

文章编号: 1002-1582(2026)01-0115-08

基于多步匹配的超声多目标跟踪方法与系统构建

朱星龙, 艾丹妮*

(北京理工大学 光电学院, 北京 100081)

摘 要: 在肝脏肿瘤治疗中, 采用超声引导放射治疗能够最大程度减少对病人正常组织的伤害。超声引导放射治疗通过超声探头实时获取病人的肿瘤位置信息, 并控制放射源跟随肿瘤移动。超声引导放射治疗技术的难点在于如何通过超声图像预测目标位置, 这个步骤也称为超声目标跟踪。文章研究了基于多步匹配的超声多目标跟踪方法, 该方法能同时跟踪肝脏超声图像序列中的多个肝脏血管。首先检测—跟踪多任务网络根据序列当前帧初步预测目标位置的若干提案, 同时卡尔曼滤波器会根据目标的历史位置信息预测目标的先验位置, 然后多步匹配方法会对目标位置提案和先验位置进行匹配, 最后利用轨迹活跃度对输出的目标进行筛选。文章构建了超声目标跟踪系统。该系统可以通过图形用户界面完成多目标跟踪演示、图像标注和评估功能。文章根据 MOT 竞赛的指标将提出的方法与其他方法进行了对比及消融实验, 实验结果证明文章方法具有更高的精确度和鲁棒性。同时, 通过功能测试和性能测试证明了超声目标跟踪系统能够稳定地在准实时条件下完成设计的各项功能。

关 键 词: 超声图像序列; 深度学习; 目标跟踪; 系统构建

中图分类号: TP394.1

文献标识码: A

Ultrasound multi-object tracking method based on multi-step matching and system construction

ZHU Xinglong, AI Danni*

(Beijing Institute of Technology, School of Optics and Photonics, Beijing 100081, China)

Abstract: In the treatment of liver tumors, the use of ultrasound-guided radiation therapy can minimize the damage to the normal tissue of the patient. Ultrasound-guided radiation therapy obtains the patient's tumor location information by the ultrasound probe in real time, and controls the radiation source to follow the tumor movement. As the essential technology of ultrasound-guided radiation therapy, ultrasound object tracking aims to obtain the positions of a specific target in the ultrasound sequences. A multi-object tracking method for ultrasound images based on multi-step matching is proposed, allowing multiple liver blood vessels to be tracked in a sequence of liver ultrasound images. The proposed method first inputs the image into the detection-tracking multi-task network to obtain the target position proposal, and filters the prior position of the target through Kalman filter at the same time. Then the target position proposal and prior position are matched by the multi-step matching method. Finally the estimated target positions are picked by their trajectory activity. This paper also builds an ultrasound object tracking system based on the proposed method. Besides tracking multiple objects, the system can also demonstrate annotating images and evaluating tracking results in a graphical user interface. The proposed method is evaluated with the indicators of MOT competition, and the ablation experiments reveal the availability of our improvement. The system are also evaluated with function tests and performance test, which prove the reliability of the proposed system.

Keywords: ultrasound image sequence; deep learning; object tracking; system construction

收稿日期: 2022-06-23; 收到修改稿日期: 2022-09-03

基金项目: 国家自然科学基金面上项目 (61971040)

作者简介: 朱星龙 (1996—), 男, 硕士研究生, 从事医学图像处理研究。xlzhu1996@163.com

艾丹妮 (1988—), 女, 特别研究员、博士生导师, 博士, 从事医学图像智能分析、手术导航、虚拟现实与增强现实、智能传感及人工智能等研究。

* 通讯作者: dannibj@163.com

0 引 言

放射治疗是治疗肝脏肿瘤的一种重要方法。在传统的放射治疗中,由于患者的呼吸会导致肿瘤随肝脏发生不规则运动,医生往往会选择扩大放射范围。这种方法虽然保证了疗效,但对健康肝脏组织造成了不必要的伤害,对患者的预后造成了不良影响。为了减少对健康组织的伤害,学界试图采用超声引导辅助放射治疗,通过超声探头实时地采集超声图像,通过超声图像预测肝脏血管位置,再进一步地预测肿瘤位置,为放射源的位置和方向提供参考^[1]。本文对超声图像序列肝脏血管跟踪进行了研究。

超声目标跟踪是超声引导放射治疗中一个不可缺少的关键环节,可简单划分为单目标跟踪和多目标跟踪两个技术方向。单目标跟踪具有较高的运算速度以及相对较高的精度,但无法处理目标在视野中短暂消失的问题,对于超声采集环境有较高的要求,并不能很好地符合临床需求。多目标跟踪首先探测图像中的所有目标,然后关联前后帧的同一目标,多目标跟踪能够更好地处理目标离开视野和重新出现的情况,因而对超声采集环境的要求不高,更加符合临床需求。

在超声图像领域,单目标跟踪已经取得了相当进展,在 MICCAI 主办的肝脏超声跟踪竞赛中^[2],跟踪精度榜首的级联 SiamFC 算法由 Liu 等在 2020 年提出。该算法在 CLUST 数据集上得到平均误差 0.69mm 的结果,已基本满足临床治疗的精度要求^[3]。但在超声图像领域,多目标跟踪的关注度相对较低,相关的研究更多集中在自然图像领域,常以自动驾驶和视频监控为研究背景^[4-6]。

