复旦大学博士研究生学位论文答辩

细粒度医疗行为识别与技能评估技术研究

Research of Fine-grained Medical Action Recognition and Skill Assessment Technologies

答辩人：王顺利 (19级本科直博生)

专　业：计算机应用技术

导　师：张立华 教授

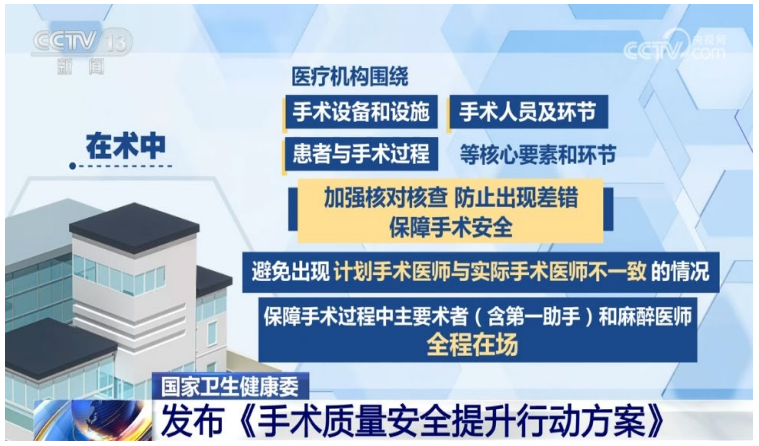
日　期：2024年5月25日

目 录

1.1 研究背景与意义

我国正面临着**医疗资源短缺**、**医疗服务系统承压过重**、**地域医疗资源分布不均匀**等问题，**填补医务人员数量缺口**的举措势在必行。

医护比指标	2020~2025目标
每千人口执业医师数	2.9➔3.2
每千人口注册护士数	3.34➔3.8
医护比	1:1.15➔1:1.20
卫生人员比	1:1.48➔1:1.62
执业医师缺口数量	43.29万人
护士缺口数量	66.37万人



- 国家卫健委2022年印发的《医疗机构设置规划指导原则（2021—2025年）》指出：

> “要强化信息化的支撑作用，切实落实医院、基层医疗卫生机构信息化建设标准与规范，推动人工智能、大数据、云计算、5G、物联网等新兴信息技术与医疗服务深度融合，构建优质均衡高效的医疗服务体系”

- 国家卫生健康委2023年印发的《手术质量安全提升行动方案》指出：

> “强化手术人员及环节核查，严防手术部位错误、手术用物遗漏、植入物位置不当、手术步骤遗漏等问题；严格执行手术室无菌技术、各项操作流程及技术规范，规范使用抗菌药物、止血药物和耗材”

- **关键问题**：如何在**保证医疗服务质量**的前提下，有效**提高医务人员的培训效率**、**同时降低培训成本**。
- **研究目标**：通过**智能医疗技能评估系统**的构建，替代传统培训与考核方式中的资深医师，实现医疗技能的自动化精准评估。

1.1 研究背景与意义

传统的医疗技能培训考核方式

由资深医师进行医疗技能的教学与考核

缺点

- 医生接诊与教学工作繁重
- 医疗技能教学考核效率较低
- 人力成本过高

基于人工智能技术的医疗技能评估系统

通过传感器技术由AI系统完成医疗技能的自动评估

优点

- 减轻一线医生工作负担
- 大幅提升技能教学考核效率
- 节约教考环节的人力成本

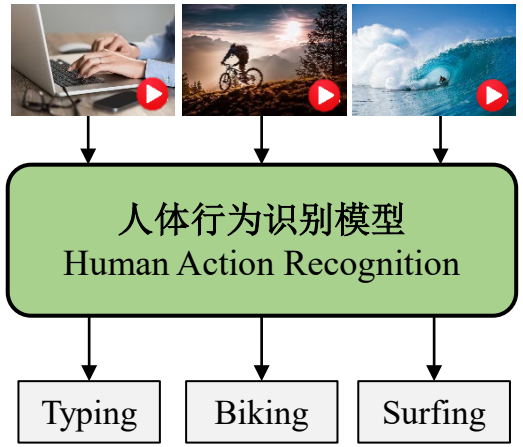
- **在微观层次**：智能的医疗技能评估系统能够**有效提升医疗培训和考核的效率**，从而显著**减轻一线医生在教考工作中的负担**，大幅**降低培训环节中的人力成本投入**；
- **在宏观层次**：智能评估系统能够**提升医院的数字化与信息化建设水平**，从一定程度上能够**减少地域医疗资源分布的差异**，有助于**提升国家的医疗系统整体服务质量与效率。**

1.2 研究现状

现有的视频理解任务中，与本文所探究的**医疗技能评估技系统**最相关的任务包括：

- **人类行为识别任务** / Human Action Recognition / HAR：对人体行为视频进行分类；
- **行为质量评估任务** / Action Quality Assessment / AQA：对人体行为的完成质量进行评估；
- **时序行为分割任务** / Temporal Action Segmentation / TAS：对未剪辑的长视频进行逐帧行为标签预测。

Kinetics-400、YouTube-8M、ActivityNet、Sports-1M、Olympic、FineGym、FineDiving 、…

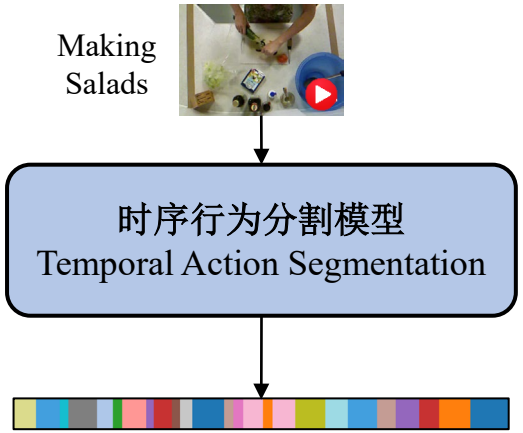


行为类别信息

MTL-AQA、AQA-7、FisV-5、FR-FS、UI-PRMD、…

行为质量评估分数

Breakfast、50Salads、GTEA、Cataract-101、Hei-Chole、…



帧级别行为类别预测结果

1.3 现存问题

医疗研究领域对计算机辅助医疗技能评估系统构建进行了初期探索，提出了一些数据集与算法：

行为质量评估任务
Action Quality Assessment

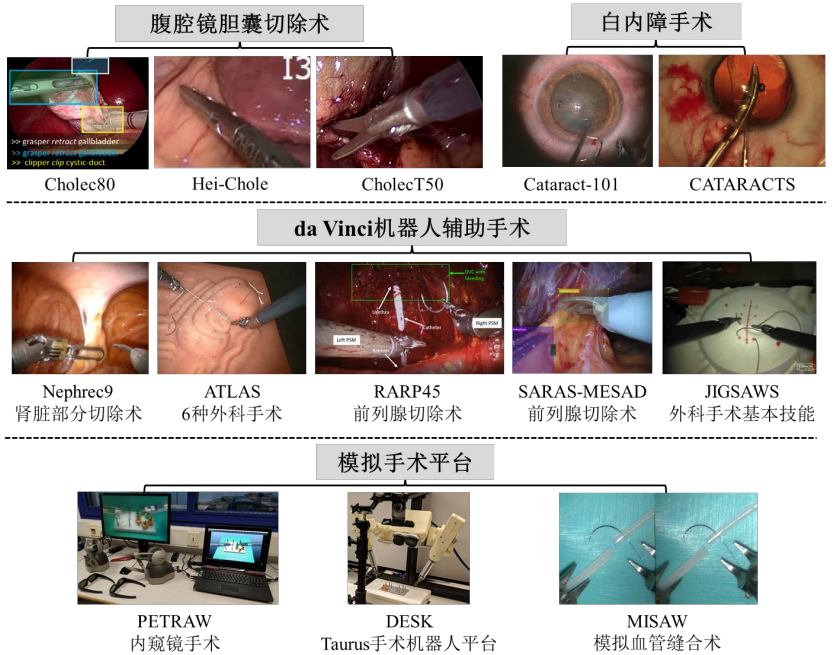
模拟外科手术操作
佐治亚理工学院

da Vinci 手术机器人操作
约翰斯霍普金斯大学

腹腔镜手术模拟平台
亚利桑那州立大学

医疗康复训练技能评估
UI-PRMD数据集

手术流程识别任务
Surgical Workflow Recognition



尽管以上研究取得了一定进展，但是医疗技能评估研究仍然面临着以下问题：

现有算法往往直接使用CV领域中的方法，未考虑算法适配

↓

×算法适配效果差

医疗行为数据采集难度较大，数据集的数量和体量不足

↓

×细粒度数据集匮乏

数据集中对医疗行为划分粒度不够细，制约算法研究

↓

×行为划分粒度粗

算法通常只有输入输出，几乎没有人机交互方式

↓

×人机交互性差

目 录

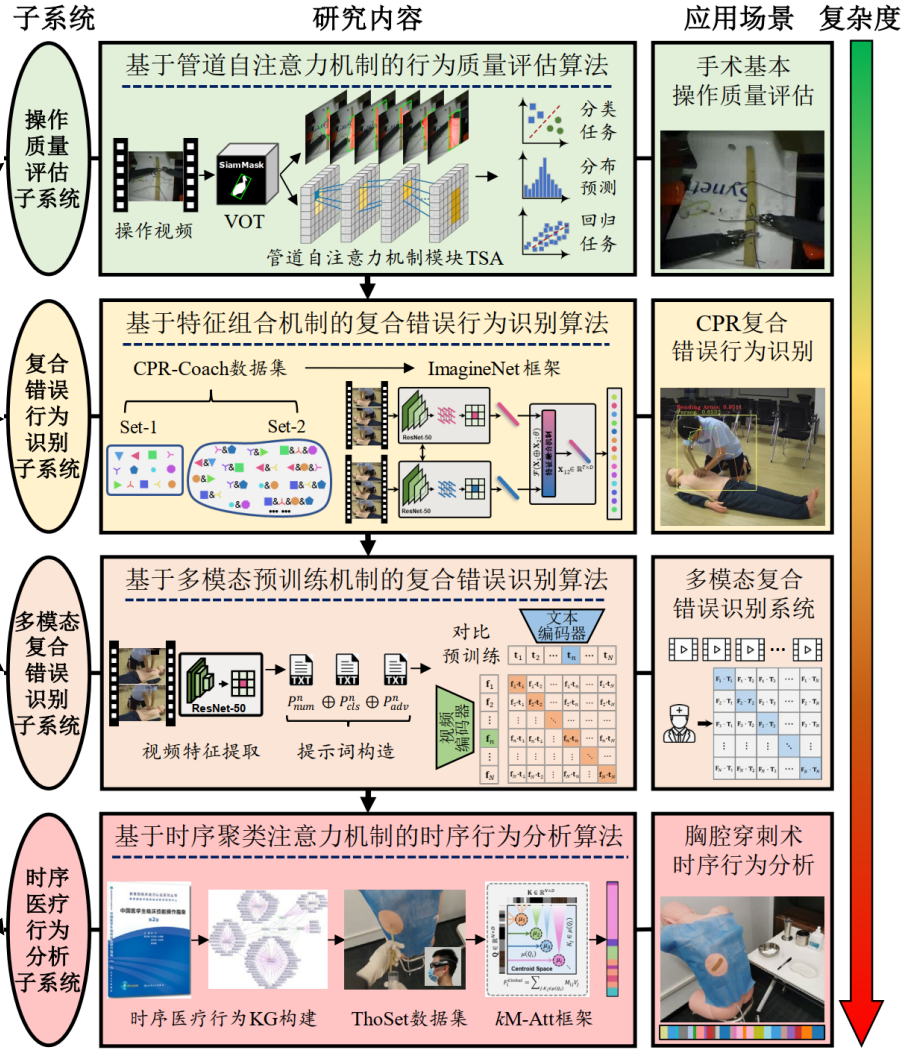
2.1 本文研究内容与技术路线

细粒度医疗行为识别与技能评估技术研究

研究难点	研究方法	医疗技能评估系统
如何在行为质量评估模型的结构设计中实现对医疗场景的适配	基于管道自注意力机制的行为质量评估算法	操作质量评估子系统
如何构建细粒度错误行为识别数据集并设计相关识别算法	基于特征组合机制的复合错误行为识别算法	复合错误行为识别子系统
如何丰富医疗技能评估系统的交互形式，使其支持语言模态的查询与检索功能	基于多模态预训练机制的复合错误行为识别算法	多模态复合错误识别子系统
如何在时序维度上实现复杂医疗行为的细粒度分割与合规性评估	基于时序聚类注意力机制的扩散时序行为分析算法	时序医疗行为分析子系统

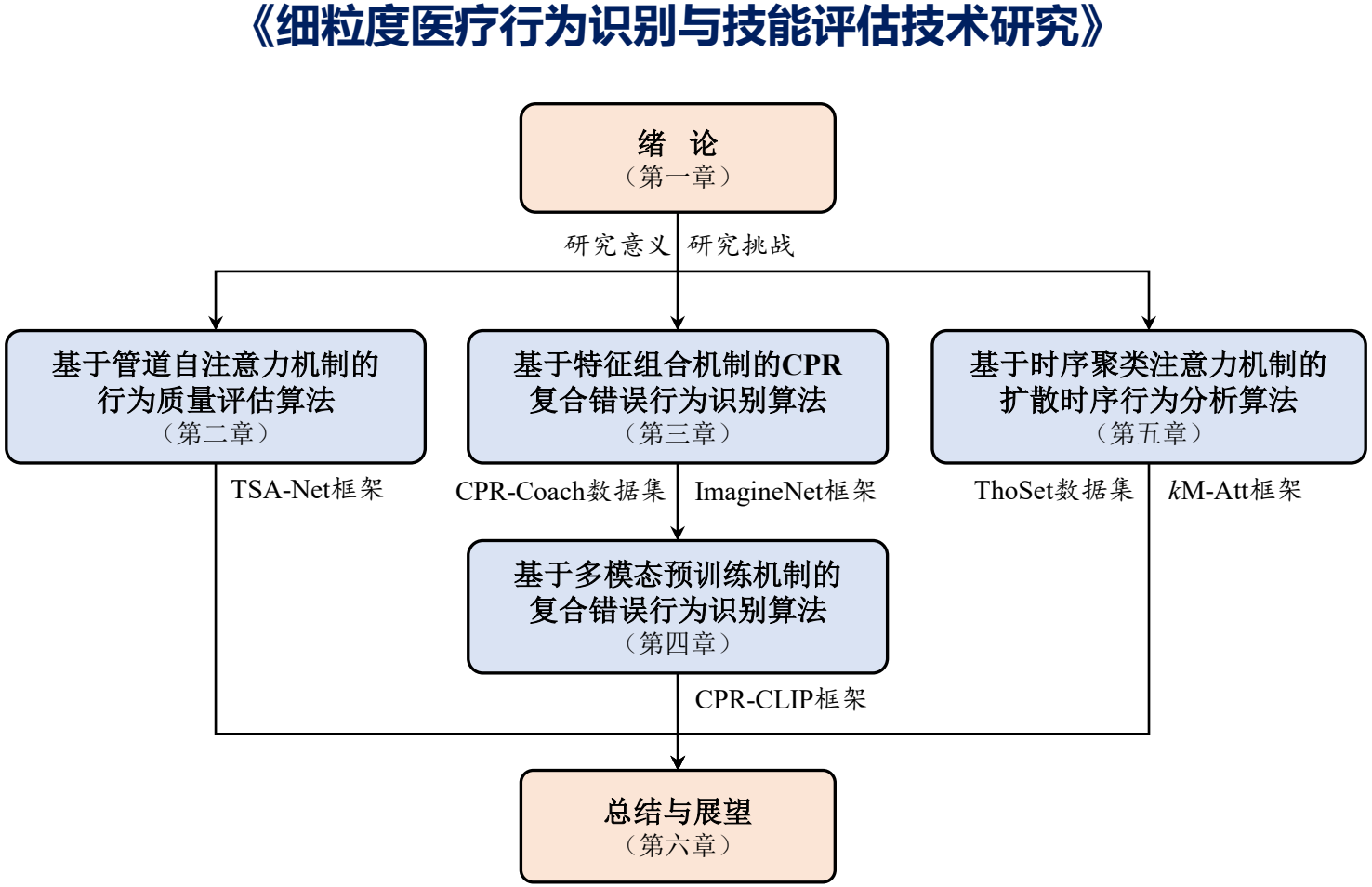
研究目标

针对医疗技能评估领域所面临的**细粒度行为识别数据集匮乏**、**行为划分粒度粗**、**算法性能受限**和**人机交互性差**问题，本文从**数据集构建**、**任务范式创新**、**算法创新**和**集成应用**方面开展探索，拟构建出一套支持多任务、多场景的医疗技能评估系统，从而填补医疗技能评估研究中的部分空白，为后续医疗技能评估系统的落地应用奠定基础。



