

# 双半球胶囊机器人自监督姿态快速矫正方法<sup>\*</sup>

马钰琳 张永顺 邢立君  
(大连理工大学机械工程学院 大连 116024)

**摘要:** 为了解决胃肠道粘弹性阻尼和胶囊质心偏移等非线性因素会使胶囊机器人姿态偏离控制目标,胶囊运动过程中大视差和运动图像模糊所导致的姿态估计误差等问题,提出一种自监督姿态快速矫正方法。在胶囊初始姿态标定方面,消除了初始自转角的影响;在胶囊姿态估计方面,基于注意力机制 SAB 和 TAM,使用深度可分离卷积替代部分标准卷积,设计出一种时空注意力机制 TSAM 并嵌入到 PoseNet 中得到注意力姿态估计网络 APEN,加强模型特征提取能力。实验表明,相较于现有胶囊姿态估计方法 Endo-SfM,APEN 可在保持推理速度几乎不变的同时,相对姿态估计精度提高 52%,并且该方法能实时、有效的对胶囊姿态进行矫正,姿态控制精度提高 38.8%,为实现胃肠道高效动态诊疗奠定了基础。

**关键词:** 双半球胶囊机器人;注意力机制;自监督学习;姿态矫正

**中图分类号:** TP242.6 **文献标识码:** A **国家标准学科分类代码:** 460.50

## Posture rapid correction for a double hemisphere capsule robot through self-supervised learning

Ma Yulin Zhang Yongshun Xing Lijun  
(School of Mechanical Engineering, Dalian University of Technology, Dalian 116024, China)

**Abstract:** To solve the problem that the posture of the double hemisphere capsule robot (DHCR) was likely to deviate from the targeted orientation due to nonlinear factors such as viscoelastic damping of the gastrointestinal (GI) tract and DHCR centroid deviation, as well as posture estimation error caused by large visual disparity and motion blur, a self-supervised learning-based rapid posture correction method was proposed. In terms of calibration of the capsule's initial posture, the influence of the initial rotation angle was eliminated; in terms of estimation of the capsule's attitude, based on spatial attention block(SAB) and temporal attention module (TAM), a spatial-temporal attention mechanism (TSAM) was designed by replacing a component of the standard convolution with the depthwise separable convolution and embedding it into PoseNet to generate an attention posture estimation network (APEN), which enhanced the model's ability to extract features. The experimental results show that APEN can increase relative posture estimate accuracy by 52% while keeping inference speed almost unchanged when compared to the current capsule posture estimation method, Endo-SfM. Moreover, the accuracy of posture control is increased by 38.8% with this method, and it can correct the capsule posture in real time, laying the foundation for successful dynamic GI tract diagnosis and therapy.

**Keywords:** dual hemisphere capsule robot;attention mechanism;self-supervised learning;posture correction

## 0 引言

随着人们对胃肠道疾病检测需求的增加,无痛胶囊内窥镜的出现给患者带来了福音。当前,临床应用胶囊内窥镜无法根据自身姿态信息进行精准的主动控制,难以实施活检和靶向给药等诊疗作业<sup>[1]</sup>。同时,受胶囊内置电源能

量限制<sup>[2]</sup>,胶囊尚不能进行胃肠道全周期检测任务。如果才能保证胶囊姿态的精确控制,可望显著提高疾病检测的全面性和有效性,因此,开展主动型胶囊机器人快速姿态矫正方法的研究具有重要意义。

目前,临床上主动胶囊主要采用体外磁铁以磁悬浮方式控制胶囊姿态与运动,原理简单,操作方便。但存在磁力

收稿日期:2022-11-04

<sup>\*</sup> 基金项目:国家自然科学基金(61773084,62173059)项目资助

和磁矩耦合,力平衡不易控制等问题,容易对胃肠道产生较大冲击,采用纯磁矩驱动式胶囊机器人能有效改善上述问题,并成为国际研究热点<sup>[3]</sup>。为消除采用磁铁存在奇异平面<sup>[4]</sup>而导致胶囊姿态控制非唯一性问题,以空间万向均匀旋转磁场为驱动源,我们提出了一种双半球胶囊机器人<sup>[5]</sup>,基于旋转磁场和永磁体之间的磁矩随动效应,采用动平衡控制能够实现快速定点调姿和滚动行走。