多目标跟踪一般分为两个步骤,首先通过目标检测算法得到当前帧视野中可能的目标位置,然后通过目标关联方法关联当前帧与上一帧的目标 id。

一些多目标跟踪方法会独立地执行目标检测和目标关联。如 Wojke N 等采用 Faster RCNN 作为检测网络,截取网络深度特征,计算两个特征向量的余弦角作为前后帧匹配的依据^[8]。Zhang Y 等则使用 YOLO 算法进行目标检测,利用交并比计算两个包围盒的距离匹配前后帧^[9]。在这些方法中,目标检测和目标关联方法的调优过程是相对孤立的,目标检测一般会采用 Faster R-CNN^[7, 8, 10-12]、SSD^[13-15]、YOLO^[9, 16]、CenterNet^[17] 等目标跟踪网络。关联算法则依据特征提取方式^[7-9, 18]、相似度计算^[7-9]和匹配方法^[7, 19, 20]的不同而有较大差异。

另一些多目标跟踪算法将目标检测和目标关联视为一个整体进行调优。联合跟踪检测网络往往会采取一个特征提取器和多个网络头的形式,通过构建一个统一的目标函数进行对网络进行端到端的训练,使得检测和关联两个步骤得到更好的耦合效果^[17, 21]。

本文采用分离的目标检测和目标关联步骤来进行多目标跟踪,通过构建卡尔曼滤波预测目标的先验位置,再使用多步匹配的方法关联前后帧目标,取得了良好的跟踪精度。在后处理阶段,本文通过轨迹活跃度筛选保证了结果的稳定性。最后,本文将算法演示、标注和评估三个功能在超声目标跟踪系统上进行了集成,并对该系统进行了测试。

1 本文方法

本文采用基于多步匹配的多目标跟踪算法对超声序列上的肝脏肿瘤进行了跟踪。整个算法流程大致可以分为检测阶段和关联阶段。在检测阶段,检测—跟踪多任务网络根据当前帧图像预测目标位置作为初步的提案;在关联阶段,卡尔曼滤波通过目标轨迹预测目标的先验位置,再利用多步匹配算法将检测—跟踪多任务网络预测的提案按照响应度依次与先验位置进行匹配,最后利用轨迹活跃度筛选输出轨迹。整个方法的流程如图 1 所示。

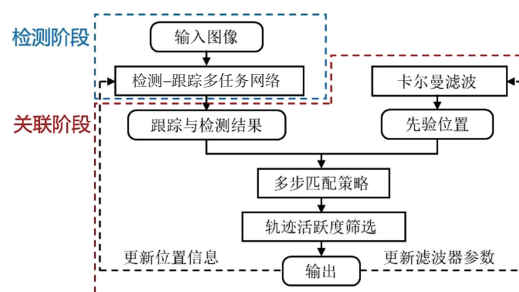


图 1 基于多步匹配的超声多目标跟踪方法流程总图

1.1 检测—跟踪多任务网络

图 2 展示了检测—跟踪多任务网络结构。特征提取器 (B) 采用了基于深层聚合结构的 DLA-34, 该结构能够通过空间融合和语义融合提升特征质量, 提高特征运算^[22]。

在目标跟踪网络中,感兴趣区域对齐 (E) 将第 t 帧的图像特征 f^t 根据第 t 帧的跟踪结果区域 R_t^i 提取特征调整尺寸至感兴趣区域特征 $f_{R_t^i}^t$, 并相应地将第 $t + \delta$ 帧的图像特征 $f^{t+\delta}$ 按照第 t 帧的跟踪结果中心不变两倍长宽变换得到的区域 S_t 提取特征调整尺寸至感兴趣区域特征 $f_{S_t}^{t+\delta}$ 。特征的大小及深度均在图 2 中进行了标识。

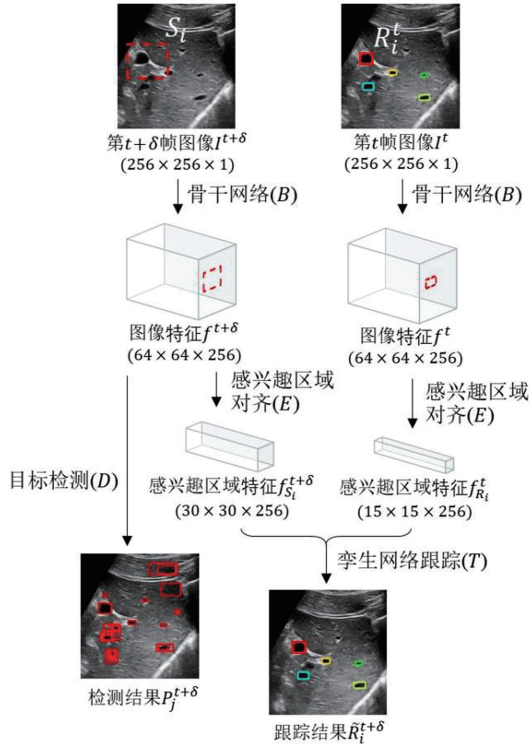


图2 检测—跟踪多任务网络结构流程图

孪生跟踪网络 (T) 通过深度互相关运算比较输入特征的相似程度以及相似目标的相对位置。互相关运算将规模较小的特征作为卷积核对规模较大的特征进行卷积计算, 这个计算等效于计算了两个图像特征信号的互相关, 因而可以定量地表达两个图像特征信号之间的相似度和两个相似信号间的相对位移。相似信号间的位移可以直接用于预测目标的相对位移, 而两个图像特征信号间的相似程度即为当前帧目标先验位置的置信度^[24]。通过上述运算, 孪生跟踪网络能够得到目标 i 的跟踪结果 $\tilde{R}_i^{t+\delta}$, 如果并行地跟踪多个目标, 就可以得到多个目标的跟踪结果集合 $\{\tilde{R}_i^{t+\delta}\}$ 。