2.2 全文组织结构

本文共由六个章节构成，章节之间的关联情况如下图所示。其中第三章为第四章的前置章节。

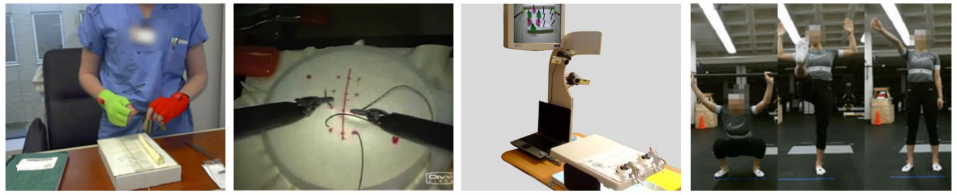


目 录

3.1 行为质量评估任务-研究现状

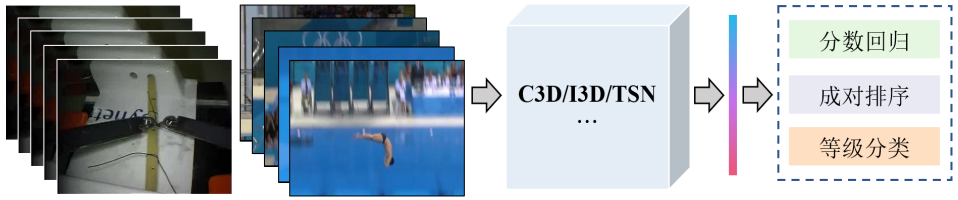
行为质量评估任务 / AQA的目标为：依据摄像头和运动学传感器记录的信息**对操作者的行为进行质量评估**。AQA技术在体育、医疗和工厂技能培训等场景中有着广阔的应用前景。

医疗AQA数据集案例展示



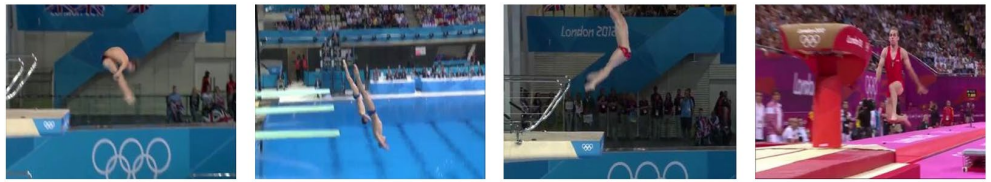
模拟外科手术操作　da Vinci 手术机器人操作　腹腔镜手术模拟平台　医疗康复训练技能评估

现有AQA模型通用架构



输入视频　视频特征提取器　神经网络头

体育AQA数据集案例展示



跳水 Diving　双人3m跳板 Sync. Diving 3m　双人10m跳台 Sync. Diving 10m　体操 Gym Vault

雪板大跳 Snowboard Big Air　滑雪大跳 Ski. Big Air　蹦床 Trampoline　花样滑冰 Figure Skating

大部分现有AQA模型在特征提取环节中直接借用了行为识别模型中的视频主干网络，因此会引起以下问题：

- **任务差异性的欠考虑**：行为识别模型需要区分不同行为之间的差异，而AQA模型需要进行优劣性判别。**AQA任务具有更高的视频表征能力需求**；
- **时空上下文信息建模能力弱**：现有模型所使用的视频主干网络通常由卷积层构成，其感受野的尺度严重依赖于卷积核的大小，因此**模型无法捕获到远距离的特征关联信息**。

3.2 TSA-Net网络框架

针对以上问题，本文提出了**TSA-Net行为质量评估框架**，其结构图如下图所示。TSA-Net可划分为五个阶段：

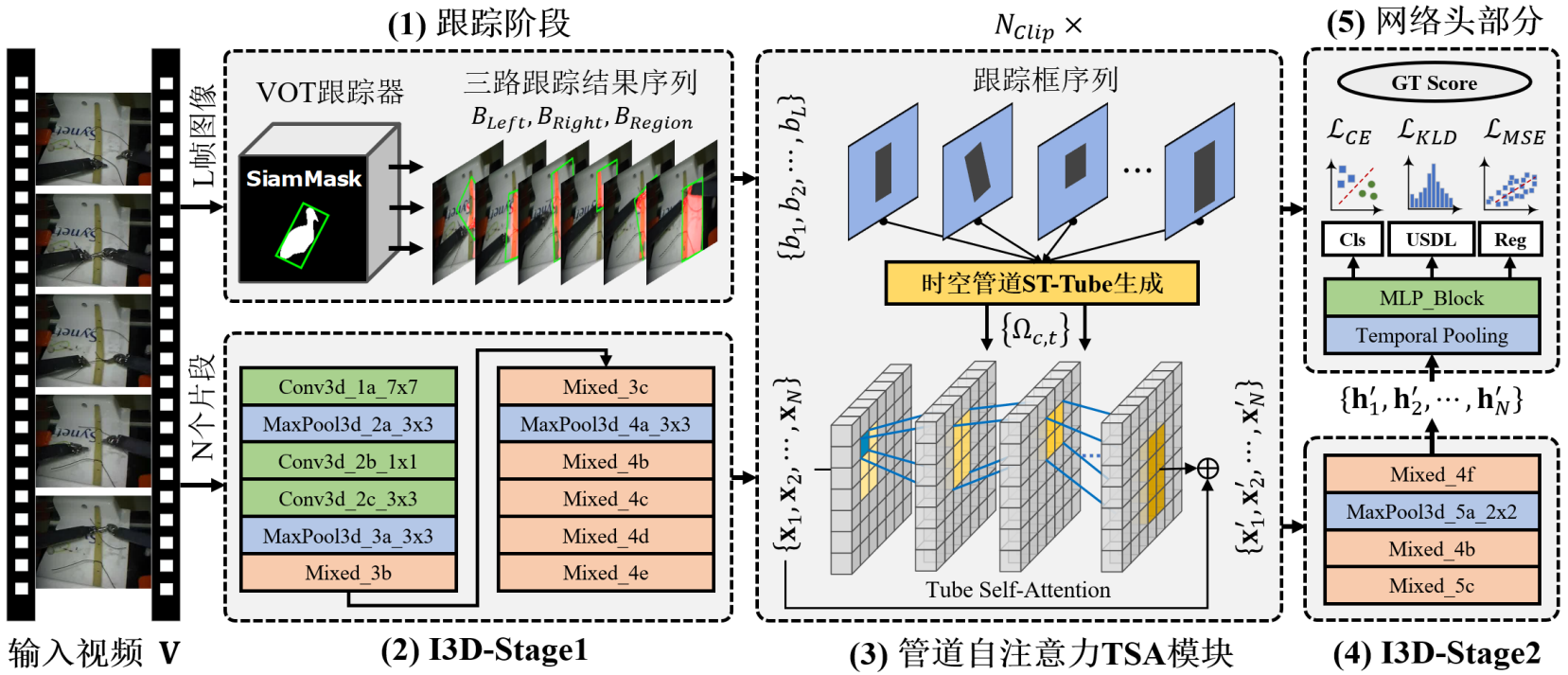
（1）**跟踪阶段**：使用SiamMask单目标跟踪器对视频中的目标物体进行跟踪，生成跟踪框集合 $B = \{b_l\}_{l=1}^{L}$

（2）**特征提取阶段**：将视频送入I3D网络的第一阶段，获取视频特征 $\mathbf{X} = \{\mathbf{x}_n\}_{n=1}^{N}, \mathbf{x}_n \in \mathbb{R}^{T \times H \times W \times C}$

（3）**特征增强阶段**：通过时序管道自注意力模块TSA对视频特征进行增强，获得特征 $\mathbf{X}' = \{\mathbf{x}'_n\}_{n=1}^{N}$

（4）**特征提取与融合**：增强后的特征送入I3D网络的第二阶段，获得特征 $\mathbf{H} = \{\mathbf{h}_n\}_{n=1}^{N}$

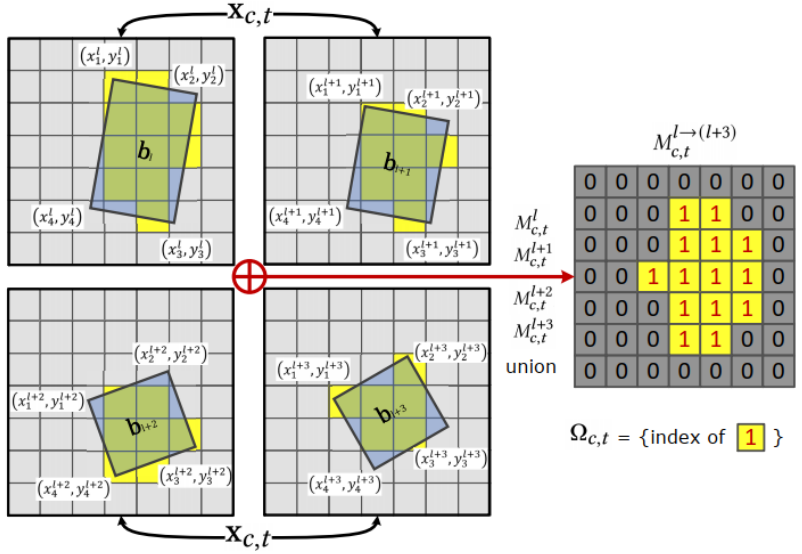
（5）**分值预测阶段**：通过时序平均池化操作融合多个片段 $\bar{\mathbf{h}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{h}_n$, $\bar{\mathbf{h}} \in \mathbb{R}^{T \times H \times W \times C}$ ，并最终通过多层感知机生成结果。



3.3 管道自注意力机制与TSA模块

管道自注意力机制依据SiamMask跟踪器生成的跟踪框序列对**特征图中的部分元素**进行选择性增强，具体步骤为：

步骤1：时空管道生成阶段
Spatio-temporal Tube Generation



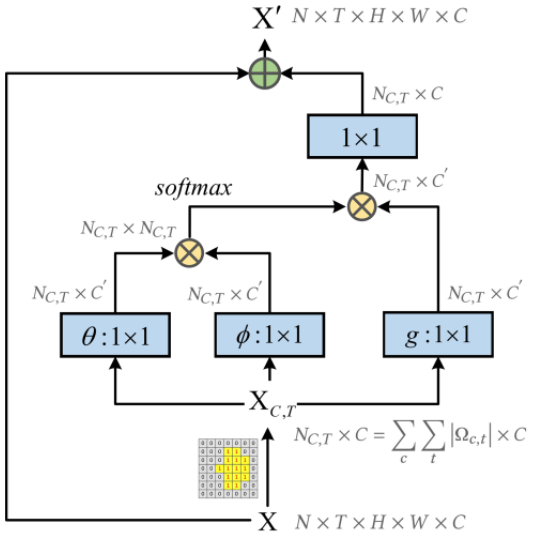
掩码生成 $$M_{c,t}^{l}(i,j)=\begin{cases}1, S(b_l,(i,j)) \geqslant \tau \\ 0, S(b_l,(i,j)) < \tau\end{cases}$$

元素级别OR操作 $$M_{c,t}^{l\to(l+3)} = Union(M_{c,t}^{l}, M_{c,t}^{l+1}, M_{c,t}^{l+2}, M_{c,t}^{l+3})$$

索引集合 $$\Omega_{c,t} = \left\{(i,j) | M_{c,t}^{l\to(l+3)}(i,j) = 1\right\}$$

- ✓ 高效性
- ✓ 有效性
- ✓ 可拓展性

步骤2：管道自注意力计算
Tube Self-attention Operation



TSA 机制

$$\mathbf{y}_p = \frac{1}{C(\mathbf{x})}\sum_{\forall c}\sum_{\forall t}\sum_{\forall(i,j)\in\Omega_{c,t}} f\left(\mathbf{x}_p, \mathbf{x}_{c,t}(i,j)\right) g(\mathbf{x}_{c,t}(i,j))$$

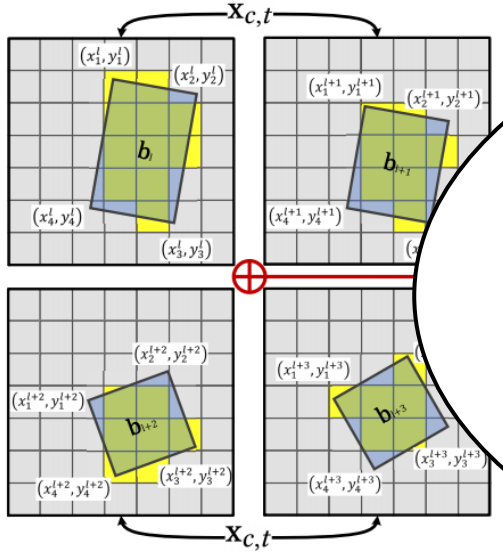
归一化参数 $$C(\mathbf{x}) = \sum_c\sum_t\left|\Omega_{c,t}\right|$$

相似度函数 $$f\left(\mathbf{x}_p, \mathbf{x}_{c,t}(i,j)\right) = \theta(\mathbf{x}_p)^T\phi(\mathbf{x}_{c,t}(i,j))$$

3.3 管道自注意力机制与TSA模块

管道自注意力机制依据SiamMask跟踪器生成的跟踪框序列对**特征图中的部分元素**进行选择性增强，具体步骤为：

步骤1：时空管道生成阶段 Spatio-temporal Tube Generation



掩码生成 $M_{c,t}^{l}(i,j)=\begin{cases}1, S(b_l,(i,j)) \geqslant \tau \\ 0, S(b_l,(i,j)) < \tau\end{cases}$

元素级别OR操作 $M_{c,t}^{l\rightarrow(l+3)}=Union(M_{c,t}^{l},M_{c,t}^{l+1},M_{c,t}^{l+2},M_{c,t}^{l+3})$

索引集合 $\Omega_{c,t}=\left\{(i,j)|M_{c,t}^{l\rightarrow(l+3)}(i,j)=1\right\}$

相较于Non-local模块, TSA模块能够显著地
将计算复杂度从

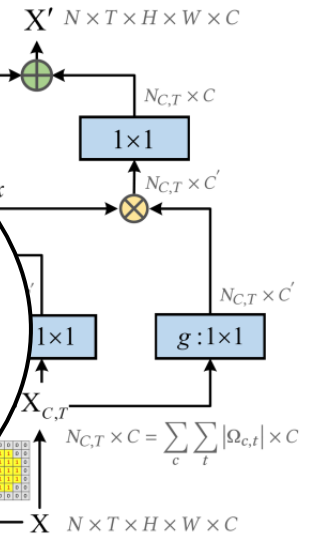
$$O\left((N\times T\times H\times W)\times(N\times T\times H\times W)\right)$$

降低到

$$O\left(\left(\sum_c\sum_t|\Omega_{c,t}|\right)\times\left(\sum_c\sum_t|\Omega_{c,t}|\right)\right)$$

