况，并讨论了顶尖算法。主要贡献总结如下：

- 我们组织了首个肺动脉分割挑战赛，旨在-

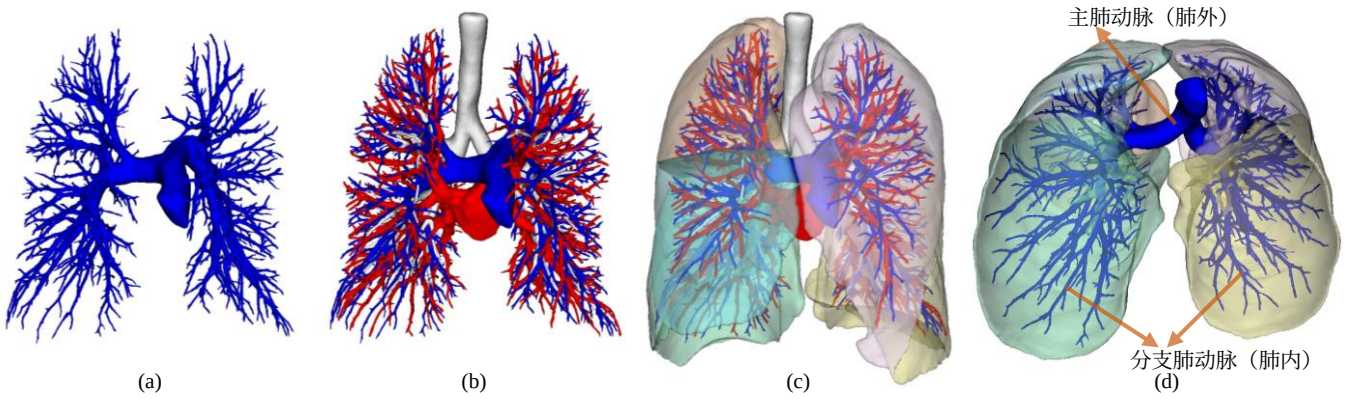


图 1. 复杂的肺动脉拓扑结构：(a) 肺动脉（蓝色）；(b) 肺动脉与肺静脉（红色）相互交织，而肺气道（灰色）使结构更加复杂；(c) 肺动脉、肺静脉和肺气道及肺部（阴影区域）；(d) 肺动脉分为两个层级：主肺动脉（肺外）和分支肺动脉（肺内）。

同时聚焦主肺动脉和分支肺动脉分割，以同时实现高准确性和高效率。

- 我们分析了提交的算法，并总结了表现最佳团队的结果。

- 我们介绍了多种提高肺动脉分割准确性和效率的算法技术，并提出了相关建议。

2. 方法

2.1. 挑战赛数据集

PARSE挑战赛提供了来自203名受试者的计算机断层肺血管造影图像集，这些受试者均已被诊断患有肺结节性疾病。平面内尺寸和分辨率分别为 512 × 512 像素和0.5 ~ 0.95毫米/像素。切片数量和厚度分别为 228 ~ 408和1毫米。所有数据均已匿名化。数据集划分及详细参数见表1。其中，年龄、切片数量和像素间距均以最小值、最大值和均值表示。这些数据来自四个中心（哈尔滨医科大学附属肿瘤医院、中国人民解放军总医院、牡丹江医科大学附属红旗医院和佳木斯大学附属第一医院）以及两家设备制造商（西门子和飞利浦）。

标注工作由 10 位具有 5 年以上临床经验的专家参与，他们被分为两组。每组包含 5 位专家，每个扫描均由 5 位专家进行标注。标注使用 MIMICS 软件，基于区域生长法半自动完成。首先，调整窗宽和窗位，以清晰显示肺动脉结构。其次，迭代并手动选择种子点，生成粗略掩膜。第三，所有专家对肺动脉结构的掩膜结果进行微调，并相互核查标注。最后，采用投票法合并每个扫描的 5 份标注。

数据被分为三个集合，即训练集、验证集和测试集，病例数分别为100、30和73。参与者可以访问带有相应标注的训练图像，以及不带标注的验证图像。验证集标注和带有相应标注的测试图像由组织- 者保管。

2.2. 挑战赛组织

在组织PARSE挑战赛和撰写本文时，我们遵循了BIAS（生物医学图像分析挑战赛）指南（Maier-Hein 等，2020）。PARSE 挑战赛包括训练阶段、验证阶段和测试阶段。在训练阶段，参与者的申请获批后，可以使用训练图像及相应的掩膜开发并交叉验证稳健的分割算法。在验证阶段，只能上传验证图像的三项分割结果

表1. 图像信息（数量，通用der、年龄、供应商、中心和物理尺寸）训练集、验证集和测试集的。所有层厚均为1 mm.

	样本性别		Age	西门子			飞利浦	切片数量	像素间距
	Nums	（女性/男性）		中心1	中心2	中心3	Center4		(mm)
训练	100	79/21	29-71（56.3）	81	16	3		228-390（303）	0.5039-0.9238（0.6740）
验证	30	25/5	29-70（54.2）	25	3	2		228-397（295）	0.5469-0.7891 (0.6651)
测试	73	52/21	23-73 (54.4)	61	9		3	229-408 (312)	0.5039-0.8770 (0.6906)

提交至官方网站，以便每天验证算法，分数也可以自动显示。参与者开发出方法后，在测试阶段结束前提交了能够通过特定命令生成分割结果的 Docker 容器。每个团队只有一个可正常运行的 Docker 容器被确认能够生成最终测试结果。每个排名靠前的团队都提交了一篇包含充分方法细节的短篇论文。

为确保公平比较，本次挑战赛不允许使用预训练模型和额外的训练数据。所有排名靠前的团队均获得了特定证书，前3名团队还获得了现金奖励。他们的最终结果公开展示。前10名团队受邀在MICCAI 挑战赛会议上展示其优秀算法，并担任挑战赛综述论文的共同作者。

2.3. 评估指标和排名方案

由于肺动脉目标具有较大的感兴趣区域（ROI）且需要分割精细结构，因此很难在确保分割准确率的同时确保分割效率。然而，临床应用需要准确性和效率指标（Ma等，2022）。因此，受 FLARE 挑战赛（Ma等，2022）的启发，我们采用了面向准确性和面向效率的评估指标。具体而言，对于面向准确性的指标，选择了广泛使用的Dice 相似系数（DSC）和95% Hausdorff距离（HD95）。对于面向效率的指标，选择了运行时间（RT）和最大使用的 GPU 内存（GPU），与 FLARE 挑战赛类似。

如图1所示，PARSE 的分割目标大致可分为两个级别，即主肺动脉和分支肺动脉。将二者一视同仁并不合适，因为分支肺动脉的提取难度要大得多。因此，我们针对多级肺动脉引入了多级准确性指标。

具体而言，我们采用了两级DSC 和 HD95 指标：(1)第一级：位于肺区内的分支肺动脉的DSC 和 HD95 指标；(2)第二级：位于肺区外的主肺动脉的 DSC 和HD95指标。值得注意的是，我们对分支肺动脉分割指标（第一级）赋予了更高的权重，因为较细的结构更难进行分割，并且对临床应用更为关键（Shahin 等人，2022年）。具体而言，第一级和第二级在最终得分中的权重分别为 80% 和 20%。

获得所有团队的最终得分后，我们采用了以下排名方案：

- 步骤 1. 分别计算每个团队所有测试病例的加权 DSC 和 HD95 得分，再取均值，得到各团队的平均 DSC 和 HD95 得分。随后，计算每个团队的平均运行时间和 GPU 得分。
- 步骤 2. 分别对所有团队的加权 DSC、加权 HD95、运行时间和 GPU 平均得分进行排名。
- 步骤 3. 对每个团队的这四项排名取平均。
- 步骤 4. 根据步骤 3 中的平均排名确定最终排名。如果两个团队的平均排名相同，则并列排名。

3. 结果

3.1. 挑战赛参与者与提交内容

官方 PARSE 挑战赛位于 grand-challenge 平台[1]。参与者应签署挑战赛规则协议，并将其发送至官方邮箱。图 2 展示了

[1] https://parse2022.grand-challenge.org/Parse2022/

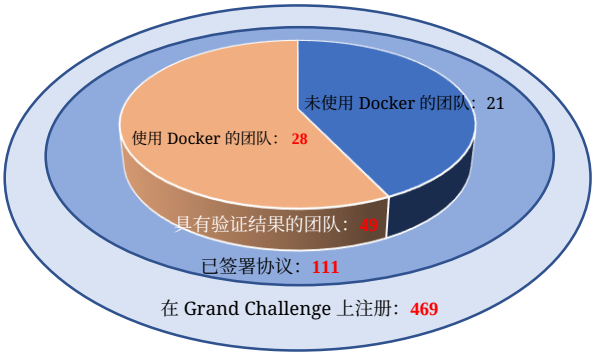


图 2. PARSE 挑战赛参与者及提交情况总结。在官方 grand-challenge 网页上共有 469 个团队注册，其中 111 个团队在训练阶段结束前获批。最终，49 个团队提交了验证结果，28 个团队提交了 Docker 容器，其中 25 个结果符合资格。