在目标检测网络 (D) 中, 第 $t+\delta$ 帧的图像特征 $f^{t+\delta}$ 会先输入区域建议网络, 得到候选结果并提取相关特征, 相关特征经过感兴趣区域对齐调整尺寸后, 输入至 Fast-RCNN 网络得到目标检测结果 $\{P_j^{t+\delta}\}$ 。

最终, 跟踪结果 $\{\tilde{R}_i^{t+\delta} | i \in [1, M]\}$ 和检测结果 $\{P_j^{t+\delta} | j \in [1, N]\}$ 将由网络输出, 并合并输入至下一步骤。

网络在多处采用了感兴趣区域对齐 (E) 对输出的特征进行了后处理, 减少了特征规模不匹配导致的信息损失。

网络的 Loss 由检测网络 (D) 和跟踪网络 (T) 的 Loss 值组成

$$L = \mathcal{L}_D + \mathcal{L}_T \quad (1)$$

式中, $\mathcal{L}_D$ 来自于检测网络内的 RPN 和 Fast-RCNN, 即

$$\mathcal{L}_D = \mathcal{L}_{\text{RPN}} + \mathcal{L}_{\text{Fast-RCNN}} \quad (2)$$

从结构上看, 跟踪网络 (T)、RPN 和 Fast-RCNN, 都由判别和回归分支构成, 其 Loss 值为判别 Loss 和回归 Loss 的加和。对于 $\mathcal{L}_i \in \{\mathcal{L}_T, \mathcal{L}_{\text{RPN}}, \mathcal{L}_{\text{Fast-RCNN}}\}$, 有

$$\mathcal{L}_i = \mathcal{L}_{\text{cls}} + \mathcal{L}_{\text{reg}} \quad (3)$$

$$\mathcal{L}_{\text{cls}} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (4)$$

$$\mathcal{L}_{\text{reg}} = \begin{cases} -\frac{1}{N} \sum_{i=1}^N |r_i - \hat{r}_i|, & \text{if } y_i = 1 \\ 0, & \text{if } y_i = 0 \end{cases} \quad (5)$$

$\mathcal{L}_{\text{cls}}$ 、 $\mathcal{L}_{\text{reg}}$ 分别表示判别和回归的 Loss 值, y_i 和 r_i 分别表示分类和回归的金标准, $\hat{y}_i$ 和 $\hat{r}_i$ 分别表示分类和回归的预测结果。由于方法仅需要跟踪血管, 在三个模块中 y_i 仅有 1 和 0 两种取值, 分别表示目标和背景。

实际训练时, 网络参数会在预训练的模型上进行迭代, 迭代时将跟踪网络头 (T)、RPN 和 Fast-RCNN 网络的网络参数同时进行迭代优化。由于标注过的超声多目标跟踪样本较少, 本文采用三维超声图像血管分割数据集进行训练, 首先将面向目标分割的以图像掩膜形式标注的金标准处理成面向目标检测的以包围盒标注的形式, 对图像进行裁剪、缩放处理为图像对, 并加入运动模糊、灰度调整增加训练难度, 以无监督学习的形式训练该网络。

1.2 基于多步匹配的关联方法

如图 1 所示的, 关联方法主要包含三个组件分别是卡尔曼滤波、多步匹配和基于轨迹活跃度的筛选。卡尔曼滤波通过目标轨迹对目标位置进行先验预测, 先验位置将输入至多步匹配, 再将整个方法输出的后验位置输入至卡尔曼滤波器用于更新滤波器的参数。在整个算法流程中, 卡尔曼滤波器提取了运动特征, 由卡尔曼滤波输出的先验位置实际包含了目标的运动特征。

多步匹配方法用以匹配提案 $\{\bar{P}_i^t\}$ 和先验目标位置 $\{\hat{R}_i^t\}$, 与 $\hat{R}_i^t$ 成功匹配的 $\bar{P}_i^t$ 记作后验位置 R_i^t , R_i^t 将被记入目标 i 的轨迹 $\mathcal{T}_i = \{R_i^t\}$ 。算法 1 为多步匹配方法的伪代码。多步匹配的思路是, 首先将提案按照响应值进行分类, 优先匹配高响应提案, 然后将未匹配的先验位置匹配到响应值较低的提案上, 再对该步骤进行重复。如果未匹配提案的响应值高, 则据此创建新的轨迹, 如果未匹配提

案的响应值较低则直接删除。多步匹配方法不会因为响应值的波动立即删除或创建有效轨迹，新创建的轨迹需要下一帧匹配到高响应提案才能被激活，如果未能激活则会被删除，轨迹当匹配连续失败五次才会被删除。

算法 1 多步匹配算法伪代码

输入：

上一帧的跟踪结果获取活跃轨迹 $\mathcal{T}^a$ 、休眠轨迹 $\mathcal{T}^d$ 和未确认轨迹 $\mathcal{T}^u$

轨迹 $\mathcal{T}^a \cup \mathcal{T}^d \cup \mathcal{T}^u$ 对应的先验位置 $\{\hat{r}_i' | i \in \mathcal{T}^a \cup \mathcal{T}^d \cup \mathcal{T}^u\}$