- ✓ 高效性
- ✓ 有效性
- ✓ 可拓展性

步骤2：管道自注意力计算 Tube Self-attention Operation



TSA 机制

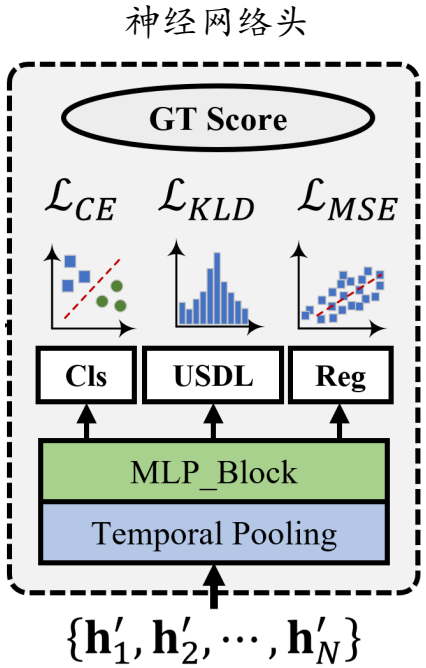
$$\mathbf{y}_p=\frac{1}{C(\mathbf{x})}\sum_{\forall c}\sum_{\forall t}\sum_{\forall(i,j)\in\Omega_{c,t}} f\left(\mathbf{x}_p,\mathbf{x}_{c,t}(i,j)\right)g(\mathbf{x}_{c,t}(i,j))$$

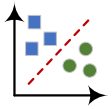
归一化参数 $C(\mathbf{x})=\sum_c\sum_t\left|\Omega_{c,t}\right|$

相似度函数 $f\left(\mathbf{x}_p,\mathbf{x}_{c,t}(i,j)\right)=\theta(\mathbf{x}_p)^T\phi(\mathbf{x}_{c,t}(i,j))$

3.4 网络头设计与损失函数

TSA-Net框架能够支持多种行为质量评估任务，包括：**分类任务**、**回归任务**和**分布预测任务**。

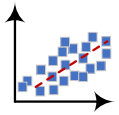




➢ 分类任务

预测概率：$S \in \mathbb{R}^M$　　真实标签：$Y \in \{0,1\}^M$

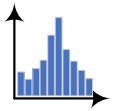
交叉熵损失函数：$\mathcal{L}_{CE} = -\sum_{i=1}^{M} Y_i \cdot \log S_i$



➢ 回归任务

预测分数：S_n　　标准分数：Y_n

均方误差损失函数：$\mathcal{L}_{MSE} = \frac{1}{N}\sum_{n=1}^{N}(Y_n - S_n)^2$



➢ 分布预测任务

预测分数分布：S_{pre}　　标准分数分布：P_c

分布间KL散度：$KL[P_c \,\|\, S_{pre}] = \sum_{i=1}^{M} P(c_i)\log\frac{P(c_i)}{S_{pre}(c_i)}$

难度系数转换：$S = S_{DD} \cdot S_{pre}$

3.5 实验结果-性能对比

本文在3个开源数据集上对TSA-Net的性能开展了探究：

- 医疗技能评估数据集：JIGSAWS
- 体育技能评估数据集：AQA-7、MTL-AQA

对比方法共分为三类：

- 当前最优方法：USDL、JRG、C3D-SVR等模型
- Non-Local特征增强：基于此模块所搭建的NL-Net模型
- TSA模块堆叠：不同堆叠数量下的TSA-Net模型性能

性能对比结果显示：

- 相较于SOTA方法**TSA-Net能够达到更优异的性能**；
- 过多的TSA模块堆叠会引起过拟合，存在**边际效益递减现象**。

表 2-1　JIGSAWS 数据集中 TSA-Net 性能对比结果

模型	Suturing	Needle Pass.	Knot Tying	Avg. Corr.
ST-GCN[101]	0.31	0.39	0.58	0.43
TSN[6]	0.34	0.23	0.72	0.46
JRG[166]	0.36	0.54	**0.75**	0.57
USDL[162]	0.64	0.63	0.61	0.63
NL-Net	0.65	0.64	0.67	0.65
TSA-Net	**0.68**	**0.65**	0.71	**0.67**

表 2-2　AQA-7 数据集中 TSA-Net 性能对比结果

模型	Diving	Gym	Skiing	Snowboard	Sync. 3m	Sync. 10m	Avg. Corr.
Pose+DCT[161]	0.5300	-	-	-	-	-	-
ST-GCN[101]	0.3286	0.5770	0.1681	0.1234	0.6600	0.6483	0.4433
C3D-LSTM[166]	0.6047	0.5636	0.4593	0.5029	0.7912	0.6927	0.6165
C3D-SVR[166]	0.7902	0.6824	0.5209	0.4006	0.5937	0.9120	0.6937
JRG[166]	0.7630	0.7358	0.6006	0.5405	0.9013	0.9254	0.7849
USDL[162]	0.8099	0.7570	0.6538	**0.7109**	0.9166	0.8878	0.8102
NL-Net	0.8296	0.7938	**0.6698**	0.6856	0.9459	0.9294	0.8418
TSA-Net	**0.8379**	**0.8004**	0.6657	0.6962	**0.9493**	**0.9334**	**0.8476**

表 2-3　MTL-AQA 数据集中 TSA-Net 性能对比结果

模型	Avg. Corr.
Pose+DCT[161]	0.2682
C3D-SVR[166]	0.7716
C3D-LSTM[166]	0.8489
C3D-AVG-STL[28]	0.8960
C3D-AVG-MTL[28]	0.9044
MUSDL[162]	0.9273
NL-Net	**0.9422**
TSA-Net	0.9393

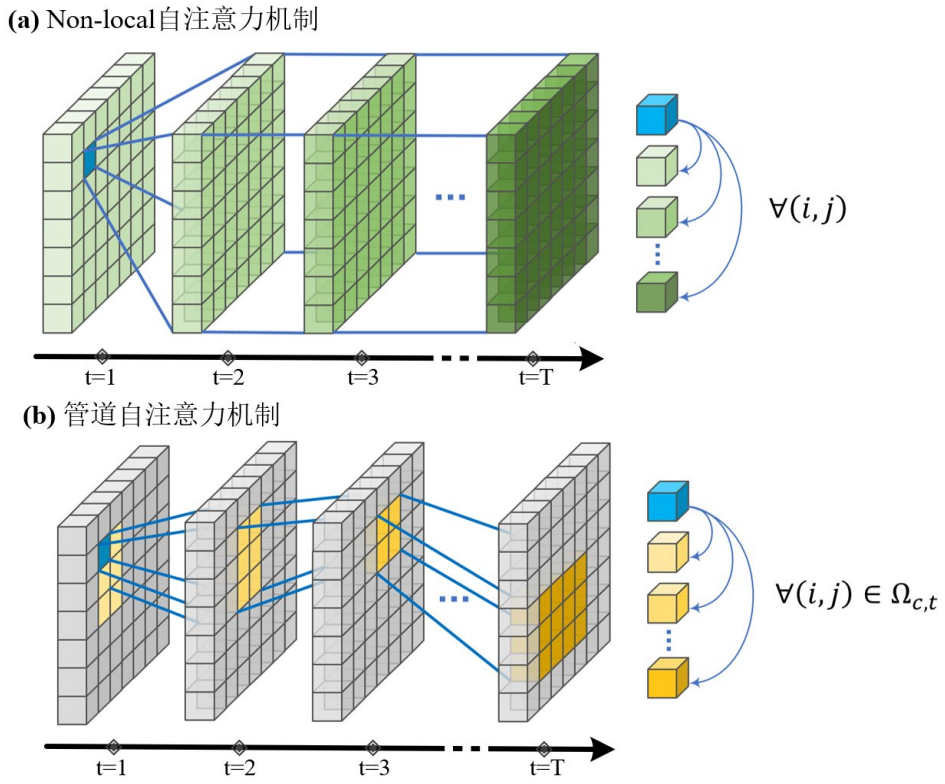
表 2-4　AQA-7 数据集中 TSA 模块堆叠实验结果

模型	Diving	Gym	Skiing	Snowboard	Sync. 3m	Sync. 10m	Avg. Corr.
TSA-Net	0.8379	0.8004	0.6657	0.6962	**0.9493**	0.9334	0.8476
TSAx2-Net	0.8380	0.7815	**0.6849**	**0.7254**	0.9483	**0.9423**	**0.8526**
TSAx3-Net	**0.8520**	**0.8014**	0.6437	0.6619	0.9331	0.9249	0.8352

3.5 实验结果-计算复杂度对比

计算复杂度与性能对比结果说明：

- 相较于Non-Local特征增强模块， TSA-Net能够**以更少的计算量达到相似甚至更优的性能**；
- 稀疏特征增强策略使管道自注意力TSA模块同时具备**高效性&有效性**。



Non-Local复杂度： $O((N \times T \times H \times W) \times (N \times T \times H \times W))$

TSA模块复杂度： $O\left(\left(\sum_c \sum_t |\Omega_{c,t}|\right) \times \left(\sum_c \sum_t |\Omega_{c,t}|\right)\right)$

表 2-5　MTL-AQA 数据集中 TSA 模块堆叠实验结果

模型	Sp. Corr.↑	MSE↓	FLOPs↓
NL-Net	**0.9422**	47.83	2.2 G
TSA-Net	0.9393	**37.90**	**1.012** G
TSAx2-Net	0.9412	46.51	2.025 G
TSAx3-Net	0.9403	47.77	3.037 G

表 2-6　TSA-Net 的计算复杂度与性能对比（AQA-7 数据集）

对比项目	NL-Net	TSA-Net	计算量节省	性能提升
Diving	2.2 GFLOPs	0.864 GFLOPs	-60.72%	↑0.0083
Gym	2.2 GFLOPs	0.849 GFLOPs	-61.43%	↑0.0066
Skiing	2.2 GFLOPs	0.283 GFLOPs	-87.13%	↓0.0041
Snowboard	2.2 GFLOPs	0.265 GFLOPs	-87.97%	↑0.0106
Sync. 3m	2.2 GFLOPs	0.952 GFLOPs	-56.74%	↑0.0034
Sync. 10m	2.2 GFLOPs	0.919 GFLOPs	-58.24%	↑0.0040
Average	2.2 GFLOPs	0.689 GFLOPs	-68.70%	↑0.0058

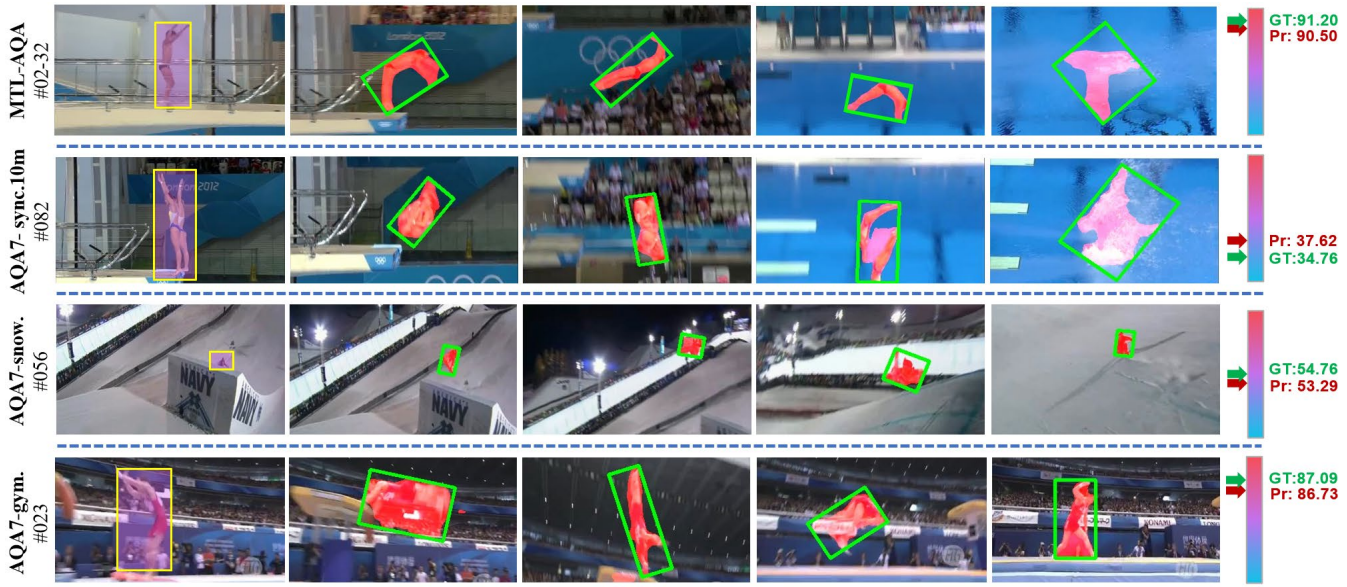
表 2-7　TSA-Net 的计算复杂度与性能对比（JIGSAWS 数据集）

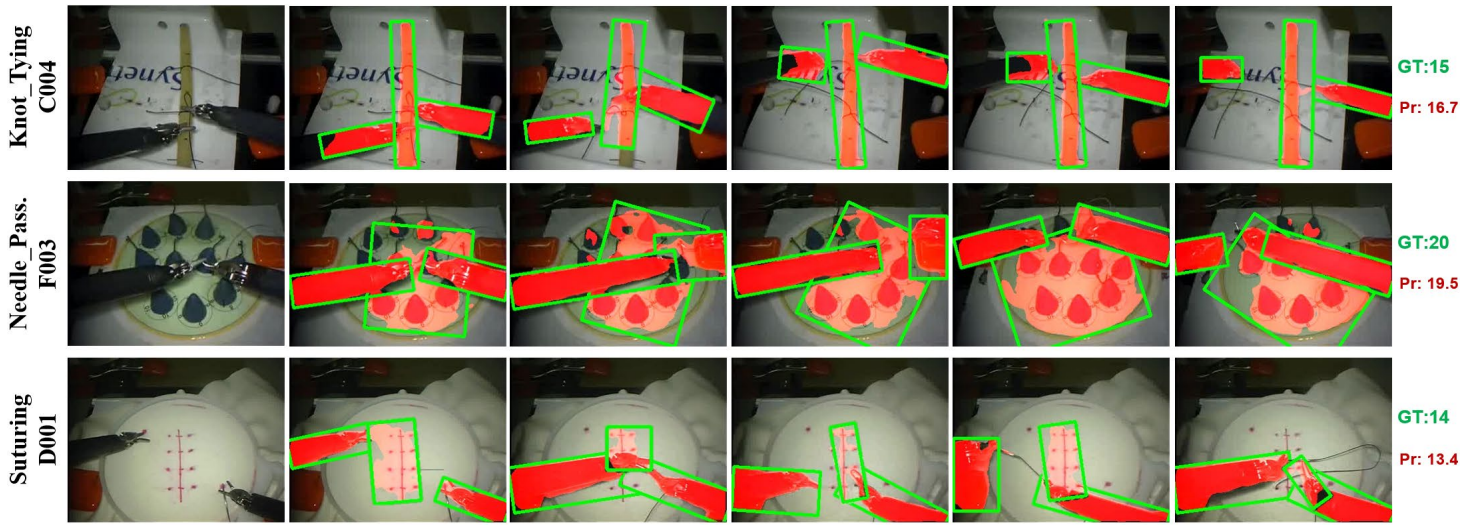
对比项目	NL-Net	TSA-Net	计算量节省	性能提升
Suturing	3.52 GFLOPs	1.87 GFLOPs	-46.89%	↑0.03
Needle Pass.	3.52 GFLOPs	1.77 GFLOPs	-49.78%	↑0.01
Knot Tying	3.52 GFLOPs	1.99 GFLOPs	-43.33%	↑0.04
Average	3.52 GFLOPs	1.86 GFLOPs	-46.67%	↑0.02