参与者和提交情况的信息。具体而言，我们在 grand-challenge 网页上收到了来自 30 多个国家的 400 多份申请，其中 111 个团队提交了完整的签署申请表并获批。在验证阶段，49 个团队提交了验证结果。在测试阶段，28 个团队提交了 Docker 容器。但其中 3 个 Docker 容器无法正常运行。最终，我们获得了 25 份合格的提交结果。

3.2. 十大团队的算法总结

表2总结了排名前十的算法的要点。它们提升分割准确率和效率的关键策略如下所示。

3.2.1. T1：*Infervision Medical TechnologyCo.,Ltd*（*IMT*）

IMT 开发了一种高效的U-Net(Ronneberger等，2015)框架，并采用了残差块(He 等，2016年)，如图 3 所示。在训练阶段，IMT从三个方面（即数据、模型和损失函数）进行了设计，以在分割准确率和效率之间实现最优的权衡。对于数据，IMT提出了以下特殊设计：(1) 使用两个窗口裁剪图像的强度，并将其构建为双通道输入；(2) 使用困难样本挖掘策略[2] 来

[2]训练模型生成预测概率图后，将预测概率介于0.3至0.5之间的正类体素，以及预测概率介于0.5至0.8之间的负类体素定义为难分类样本

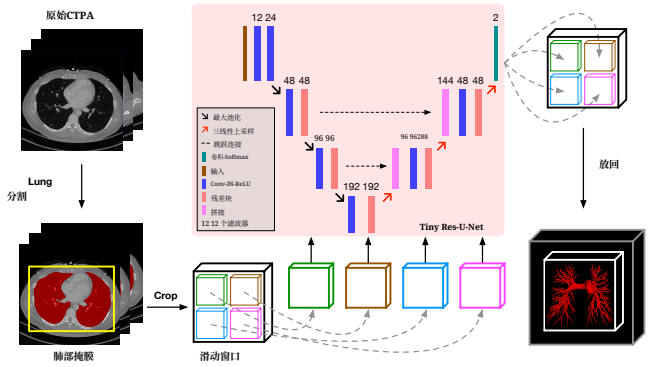


图3. 排名前一团队的框架流程。该团队设计了一种名为 Tiny Res-U-Net 的轻量级模型。

对训练数据进行采样，使难以区分的区域得到更多关注；(3)通过阈值法进行肺部区域分割，以提取感兴趣区域，这将有助于提高推理速度和准确性。对于模型，IMT构建了一个名为TinyRes-U-Net 的轻量级模型，如图 3 所示，用于执行肺动脉分割。具体而言，TinyRes-U-Net仅进行三次下采样，并且未针对原始分辨率使用跳跃连接。此外，其通道数较少，因此可以同时保证运行时间和 GPU 内存消耗。进一步地，IMT 通过实验发现，更深或更宽的网络并不能改善性能。对于损失函数，IMT修改了广义并集损失（Zheng等，2021年），并使用加权交叉熵（CE）损失作为辅助损失，该损失仅计算肺部区域内体素的损失。在推理过程中，IMT 将模型和数据转换为半精度，这可以有效减少GPU内存消耗，同时几乎不会影响准确性。此外，IMT 对模型集成使用了参数平均。IMT发现，尽管其表现略逊于多个模型集成后的预测结果，但不会引入额外的计算。除此之外，还进行了连通区域分析，以减少假阳性结果。

3.2.2. T2：商汤科技（*ST*）

ST 从多个方面改进了 nnU-Net（Isensee 等人，2021年）。（1）数据增强：在原始 nnU-Net 的数据增强基础上加入弹性变形和全局强度扰动，以提升模型对 PARSE 数据的泛化能力。具体而言，弹性变形的概率设为

表2。前十名团队基准方法汇总。

Team	框架网络		重点方法
IMT	单阶段	带残差块的 U-Net	通过阈值法提取的肺区; 改进的广义并集损失和加权方式 交叉熵损失; 困难样本挖掘策略; 参数平均.
ST	单阶段	nnU-Net	通过阈值法提取的肺区; 适当的数据增强; 过采样机制; 选定的测试时视图增强。
BIT	两阶段	带有残差块的 U-Net	由训练好的网络提取的肺区; 标签噪声应对; 加权斑块组装。
PKU	单阶段	骨架感知型 U-Net	通过阈值法提取的肺区及 连通分量发现; 添加骨架解码器。
IR	单阶段	U-Net	自适应滑动窗口策略; 移除空气操作; 深度监督。
NEU	单阶段	双U-Net	通过阈值法提取的肺区; 具有不同输入尺寸的两个 U-Net 结构和 下采样次数.
UCL	单阶段	nnU-Net	形态学闭运算; 模型集成 .
THU	单阶段	U-Net	两个训练阶段; 使用 clDice 进行微调。
CQU	两阶段	U-Net	由粗到细策略。
IITBHU	单阶段	nnU-Net	图像方向策略。

0.2 使用 $\alpha \in [0, 1000]$ 和 $\sigma \in [9, 13]$，而服从高斯分布N(0,0.1)的全局强度扰动概率设为0.15。（2）过采样机制：将分支肺动脉的采样概率提高至 0.7，使网络更加关注该分支。（3）特定的测试时增强（测试时增强）视图：评估每个增强视图（即翻转显式轴）的分割性能，仅保留四个关键的测试时增强视图，以平衡准确性和效率。此外，通过阈值法提取肺区，以加快推理并减少假阳性。

3.2.3. T3：北京理工大学（*BIT*）

BIT提出了一种3D Res-Unet，采用带有残差块的U-Net骨干网络（He等，2016年）。BIT采用实例归一化（Loshchilov和Hutter，2018年）（IN）代替批归一化（BN），因为前者对小批量大小更加友好。为了缓解主肺动脉中的标签噪声，BIT将五折交叉验证的预测结果与原始真实标注进行整合，以重新构建标签。随后，使用重新构建的标签训练最终网络。在推理过程中，使用由传统方法提取肺部掩膜训练的神经网络来裁剪肺区。BIT认为，该方法提取肺部区域的速度比传统方法更快。由于读取和写入Nifty文件（.nii.gz）耗时较长，BIT使用异步线程与主处理过程并行进行读写，以提高推理效率。

3.2.4. T4：北京大学（*PKU*）

与肺动脉掩膜相比，PKU 认为肺动脉骨架具有以下优势：（1）骨架结构不受血管厚度影响；（2）骨架能更准确地反映肺动脉的拓扑结构。因此，PKU 在 Res-Unet（Zhang et al., 2018）（图 4）中增设了一个骨架解码器，用于在训练阶段提取血管骨架，并在推理阶段将其弃用。具体而言，骨架解码器的输入是来自 Res-Unet 解码器的 $4\times$ 下采样特征图。研究引入了三个线性层，分别使用焦点损失（Lin et al., 2017）、二元交叉熵损失和 L1 损失，预测各位置是否存在肺动脉、目标性分数以及相对于区域中心点的坐标偏移，以此得到目标性分数和相对于区域中心点的坐标偏移。此外，通过阈值法和连通分量分析提取肺部掩膜，从而有效且高效地消除假阳性预测。

3.2.5. T5: 独立研究者 *(IR)*

IR 采用 U-Net（Ronneberger 等人，2015）作为骨干网络，并参考了 nnU-Net（Isensee 等人，2021）的实现经验。在将图像输入网络之前，使用自适应滑动窗口策略并去除图像中的空气，以减少图像块数量并提高效率。此外，还采用了深度监督来缓解假阳性问题。

3.2.6. T6：东北大学（东北大学）

东北大学构建了一个三维双U-Net，由两个输入尺寸（分别为 [128, 160, 192]和 [112, 128, 160], ）和下采样次数（分别为 5 和 4）不同的 3D U-Net（C¸ic¸ek 等，2016年）结构组成。东北大学认为，输入尺寸较大的 U-Net 更擅长提取全局特征，从而减少假阳性；而输入尺寸较小的 U-Net 在提取局部特征方面具有优势，因此有助于分割前景。具体而言，首先通过阈值法提取肺部区域，并根据肺部边界框裁剪图像以获得感兴趣区域。随后，将两种尺寸的图像块（[128, 160,192] 和 [112, 128, 160]）分别输入两个网络。最后，将两个网络最后一层的输出逐元素相加，以预测肺动脉掩膜。此外，推理过程中还使用肺部掩膜过滤假阳性，以提高准确性和效率。