由检测—跟踪多任务网络生成的提案 $\{\bar{p}_j'\}$

输出：

后验位置 $\{r_i'\}$

变更后的活跃轨迹 $\mathcal{T}^a$ 、休眠轨迹 $\mathcal{T}^d$ 和未确认轨迹 $\mathcal{T}^u$

for $p' \in \{\bar{p}_j'\}$

 获取提案 p' 的响应值 s

 If $0 \leq s < 0.1$: 把提案 p' 加入集合 $\{p_i'\}$

 If $0.1 \leq s < 0.5$: 把提案 p' 加入集合 $\{p_i^m\}$

 If $0.1 \leq s \leq 1.0$: 把提案 p' 加入集合 $\{p_i^h\}$

 匹配 $\{p_i^h\}$ 与 $\{\hat{r}_i' | i \in \mathcal{T}^a \cup \mathcal{T}^d\}$

 与 $\hat{r}_i'$ 成功匹配的 p_i^h , $\hat{r}_i' \leftarrow p_i^h$, 记作轨迹 $\{r_i' | i \in \mathcal{T}^a\}$

 未成功匹配的 p_i^h , 加入集合 $\{p_h'\}$

 匹配 $\{p_h'\}$ 与 $\{\hat{r}_i' | i \in \mathcal{T}^u\}$

 与 $\hat{r}_i'$ 成功匹配的 p_h'

 If $s \geq 0.6$: 将 $\hat{r}_i'$ 加入集合 $\{r_i' | i \in \mathcal{T}^u\}$

 未成功匹配的 p_h'

 If $s \geq 0.9$: 创建新轨迹标号 $i \in \mathcal{T}^u$, 将 p_h' 加入集合 $\{r_i' | i \in \mathcal{T}^u\}$

 未成功匹配的 $\hat{r}_i'$

 If $i \in \mathcal{T}^a$: 将 $\hat{r}_i'$ 加入集合 $\{\hat{r}_i' | i \in \mathcal{T}_1^a\}$

 匹配 $\{p_i^m\}$ 与 $\{\hat{r}_i' | i \in \mathcal{T}_1^a\}$

 与 $\hat{r}_i'$ 成功匹配的 p_i^m , $\hat{r}_i' \leftarrow p_i^m$, 记作轨迹 $\{r_i' | i \in \mathcal{T}^a\}$

 未成功匹配的 p_i^m , 删除

 未成功匹配的 $\hat{r}_i'$, 将 $\hat{r}_i'$ 加入集合 $\{\hat{r}_i' | i \in \mathcal{T}_2^a\}$

 匹配 $\{p_i^h\}$ 与 $\{\hat{r}_i' | i \in \mathcal{T}_1^a\}$

 与 $\hat{r}_i'$ 成功匹配的 p_i^h , $\hat{r}_i' \leftarrow p_i^h$, 记作轨迹 $\{r_i' | i \in \mathcal{T}^u\}$

 未成功匹配的 p_i^h , 删除

 未成功匹配的 $\hat{r}_i'$, 将 $\hat{r}_i'$ 加入集合 $\{\hat{r}_i' | i \in \mathcal{T}_3^a\}$

for $i \in \mathcal{T}_3^a$

 轨迹活跃计数 $n_i = n_i - 1$

 If $n_i \leq 0$

 从 $\mathcal{T}^a$ 中剔除 i , 将 i 加入休眠轨迹 $\mathcal{T}^d$

return $\{r_i'\}$ 、 $\mathcal{T}^a$ 、 $\mathcal{T}^d$ 、 $\mathcal{T}^u$

最后，计算当前活跃轨迹的轨迹活跃度。轨迹活跃度为活跃帧数比当前输入总帧数，输出轨迹活跃度大于 0.8 的轨迹，活跃度低于 0.8 的轨迹将被隐藏。轨迹活跃度筛选保证了输出轨迹的连续性和置信度。

1.3 超声目标跟踪系统的构建

本文将多目标跟踪算法的演示与标注、评估功能整合到了超声目标跟踪系统上。这三项功能必须依赖于相应的超声图像数据和标注数据，同时也必须通过可视化模块对用户进行展示，整个系统在时序上必须满足一定的先后关系，整个系统的功能流程图如图 3 所示。

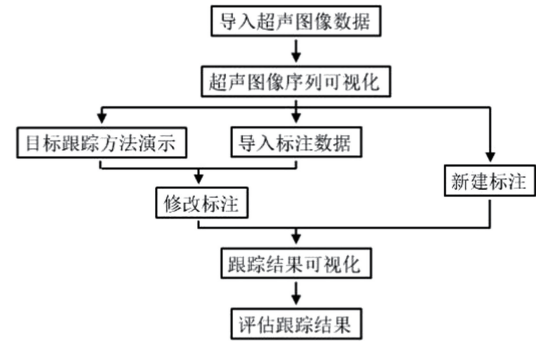


图 3 超声目标跟踪系统流程图

超声跟踪系统对于超声图像的可视化依赖于超声图像数据，而跟踪和标注必须在图像数据可视化之后才能由用户控制完成。标注是基于图像本身的，在确定图像和标注之后才能完成标注的可视化。评估则是对于跟踪结果的总结，因而置于流程的最后。