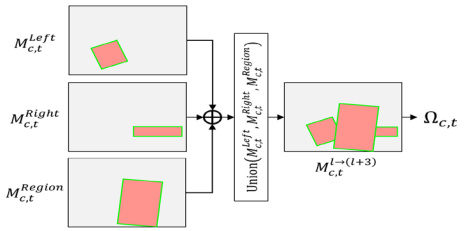
3.5 实验结果-可视化对比

MTL-AQA与AQA-7数据集案例展示：

- SiamMask跟踪器能够在各种运动中生成**稳定的跟踪结果**；
- TSA-Net能够在体育技能评估任务中**生成精准的评分结果**。







JIGSAWS数据集案例展示：

- 三路SiamMask跟踪信息**汇总策略能够获取到重要位置信息**；
- TSA-Net能够在医疗技能评估数据集的**各个子项目中生成精准的评分结果**。

目 录

4.1 行为识别技术-研究现状

人体行为识别技术（Human Action Recognition, HAR）旨在让计算机理解视频中人体的动作和行为。

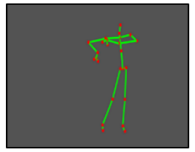
现有行为数据集使用的模态信息

视觉模态信息

RGB

弹钢琴

骨架信息



穿衣服

深度信息

握手

红外信息

打架

点云信息

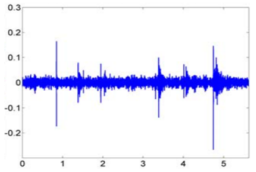
弯腰

事件流

跑步

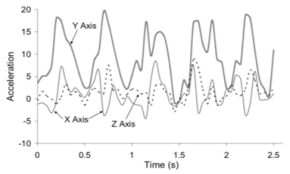
非视觉模态信息

音频



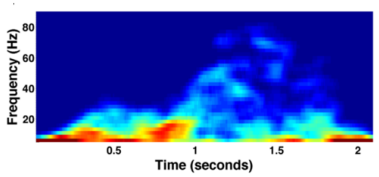
跳跃

加速度计



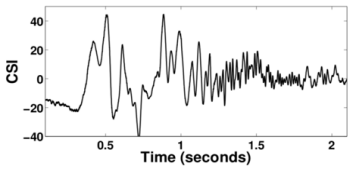
行走

雷达



跌倒

WiFi CSI信息



跌倒

现有行为识别算法的分类

- 基于双流网络Two-Stream模型：Two-Stream, TSN, SiameseNet
- 基于循环神经网络RNN模型：LRCNs, ARNet, DB-LSTM
- 基于3D卷积网络的模型：C3D, R(2+1)D X3D, 3DResNet
- 基于Transformer框架的模型：ASFormer, ViViT Video Swin Transformer

尽管HAR研究取得了一定进展，但仍然面临着一些**挑战**：

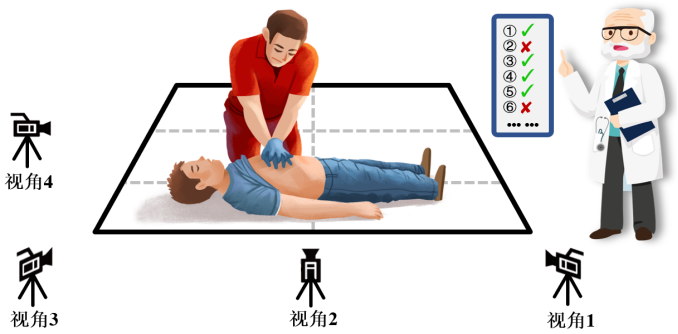
- **行为标签体系划分的细粒度较粗**
- **数据集复杂度低导致模型性能饱和**
- **算法停留在理论层面难以实际应用**

4.2 CPR-Coach数据集构建

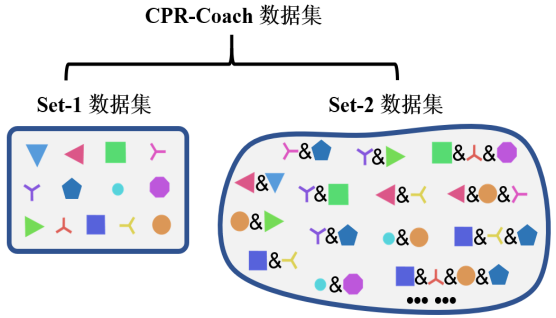
- **背景**：现有医疗技能评估数据集只对操作技能进行了评分与定级，**并未考虑到操作错误情况**。而在真实的医疗技能评估中，错误行为识别占据更重要的地位。

- **研究对象**：对心肺复苏术CPR中的胸外按压行为进行了探究。搭建了**多视角行为采集平台**，在医生的指导下总结了**13类错误行为和74类复合错误行为**，构建了**CPR-Coach数据集**，提出了**复合错误行为识别任务**。

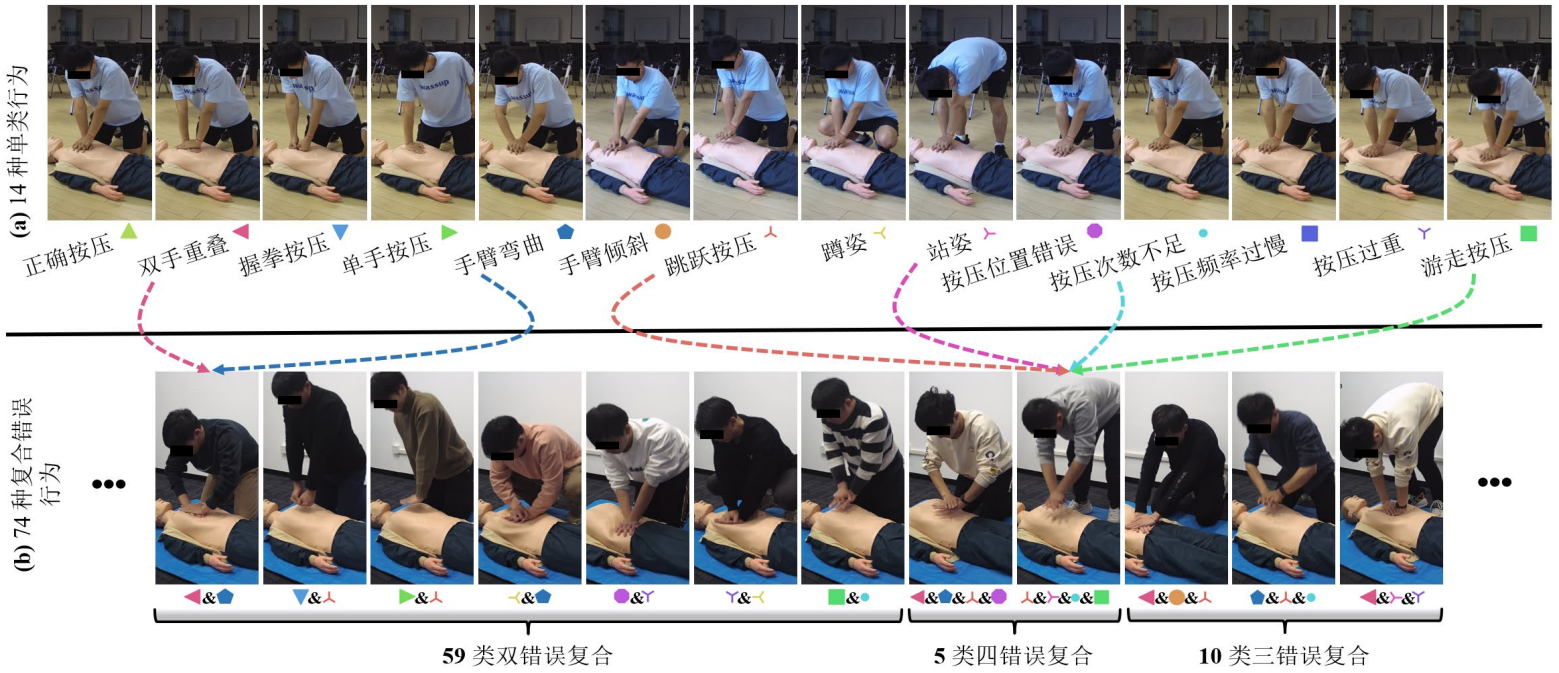
多视角采集平台



CPR-Coach数据集结构



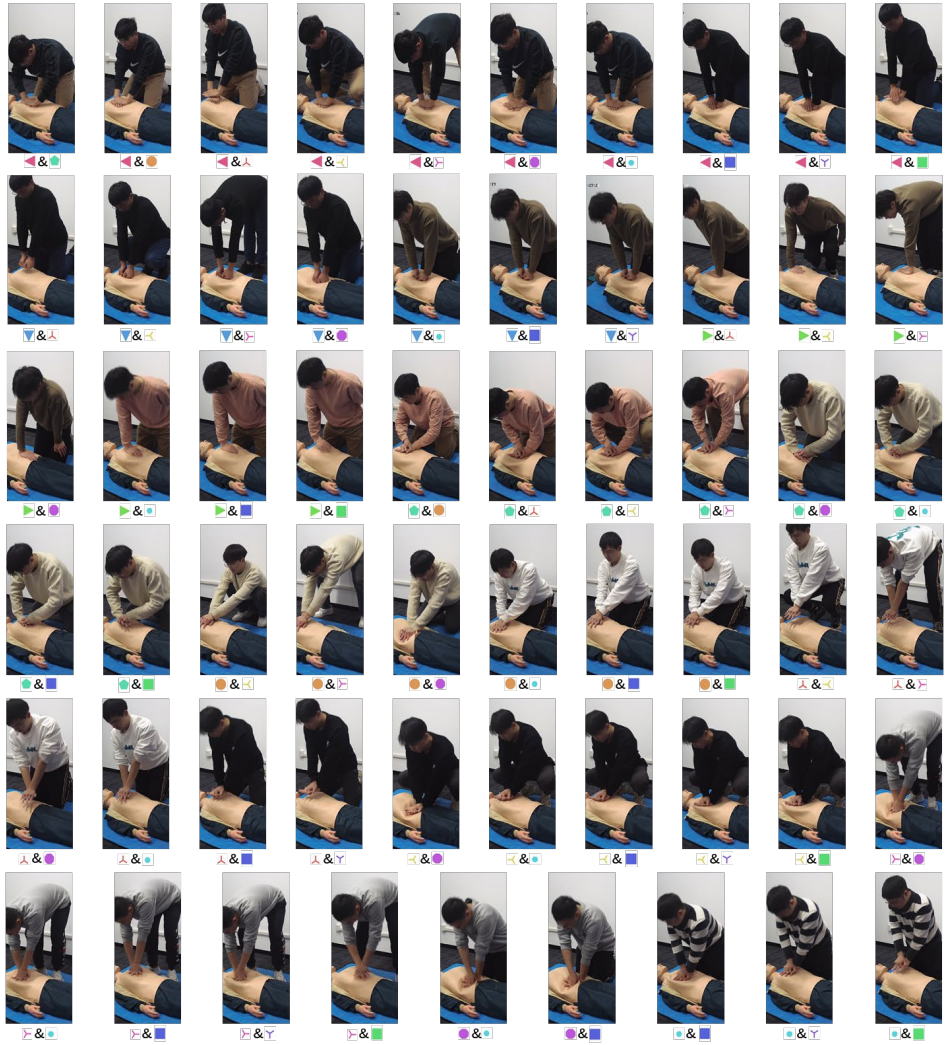
单类错误与复合错误案例展示



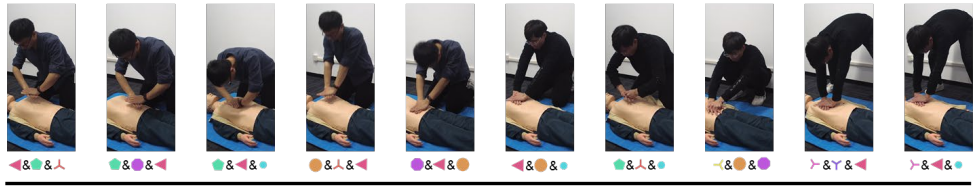
4.2 CPR-Coach数据集

➢ **CPR-Coach数据集构成**：**59类双错误复合** + **10类三错误复合** + **5类四错误复合**，错误组合遵循不矛盾的规则。

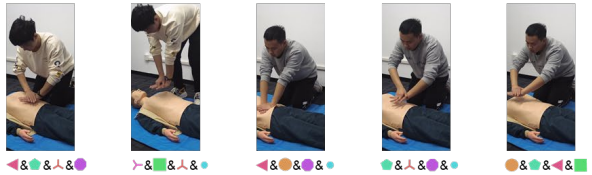
(a) 59类双错误复合行为

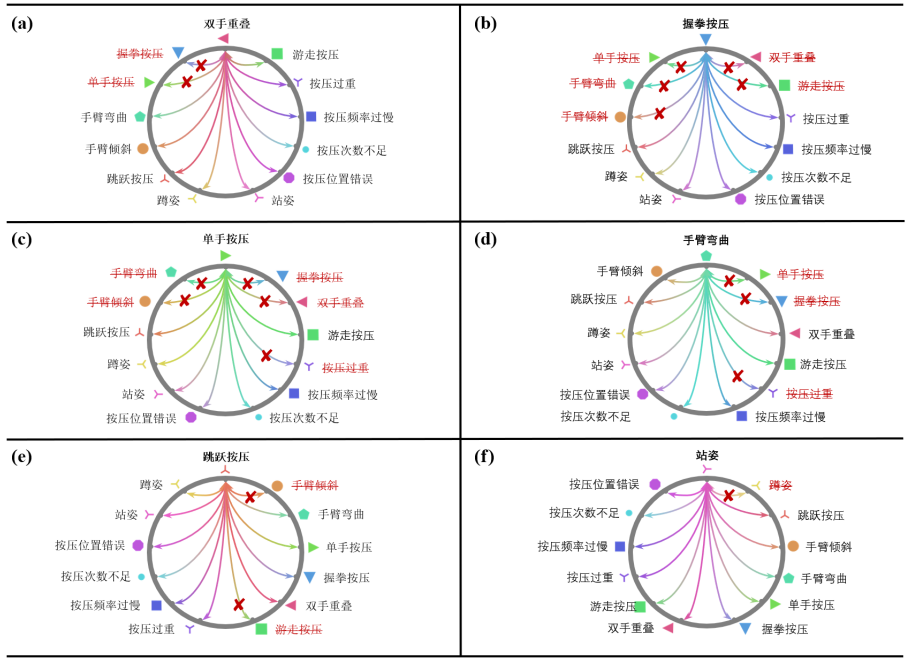


(b) 10类三错误复合行为



(c) 5类四错误复合行为

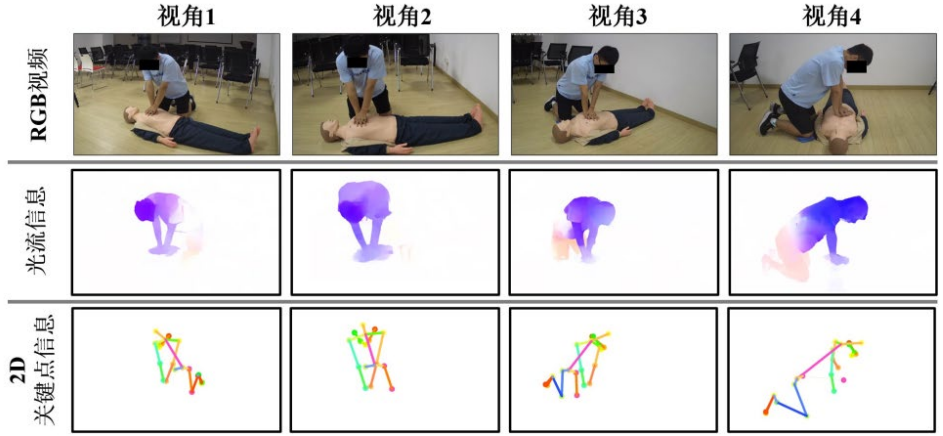




4.2 CPR-Coach数据集

➢ **CPR-Coach数据集**：招募12名被试者参与CPR-Coach数据集构建，共含有视频5664条，提供**RGB**、**光流**、**2D Pose**三种模态信息。