现有胶囊姿态信息的获取主要采用传感器和视觉估计两种检测方法。使用体外磁阵列传感器<sup>[7]</sup>或胶囊内嵌加速度计<sup>[8]</sup>可以实现胶囊姿态高精度实时检测。上述方法尽管精度高、可靠性和实时性好,但受限于胶囊体积,对硬件的微型化要求苛刻,尚难以满足要求。视觉姿态估计方法功耗低,硬件资源需求少,应用前景良好。其中,胶囊视觉姿态估计方法主要有两类:一类是基于特征点匹配法<sup>[9]</sup>,不依赖特定工作环境,能对存在较大位移的两帧图像进行姿态估计,但在计算速度和估算精度上难以权衡,不适用于缺乏特征的低纹理光滑区域;另一类则是基于自监督学习方法<sup>[10]</sup>,无需真实姿态标签即可训练,在图像的低纹理区域中依然具有较高推理速度和估计精度,但在快速运动图像数据中效果较差。并且,胶囊在运动过程中拍摄得到的视频图像数据有着较为复杂的时序动态特性<sup>[11]</sup>。

因此,对于临床中胶囊机器人难以实现准确的姿态控制,和在高效诊疗任务中胶囊快速运动时的大视差和运动图像模糊<sup>[6]</sup>等问题,本文针对双半球胶囊机器人提出了一种自监督姿态快速矫正方法。该方法通过空间注意力模块(spatial attention block, SAB)<sup>[10]</sup>强调空间维度的特征,通过时域自适应模型(temporal attention module, TAM)<sup>[11]</sup>自适应的聚合时序信息,并采用深度可分离卷积替换部分标准卷积的方式,将二者加权融合得到一种时空注意力机制(temporal-spatial attention mechanism, TSAM),使得姿态估计网络在不增加额外计算量的前提下,可动态捕捉胶囊图像数据的中运动信息,为胶囊胃肠道疾病高效检测和精准诊疗奠定了基础。

## 1 双半球胶囊机器人

### 1.1 机器人结构及工作原理

双半球胶囊机器人包括主动半球和被动半球两部分<sup>[5]</sup>,如图1所示。主动半球与磁环固联,被动半球包含图像采集元件、LED照明电路、电池、发射天线和射频电路等元件,主动半球和被动半球由轴承连接,并可以相对转动。

双半球胶囊机器人工作原理如图2所示,被动半球处于欠驱动状态,与磁铁静态平衡控制方式不同,在空间万向旋转磁场的作用下,内嵌磁环的主动半球处于旋转的动平衡状态,基于随动效应<sup>[5]</sup>,通过控制旋转磁矢量 $\mathbf{B}$ 轴线方向即可获得胶囊初始姿态(a)被动模态(b)和主动模态(c)3种运动状态由于采用磁陀螺进动态平衡控制,胶囊姿态具有良好的抗干扰能力。