根据系统的功能流程，本文对系统进行了概要设计，确定了如图 4 所示的系统的组织结构，整个系统包括若干子模块和一个信号控制台，子模块可以分为控制器、视图和系统内部逻辑三个类型，用以实现具体功能，而信号控制台用于控制不同子模块之间通信。

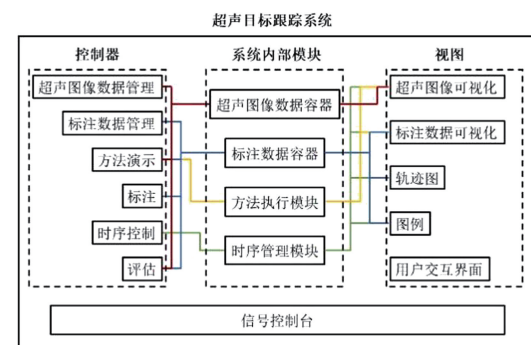


图 4 超声目标跟踪系统组织结构图

控制器用于用户对系统进行操作，具体体现为按钮、滑块、下拉框等用户可以操作的控件。包括两个用于管理数据的子模块，实现方法演示、标注和评估的三个核心功能子模块，以及用于控制图像序列播放的时序控制模块。

视图模块用于向用户提供数据的可视化，包括用于渲染超声图像和标注的两个可视化子模块，用于展示各目标横纵坐标随时间变化的轨迹图，

以及方便用户查找对应目标的图例。用户交互界面则包含各种对话框,对话框中既包含控制器也包含视图,方便用户完成一些短平快的操作。

系统内部模块封装了系统内部的功能逻辑,包括两个与数据直接交互的数据容器,封装了多目标跟踪方法的方法执行模块,和时序管理模块。时序管理模块相当于一个受用户控制的时钟,向视图发布时序变更的信号。

2 实验数据与评估指标

2.1 实验数据与参数配置

本文用自采三维超声数据集训练检测-跟踪多任务网络。图5展示了该数据集的部分图像切片及标注。该数据集采集自肝脏部分,用掩膜的方式标注了肝脏血管,如图5第二行的红色掩模所示。在训练时,计算 xOy 、 yOz 、 xOz 三个切面下的掩模包围盒,作为目标检测的金标准,如图5第三行的绿色包围盒所示。在跟踪方面,本文采用无监督训练的方法,通过几何变换和拓扑变换生成图像对。

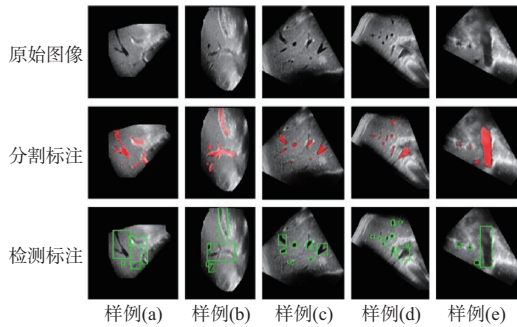


图5 训练数据集示例

在方法参数方面,设置感兴趣区域对齐后的搜索区域特征尺寸为 30×30 ,目标区域特征尺寸为 15×15 。在训练模型时,本文将SiamMOT的网络权重迁移至本文网络,然后训练25000轮。设置学习率为0.0005,权重衰减为0.0001。

超声目标跟踪系统部署在Ubuntu上。系统采用Python语言开发,用户图形界面使用第三方库PyQt,在深度学习和计算机视觉方面主要使用了PyTorch和OpenCV。

2.2 实验评估指标

为了评估多目标跟踪方法,本文采用了MOT竞赛的评价标准。按照评价指标的计算方式不同,可将评价指标分为局部指标^[25]和全局指标^[26]。

局部指标会分别计算每一帧图像中预测目标与金标准的交并比,用于统计预测目标和金标准的匹配情况,再根据匹配情况计算局部指标。当交并比大于50%时,预测结果与金标准匹配,记

作真阳性。未能与预测的结果匹配的金标准和未能与金标准匹配的预测结果,分别被称为假阴性和假阳性。用ID误匹配(ID-Swath)将记录交并比大于50%但ID匹配错误的情况,只记录误匹配发生的第一帧,不重复计算。

在预测结果和金标准标定完毕之后,计算局部指标MOTA和MOTP,两个指标分别反映了多目标跟踪的准确度和精度

$$MOTA = 1 - \frac{FP + FN + IDw}{gt_num} \quad (6)$$

$$MOTP = \sum_i \frac{d_i}{TP} \quad (7)$$

真阳性、假阳性、假阴性、ID误匹配的数目和金标准标注总数分别用TP、FP、FN、IDw和gt_num表示, d_i 表示一对真阳性匹配中目标与金标准的误差,计算公式为 $d = 1 - IoU$,其中IoU为区域交并比。

全局指标需要先将跟踪结果中某一个ID唯一确定地匹配至金标准中的一个ID,并且令全局上匹配数目最大,然后取该匹配下的真阳性、假阳性和假阴性,计算相应的指标。在这种匹配模式下,需要在全局匹配中ID一致且包围盒交并比大于50%才可以判断为真阳性idtp。交并比大于50%但ID未正确匹配的记作假阳性idfp, ID正确匹配但交并比小于等于50%的记作假阴性idfn,再分别计算ID精确率idp、ID召回率idr、IDF1分数idf1

$$idp = \frac{idtp}{idtp + idfp} \quad (8)$$

$$idr = \frac{idtp}{idtp + idfn} \quad (9)$$

$$idf1 = \frac{2 \times idtp}{2 \times idtp + idfp + idfn} \quad (10)$$

由于idtp的判定较TP更加严格,有 $idtp \leq TP$ 。当且仅当ID误匹配数目为0时, $idtp = TP$ 。全局指标相比局部指标更加关注多目标跟踪算法的前后帧目标关联,对于目标误关联更加敏感。