多模态信息展示



CPR-Coach数据集统计信息

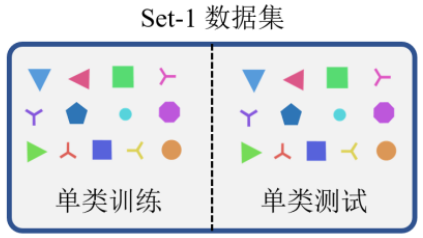
数据集统计项	数据
视角	4
帧率/FPS	25
分辨率	4096×2160 (4K)
参与人数	12
单类行为数量	1+13=14
复合错误行为数量	59+10+5=74
帧数（RGB）	2,217,756
帧数（RGB+光流）	6,644,596
视频数量	5,664
平均视频时长	19.52s
存储空间	450GB

CPR-Coach与其他医疗基准对比

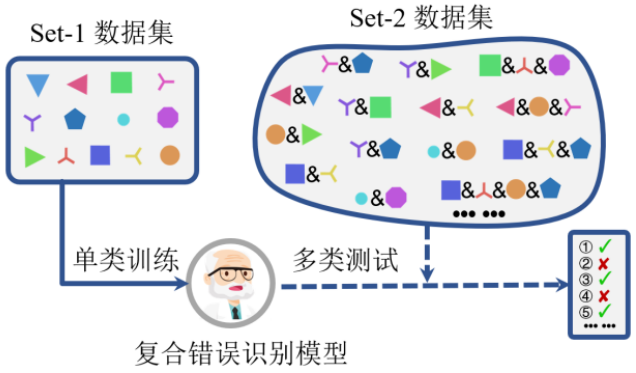
研究主题	数据集	#行为	数据模态	#视频	#视角	评估类型
腹腔镜手术技能评估	FLS-ASU[172]	1	RGB	28	2	技能排序
	Zhang *et al.*[128]	1	RGB	546	1	技能排序
	Chen *et al.*[127]	3	RGB	720	2	技能排序
基本手术技能评估	Sharma *et al.*[39]	2	RGB	33	1	OSATA 分数
	Bettadapura *et al.*[218]	3	RGB	64	2	技能排序
	Zia *et al.*[36]	2	RGB	104	1	技能排序
da Vinci 手术机器人系统操作评估	MISTIC-SL[171]	4	RGB+Kinematics	49	1	技能排序
	JIGSAWS[35]	3	RGB+Kinematics	103	1	技能排序
康复训练评估	UI-PRMD[130]	10	RGB+Kinematics	1,000	4	技能排序
外科手术流程识别	Cataract-101[41]	10	RGB	101	1	手术流程识别
	Hei-Chole[203]	7	RGB	33	1	手术流程识别
	HeiCo[135]	20	RGB	30	1	手术流程识别
	RARP45[140]	8	RGB	45	1	手术流程识别
	Cholec80[136]	7	RGB	80	1	手术流程识别
	GastricBypass[219]	10	RGB	337	1	手术流程识别
	Gastrectomy[220]	8	RGB	461	1	手术流程识别
	Nephrec9[42]	10	RGB	1,262	1	手术流程识别
	CATARACTS[138]	21	RGB	50	1	手术器具识别
	CholecT50[40]	10	RGB	50	1	三元组识别
	Laparo425[221]	9	RGB	425	1	手术阶段预测
	PETRAW[33]	6	RGB+Kinematics	90	1	手术流程识别
	DESK[141]	7	RGB+Kinematics	2,897	1	手术流程识别
心肺复苏	CPR-Coach	14+74	RGB+Flow+Pose	5,664	4	复合错误识别

➢ **复合错误行为识别任务**：**训练集只含有单类错误，而测试集包含多类复合错误**。模型需要在这种监督信息极度受限的情况下实现复合错误的精准识别。实际医疗技能评估场景中经常面临负面案例缺乏的问题，所以此问题对实际应用有重要意义。

(a) 任务**1**：单类错误行为识别任务



(b) 任务**2**：复合错误行为识别任务

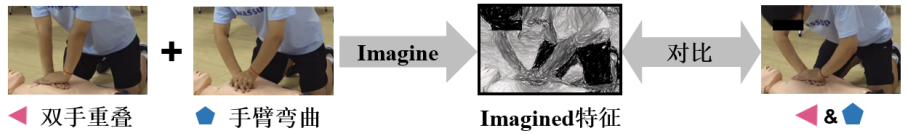


行为复合案例展示

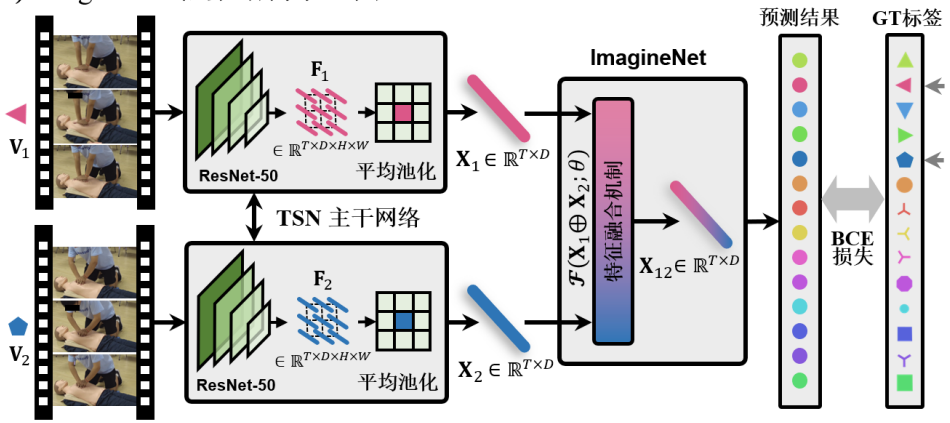
4.3 ImagineNet框架

- **启发**：人类可以**依据极少参考案例**进行复合错误案例判断，这是因为人具有极强的知识组合与推理能力。
- **ImagineNet框架**：使用**视觉特征组合训练机制**，充分地**利用单类错误样本进行特征组合训练**，从而最终有效地提升模型对**复合错误样本**的识别精度。

(a) ImagineNet 框架主体思路

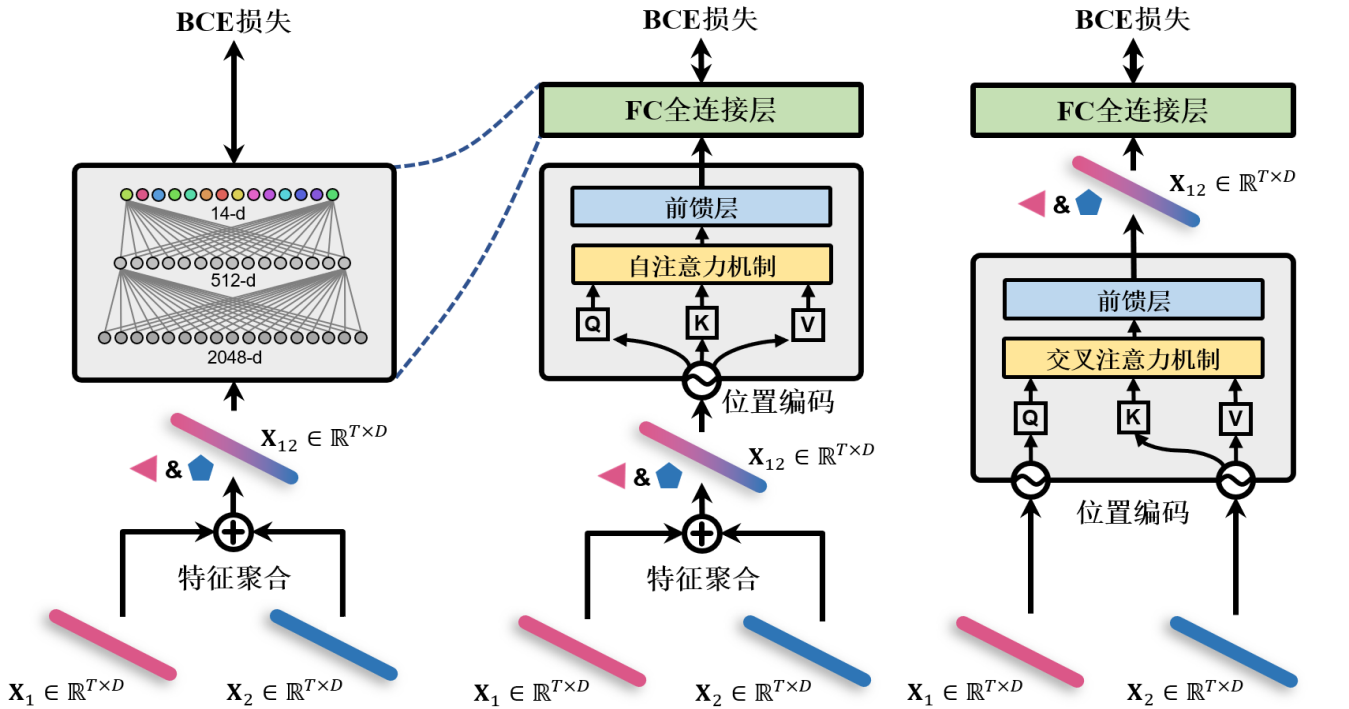


(b) ImagineNet 框架结构示意图



ImagineNet框架实例化的三种模型

(a) ImagineNet-FC融合网络 **(b)** ImagineNet-SA融合网络 **(c)** ImagineNet-CA融合网络



$$S_{FC}=\mathcal{F}_{FC}(\mathbf{X}_1\oplus\mathbf{X}_2,\boldsymbol{\theta}_{FC})$$

$$\boldsymbol{\theta}^{*}_{FC}=\arg\min_{\boldsymbol{\theta}_{FC}}\mathcal{L}_{BCE}(S_{FC},\ GT)$$

$$GT=\mathrm{Onthot}(C_1)\cup\mathrm{Onthot}(C_2)$$

$$S_{SA}=\mathcal{F}_{FC}(\mathcal{F}_{SA}(\mathbf{X}_1\oplus\mathbf{X}_2))$$

$$\mathbf{X}'_{SA}=LN\left[\mathbf{X}_{12}+softmax\left(\frac{\mathbf{X}_{12}\mathbf{X}_{12}^{T}}{\sqrt{D}}\right)\mathbf{X}_{12}\right]$$

$$\mathbf{X}_{FFN}=LN[\mathbf{X}'_{SA}+\mathcal{F}_{FFN}(\mathbf{X}'_{SA})]$$

$$S_{SA}=\mathcal{F}_{FC}(\mathbf{X}_{FFN})$$

$$S_{CA}=\mathcal{F}_{FC}(\mathcal{F}_{CA}(\mathbf{X}_1,\mathbf{X}_2))$$

$$\mathbf{X}'_{CA}=LN\left[\mathbf{X}_1+softmax\left(\frac{\mathbf{X}_1\mathbf{X}_2^{T}}{\sqrt{D}}\right)\mathbf{X}_2\right]$$

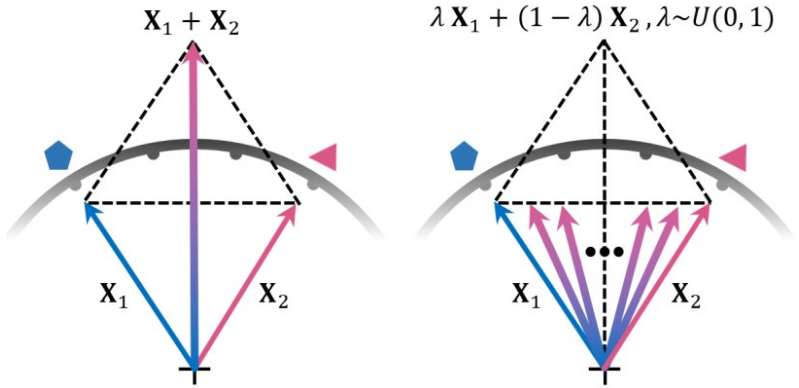
$$\mathbf{X}_{FFN}=LN[\mathbf{X}'_{CA}+\mathcal{F}_{FFN}(\mathbf{X}'_{CA})]$$

$$S_{SA}=\mathcal{F}_{FC}(\mathbf{X}_{FFN})$$

4.3 ImagineNet框架

- **随机线性组合的特征聚合策略**：本文将**随机线性加权机制**引入到复合错误行为识别任务中，并将融合机制拓展到更宽泛的多输入形式。
- **训练过程**：为在不影响模型单错误识别性能的前提下，提升复合错误识别性能，本文同时使用**四种采样配置**进行训练。

随机线性加权机制



双错误复合：$\mathbf{X}_{12}=\lambda\mathbf{X}_1+(1-\lambda)\mathbf{X}_2,\ \lambda\sim U(0,1)$

四错误复合：$\begin{cases}\mathbf{X}_{Fuse}=\lambda_1\mathbf{X}_1+\lambda_2\mathbf{X}_2+\lambda_3\mathbf{X}_3+\lambda_4\mathbf{X}_4\\ \lambda_1+\lambda_2+\lambda_3+\lambda_4=1\end{cases}$

训练案例构建

- 单错误输入案例 $C_{13}^1=13$ ➡ $s\sim\mathcal{S}$
- 双错误输入案例 $C_{13}^2=78$ ➡ $p\sim\mathcal{P}$
- 三错误输入案例 $C_{13}^3=286$ ➡ $t\sim\mathcal{T}$
- 四错误输入案例 $C_{13}^4=715$ ➡ $q\sim\mathcal{Q}$

损失函数：$\mathcal{L}_{Total}=\underset{s\sim\mathcal{S}}{\mathbb{E}}[\mathcal{L}_{BCE}^s]+\underset{p\sim\mathcal{P}}{\mathbb{E}}[\mathcal{L}_{BCE}^p]+\underset{t\sim\mathcal{T}}{\mathbb{E}}[\mathcal{L}_{BCE}^t]+\underset{q\sim\mathcal{Q}}{\mathbb{E}}[\mathcal{L}_{BCE}^q]$

4.4 实验结果-单类错误行为识别结果

- **单类错误行为识别结果**：现有行为识别模型**能够妥善处理CPR场景下的单错误行为识别任务**；
- **直接迁移策略实验结果**：三种损失函数均无法妥善处理**单类错误识别**与**复合错误识别**两个任务之间的巨大差异；