3.2.7. T7：伦敦大学学院（*UCL*）

UCL 使用 nnU-Net（Isensee 等人，2021年）构建了多个模型，并对其预测结果进行集成，以生成最终输出。具体而言，共有四种模型：仅二维、三维低分辨率、三维全分辨率和三维级联模型。此外，还对最终输出进行形态学闭运算

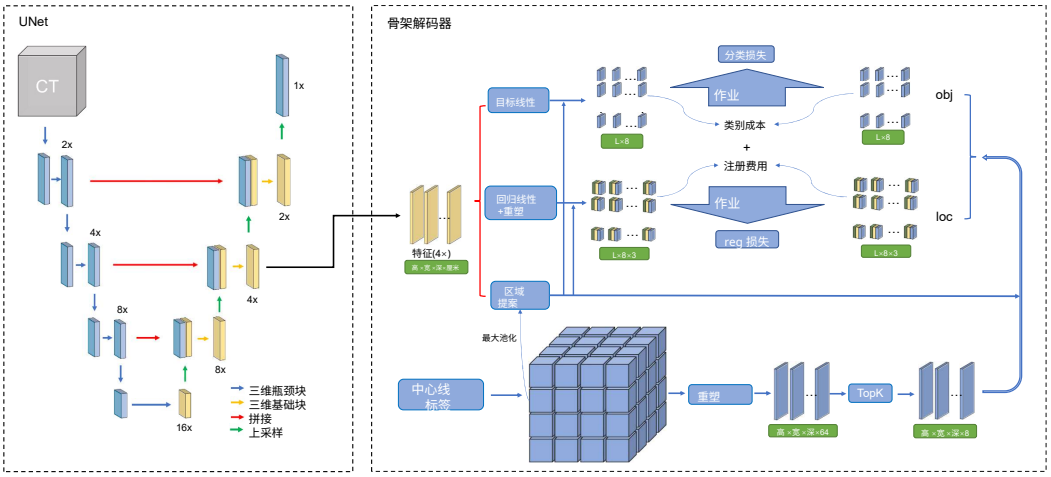


图4. T4团队的框架概览。该模型由一个用于执行分割的U-Net架构，以及一个在训练阶段用于预测血管骨架的骨架解码器组成。

并结合强度阈值处理，以改善连通性。

3.2.8. T8：清华大学 *(THU)*

THU 使用两个训练阶段开发了一个 3D U-Net（Ronneberger 等，2015）。模型首先使用 Dice 损失和 CE 损失进行训练，以快速定位肺动脉区域。然后，使用 clDice 损失（Shit 等，2021年）对模型进行微调，使其专注于肺动脉真实标注的中心线。

3.2.9. T9：重庆大学（*CQU*）

CQU提出了一种基于U-Net（Ronneberger 等，2015）的由粗到细框架。粗模型和细模型采用相同的网络架构，但输入尺寸不同，分别为 $160 \times 160 \times 160$ 和 $96 \times 96 \times 96$。图像首先经过调整大小，然后输入粗模型以定位感兴趣区域（ROI）。随后，根据粗模型的输出对图像进行裁剪，并将裁剪后的图像输入细模型，以实现精细分割。

3.2.10. T10：阿尔伯塔大学 - 印度理工学院（*BHU*）瓦拉纳西分校（*UA-IITBHU*）

UA-IITBHU 开发了一种基于 nnU-Net（Isensee 等人，2021年）的图像方向策略。在训练阶段，首先通过移除 NIFTI 图像头文件中的位置信息和方向信息，将图像及其对应的标签调整为与坐标轴平行，然后将它们送入网络进行训练，按照 the nnU-Net 框架的标准设置。在推理过程中，首先按照上述描述对图像进行定向，然后将其输入训练好的模型以获取预测结果，随后重新定向以匹配原始图像，从而得到最终预测结果。

3.3. 评估结果与排名分析

表3展示了25支合格队伍在主肺动脉、分支肺动脉和加权PA上的平均DSC和HD95，以及平均RT和GPU指标。在接下来的小节中，我们基于ChallengeR（Wiesenfarth 等，2021）对表3中的DSC和HD95指标进行结果分析，采用点图和箱线图可视化以及统计显著性图（图5和图6）进行展示。ChallengeR是一种用于挑战赛结果分析和可视化的专业工具。统计显著性图采用显著性水平为5%的单侧 Wilcoxon 符号秩检验进行分析，该检验已广泛应用于许多挑战赛结果分析中（Wiesenfarth 等，2021；Ma 等，2022）。随后，我们对RT和GPU指标进行比较分析。为确保描述简洁且明确，我们主要关注前20名团队。

3.3.1. DSC 指标分析

如表3和图5（a - c）所示，所有团队在主肺动脉上的 DSC 得分均优于分支肺动脉。加权后，由于分支肺动脉获得了更高的权重，得到了偏向分支肺动脉的平衡得分。同样，从表3可以看出，T2、T3 和 T8

表3. 符合条件的25个团队的定量评估结果，指标包括（均值 ± 标准差）Dice 相似系数（DSC）、95% Hausdorff 距离（HD95）、每例运行时间（运行时间）和 GPU 最大内存消耗（GPU）。

团队	DSC（%）↑			HD95（mm）↓			RT ↓	GPU ↓
	主肺动脉	分支肺动脉	加权PA	主肺动脉	分支肺动脉	加权PA	（秒/病例）	(MB)
T1	89.70 ± 6.46	**77.19 ± 8.32**	**79.69 ± 6.95**	7.08 ± 6.62	**4.80 ± 4.12**	5.26 ± 3.54	7.92	**1674**
T2	**91.71 ± 4.00**	76.67 ± 8.39	**79.68 ± 6.94**	**4.83 ± 4.97**	**4.74 ± 4.17**	**4.75 ± 3.66**	18.61	3658
T3	**91.57 ± 4.25**	**76.86 ± 7.94**	**79.80 ± 6.61**	**4.99 ± 5.11**	5.46 ± 4.77	5.36 ± 4.11	14.53	6300
T4	89.50 ± 6.58	75.53 ± 8.85	78.33 ± 7.49	6.76 ± 6.42	4.92 ± 4.00	5.29 ± 3.57	**6.63**	3326
T5	89.98 ± 4.94	**76.76 ± 7.88**	79.40 ± 6.38	7.74 ± 7.72	5.70 ± 4.87	6.10 ± 3.72	63.11	2828
T6	90.05 ± 5.65	76.76 ± 8.15	79.42 ± 6.89	8.38 ± 11.32	5.44 ± 4.61	6.03 ± 4.53	25.55	8176
T7	90.49 ± 5.13	76.57 ± 8.29	79.36 ± 6.80	6.37 ± 6.30	**4.56 ± 3.85**	**4.92 ± 3.35**	263.24	3658
T8	**90.83 ± 4.68**	76.25 ± 8.29	79.17 ± 6.80	9.19 ± 19.91	5.17 ± 4.13	5.98 ± 5.22	57.57	4120
T9	90.29 ± 5.04	76.53 ± 8.33	79.28 ± 6.87	6.50 ± 6.13	5.36 ± 4.51	5.58 ± 3.86	55.71	9303
T10	90.45 ± 5.20	76.10 ± 8.71	78.97 ± 7.16	6.18 ± 6.30	4.93 ± 4.34	5.18 ± 3.60	380.57	3472
T11	90.08 ± 5.37	75.71 ± 8.68	78.58 ± 7.29	9.19 ± 15.00	4.94 ± 3.72	5.79 ± 4.19	88.93	3658
T12	89.69 ± 5.30	75.46 ± 8.40	78.31 ± 6.94	31.65 ± 57.21	5.91 ± 4.55	11.06 ± 11.81	**7.74**	4298
T13	90.12 ± 4.82	74.30 ± 9.06	77.46 ± 7.53	18.43 ± 40.91	5.52 ± 4.24	8.10 ± 9.07	51.14	2800
T14	90.55 ± 4.63	68.86 ± 9.70	73.20 ± 7.97	6.24 ± 6.17	17.26 ± 6.32	15.06 ± 5.25	8.11	**1894**
T15	90.55 ± 5.13	76.29 ± 8.58	79.15 ± 7.12	**6.05 ± 6.58**	4.71 ± 4.05	**4.98 ± 3.46**	218.82	10242
T16	90.34 ± 4.86	74.98 ± 8.78	78.05 ± 7.17	6.50 ± 6.14	5.18 ± 4.18	5.44 ± 3.58	116.40	4988
T17	89.87 ± 5.04	70.00 ± 9.79	73.98 ± 8.08	6.96 ± 7.32	8.81 ± 7.09	8.44 ± 5.77	**4.21**	8230
T18	90.19 ± 4.43	71.30 ± 8.83	75.08 ± 7.28	6.13 ± 5.56	7.84 ± 3.09	7.50 ± 2.74	224.22	2328
T19	89.74 ± 4.85	70.41 ± 7.91	74.28 ± 6.63	7.20 ± 6.04	15.87 ± 4.34	14.13 ± 3.69	172.93	2640
T20	89.84 ± 4.85	75.29 ± 7.52	78.20 ± 6.19	20.65 ± 37.27	8.28 ± 4.74	10.75 ± 7.58	80.11	9424
T21	89.80 ± 5.24	74.76 ± 8.56	77.77 ± 7.14	11.77 ± 27.81	6.10 ± 4.24	7.24 ± 6.32	76.69	12288
T22	90.55 ± 4.94	76.05 ± 8.31	78.95 ± 6.89	9.14 ± 18.45	5.22 ± 4.26	6.00 ± 5.06	3178.75	**0**
T23	80.65 ± 9.56	62.53 ± 7.19	66.15 ± 6.42	140.75 ± 69.78	11.24 ± 2.84	37.15 ± 14.5	48.22	21996
T24	82.65 ± 25.41	69.37 ± 22.35	72.02 ± 22.61	–	–	–	419.20	3349
T25	85.79 ± 18.82	69.78 ± 16.4	72.98 ± 16.52	65.12 ± 69.46	–	–	111.50	9784