对于超声目标跟踪系统,本文首先测试了系统各项功能能否在指定环境下正常运行,然后通过编写程序脚本,测试系统各功能步骤的运行时间,并将平均值作为测试结果。

3 实验结果与分析

3.1 多目标跟踪方法对比实验

本文在半自动标注的自采数据上对本文提出的方法与多目标跟踪方法SiamMOT、ByteTrack、

单目标跟踪方法 SiamRPN++ 进行了评估和对比。图 6 从三个图像序列中各抽取了一帧，对比展示了各方法的跟踪结果。在展示的全部示例中，SiamMOT 均体现出较高假阳性。在示例 a4 中，ByteTrack 输出的目标 ID 值明显高于本文所提出的方法，这表明 ByteTrack 算法容易发生 ID 误切换的情况。在示例 b4 中，ByteTrack 则未能成功跟踪视野中的小目标。在 a、b、c 三个示例中，相比 SiamRPN++ 方法，本文提出的方法能更准确地预测目标形状。同时在示例 b5 中，SiamRPN++ 跟踪的目标 2 出现在了目标 1 的位置，这是由于跟踪器受到了相似目标的干扰，而本文提出的方法并无此种情况，表明本文提出的方法能够很好地排除相似目标的干扰。

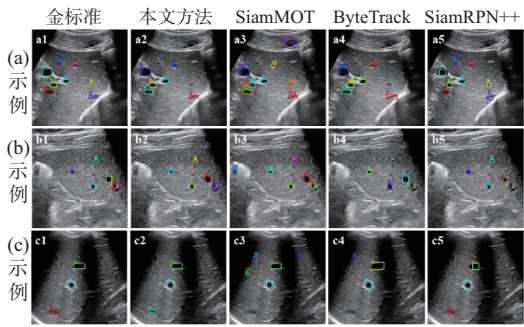


图 6 各跟踪方法在自采数据集上的结果展示

另外，本文还定量对比了本文方法、SiamMOT、ByteTrack 和 SiamRPN++ 的 MOTP、MOTA、IDP、IDR、IDF1，以及运行速度 FPS，共 6 项指标。表 1 展示了定性对比实验的结果。

表 1 各跟踪方法在自采数据集上的评估结果

方法	MOTP	MOTA/%	IDP/%	IDR/%	IDF1/%	FPS
本文方法	0.255	76.00	94.40	80.80	87.10	15.06
SiamMOT	0.243	41.50	95.80	43.40	59.80	15.17
ByteTrack	0.304	70.80	54.00	46.70	50.10	10.75
SiamRPN++	0.380	17.90	58.40	60.10	60.10	11.56

从表 1 中可以得知，本文方法的 MOTA、IDR 和 IDF1 均为最高，在 MOTP、IDP 和 FPS 三项指标上排名第二。表 1 中，SiamMOT 的 IDP 与本文方法基本相当，IDR 与 IDF1 明显偏低，这印证了 SiamMOT 假阳性较高的结论。另外表 1 中 ByteTrack 方法的 MOTA 指标与本文方法基本相当，但 IDF1 较本文方法明显较低，这也印证了 ByteTrack 方法容易发生 ID 误切换。SiamRPN++ 速度偏慢，各项指标因为无法预测目标形状而明显偏低。本文提出的方法相较于其他方法具有显著更高的鲁棒性、较高的运行速度和跟踪精度。

3.2 多目标跟踪方法消融实验

本文对多目标跟踪方法进行了消融实验，在自采数据集上测试了本文方法在缺少部分步骤时的表现。图 7 从三个图像序列中各抽取一帧，对比展示了本文方法在缺少某一步骤时的跟踪结果。比较示例 a2 和 a3，缺少跟踪网络时，检测到的目标减少，目标响应率下降；在缺少多步匹配的情况下，在展示的全部示例中跟踪成功的目标普遍减少，方法中部分已经检测到的提案未能关联至前一帧，这是由于方法未能很好地利用低响应值提案中的信息所致；在缺少轨迹活跃度筛选时，对比示例 a、b、c，可以看到方法的假阳性数目明显增加。

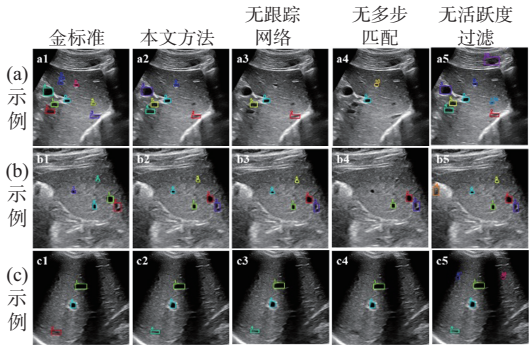


图 7 本文方法在缺失各算法组件的情况下在自采数据集上的结果展示

表 2 对比展示了在缺少各步骤时，本文方法在自采数据集上的各项指标。由表可知，多步匹配和跟踪网络均明显提升了方法的 MOTA 和 ID 召回率，因为这两个步骤通过挖掘低响应提案信息和增加检测手段提升了真阳性数目和占比，但由于两个步骤的加入跟踪精度有轻微的下降。而轨迹活跃度筛选显著提升了 MOTA 和 ID 精确率，ID 召回率稍有下降，因为轨迹活跃度筛选减少了假阳性的数目和占比，因此 ID 精确率提升较高，ID 召回率影响不显著。