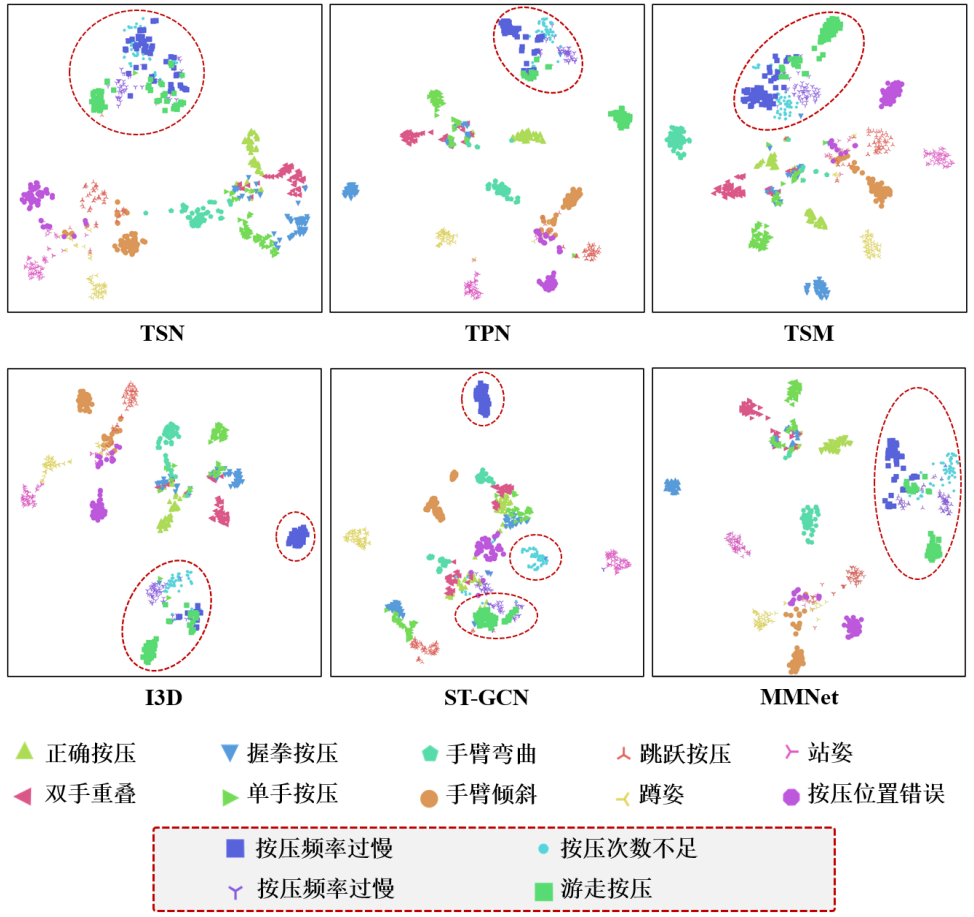
表3-3. 单类错误行为识别结果

模型	模态	主干	配置	轮次	CE Loss		BCE Loss		Multi-margin Loss	
					Top-1	Top-3	Top-1	Top-3	Top-1	Top-3
TSN[6]	RGB	ResNet-50	1x1x8	50	0.8879	0.9940	0.8829	0.9960	0.8502	0.9901
	RGB	ResNet-50*	1x1x8	50	0.9067	0.9921	0.8919	0.9940	0.8690	0.9901
	Flow	ResNet-50	1x1x8	50	0.7907	0.9603	0.8304	0.9851	0.7073	0.9355
TSM[116]	RGB	ResNet-50	1x1x8	50	0.9067	0.9901	0.9325	0.9950	0.8433	0.9881
I3D[21]	RGB	ResNet-50	32x2x1	50	0.9692	0.9960	0.9117	0.9940	0.8591	0.9861
TPN[224]	RGB	ResNet-50	8x8x1	50	**0.9802**	0.9960	0.9087	**0.9980**	0.8720	0.9901
C3D[76]	RGB	C3D*	16x1x1	50	0.9722	0.9931	0.9702	0.9931	0.8621	0.9802
TIN[225]	RGB	ResNet-50	1x1x8	50	0.8800	0.9901	0.7192	0.9335	0.8393	0.9861
SlowFast[202]	RGB	ResNet-50	4x16x1	256	0.8695	0.9734	0.8719	0.9781	0.8625	0.9688
TimeSFormer[85]	RGB	ViT	8x32x1	50	0.8879	0.9921	0.8998	0.9940	0.8462	0.9762
ST-GCN[101]	Pose	ST-GCN	1x1x300	50	0.9246	**0.9970**	0.9187	0.9881	0.9196	**0.9970**
PoseC3D[226]	Pose	ResNet3D-50	1x1x300	240	0.9208	0.9922	0.9035	0.9715	0.8837	0.9606
Two-Stream[4]	RGB+Flow	TSN+TSN_Flow	Late-Fusion	50	0.9533	0.9891	0.9479	0.9825	0.9296	0.9802
	RGB+Pose	TSN+ST-GCN	Late-Fusion	50	0.9782	0.9962	0.9608	0.9941	**0.9692**	0.9960
MMNet[227]	RGB+Pose+RoI	MS-G3D+Incep.-v3	Late-Fusion	80	0.9756	0.9960	**0.9772**	0.9940	0.9512	0.9876

表3-4. 朴素迁移方法的复合错误行为识别性能

模型	配置	模态	预训练	CE Loss		BCE Loss		Multi-Margin Loss	
				mAP	mmit mAP	mAP	mmit mAP	mAP	mmit mAP
TSN[6]	1x1x8	RGB	K-400	0.5598	0.6143	0.4627	0.5629	0.4838	0.5579
TSM[116]	1x1x8	RGB	✗	0.5662	0.6618	0.5721	0.6688	0.5470	0.6255
ST-GCN[101]	1x1x300	Pose	✗	0.5776	0.6692	**0.5868**	0.6865	0.5874	0.6719
PoseC3D[226]	1x1x300	Pose	✗	0.5498	0.6393	0.5556	0.6241	0.5358	0.6142
MMNet[227]	Late-Fusion	RGB+Pose+RoI	✗	**0.5948**	**0.6735**	0.5871	**0.6973**	**0.5894**	**0.6830**

图3-12. 单分类模型t-SNE特征可视化结果



4.4 实验结果-复合错误行为识别结果

- **ImagineNet复合错误识别性能**：本文所提出的**组合特征训练机制**能够有效提升视频主干网络的复合错误识别精度；
- **多模态信息识别结果**：实验结果对比显示ImagineNet-CA能够**有效融合多模态信息**，生成更精准的识别结果；
- **Set-2各子集实验结果**：**随着错误种类数量的上升，识别精度逐步下降**。说明符合错误识别的辨识难度会随着错误数量增长。

表3-5. 朴素迁移方法与 ImagineNet-FC 性能对比

模型	mAP	Δ	mmit mAP	Δ
TSN[6]	0.5598	—	0.6143	—
w/ ImagineNet-FC	**0.6259**	↑6.61%	**0.6893**	↑8.50%
TSM[116]	0.5662	—	0.6618	—
w/ ImagineNet-FC	**0.7053**	↑13.91%	**0.7566**	↑9.48%
ST-GCN[101]	0.5776	—	0.6692	—
w/ ImagineNet-FC	**0.6404**	↑6.28%	**0.7115**	↑4.23%
MMNet[227]	0.5948	—	0.6735	—
w/ ImagineNet-FC	**0.6927**	↑9.79%	**0.7478**	↑7.43%

图3-13. Set-2各种类复合错误行为识别结果

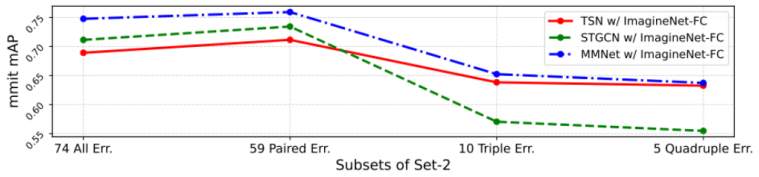


表3-13. 多模态模型性能对比

Model	Modality	Latency (ms)↓	mAP	mmit mAP
TSM[116]	RGB	---	0.5662	0.6618
ST-GCN[101]	Pose	---	0.5776	0.6692
Two-Stream[4]	RGB+Pose	**0.1501**	0.6003	0.6815
CBP[229]	RGB+Pose	0.3043	0.7089	0.7506
BLOCK[230]	RGB+Pose	1.294	0.7107	**0.7675**
MMNet[227]	RGB+Pose+RoI	0.2479	0.6927	0.7478
w/ ImagineNet-CA	RGB+Pose	0.1642	**0.7110**	0.7515

表3-6. SOTA 视频主干的单错误分类与复合错误分类性能

模型	配置	预训练	单类识别结果	
			Top-1	Top-3
Vi-ViT[84]	base-16x2	Kinetics-400	0.9814	1.0000
MViTv2[228]	base-32x3x1	Kinetics-400	0.9867	0.9980
Video Swin[86]	base-32x2x1	Kinetics-400	0.9918	1.0000
模型	**配置**	**预训练**	**直接迁移策略结果**	
			mAP	**mmit mAP**
Vi-ViT[84]	base-16x2	Kinetics-400	0.5582	0.6651
MViTv2[228]	base-32x3x1	Kinetics-400	0.5715	0.6740
Video Swin[86]	base-32x2x1	Kinetics-400	0.5696	0.6701

表3-7. SOTA 视频主干的朴素迁移方法与 ImagineNet-FC 对比

模型	mAP	Δ	mmit mAP	Δ
Vi-ViT[84]	0.5582	—	0.6651	—
w/ ImagineNet-FC	**0.6587**	↑10.05%	**0.7523**	↑8.72%
MViTv2[228]	0.5715	—	0.6740	—
w/ ImagineNet-FC	**0.6869**	↑11.54%	**0.7461**	↑7.21%
Video Swin[86]	0.5696	—	0.6701	—
w/ ImagineNet-FC	**0.7082**	↑13.86%	**0.7638**	↑9.37%

4.4 实验结果-视角对识别性能的探究

- **t-SNE特征的可视化对比**：ImagineNet-FC生成的特征相较于原始网络特征具有更强的分辨性，**特征组合训练机制能够有效地扩大类间间距，缩小类内间距**；
- **识别结果可视化**：ImagineNet框架能够在各种复杂的错误组合情况下，在各个视角中实现**精准的错误行为识别**。

图3.17 ImagineNet-FC 生成的t-SNE特征的可视化对比

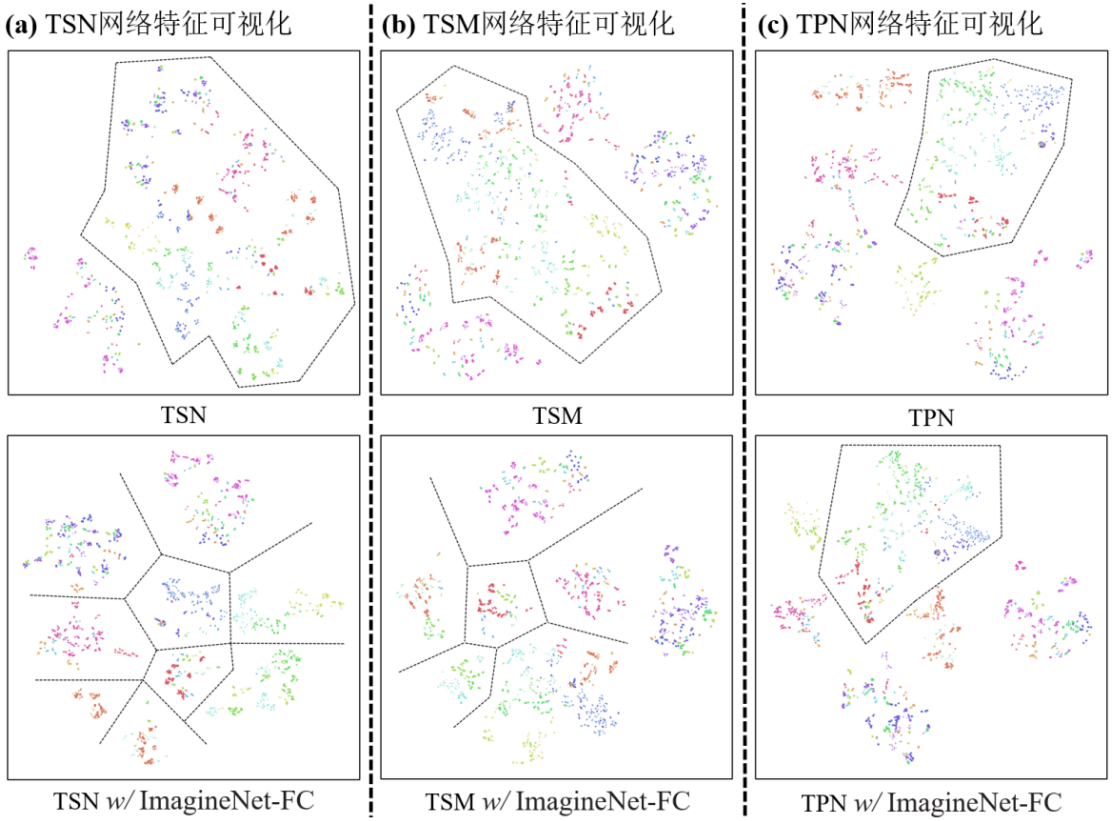
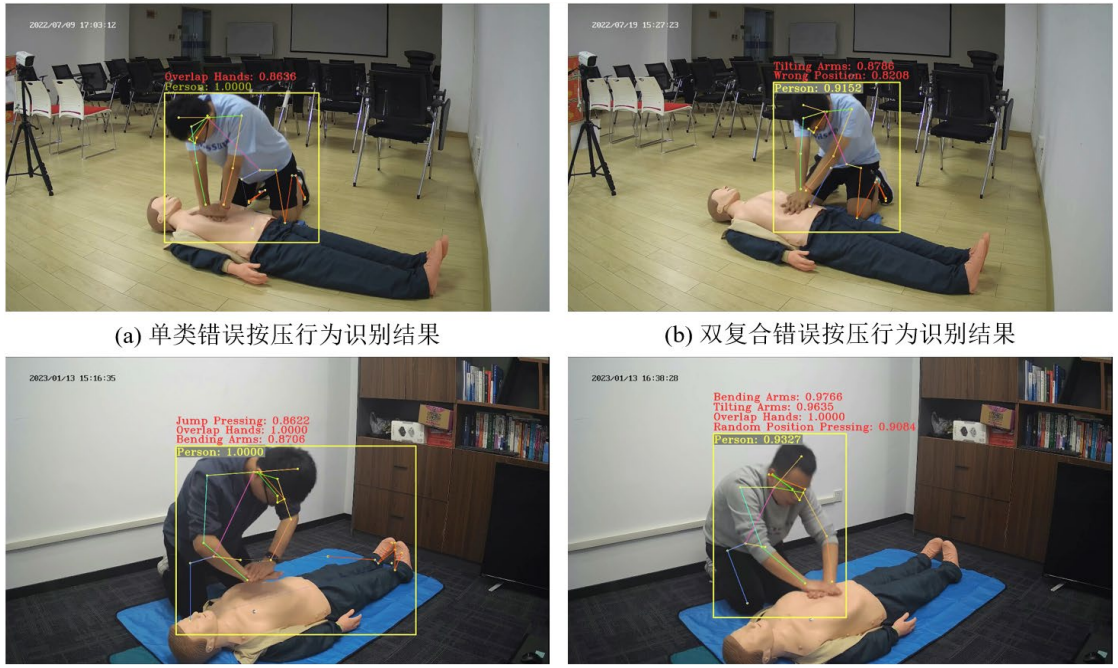