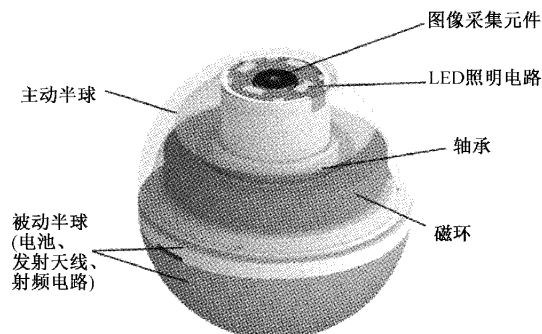


图1 双半球胶囊机器人结构

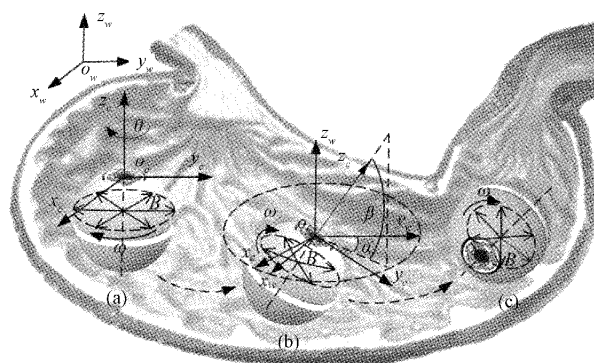


图2 双半球胶囊机器人工作原理示意图

如图2(a)所示,机器人初始姿态和磁场方位都竖直向上,此时,磁场方位误差为零且胶囊初始姿态不受胶囊重力偏心矩影响<sup>[9]</sup>。此外,因相机与被动半球固联,相机和胶囊轴线方向相同,因此相机姿态即为胶囊姿态。自转角 $\theta$ 为 $0^\circ$ 时,相机坐标系 $o_c-x_c y_c z_c$ 与固定坐标系即磁场线圈坐标系 $o_w-x_w y_w z_w$ 轴线方向重合;如图2(b)所示,处于被动模态的胶囊机器人执行定点全方位观察任务,此时,旋转磁矢量 $\mathbf{B}$ 轴线以一定夹角绕 $z_w$ 轴执行锥体旋转运动,实现胶囊定点全方位控制。 $z_c$ 轴在固定坐标系 $o_w-x_w y_w z_w$ 平面上的投影与 $y_w$ 轴的夹角为侧摆角 $\alpha$ ,与 $z_c$ 轴的夹角为俯仰角 $\beta$ ;如图2(c)所示,主动模态胶囊执行滚动行走任务,此时旋转磁矢量 $\mathbf{B}$ 轴线方向水平。

### 1.2 机器人图像初始自转角的标定

图像采集元件嵌入被动半球,由于双半球胶囊被动半球为欠驱动,因此,在体内调姿过程中胶囊自转角具有随机性,则胶囊在连续运动过程中拍摄的视频也会旋转,进而影响胶囊视觉姿态估计结果,医生无法直接由视频图像获取信息。因此,本文提出如下自转角修正策略,分为自转角初始修正和动态调姿过程自转角修正两部分。初始姿态自转角的标定,即修正过程如图3所示。

在固定坐标系 $o_w-x_w y_w z_w$ 中,为了减小自转角的响应对姿态估计的影响,初始自转角修正过程为:1)在胶囊初始姿态下拍摄图像 $I_0$ 。后,将胶囊偏转较小的俯仰角 $\beta$ ,得到图像 $I_1$ ;2)使胶囊回归初始姿态,通过初始图像 $I_0$ 、偏转后图像 $I_1$ 和消除初始自转角 $\theta$ 的图像 $I_0'$ 三者相机坐标系 $o_c$ -

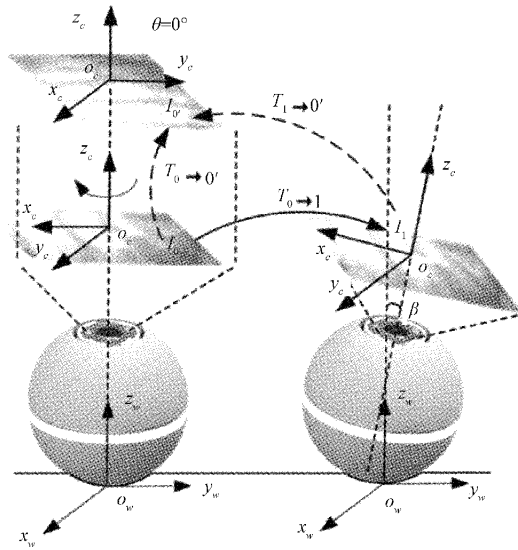


图 3 机器人图像初始自转角修正过程

$x_c, y_c, z_c$  之间的变换关系可得初始自转角  $\theta$  的旋转矩阵为:

$$T_{0 \rightarrow 0'} = T_{1 \rightarrow 0'} T_{0 \rightarrow 1} \quad (1)$$

式中:  $T_{0 \rightarrow 1}$  是图像  $I_0$  到图像  $I_1$  的旋转矩阵, 由视觉算法估计得到,  $T_{1 \rightarrow 0'}$  是图像  $I_1$  到图像  $I_{0'}$  的旋转矩阵,

$$T_{1 \rightarrow 0'} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sin\beta & -\cos\beta \\ 0 & \cos\beta & \sin\beta \end{pmatrix}^{-1},$$