表 2 缺失各算法组件的情况下，本文方法在自采数据集上的评估结果

方法	MOTP	MOTA/%	IDP/%	IDR/%	IDF1/%
本文方法	0.255	76.00	94.40	80.80	87.10
无跟踪网络	0.249	61.30	96.90	63.00	76.40
无多步匹配	0.243	41.50	95.80	43.40	59.80
无活跃度过滤	0.257	61.10	69.00	81.50	74.70

综上所述，多目标跟踪方法的多步匹配、跟踪网络和轨迹活跃度筛选三个步骤，通过不同的方式提升了方法的整体性能。

3.3 超声目标跟踪系统测试

超声目标跟踪系统核心功能包括方法演示、

标注和评估。首先，超声目标跟踪系统能够进行多目标跟踪的方法演示。方法演示功能主要对单目标跟踪和多目标跟踪功能进行了测试。分别点击方法下拉框选择单目标和多目标跟踪，再点击播放按钮，系统的单目标和多目标跟踪功能均运行正常，系统功能实机演示如图 8 所示。

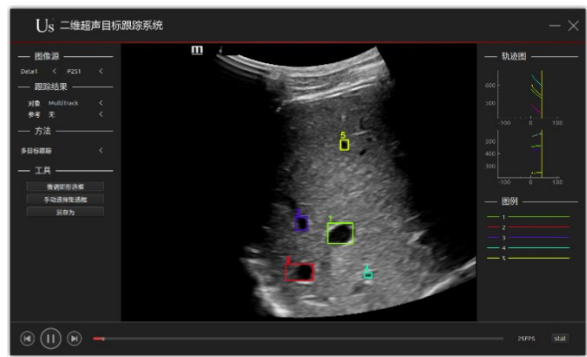


图 8 多目标跟踪功能系统实机演示

修改标注可以通过点击标注工具控件的两个按钮进行测试，另外可以在图例中测试对标注进行删除、添加和重选等操作。点击标注工具控件的“微调矩形选框”，可以在图例中点击修改的对象，然后通过键盘对包围盒进行像素级的修改。点击标注工具控件的“手动选择矩形框”的功能，能够通过鼠标拖拽标注新的目标或修改已有目标的跟踪结果，相应的实机功能演示如图 9 所示。

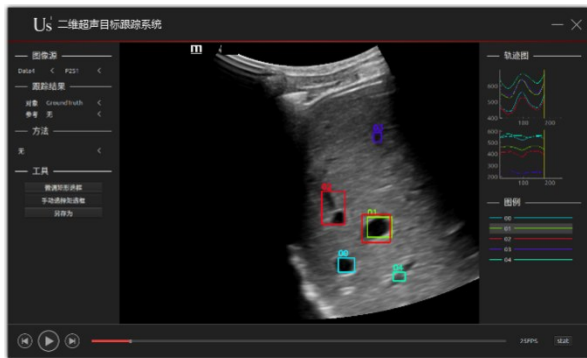


图 9 标注功能系统实机演示

测试评估功能可以点击评估按钮，进入评估对话框。通过下拉框选择当前数据集、需要评估的跟踪结果和评估所采用的金标准后，点击“评估”，将在右侧的显示区显示评估结果，上文提及的评估指标全部集成在“主要统计指标”标签下，而“各目标跟踪成功率”可以显示各序列中各目标的跟踪情况，方便研究者检查当前跟踪结果中的不足。评估功能的实机功能演示如图 10 所示。

本文测试了在系统上执行各个操作的平均响应时间。由于在临床场景下超声图像大小不超过 1000×1000 ，视野内的血管数目通常不超过 10，

本文在测试时设定对图像大小和标注目标个数进行了相应的限制，表 3 展示了性能测试的测试结果。



图 10 评估功能系统实机演示

表 3 系统性能测试表

测试内容	平均用时/s
刷新超声图像	9.23×10^{-3}
刷新图例控件	6.72×10^{-3}
刷新轨迹图控件	2.69×10^{-2}
初始化单目标跟踪方法	1.98
对单帧图像进行单目标跟踪	7.92×10^{-2}
初始化多目标跟踪方法	2.11
对单帧图像进行多目标跟踪	1.83×10^{-1}
对单个数据集进行统计	6.42

从测试结果来看，功能测试均获通过，表明系统具备在指定环境下平稳运行的能力，性能测试中，多目标跟踪算法未达到实时水平，还未能满足临床治疗的需要，系统还存在优化的空间。

4 结 论

本文提出了基于多步匹配的多目标跟踪方法，并基于该方法构建了超声目标跟踪系统。本文还测试了超声目标跟踪系统的功能和性能，系统运行平稳，能够实现设计的各项功能，系统的各项实时性水平达标，但多目标跟踪方法还需要在运算速度方面进行一定程度的优化。