图3-16. 单错误与复合错误识别结果展示



(a) 单类错误按压行为识别结果

(b) 双复合错误按压行为识别结果

(c) 三错误复合识别结果

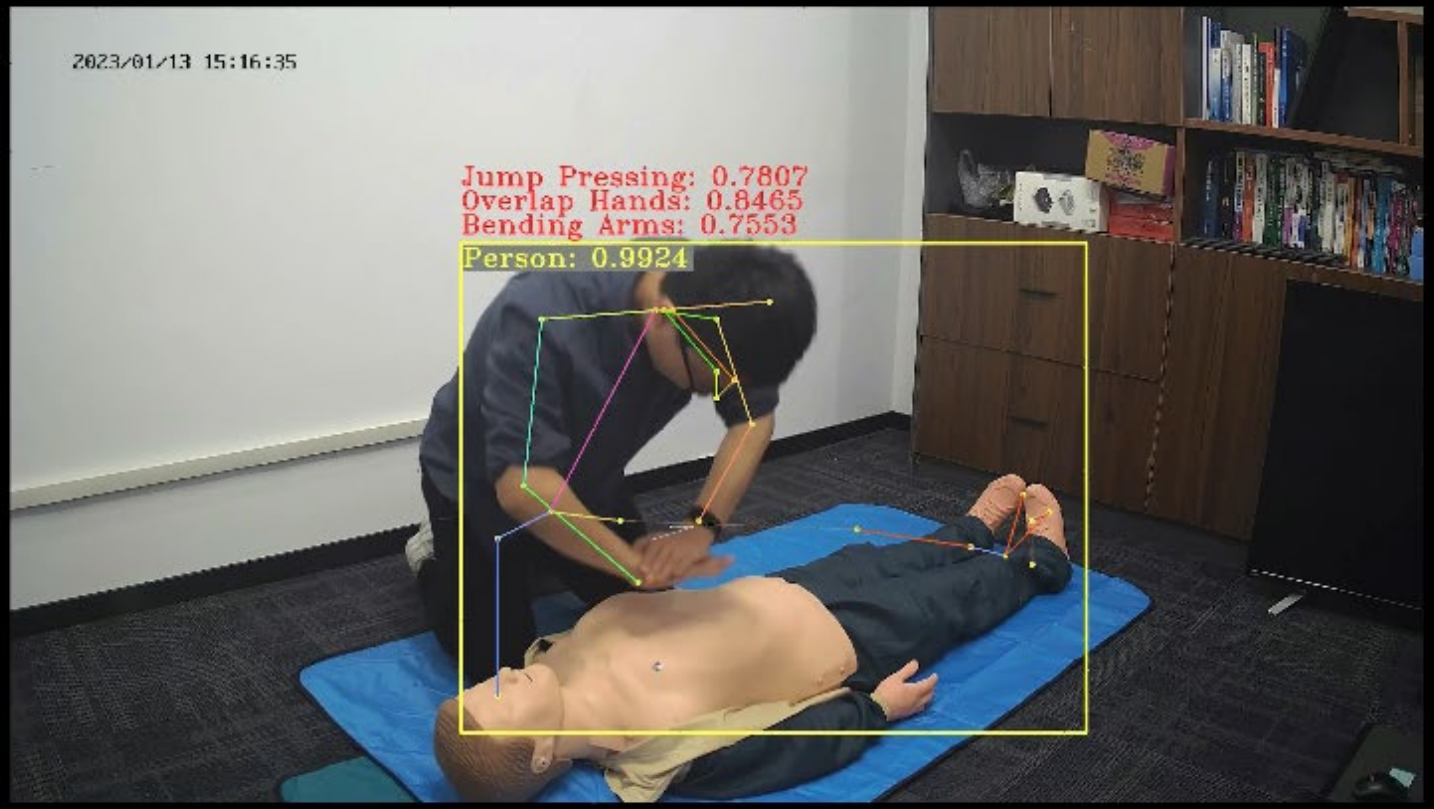
(d) 四错误复合识别结果

4.5 复合错误识别系统展示

Part 3: Visualization & System Demonstration

Triple-composite Error Actions Recognition Results

- Correct
- Overlap Hands
- Clenching Hands
- Single Hand
- Bending Arms
- Tilting Arms
- Jump Pressing
- Squatting
- Standing
- Wrong Position
- Insufficient Pressing
- Slow Frequency
- Excessive Pressing
- Random Position Pressing



目 录

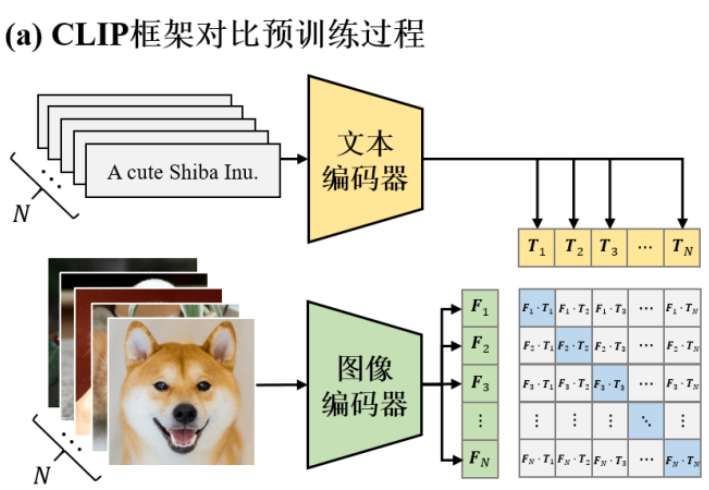
5.1 多模态学习-研究现状

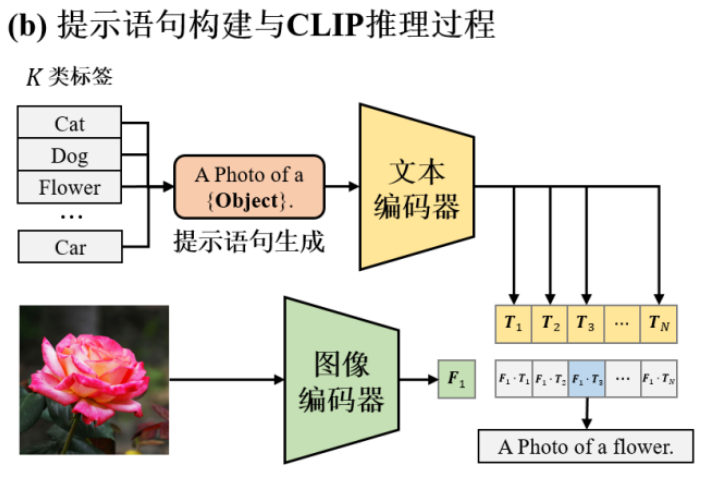
现有医疗技能评估模型输入输出通常采用“**输入视频—输出结果**”的简单映射形式，存在以下问题：

- **人机交互性差**：硬性输入输出形式的模型通常难以直接应用在真实的医疗技能评估场景中；
- **技能评估模态单一**：尽管CPR-Coach数据集提供了三种视觉模态信息，但并没有包含语言模态的信息。

- **探究目标**：通过语言模态信息与多模态预训练框架的引入，在**提升模型识别性能**的同时**使模型具备初步的人机交互能力。**

- **多模态学习**：目前多模态学习方法已被广泛应用于**语言—图像生成**、**视觉问题回答**、**多模态感知融合**等各类学习任务；
- **CLIP框架**：由OpenAI公司于2021年提出的一种**多模态对比预训练框架**，在零样本图像识别中发挥了强大的能力。





训练

$$\mathcal{L}_{img} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{\exp\left(\mathrm{sim}\left(\mathbf{F}_i,\mathbf{T}_j\right)/\tau\right)}{\sum_{j=1}^{N}\exp\left(\mathrm{sim}\left(\mathbf{F}_i,\mathbf{T}_j\right)/\tau\right)}$$

$$\mathcal{L}_{txt} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{\exp\left(\mathrm{sim}\left(\mathbf{T}_i,\mathbf{F}_j\right)/\tau\right)}{\sum_{j=1}^{N}\exp\left(\mathrm{sim}\left(\mathbf{T}_i,\mathbf{F}_j\right)/\tau\right)}$$

$$\mathcal{L} = \frac{1}{2}\left(\mathcal{L}_{img} + \mathcal{L}_{txt}\right)$$

推理

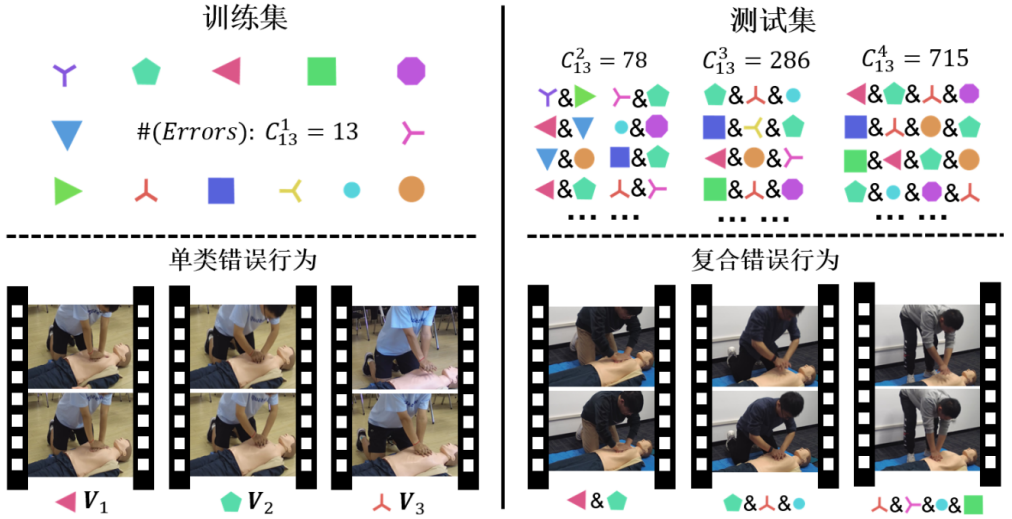
$$P(Y=i|\mathbf{f}) = \frac{\exp\left(sim\left(\mathbf{f},\mathbf{T}_i\right)/\tau\right)}{\sum_{j=1}^{K}\exp\left(sim\left(\mathbf{f},\mathbf{T}_j\right)/\tau\right)}$$

$$Y = \arg\max_i P(Y=i|\mathbf{f})$$

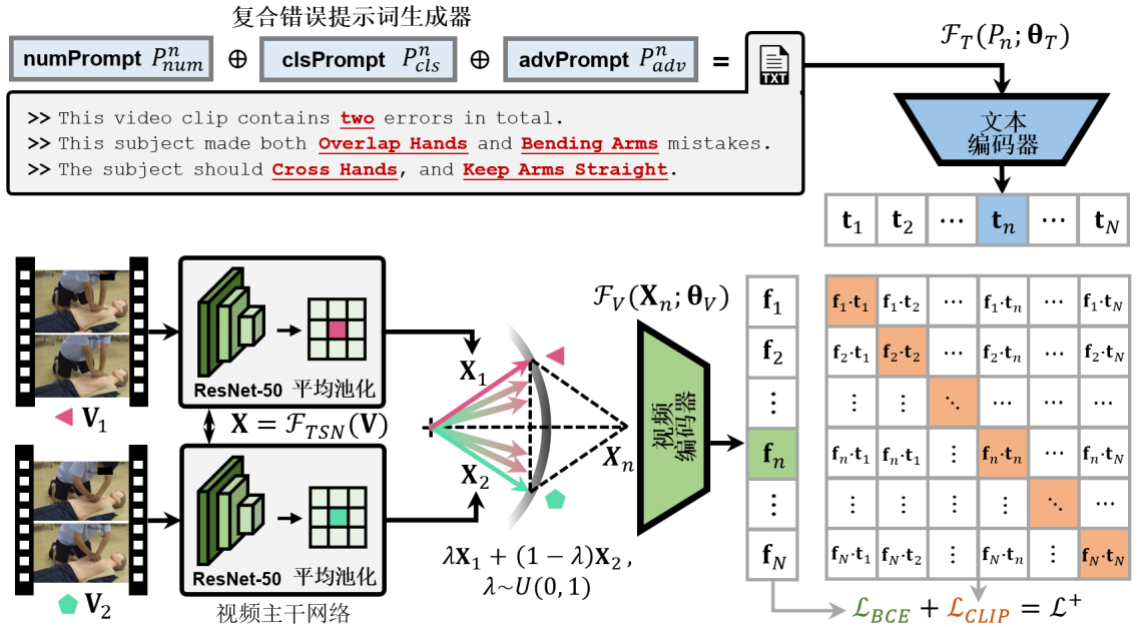
5.2 基于多模态预训练机制的复合错误识别框架CPR-CLIP

- **动机**：为解决 ImagineNet 算法所面临的**模态信息单一**、**人机交互性能差**等问题；
- **提出框架**：本章将**多模态对比预训练框架 CLIP**与**提示词工程**引入到复合错误行为识别任务中，提出了CPR-CLIP框架。

复合错误行为识别任务形式



CPR-CLIP框架结构



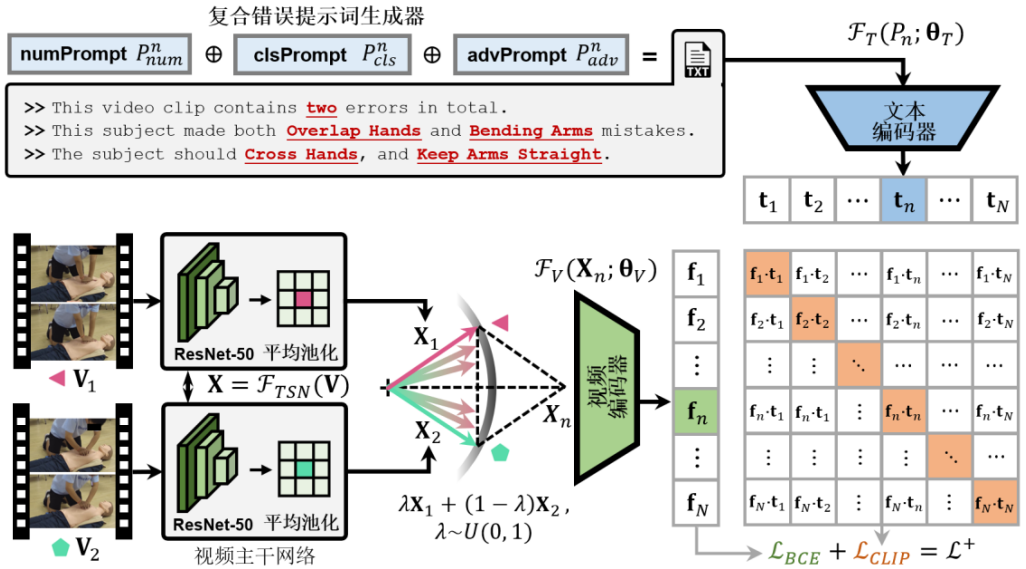
- **视觉通路 / Visual Pathway**：从单错误样本数据集中采样，并完成视频特征提取与特征映射；
- **语言通路 / Language Pathway**：通过描述模板对心肺复苏中的复合错误信息进行描述；
- **损失通路 / Loss Computation**：使用对比预训练损失进行模型训练。

$$\mathbf{f}_n = \mathcal{F}_V(\mathbf{X}_n;\boldsymbol{\theta}_V), \mathbf{f}_n \in \mathbb{R}^D$$

$$\mathbf{t}_n = \mathcal{F}_T(P_n;\mathbf{t}), \mathbf{t}_n \in \mathbb{R}^D$$

$$(\boldsymbol{\theta}_V^*, \boldsymbol{\theta}_T^*) = \arg\min_{(\boldsymbol{\theta}_V, \boldsymbol{\theta}_T)} \mathcal{L}_{CLIP}$$

5.2 CPR-CLIP框架构建过程



1、视频特征提取：使用视频主干网络进行特征提取；

视频网络主干特征提取：$\mathbf{X} = \mathcal{F}_{TSN}(\mathbf{V};\boldsymbol{\theta}_{TSN}), \mathbf{X} \in \mathbb{R}^{T\times D}$

时序平均池化操作：$\mathbf{X}_n = \frac{1}{T}\sum_{t=1}^{T}(\lambda \mathbf{X}_1^t + (1-\lambda)\mathbf{X}_2^t),\ \mathbf{X}_n \in \mathbb{R}^D$

视频特征编码：$\mathbf{f}_n = \mathcal{F}_V(\mathbf{X}_n;\boldsymbol{\theta}_V), \mathbf{f}_n \in \mathbb{R}^D$

2、提示语句构造与嵌入：提示语句模板分别从**错误数量**、**错误种类**和**改正建议**三个方面对复合错误进行了描述；

提示词模板

数量提示词：$P^n_{num} =$ *"This video clip contains* $\{\boldsymbol{cnt}\}$ *errors in total."*

种类提示词：$P^n_{cls} =$ *"This subject made both* $\{\boldsymbol{C}_1\}$ *and* $\{\boldsymbol{C}_2\}$ *mistakes."*

建议提示词：$P^n_{adv} =$ *"This subject should* $\{\boldsymbol{A}_1\}$ *and* $\{\boldsymbol{A}_2\}$*."*

提示词拼接：$P_n = P^n_{num} \oplus P^n_{cls} \oplus P^n_{adv}$

↓

提示词嵌入：$\mathbf{t}_n = \mathcal{F}_T(P_n;\mathbf{t}), \mathbf{t}_n \in \mathbb{R}^D$