其中,  $\beta$  是侧摆角  $\alpha$  为零时, 图像  $I_1$  所对应机器人的俯仰角; 3) 基于欧拉角的旋转变换由  $T_{0 \rightarrow 0'}$  可得相机坐标系  $o_c - x_c, y_c, z_c$  绕  $z_c$  轴的初始自转角  $\theta$ ; 4) 将初始图像  $I_0$  旋转负  $\theta$  角即为无自转角的理想初始图像  $I_{0'}$ , 完成初始图像自转角的标定过程。此外, 根据旋转矩阵  $T$  的一般形式

$$T = \begin{pmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{pmatrix},$$

由文献[9]可知机器人姿态角计算公式为:

$$\alpha = \begin{cases} -\arctan(R_{13}/R_{23}), & (R_{13}/R_{23}) \leq 0 \\ \pi + \arctan(R_{13}/R_{23}), & (R_{13}/R_{23}) > 0 \end{cases} \quad (2)$$

$$\beta = \begin{cases} \arccos(R_{33}), & (R_{33} \geq 0) \\ \pi - \arccos(R_{33}), & (R_{33} < 0) \end{cases}$$

动态调姿过程自转角修正中, 尽管消除了初始姿态角的影响, 但在肠道非结构环境等因素影响下, 自转角  $\theta$  还会发生变化, 也应该消除自转角  $\theta$  变化对姿态检测结果的影响, 自转角  $\theta$  对胶囊连续轨迹姿态估计任务的影响可通过逐次估计任意两帧连续图像转角变化值的补偿过程来消除, 进而实现无自转角影响的连续图像姿态估计。

## 2 胶囊姿态矫正方法

### 2.1 机器人姿态控制系统

基于视觉的双半球胶囊机器人姿态控制系统概念图如

图 4 所示, 由上位机输入目标姿态角度信息, 经控制器和变压器转换为电流信号, 驱动三轴亥姆霍兹线圈形成万向旋转磁场, 基于随动效应实现胶囊调姿。为实时估计胶囊姿态, 我们将姿态估计网络嵌入控制系统。姿态估计网络以机器人拍摄的视频连续帧为输入, 可实时获取胶囊姿态角, 将其与目标姿态角度比较可得姿态偏差, 补偿后的姿态角输入为:

$$P'(\alpha', \beta') = 2P_i(\alpha_i, \beta_i) - \hat{P}(\alpha', \beta') \quad (3)$$

其中,  $P'$  为补偿后的输入姿态角,  $P_i$  为目标姿态角,  $\hat{P}$  为姿态网络估计的姿态角。通过改变三轴亥姆霍兹线圈内的旋转磁场方向, 来矫正机器人姿态, 进而实现对机器人姿态的精准控制。

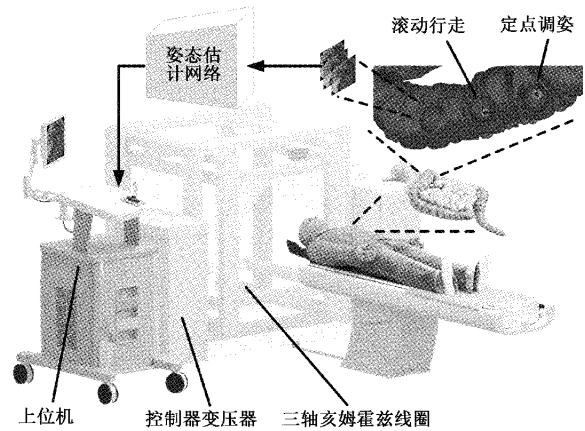


图 4 双半球胶囊机器人姿态控制系统概念图

### 2.2 APEN 网络结构

本文提出的胶囊姿态估计网络 (attention posture estimation network, APEN) 主要是在 SC-SfMLearner<sup>[12]</sup> 的基础上进行改进。网络的输入为两帧相邻的胃肠道图像, 输出为两图像间的相对姿态, 网络模型的结构如图 5 所示, 姿态估计网络包含编码层 Encoder 和解码层 Decoder 两部分。为了更好的适应胶囊机器人获取图像的时序动态特性<sup>[11]</sup>, 我们在以 ResNet18 作为骨干网络的编码层嵌入轻量级的时空注意力机制 TSAM。同时, 将 TSAM 中的部分卷积核为  $1 \times 1$  普通卷积替换为  $3 \times 3$  的深度可分离卷积, 在增大感受野的同时有效减少注意力模型的参数量。