在主肺动脉方面取得了 DSC 性能前三名，而 T1、T3 和 T5 在分支肺动脉方面取得了 DSC 性能前三名。因此，在主肺动脉方面表现更好的团队，在分支肺动脉方面可能得分较低，反之亦然。这表明，有必要分别评估多级肺动脉（主肺动脉和分支肺动脉）。

统计显著性分析（图5 (d)）显示，T2 在主肺动脉方面的 DSC 得分与 T3 相比无显著差异。然而，T2 和 T3 在主肺动脉方面的 DSC 得分均显著优于其他团队。对于分支肺动脉，统计显著性分析（图5 (e)）显示，T1 的最高得分在统计学上显著高于所有其他团队，而 T3 和 T5 的最佳得分则显著高于部分其他团队。加权后，排名前三的 DSC 得分（T3、T1 和 T2）彼此之间没有明显差异，但在统计学上显著优于大多数其他团队的得分（图5 (f)）。因此，采用加权方法对多级肺动脉（主肺动脉和分支肺动脉）进行综合评估具有重要意义。

具体而言，团队T1 在分支肺动脉上取得了最佳DSC得分（表3），并且显著优于所有其他团队（图5（e））。图7中团队T1 的误差图表明，T1 在分支肺动脉区域的分割误差最小。然而，T1在主肺动脉上的得分表现不佳（表3），尤其是在肺动脉根部区域（参见图 7中的误差图）。此外，T2 在主肺动脉上取得了最佳DSC 得分（表3），并且显著优于大多数团队（图5（d））。图7中团队 T2 的误差图显示，T2 在主肺动脉区域的分割误差最小。

*3.3.2. HD95*指标分析

如表3和图6（a - c）所示，与DSC分数不同，大多数团队在分支肺动脉上的HD95分数优于主肺动脉，表现为平均值更低、离散程度更小。这一结果可能归因于分支肺动脉结构被肺组织包围，且HD95的计算范围局限于肺内部。此—

此外，主肺动脉、分支肺动脉和加权PA的 HD95 得分排名前三的团队分别为 [T2、T3、T10], [T7、T2、T1] 和 [T2、T7、T10], 。因此，与 DSC 得分较高的团队类似，各团队并不总能在主肺动脉和分支肺动脉的 HD95 上同时取得高分。此外，DSC 和 HD95 得分排名靠前的团队截然不同。因此，DSC 和 HD95 是互补指标，有必要同时引入这两项指标来评估算法。

具体而言，如表3以及图6(b)和(e)所示，T1在分支肺动脉方面获得了最高的HD95得分，并且显著优于大多数团队。同时，T7的DSC得分并不低。因此，T7在分支肺动脉区域的分割误差（图7）相对较小。相比之下，T10的DSC和HD95得分均不太理想（表3）。因此，T10在主肺动脉和分支肺动脉区域存在更多分割错误（图 7）。

如统计显著性图（图5（d - f）和图6（d - f））所示，只有排名前三的团队在主肺动脉的 DSC 和 HD95 评分上取得了显著更好的结果，而大多数团队在分支肺动脉上取得了此类结果。这些结果表明，评估分支肺动脉是一种更稳健的方法，并且给予分支肺动脉更高的权重是合理的。

3.3.3. 运行时间和 *GPU* 内存最大使用量分析

运行时间和 GPU 是注重效率的指标，与临床应用密切相关。图8展示了前20名团队的运行时间和 GPU 测评分数，以便更好地呈现结果。一些团队，如 T3、T6 和 T17，更关注运行时间指标，而忽略了 GPU 指标。相反，另一些团队，如 T19、T18、T7 和 T10，则更关注 GPU 指标，而忽略了运行时间指标。值得注意的是，一些团队，如 T1 和 T14，能够同时取得出色的运行时间和 GPU 分数。因此，如果处理得当，临床应用中的肺动脉分割推理速度可以加快，同时减少 GPU 使用量。

4. 讨论

4.1. 排名稳定性分析

为了在分割准确率和效率之间取得更好的平衡，四项互补指标（DSC、HD95、运行时间和

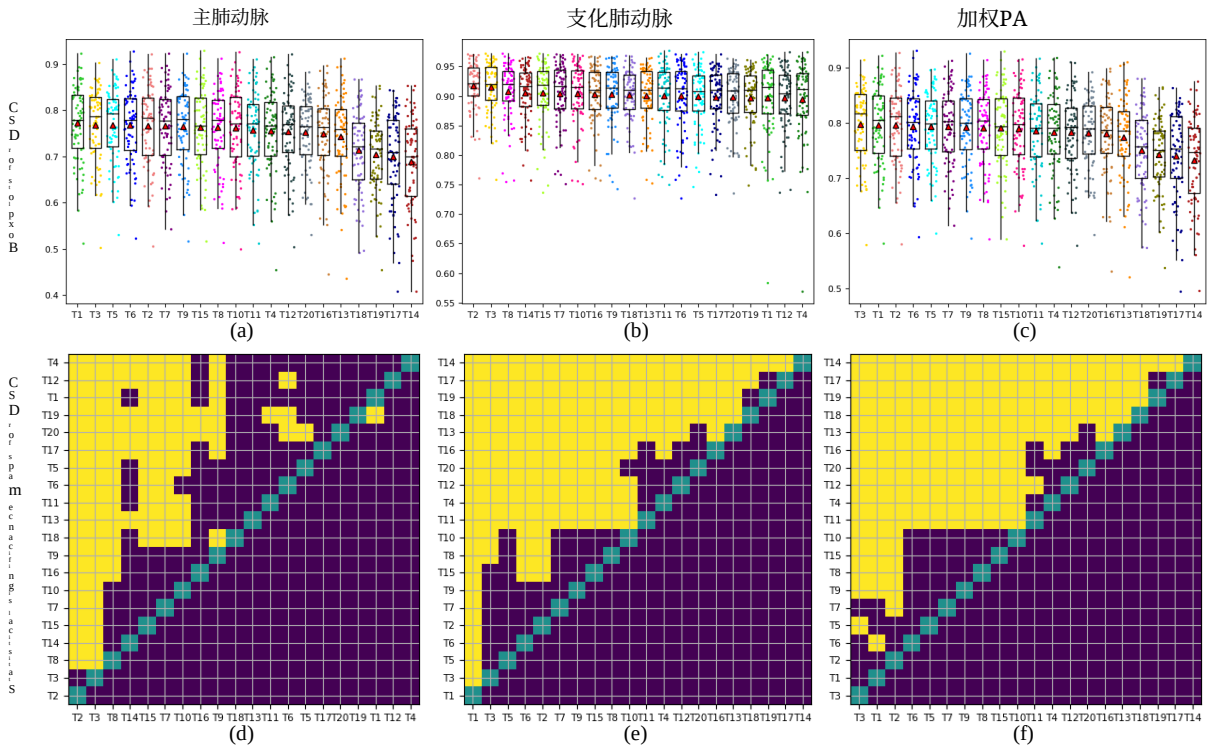


图5. 前20名团队的 DSC 指标点图和箱线图可视化（第一行）以及统计显著性图（第二行）。(a) 和 (d)、(b) 和 (e)、(c) 和 (f) 分别为主肺动脉、分支肺动脉和加权PA的结果。在每个子图中，团队根据相应的均值分数从左到右排列。在统计显著性图中，浅红色阴影表示根据单侧 Wilcoxon 符号秩检验（*p* 值 <5%），x 轴上的团队 DSC 分数显著高于 y 轴上的团队分数；深红色阴影表示其分数并未显著高于对方。