从自然图像目标跟踪的发展来看，从单目标跟踪到多目标跟踪是算法实用化过程中重要的一步。临床上，对肝脏血管跟踪的运行速度要求为超过 24FPS，单目标位置误差低于 1mm，本文提出的方法还未能在运行速度和精度上达到临床需求。未来的超声多目标跟踪方法需要更加稳定的目标检测手段以减少目标关联阶段处理大量无效提案的压力。同时，简化算法流程和网络结构，对算法运行速度的改进提升，也无疑将是后续工作的主要改进方向。

参考文献：

[1] De Luca V, Banerjee J, Hallack A, et al. Evaluation of 2D and 3D ultrasound tracking algorithms and impact on ultrasound

- guided liver radiotherapy margins[J]. *Medical Physics*, 2018, 45 (11):4986—5003.
- [2] De Luca V, Benz T, Kondo S, et al. The 2014 liver ultrasound tracking benchmark[J]. *Physics in Medicine & Biology*, 2015, 60(14):5571.
- [3] Liu F, Liu D, Tian J, et al. Cascaded one-shot deformable convolutional neural networks: Developing a deep learning model for respiratory motion estimation in ultrasound sequences[J]. *Medical Image Analysis*, 2020, 65:101793.
- [4] Leal-Taixé L, Milan A, Reid I, et al. MOTChallenge 2015: Towards a benchmark for multi-target tracking[J/OL]. *arXiv preprint arXiv:1504.01942*, 2015.
- [5] Milan A, Leal-Taixé L, Reid I, et al. MOT16: A benchmark for multi-object tracking[J/OL]. *arXiv preprint arXiv:1603.00831*, 2016.
- [6] Dendorfer P, Rezatofighi H, Milan A, et al. MOT20: A benchmark for multi object tracking in crowded scenes[J/OL]. *arXiv preprint arXiv:2003.09003*, 2020.
- [7] Bewley A, Ge Z, Ott L, et al. Simple online and realtime tracking[C]//2016 IEEE International Conference on Image Processing (ICIP). Phoenix, USA:IEEE, 2016:3464—3468.
- [8] Wojke N, Bewley A, Paulus D. Simple online and realtime tracking with a deep association metric[C]//2017 IEEE International Conference on Image Processing (ICIP). Beijing, China: IEEE, 2017:3645—3649.
- [9] Zhang Y, Sun P, Jiang Y, et al. ByteTrack: Multi-object tracking by associating every detection box[C]//European Conference on Computer Vision (ECCV). Tel Aviv, Israel: Springer, 2022:1—17.
- [10] Bergmann P, Meinhardt T, Leal-Taixé L. Tracking without bells and whistles[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, South Korea: IEEE, 2019:941—951.
- [11] Tang S, Andriluka M, Andres B, et al. Multiple people tracking by lifted multicut and person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017: 3539—3548.
- [12] Shuai B, Berneshawi A, Li X, et al. SiamMOT: Siamese multi-object tracking[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE, 2021:12372—12382.
- [13] Kieritz H, Hubner W, Arens M. Joint detection and online multi-object tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Salt Lake City, USA:IEEE, 2018:1459—1467.
- [14] Pearson K. LIII. On lines and planes of closest fit to systems of points in space[J]. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 1901, 2(11):559—572.
- [15] Wang M, Liu Y, Huang Z. Large margin object tracking with circulant feature maps[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA:IEEE, 2017:4021—4029.
- [16] Kim S J, Nam J Y, Ko B C. Online tracker optimization for multi-pedestrian tracking using a moving vehicle camera[J]. *IEEE Access*, 2018, 6:48675—48687.
- [17] Zhou X, Koltun V, Krähenbühl P. Tracking objects as points[C]//European Conference on Computer Vision (ECCV). Glasgow, UK:Springer, 2020:474—490.
- [18] Feng W, Hu Z, Wu W, et al. Multi-object tracking with multiple cues and switcher-aware classification[J/OL]. *arXiv preprint arXiv:1901.06129*, 2019.
- [19] Weng X, Wang J, Held D, et al. AB3DMOT: A baseline for 3D multi-object tracking and new evaluation metrics[C]//IEEE International Conference on Robotics and Automation (ICRA). Xi'an, China:IEEE, 2021:10942—10948.
- [20] Kim A, Ošep A, Leal-Taixé L. EagerMOT: 3D multi-object tracking via sensor fusion[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Prague, Czech Republic: IEEE, 2021:8148—8155.
- [21] Wang Z, Zheng L, Liu Y, et al. Towards real-time multi-object tracking[C]//European Conference on Computer Vision (ECCV). Glasgow, UK:Springer, 2020:107—122.
- [22] Yu F, Wang D, Shelhamer E, et al. Deep layer aggregation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, USA:IEEE, 2018:2403—2412.
- [23] He K, Gkioxari G, Dollár P, et al. Mask R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017:2961—2969.
- [24] Li B, Yan J, Wu W, et al. High performance visual tracking with siamese region proposal network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, USA:IEEE, 2018:8971—8980.
- [25] Bernardin K, Stiefelhagen R. Evaluating multiple object tracking performance: the CLEAR MOT metrics[J]. *EURASIP Journal on Image and Video Processing*, 2008, 2008:1—10.
- [26] Ristani E, Solera F, Zou R, et al. Performance measures and a data set for multi-target, multi-camera tracking[C]//European Conference on Computer Vision (ECCV). Amsterdam, Netherlands: Springer, 2016:17—35.
- [27] Li B, Wu W, Wang Q, et al. SiamRPN++: Evolution of siamese visual tracking with very deep networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA:IEEE, 2019:4282—4291.