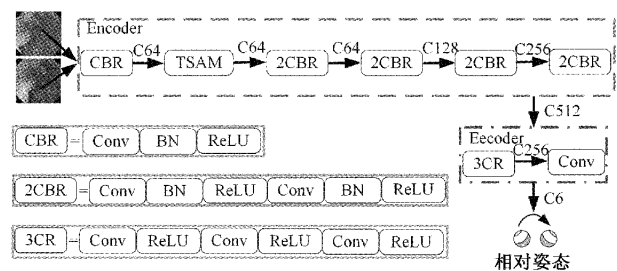


图 5 姿态估计网络 APEN 结构

### 2.3 时空注意力机制 TSAM

在姿态估计时,为了消除输入图像不同位置的局部特征对估计结果的影响,空间注意力机制能够对网络重要区域设置更多的权重。由于姿态估计是一个时序任务,姿态网络以两帧图像为输入,需要时序注意力机制来学习相邻帧的时序信息,因此提出一种时空注意力机制模块 TSAM,如图 6 所示。

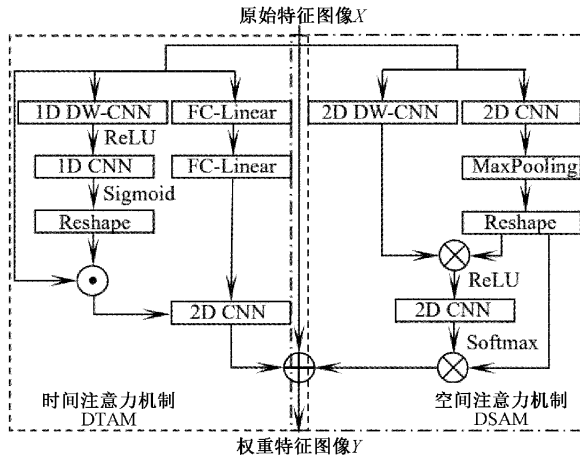


图 6 TSAM 注意力机制流程

TSAM 采用并行结构将输入特征图与时序注意力机制 (depthwise temporal attention mechanism, DTAM) 和空间注意力机制 (depthwise spatial attention mechanism, dsam) 输出的特征信息加权融合。其中, dsam 采用非局部卷积<sup>[13]</sup>, 学习局部空间图像特征, 强调图像中的纹理细节和梯度差异, 提取特征图的空间信息, 而 DTAM 利用一维时序卷积和全连接层分别学习局部和全局时序信息, 随后通过逐通道时序卷积得到具有时序信息的特征图。最后将原输入特征图与带有空间信息和时序信息的特征图加权融合, 使网络同时关注空间信息和时序信息, 进而更准确的估计运动胶囊的姿态信息。

### 2.4 自监督学习训练

胶囊机器人在胃肠道环境中难以获得带姿态标签的大量图像数据, 难以满足监督或半监督学习条件, 因此, 采用无需标签自监督学习训练框架<sup>[12]</sup>, 通过联合训练深度估计网络和姿态估计网络的方式对模型进行训练。同时为加快模训练时的收敛速度, 在 EndoSLAM 数据集<sup>[10]</sup>中训练好的模型权重应用到新模型中, 后对网络进行训练微调。

## 3 胶囊姿态估计算法评估

### 3.1 实验环境及评价指标

模型训练采用 EndoSLAM 数据集由 Unity3D 合成 2 000 帧虚拟胃肠道图像数据。模型测试采用 EndoSLAM 的子集 LowCam(Stomach-III, Traj-III) 中的图像数据(胶囊运动速度和加速度分别为 25.97 mm/s, 519.361 mm/s<sup>2</sup>, 图像分辨率为 640×480)。算法采用 Pytorch 框架构建, 在