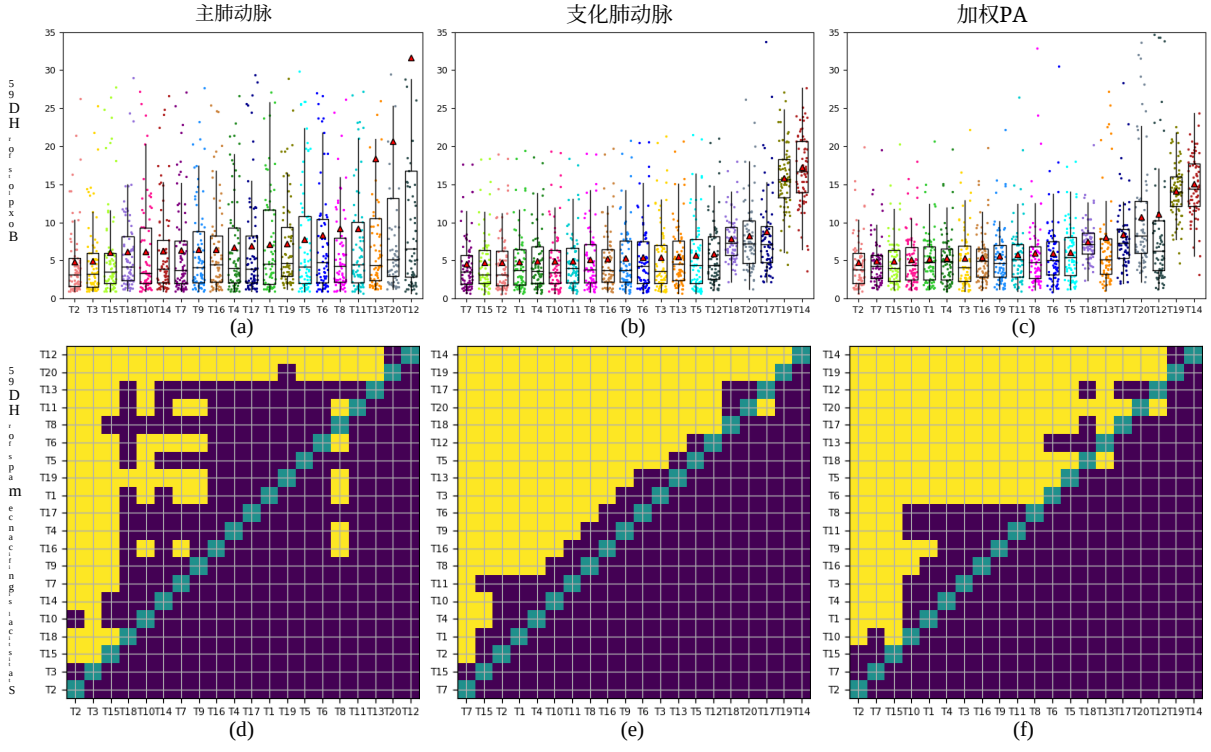


图6。前20名团队HD95指标的点图和箱线图可视化（第一行）以及统计显著性图（第二行）。该图示与上一图中的图示相似，但本图使用HD95指标。

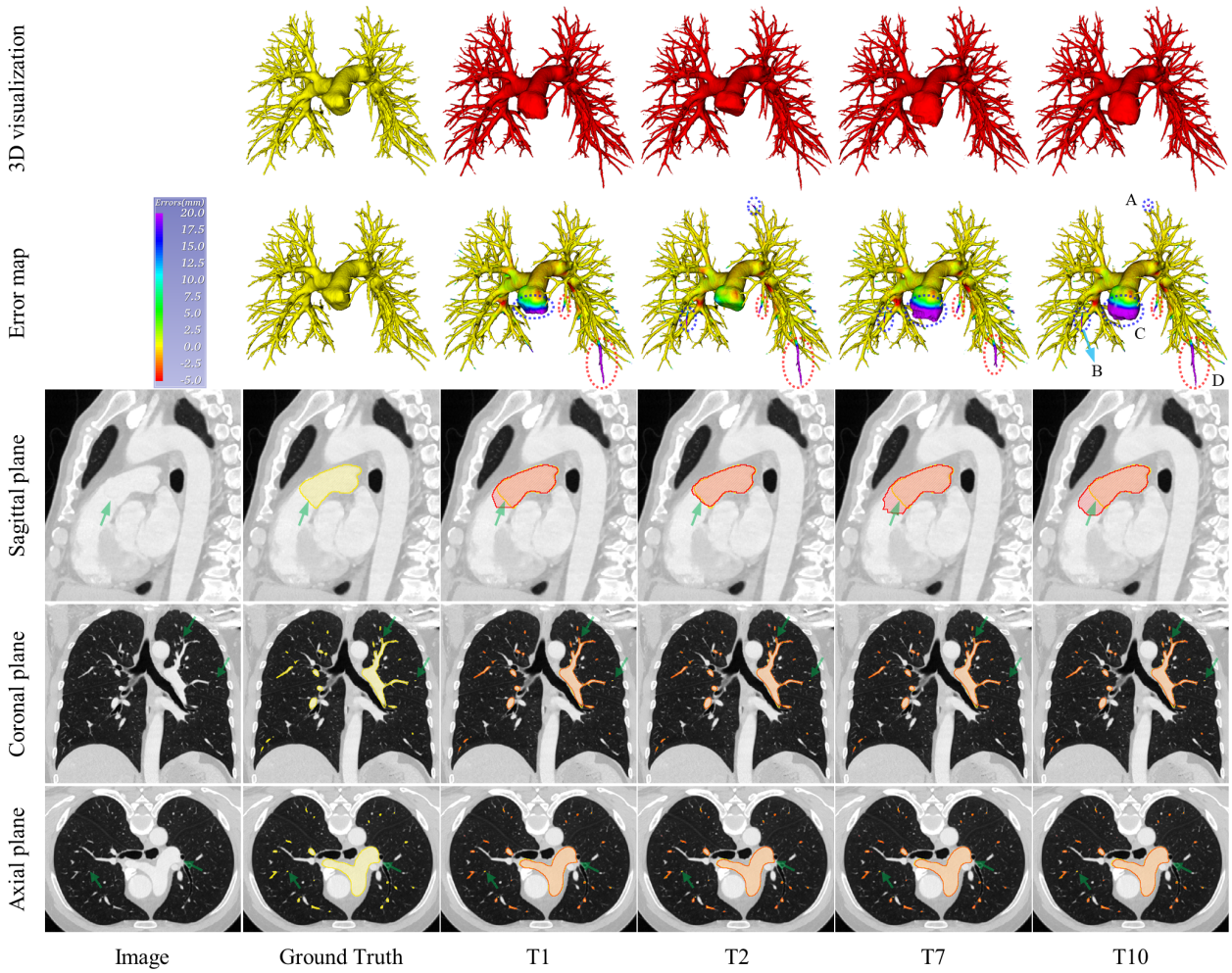


图 7. 代表性团队的分割结果三维可视化（第一行）、预测结果与真实标注之间的误差图（第二行），以及三视图分割结果（最后三行）。从左到右：原始图像、肺动脉（PA）的真实标注，以及团队 T1、T2、T7 和 T10 的分割结果。误差图表示分割结果与真实标注最近位置之间的距离。第二行中的蓝色虚线区域容易出现分割假阳性结果。此外，红色点状区域预计会出现在大多数团队的结果中。浅绿色箭头指向最后三行中的易出错区域。

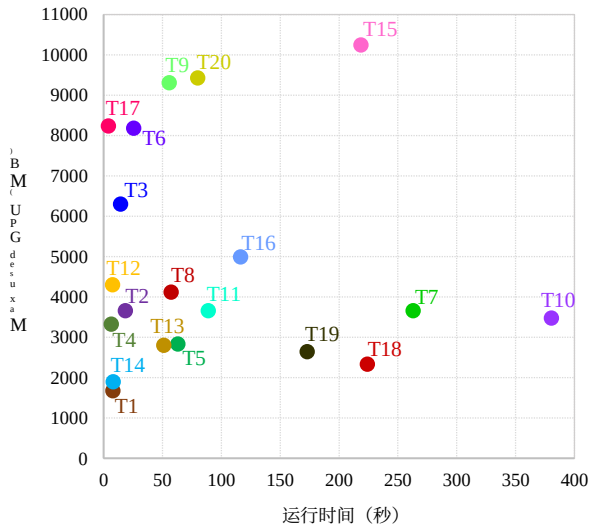


图8。前20名团队的运行时间和最大使用 GPU 指标的比较分析。

GPU 内存）在最终排名方案中被赋予了相同的权重。进一步地，为了重点关注分支肺动脉的分割准确率，分支肺动脉和主肺动脉的权重分别设为80%和20%。为了考察单项指标和感兴趣区域（主肺动脉和分支肺动脉）如何影响最终排名，图9展示了前20名团队的排名稳定性。对于指标（带阴影）和感兴趣区域（不带阴影），结果均出现了较大波动。这是因为不同团队的侧重点不同。例如，T17侧重于运行时间指标，并在运行时间方面获得了第一名，而其他指标的排名则不太理想。此外，T1专注于分支肺动脉分割，在DSC-B方面排名第一，但在DSC-M方面仅排名第18，在DSC-W方面排名第2。相比之下，T3对主肺动脉和分支肺动脉给予了相同的优先级，在DSC-M和DSC-B方面均排名第2，同时在DSC-W方面排名第1。因此，更全面地考虑各种指标和感兴趣区域将带来更好的性能。这也是未来算法开发的一个合适方向。

4.2. 分割准确性与效率之间的平衡

前20名团队的四项指标得分如图10所示，以三维气泡图呈现。图10中，DSC较高、HD95较低、运行时间较短且气泡较小的团队（如T1和T2），能够在准确性和效率之间取得适当的平衡。然而，有些团队（如T14、T17、T18和T19）倾向于牺牲准确性（DSC和HD95表现相当差），来优化效率（GPU或运行时间表现更好）。有些团队（如T15）则倾向于牺牲效率（GPU和运行时间得分较低），来优化准确性（DSC和HD95得分较高）。另一个例子是，尽管T7取得了最高的HD95得分，且DSC表现良好，却忽视了效率。

4.3. 提高分割准确性的策略

如图 7 所示，主肺动脉的分割易出错区域位于主肺动脉根部。对于分支肺动脉，分割易出错区域位于某些分支肺动脉节段及其末端。此外，与肺静脉或其他血管交织的肺动脉也更难以分割（图 7）。因此，肺动脉复杂的拓扑结构以及较大的类内和类间不平衡，使肺动脉难以得到有效且高效的分割。我们在表 4 中概述了前10名团队采用的策略，并分析如下常见方法。

(1) 数据预处理 大多数团队采用强度裁剪（IC）方法，并将强度归一化（N）至 [0,1] 或零均值、单位标准差。这有助于减少不同样本之间的强度差异。由于原始图像的平面内分辨率达到 512× 512，直接输入网络会消耗大量内存，因此很难直接使用。大多数团队按肺部区域裁剪（C）图像以获取感兴趣区域，然后重采样（RS）至网络输入尺寸。具体而言，顶尖团队将肺区提取（LAE）作为数据预处理的一部分，采用阈值法和连通区域分析。请注意，T3 利用预训练网络实现 LAE，而 T9 则采用由粗到细的策略，使 LAE 参与肺动脉分割的训练。LAE 有助于网络聚焦于肺动脉区域，从而提高准确性和效率。

(2) 数据增强 大多数顶尖团队都采用了大量数据增强方法（例如旋转（R）、翻转（F）、缩放（S）、形变（D）、强度变换（I）、镜像（M）、随机裁剪（RC）、随机噪声（RN）等）来提高分割准确率。例如，T2 通过实验确认，采用更优的弹性形变和全局强度扰动参数有助于提高 PARSE 数据的泛化能力。

(3) 网络设计 前10名团队均使用 U-Net（Ronneberger 等人，2015年）或其改进框架，例如 nnU-Net（Isensee 等人，2021年）。这也是其他分割挑战赛获胜算法的首选方案（Isensee 等人，2021年；Oreiller 等人，2022年；Timmins 等人，2021年；Heller 等人，2021年；Ma 等，2022年）。值得注意的是，直接复制并应用原始 nnU-Net 仍能获得可接受的分割准确率，但运行时间得分较低，例如 T7 和 T10。这是因为 nnU-Net 使用了许多复杂的预处理和后处理技术（Isensee 等人，2021年），这些技术可能会耗费大量时间。

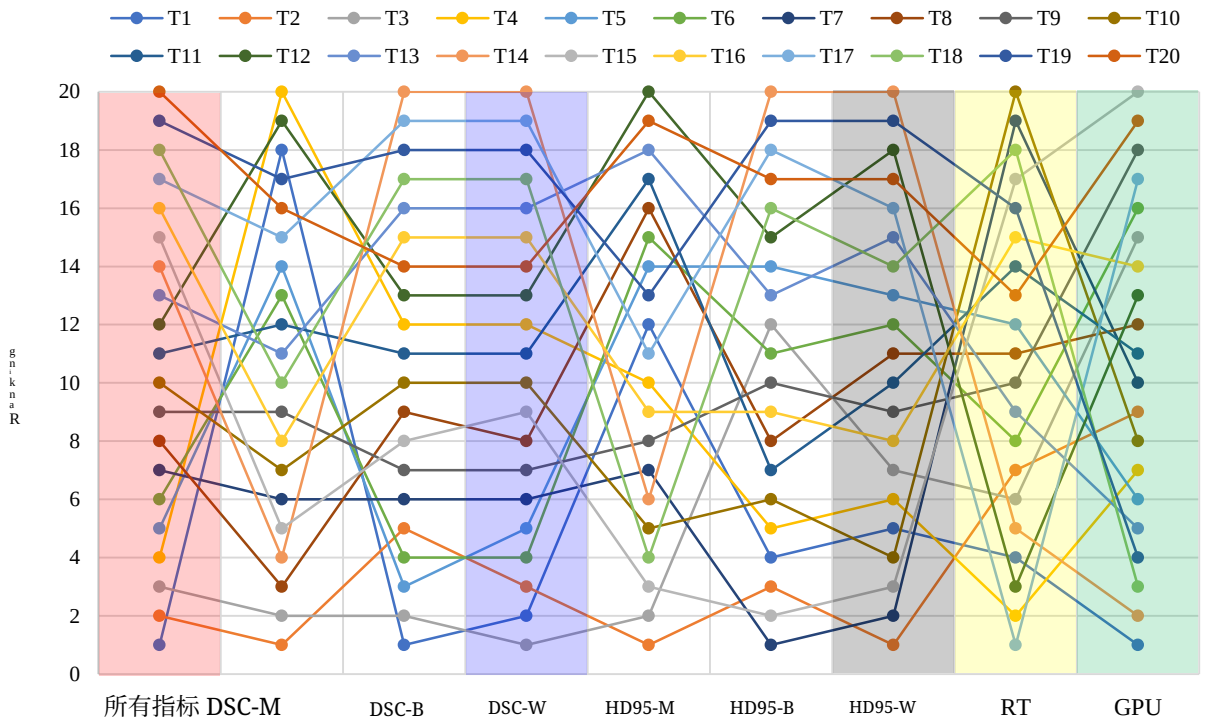


图9. 前20名团队在单项指标和集成指标下的排名稳定性分析。根据官方排名方案，“所有指标”（红色阴影）是四项指标的集成（加权肺动脉的 DSC、加权肺动脉的 HD95、运行时间和 GPU 内存，分别以蓝色、灰色、黄色和绿色阴影表示）。M、B 和 W 分别表示主肺动脉、分支肺动脉和加权肺动脉。