台式机(内存:8 GB, 操作系统:Ubuntu18.04LTS)和 8 G 显存的 NVIDIA GeForce RTX 3060Ti 上进行训练和测试。

算法性能评估包涵姿态估计精度和姿态网络推理速度两部分, 其中姿态估计精度采用相对姿态误差的旋转部分 (RPE. rot)<sup>[14]</sup>, 姿态网络的推理速度则采用每秒处理的图像帧数 FPS 来衡量。相对姿态误差为:

$$Rot RPE_i(\Delta) = \angle \left( rot \left( \begin{pmatrix} (Q_i^{-1} Q_{i+\Delta})^{-1} \\ (P_i^{-1} P_{i+\Delta}) \end{pmatrix} \right) \right) \quad (4)$$

其中, Rot 为相对位姿误差 (relative pose error, RPE) 的旋转部分,  $i$  为时序或帧,  $\Delta$  为图像帧时间间隔,  $\angle(rot(\cdot))$  为旋转角度,  $Q_1, \dots, Q_2, \dots, Q_n \in SE(3)$  分别为图像帧对应的真实位姿,  $P_1, \dots, P_2, \dots, P_n \in SE(3)$  分别为算法估计位姿。

### 3.2 注意力机制对比及消融实验

本文 TSAM 模块消融因素包括 DSAM 和 DTAM。消融对比实验过程为: 1) 将 SC-SfMlearner<sup>[12]</sup> 作为 Baseline, 主干网络采用 ResNet18。2) 加入 SAB<sup>[10]</sup> 模块和 TAM<sup>[11]</sup> 模块, 并引入卷积核大小为 3 的深度可分离卷积, 得到 DSAM 和 DTAM。3) 加入 TSAM 进行消融分析, 将本算法同其他单目视觉无监督姿态估计算法进行对比, 实验结果如表 1 所示。表中, Baseline 标注为“B”。

表 1 姿态估计算法消融及对比分析

| 方法                              | LowCam(Stomach-III, Traj-III) |               |
|---------------------------------|-------------------------------|---------------|
|                                 | Rot. RPE                      | 推理速度/FPS      |
| B <sup>[12]</sup>               | 2.041(0.839)                  | 35.070        |
| SfMlearner <sup>[15]</sup>      | 0.802(0.424)                  | 34.960        |
| Monodepth2 <sup>[16]</sup>      | 0.735(0.835)                  | 36.740        |
| Endo-SfM(B+SAB) <sup>[10]</sup> | <b>0.671(0.382)</b>           | <b>35.356</b> |
| B+DSAM                          | 0.466(0.230)                  | 35.320        |
| B+ASPP <sup>[17]</sup>          | 0.426(0.282)                  | 34.226        |
| B+TAM <sup>[11]</sup>           | 0.423(0.239)                  | 35.003        |
| B+DTAM                          | 0.384(0.255)                  | 35.018        |
| B+TSAM(APEN)                    | <b>0.349(0.235)</b>           | <b>35.007</b> |

表 1 说明: 1) 在 SAB 和 TAM 模块基础上, 采用大卷积核深度可分离卷积得到的 DSAM 和 DTAM 不增加额外计算量依然能有效降低相对姿态误差。原因是较大的卷积核增大了网络的感受野, 深度可分离卷积降低了参数量; 2) 所提出的 B+TSAM 与 B+SAB、B+DSAM、B+TAM 和 B+DTAM 相比, 误差更小, 但推理速度相近。说明 DSAM 可优化网络在空间域的特征图权重配比, 而 DTAM 可学习图像的时序信息; 3) 与 B+TSAM 相比, SfMlearner<sup>[15]</sup> 和 Monodepth2<sup>[16]</sup> 由于没考在损失函数中虑环境光照变化对训练过程的影响且忽略了相邻图像之间的全局特征, 导致相对姿态误差较高; 4) B+ASPP<sup>[17]</sup> 则说明通过堆叠空洞卷积能够有效增大网络感受野, 提高估计精度, 但引入额外的卷积运算, 推理速度明显降低。

综上所述, APEN 在快速运动场景下, 在保持实时推理速度基本不变的同时, 姿态估计精度最优, 可满足双半球胶囊机器人运动过程中实时姿态估计及矫正的需求。

### 4 胶囊姿态矫正试验

为验证姿态矫正效果, 设计了三轴亥姆霍兹线圈和双半球胶囊机器人样机。胶囊可在磁场区域内定点调姿, 姿态估计算法推理速度为 35 FPS, 拍摄图像的分辨率为  $640 \times 360$ , 帧率为 25 FPS。试验中, 在胃部模型内部预先铺设肠道组织来模拟非结构性的粘弹性环境, 主要有变压器、控制器、上位机控制系统、三轴亥姆霍兹线圈、双半球胶囊机器人样机、个人电脑和离体猪大肠等, 如图 7 所示。