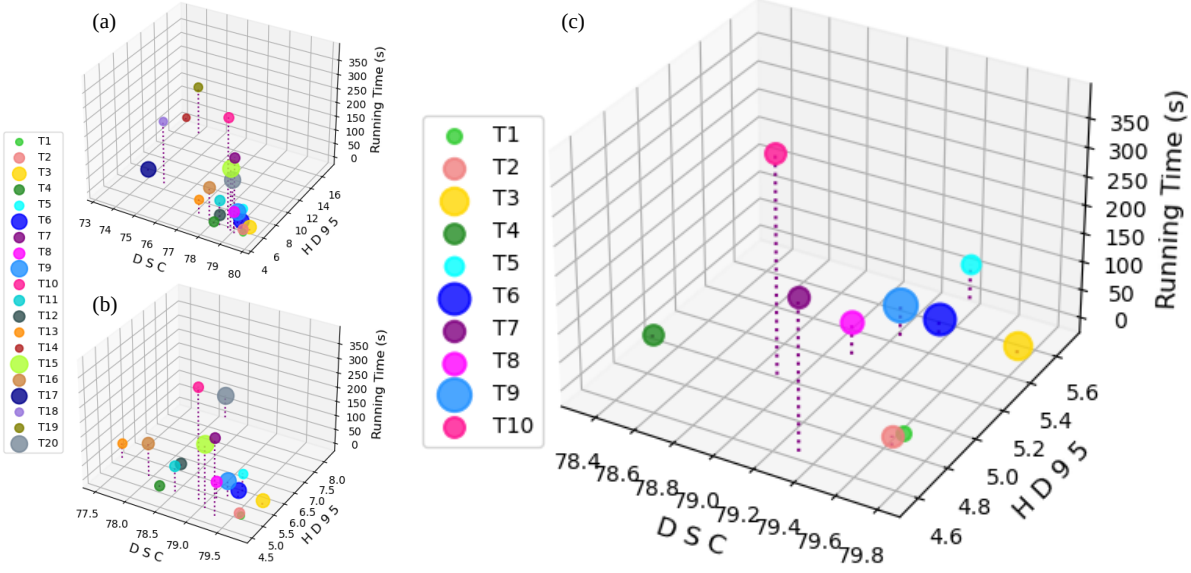


图10. 前20名团队（a）、剔除 DSC 和 HD95 得分明显较低的四个团队后的前20名团队（b）以及前10名团队（c）的得分三维气泡图，其中平均 DSC（x轴）、HD95（y轴）、运行时间（z轴）和 GPU 消耗（不同球体大小）均有体现。球体大小与 GPU 消耗成正比。不同颜色表示不同团队。由于这些团队的 DSC 较高、HD95 较低、运行时间较短且球体较小，因此在准确性和效率之间实现了良好平衡。

表 4. 前10名团队采用的关键技术。缩写：a) 预处理：强度截断（IC）、归一化（N）、裁剪（C）、重采样（RS）、肺区提取（LAE）；b) 数据增强：旋转（R）、翻转（F）、缩放（S）、变形（D）、强度变换（I）、镜像（M）、随机裁剪（RC）、随机噪声（RN）；c) 网络设计：由粗到细框架（CF）、残差块（RB）、深度监督（DS）、辅助损失（AL）；d) 推理：肺区提取（LAE）、测试时增强（TTA）、多模型集成（ME）；e) 后处理：连通域分析（CCA）、利用肺区减少假阳性（RFP）、形态学闭运算（MC）。

方法	预处理							数据增强										网络设计			推理		后处理		
	IC	N	C	RS	LC	R	F	S	D	I	M	RC	RN	CF	RB	DS	AL	LC	测试时增强		ME	CCA	RFP	MC	
T1	✓	✓	✓			✓		✓	✓				✓	✓	✓	✓				✓	✓	✓	✓		✓	✓	✓	
T2	✓	✓	✓	✓		✓		✓	✓	✓	✓	✓	✓	✓	✓	✓		✓			✓		✓	✓				
T3	✓	✓	✓	✓		✓		✓	✓	✓	✓	✓	✓	✓	✓		✓	✓			✓		✓	✓				
T4		✓	✓			✓				✓							✓				✓			✓		✓		
T5	✓	✓	✓	✓				✓	✓				✓	✓	✓						✓							
T6	✓	✓	✓			✓		✓	✓				✓		✓	✓					✓		✓		✓		✓	
T7	✓	✓	✓	✓				✓	✓				✓		✓	✓		✓	✓		✓				✓	✓		✓
T8		✓		✓				✓	✓	✓	✓		✓		✓	✓					✓	✓						
T9		✓	✓	✓						✓	✓		✓				✓	✓	✓						✓	✓	✓	
T10		✓	✓	✓				✓	✓			✓	✓	✓		✓										✓		

此外，T2 通过数据增强、过采样机制以及减少测试时增强视图，对 nnU-Net（Isensee 等人，2021年）进行了改进，从而实现了最佳的分割准确率和效率。此外，T1设计了一个带有残差块的轻量级 U-Net框架，也取得了更优的分割准确率和效率。此外，深度监督（DS）是顶尖团队用于提高肺动脉分割准确性的常用技术。

(4) 由粗到细框架（C2F）C2F 包括一个粗略定位网络和一个精细分割网络，这是许多分割任务中常用的方法（Ma 等，2022）。T9 明确采用了 C2F方法。实际上，大多数团队在数据预处理阶段通过 LAE 隐式采用了 C2F 方法，尤其是排名前 5 的团队。

(5) 辅助损失（AL） 除了常用的 Dice 损失或 CE 损失外，一些团队还使用额外的损失来辅助训练。例如，T1 修改了广义并集损失（Zheng 等，2021年），并使其与加权 CE 损失一起训练。另一个例子是，T4 使用焦点损失、BCE 损失和 L1 损失。

(6) 推理 在推理阶段，LAE 仍是加快预测速度的有效技术，排名前四的团队都使用了 LAE。在 nnU-Net（Isensee 等人，2021年）中，通过沿所有轴进行镜像来应用测试时增强（TTA），并已证明能够提高分割准确率。然而，TTA 会降低分割效率。T2 评估了 the 每个增强视图的分割准确率（即翻转明确的轴），并集成准确率最高的四个视图，以在分割准确率和效率之间建立权衡。模型集成（ME）是医学图像分割挑战赛中常用的方法。例如，在 10 项 MICCAI 2020 分割挑战赛中，所有获胜算法都采用了 ME 方法（Ma 等，2022）。然而，ME 会导致模型规模较大并消耗计算资源，不适用于本次挑战赛的效率指标。因此，排名前五的团队未使用 ME 方法。值得注意的是，T1 引入了多个模型的“参数平均”，而不是简单的 ME；这不会增加推理期间的内存消耗和计算负担。

(7) 后处理 在后处理中，采用连通分量分析（CCA）和 LAE 来减少假阳性（RFP）预测。前者由 T1、T7、T9 和 T10 使用，后者由 T1、T4 和 T6 使用。此外，T7 还利用形态学闭运算（MC）操作填充 LA 中的孔洞，以减少假阴性预测。

4.3.1. 提高分支肺动脉分割准确率的策略

为了提高分支肺动脉的分割准确率，一些团队（例如 T1 和 T2 ）为分支肺动脉提供了更多采样概率，以使网络更加关注其

分支。

4.4. 提高分割效率的策略*ffi*

与 FLARE 挑战赛（Ma 等，2022）类似，GPU内存和运行时间被选为两个面向效率的指标，因为前者是极具价值的硬件指标，且二者对于临床应用都非常重要。为了节省GPU 内存占用并实现快速推理，许多团队对其算法进行了优化，如表 4 所示。我们还概述了前10名团队提高分割效率的方法。

(1) 数据预处理从原始图像中提取肺部区域，并在将其输入网络之前对其进行裁剪以获取感兴趣区域，是排名前四的团队所采用的策略。这有助于提高准确性和效率。

(2)网络架构该网络的轻量化设计在GPU内存消耗和运行时间指标方面都更加友好。T1设计了带有残差块的轻量级U-Net，其成功表明，只要训练得当，轻量级网络就能在分割准确率和效率之间实现良好的权衡。

(3)实现技术 TTA 通常有助于提高分割准确率，但会延长推理时间。因此，T2 减少了TTA的视图数量，而T5放弃使用TTA。此外，为了提高推理效率，T3 使用异步线程与主处理过程并行进行读写。进一步地，T1 在推理过程中将模型和数据转换为半精度，从而在几乎不损失准确性的情况下减少了GPU内存消耗。此外，大多数顶尖团队使用基于图块的输入，通过滑动窗口以固定体素数从整个感兴趣区域中采样三维图块。正如FLARE挑战赛（Ma 等，2022）所指出的，与基于整个体积的策略相比，基于图块的输入需要更多时间来预测最终结果。因此，基于整个体积的策略值得在未来的算法中加以考虑。