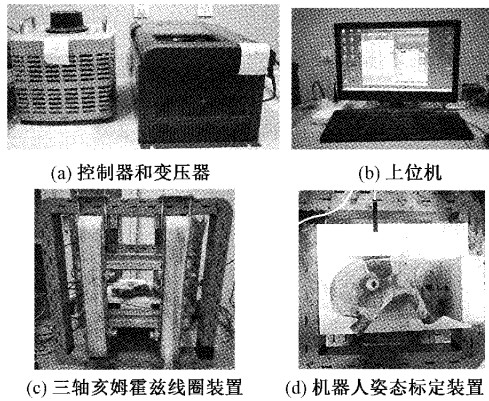


图 7 胶囊姿态矫正试验装置图

其中, 沿固定坐标系  $x_w$  轴和  $y_w$  轴方向分别放置微型摄像头, 以便在固定磁场坐标系内对胶囊姿态进行实时标定, 原理如图 8 所示。

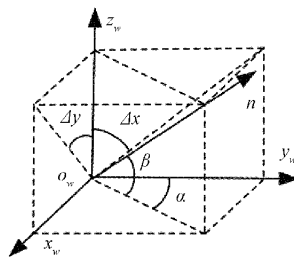


图 8 胶囊姿态标定原理

胶囊轴线  $n$  在  $x_w$  轴和  $y_w$  轴方向上的转角  $\sigma_x$ 、 $\sigma_y$  和机器人姿态角  $\alpha$ 、 $\beta$  之间的变换关系为:

$$\begin{aligned} \alpha &= \arctan(\tan(\sigma_x)/\tan(\sigma_y)) \\ \beta &= \arctan(\sqrt{\tan^2(\sigma_x) + \tan^2(\sigma_y)}^{-1}) \end{aligned} \quad (5)$$

由于姿态网络估计侧摆角趋近于  $0^\circ$ , 因此我们只对俯仰角进行补偿, 试验结果如表 2 所示, 在不同角度范围内, 该方法只进行一次姿态方位补偿便可有效降低姿态角偏差。

表 2 胶囊姿态矫正试验结果 ( $^\circ$ )

| 输入姿态角( $\alpha, \beta$ ) |          |          | 矫正前偏差 |     | 矫正后偏差 |     |
|--------------------------|----------|----------|-------|-----|-------|-----|
| 初始值                      | 目标值      | 补偿值      | 标定值   | 均值  | 标定值   | 均值  |
|                          | (0,74.1) |          | 8.3   | 8.4 | 2.4   | 3.5 |
|                          | (0,80)   | (0,75.6) | 9.2   | 7.0 | 4.3   | 1.4 |
|                          | (0,75.9) |          | 7.6   | 9.7 | 3.7   | 5.3 |
|                          | (0,62.7) |          | 8.2   |     | 1.3   |     |
| (0,90)                   | (0,70)   | (0,64.4) | 8.7   |     | 2.6   |     |
|                          | (0,63.7) |          | 4.1   |     | 0.3   |     |
|                          | (0,57.3) |          | 12.1  |     | 5.9   |     |
|                          | (0,60)   | (0,58.3) | 8.3   |     | 2.4   |     |
|                          | (0,58.9) |          | 8.7   |     | 7.7   |     |

### 5 结 论

针对快速运动胶囊由大视差和运动图像模糊问题引起的姿态估计误差难题, 本文的创新点是: 1) 提出了双半球胶囊初始姿态标定方法, 不仅消除了自转角度对姿态估计的影响, 为提高胶囊姿态估计精度奠定了基础; 2) 提出了一种胶囊姿态估计方法, 利用深度可分离卷积替换 SAB 和 TAM 中的部分普通卷积, 将其加权融合得到时空注意力机制 TSAM, 并在 PoseNet 中嵌入 TSAM 加强了网络对时序信息的提取能力。最终 EndoSLAM 胃肠道环境胶囊姿态数据集测试结果表明, 本算法在保证推理速度的前提下, 显著提高快速运动胶囊的姿态估计精度。胃部模型内姿态矫正试验结果也表明, 该方法通过实时有效的矫正, 在不增加硬件成本的前提下, 提高了双半球胶囊机器人姿态的控制精度, 为实现胃肠道病变区域的高效检测、定点活检和精准施药等诊疗操作奠定了基础。

### 参考文献

- [1] VALDIVIA P C, ROBERTSON A R, DE BOER N K H, et al. An overview of robotic capsules for drug delivery to the gastrointestinal tract[J]. Journal of Clinical Medicine, 2021, 10(24): 5791-5791.
- [2] LIU L, TOWFIGHIAN S, JIN Z. A cylindrical triboelectric energy harvester for capsule endoscopes[C]. 11th IEEE Annual Biomedical Circuits and Systems Conference(BioCAS), 2015: 1-4.
- [3] KIM J, NAM S. Capsule endoscopy for gastric evaluation [J]. Diagnostics, 2021, 11 ( 10 ): 1792-1792.
- [4] TADDESE A Z, SLAWINSKI P R, PIROTTA M, et al. Enhanced real-time pose estimation for closed-loop robotic manipulation of magnetically actuated capsule endoscopes [J]. International Journal of Robotics Research, 2018, 37(8): 890-911.
- [5] 张永顺, 周华涛, 张林霞, 等. 一种新型双半球形胶囊

- 机器人[J]. 机械工程学报, 2017, 53(15): 110-118.
- [6] XU Y F, DE CROON G. CNN-based ego-motion estimation for fast MAV maneuvers [C]. IEEE International Conference on Robotics and Automation (ICRA), 2021: 7606-7612.
- [7] XU Y X, LI K Y, ZHAO Z Q, et al. A novel system for closed-loop simultaneous magnetic actuation and localization of WCE based on external sensors and rotating actuation [J]. IEEE Transactions on Automation Science and Engineering, 2021, 18(4): 1640-1652.
- [8] SPERRY A J, CHRISTENSEN J J, ABBOTT J J, et al. Six-degree-of-freedom localization with a 3-axis accelerometer and a 2-axis magnetometer for magnetic capsule endoscopy[J]. IEEE Robotics and Automation Letters, 2022, 7(2): 2110-2115.
- [9] 张永顺, 王智博, 刘旭, 等. 磁控双半球胶囊机器人姿态的图像检测方法[J]. 华中科技大学学报(自然科学版), 2020, 48(10): 14-19.
- [10] OZYORUK K B, GOKCELER G I, BOBROW T L, et al. EndoSLAM dataset and an unsupervised monocular visual odometry and depth estimation approach for endoscopic videos [J]. Medical Image Analysis, 2021, 71: 102058.
- [11] LIU Z, WANG L, WU W, et al. TAM: Temporal adaptive module for video recognition[C]. IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 13688-13698.
- [12] BIAN J W, LI Z C, WANG N Y, et al. Unsupervised scale-consistent depth and ego-motion learning from monocular video [C]. 33rd Conference on Neural Information Processing Systems (NIPS), 2019: 35-45.
- [13] WANG X, GIRSHICK R, GUPTA A, et al. Non-local neural networks[C]. 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 7794-7803.
- [14] STURM J, ENGELHARD N, ENDRES F, et al. A benchmark for the evaluation of RGB-D SLAM systems[C]. IEEE/RSJ International Conference on Intelligent Robots and Systems, 2012: 573-580.
- [15] ZHOU T H, BROWN M, SNAVELY N, et al. Unsupervised learning of depth and ego-motion from video[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 6612-6619.
- [16] GODARD C, MAC AODHA O, FIRMAN M, et al. Digging into self-supervised monocular depth estimation[C]. IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 3827-3837.
- [17] CHEN L, PAPANDREOU G, KOKKINOS I, et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.

#### 作者简介

马钰璘, 硕士研究生, 主要研究方向机器人视觉姿态估计。

E-mail: myl\_apply@163.com

张永顺(通信作者), 博士, 教授, 主要研究方向为机器人机电一体化技术等。

E-mail: zyshun@dlut.edu.cn