4.5. 局限性与未来工作

随着挑战赛的进行，我们发现训练集中的主肺动脉存在一些标签噪声。不过，数据和标签对所有挑战赛参与者都是相同的。此外，我们修正了测试集中的相关噪声，因此测试集更加准确。这对参与者而言是公平的。尽管我们已尽最大努力，分支肺动脉（尤其是其末端）的准确标注仍然很困难，因此标签中可能存在一些错误。图 7 中红色虚线区域展示了一个例子，大多数团队都能分割出该区域。经人工检查，这些区域确实是肺动脉所在的位置。因此，深度网络可能会分割出更多肺动脉。未来的研究可以将传统分割方法与深度网络相结合，从而标注更多分支肺动脉末端，并提高标注准确性。因此，分支肺动脉的标注是未来亟待解决的问题。

在本次挑战的数据集中，仅包含一种疾病（即肺结节）。在未来版本的 PARSE 挑战中，我们将扩大数据集规模，并纳入更多疾病类型以及健康人群。此外，本数据集仅考虑了肺动脉分割。在后续挑战中，我们还将考虑肺静脉甚至肺气道的分割。此外，本次挑战数据集仅分析了两个层级的肺动脉。未来的挑战可以考虑更多层级，如文献所述（Poletti 等人，2022年）。

5. 结论

总之，我们组织了首个国际公开肺动脉分割挑战赛，重点关注主肺动脉和分支肺动脉的分割，以实现高准确性和高效率。结果表明，数据预处理、数据增强、网络架构和附加损失函数设计、推理以及数据后处理技术，有望在更短的推理时间和更低的GPU 内存消耗下实现准确的肺动脉分割。此外，我们注意到肺动脉的分割准确性仍需提高，尤其是分支肺动脉的分割准确性。另外，较高的指标得分（例如推理运行时间）并不意味着其他指标得分（例如 DSC 和 GPU 内存消耗）也更好。因此，更-

未来的研究需要进一步权衡分割准确率与效率。

参考文献

Charbonnier, J.P., Brink, M., Ciompi, F., Scholten, E.T., Schaefer-Prokop, C.M., Van Rikxoort, E.M., 2015. 使用树划分和外周血管匹配，在计算机断层扫描中自动分离并分类肺动脉与肺静脉。《IEEE医学影像汇刊》35, 882 892– . C.ic¸ek, ¨O., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O., 2016. 3d u-net：通过稀疏标注学习密集体积分割，载于：医学图像计算与计算机辅助干预国际会议（MICCAI），Springer。第424–432页。Est´epar, R.S.J., Kinney, G.L., Black-Shinn, J.L., Bowler, R.P., Kindlmann, G.L., Ross, J.C., Kikinis, R., Han, M.K., Come, C.E., Diaz, A.A., 等，2013. 吸烟者肺血管形态的计算机断层扫描测量及其临床意义。《美国呼吸与重症监护医学杂志》188, 231–239。Gali`e, N., Humbert, M., Vachiery, J.L., Gibbs, S., Lang, I., Torbicki, A., Si- monneau, G., Peacock, A., Vonk Noordegraaf, A., Beghetti, M., 等，2016. 2015年ESC/ERS肺动脉高压诊断与治疗指南：欧洲心脏病学会（ESC）与欧洲呼吸学会（ERS）肺动脉高压诊断与治疗联合工作组制定；获欧洲儿科与先天性心脏病学会（AEPC）、国际心肺移植学会（ISHLT）认可。《欧洲心脏杂志》37, 67–119。He, K., Zhang, X., Ren, S., Sun, J., 2016. 用于图像识别的深度残差学习，载于：IEEE计算机视觉与模式识别会议论文集（CVPR），第770–778页。Heller, N., Isensee, F., Maier-Hein, K.H., Hou, X., Xie, C., Li, F., Nan, Y., Mu, G., Lin, Z., Han, M., 等，2021. 增强CT成像中肾脏及肾肿瘤分割的最新进展：KiTS19挑战赛结果。《医学图像分析》67, 101821。Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H., 2021. nnU-Net：一种用于基于深度学习的生物医学图像分割的自配置方法。《自然·方法》18, 203–211。Konstantinides, S.V., Meyer, G., Becattini, C., Bueno, H., Geersing, G.J., Har- jola, V.P., Huisman, M.V., Humbert, M., Jennings, C.S., Jim´enez, D., 等，2020. 2019年ESC急性肺栓塞诊断与管理指南：与欧洲呼吸学会（ERS）合作制定，由欧洲心脏病学会（ESC）急性肺栓塞诊断与管理工作组编写。《欧洲心脏杂志》41, 543–603。Lin, T.Y., Goyal, P., Girshick, R., He, K., Doll´ar, P., 2017. 用于密集目标检测的焦点损失，载于：IEEE国际计算机视觉大会论文集（ICCV），第2980–2988页。Loshchilov, I., Hutter, F., 2018. 解耦权重衰减正则化，载于：国际学习表征会议（ICLR）。Ma, J., Zhang, Y., Gu, S., An, X., Wang, Z., Ge, C., Wang, C., Zhang, F., Wang, Y., Xu, Y., 等，2022. 快速且低GPU内存占用的腹部CT器官分割：FLARE 挑战赛。《医学图像分析》82, 102616。Maier-Hein, L., Reinke, A., Kozubek, M., Martel, A.L., Arbel, T., Eisenmann, M., Hanbury, A., Jannin, P., M¨uller, H., Onogur, S., 等，2020. BIAS：生物医学图像分析挑战赛的透明报告。《医学图像分析》66, 101796。Oreiller, V., Andrearczyk, V., Jreige, M., Boughdad, S., Elhalawani, H., Castelli, J., Valli`eres, M., Zhu, S., Xie, J., Peng, Y., 等，2022. PET/CT中的头颈部肿瘤分割：HECKTOR挑战赛。《医学图像分析》77, 102336。Poletti, J., Bach, M., Yang, S., Sexauer, R., Stieltjes, B., Rotzinger, D.C., Bre- merich, J., Sauter, A.W., Weikert, T., 2022. 自动肺血管分割揭示病毒性肺炎中的血管容积分布重构。《欧洲放射学杂志》150, 110259。Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net：用于生物医学图像分割的卷积网络，载于：医学图像计算与计算机辅助干预国际会议（MICCAI），Springer。第234–241页。Shahin, Y., Alabed, S., Alkhanfar, D., Tschirren, J., Rothman, A.M., Condliffe, R., Wild, J.M., Kiely, D.G., Swift, A.J., 2022. 定量CT评估小肺血管在肺动脉高压中具有功能和预后价值。《放射学》305, 431–440。Shit, S., Paetzold, J.C., Sekuboyina, A., Ezhov, I., Unger, A., Zhylka, A., Pluim, J.P., Bauer, U., Menze, B.H., 2021. clDice——一种用于管状结构分割的新型保拓扑损失函数，载于：IEEE/CVF计算机视觉与模式识别会议论文集（CVPR），第16560–16569页。Tan, W., Zhou, L., Li, X., Yang, X., Chen, Y., Yang, J., 2021. 基于深度神经网络的肺部CT和CTA图像自动血管分割。《X射线科学与技术杂志》，1–15。Timmins, K.M., van der Schaaf, I.C., Bennink, E., Ruigrok, Y.M., An, X., Baumgartner, M., Bourdon, P., De Feo, R., Di Noto, T., Dubost, F., 等，2021. 比较TOF-MRA中未破裂颅内动脉瘤的检测与分割方法：ADAM挑战赛。《神经影像》238, 118216。Van Ginneken, B., Armato III, S.G., de Hoop, B., van Amelsvoort-van de Vorst, S., Duindam, T., Niemeijer, M., Murphy, K., Schilham, A., Retico, A., Fantacci, M.E., 等，2010. 比较并结合用于计算机断层扫描中肺结节计算机辅助检测的算法：ANODE09研究。《医学图像分析》14, 707–722。Wiesenfarth, M., Reinke, A., Landman, B.A., Eisenmann, M., Saiz, L.A., Car- doso, M.J., Maier-Hein, L., Kopp-Schneider, A., 2021. 用于分析和可视化挑战赛结果的方法与开源工具包。《科学报告》11, 1–15。Zhang, Z., Liu, Q., Wang, Y., 2018. 基于深度残差U-Net的道路提取。《IEEE地球科学与遥感快报》15, 749–753。Zheng, H., Qin, Y., Gu, Y., Xie, F., Yang, J., Sun, J., Yang, G.Z., 2021. 缓解肺气道分割中的类别梯度失衡。《IEEE医学影像汇刊》40, 2452–2462。