3、损失函数设计：通过多模态预训练对比损失实现语义空间中的视觉特征和语言特征对齐；

跨模态余弦相似度计算：$sim(\mathbf{f}_n,\mathbf{t}_n) = \frac{\mathbf{f}_n \cdot \mathbf{t}_n}{\|\mathbf{f}_n\|\,\|\mathbf{t}_n\|}$

训练批次内部的相似度矩阵：

$$S(\mathbf{F},\mathbf{T}) = \begin{bmatrix} sim(\mathbf{f}_1,\mathbf{t}_1) & \cdots & sim(\mathbf{f}_1,\mathbf{t}_N) \\ \vdots & \ddots & \vdots \\ sim(\mathbf{f}_N,\mathbf{t}_1) & \cdots & sim(\mathbf{f}_N,\mathbf{t}_N) \end{bmatrix} \in \mathbb{R}^{N\times N}$$

KL散度定义：$KL[S_T(\mathbf{F},\mathbf{T}) \| M_{GT}] = \frac{1}{N^2}\sum_{j=1}^{N}\sum_{i=1}^{N} S_T(\mathbf{F},\mathbf{T})[i,j]\log\frac{S_T(\mathbf{F},\mathbf{T})[i,j]}{M_{GT}[i,j]}$

↓

CLIP损失函数：$\mathcal{L}_{CLIP} = \frac{1}{2}\left(KL[S_T(\mathbf{F},\mathbf{T}) \| M_{GT}] + KL[S_V(\mathbf{F},\mathbf{T}) \| M_{GT}]\right)$

↓

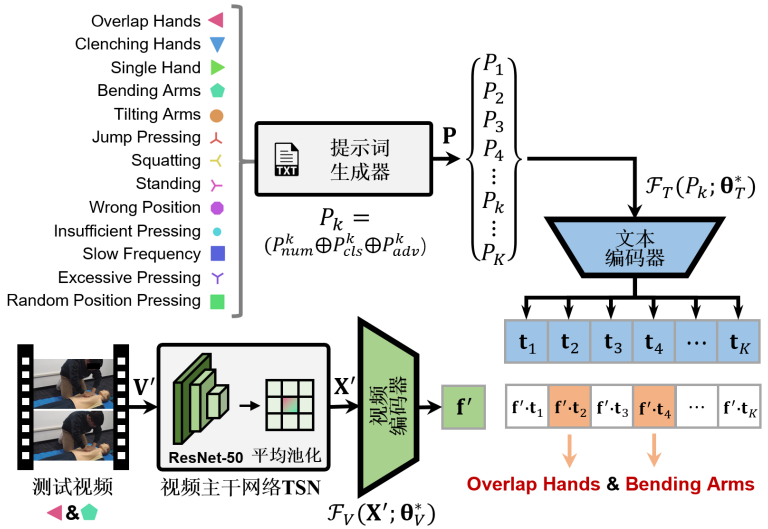
优化过程：$(\boldsymbol{\theta}_V^*, \boldsymbol{\theta}_T^*) = \arg\min_{(\boldsymbol{\theta}_V,\boldsymbol{\theta}_T)} \mathcal{L}_{CLIP}$

↓

组合损失：$\mathcal{L}^+ = \mathcal{L}_{CLIP} + \mathcal{L}_{BCE}$

5.3 CPR-CLIP框架推理模式

➢ **单视频预测推理**：对单个视频中所含有的错误行为进行预测，功能与ImagineNet相同；



单类提示语句集合：$\mathbf{P} = \{P_k\}_{k=1}^{K}$

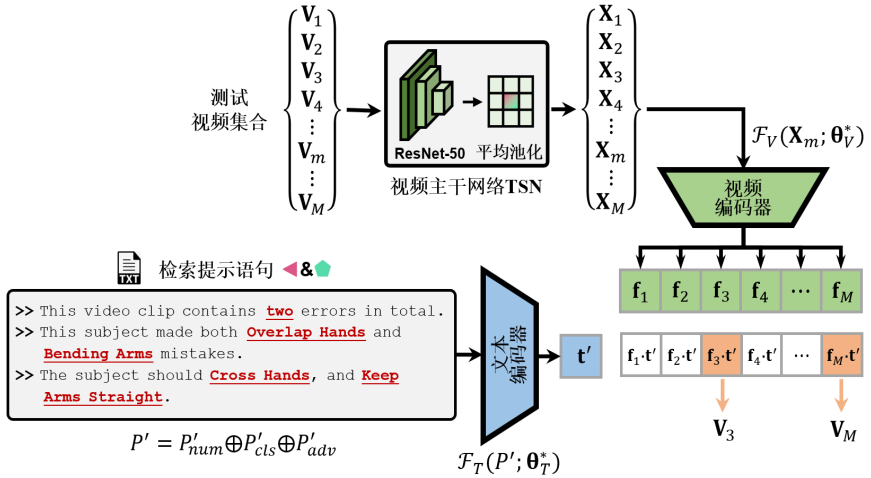
提示语句构造过程：$P_k = P^k_{num} \oplus P^k_{cls} \oplus P^k_{adv}$

相似度计算：$\mathbf{t}_k = \mathcal{F}_T(P_k;\theta^*_T), \mathbf{t}_k \in \mathbb{R}^D$

视觉特征提取：$\mathbf{f}' = \mathcal{F}_V(\mathbf{X}';\theta^*_V), \mathbf{f}' \in \mathbb{R}^D$

相似度计算：$S_V(\mathbf{f}',\mathbf{T}') = [sim(\mathbf{f}',\mathbf{t}_1),\cdots,sim(\mathbf{f}',\mathbf{t}_K)]^T$

➢ **特定类别视频检索推理**：给定视频库，通过语言描述实现特定类别错误视频的检索；



视频映射过程：$\mathbf{f}_m = \mathcal{F}_V(\mathbf{X}_m;\theta^*_V), \mathbf{f}_m \in \mathbb{R}^D$

视频特征集合：$\mathbf{F} = \{\mathbf{f}_1,\mathbf{f}_2,\cdots,\mathbf{f}_M\}$

查询提示语句：$P' = P'_{num} \oplus P'_{cls} \oplus P'_{adv}$

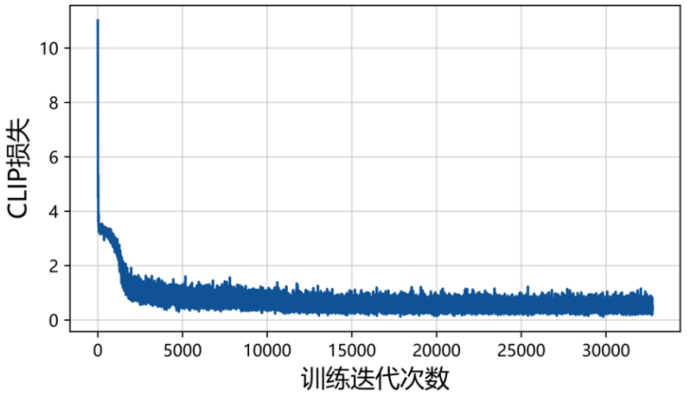
文本特征提取：$\mathbf{t}' = \mathcal{F}_T(P';\theta^*_T), \mathbf{t}' \in \mathbb{R}^D$

相似度计算：$S_V(\mathbf{F},\mathbf{t}') = [sim(\mathbf{f}_1,\mathbf{t}'),\cdots,sim(\mathbf{f}_M,\mathbf{t}')]$

5.4 实验结果-性能对比结果

- **性能对比实验**：相较于朴素迁移策略，多模态对比预训练机制的引入能够**显著提升复合错误识别性能**；
- **组合损失函数性能对比**：**同时引入两个损失的ImagineNet框架**能够在各种错误组合情况下实现精准识别。
- **训练过程Loss与mAP记录说明**：随着训练轮次的加深，**对比损失逐步下降，复合错误的识别精度逐步攀升**，说明了预训练过程的有效性。

(a) CPR-CLIP训练过程中的$\mathcal{L}_{CLIP}$记录



(b) CPR-CLIP训练过程中的平均精度记录

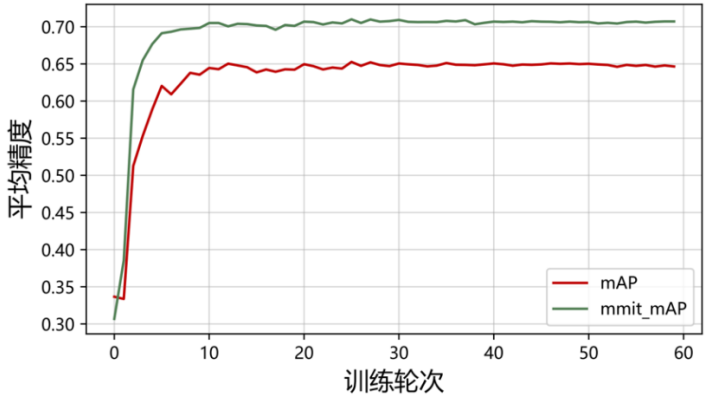


表4-1. 多模态预训练机制的性能对比实验

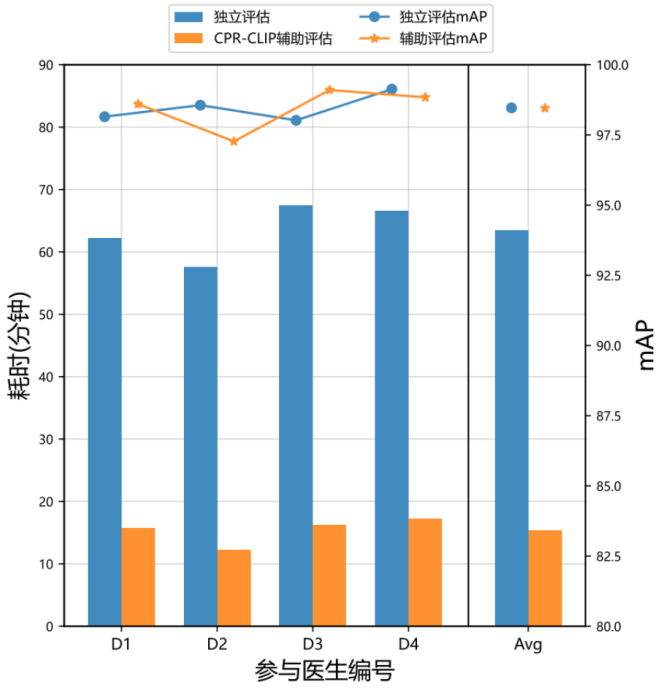
模型	mAP	Δ	mmit mAP	Δ
TSN[6]	0.5598	—	0.6143	—
CPR-CLIP	0.6034	↑4.36%	0.6727	↑5.84%
CPR-CLIP+	**0.6417**	↑8.19%	**0.7030**	↑8.87%
TSM[116]	0.5662	—	0.6618	—
CPR-CLIP	0.6401	↑7.39%	0.7074	↑4.56%
CPR-CLIP+	**0.7076**	↑14.14%	**0.7602**	↑9.84%
ST-GCN[101]	0.5776	—	0.6692	—
CPR-CLIP	0.6028	↑2.52%	0.6831	↑1.39%
CPR-CLIP+	**0.6358**	↑5.82%	**0.7127**	↑4.35%
ViViT[84]	0.5582	—	0.6651	—
CPR-CLIP	0.6503	↑9.21%	0.7494	↑8.43%
CPR-CLIP+	**0.7251**	↑16.69%	**0.7754**	↑11.03%
Video Swin[86]	0.5696	—	0.6701	—
CPR-CLIP	0.6685	↑9.89%	0.7567	↑8.66%
CPR-CLIP+	**0.7439**	↑17.43%	**0.7924**	↑12.23%

表4-2. CPR-CLIP+框架性能对比实验

模型	视频主干网络	mAP	mmit mAP
CBP[229]	TSN[6]	0.6285	0.6812
BLOCK[230]		0.6225	0.6965
ImagineNet-FC		0.6259	0.6893
CPR-CLIP+		**0.6417**	**0.7030**
CBP[229]	TSM[116]	0.6864	0.7487
BLOCK[230]		0.6651	0.7222
ImagineNet-FC		0.7053	0.7566
CPR-CLIP+		**0.7076**	**0.7602**
CBP[229]	Video Swin[86]	0.6951	0.7524
BLOCK[230]		0.6801	0.7322
ImagineNet-FC		0.7082	0.7638
CPR-CLIP+		**0.7439**	**0.7924**

5.4 实验结果-消融实验&系统部署

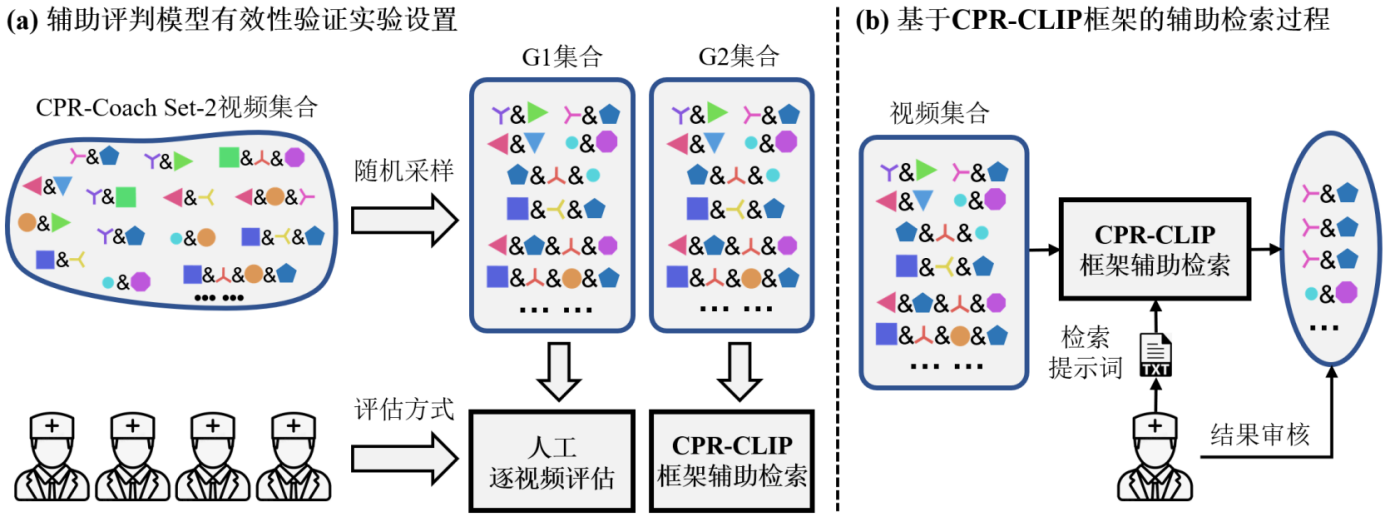
- **三种提示语句消融实验**：**种类提示语句**P_{cls} 占有最高的权重，**数量提示语句**P_{num}权重较小；
- **随机对照试验结果**：基于 CPR-CLIP 框架的智能检索系统能够在**不降低评估精度的前提下，节约 4 倍左右的评估耗时。**



开展随机对照试验验证辅助评估系统的有效性

表4-3. 三种提示语句类型的消融实验结果

主干网络	模型变体	P_{num}	P_{cls}	P_{adv}	mAP	mmit mAP
TSN[6]	CPR-CLIP	✓	✓	✓	0.6034	0.6727
		×	✓	✓	0.5493	0.6443
		✓	×	✓	0.4364	0.5226
		✓	✓	×	0.5306	0.6604
	CPR-CLIP+	✓	✓	✓	0.6417	0.7030
TSM[116]	CPR-CLIP	✓	✓	✓	0.6401	0.7074
		×	✓	✓	0.6298	**0.7147**
		✓	×	✓	0.4498	0.5480
		✓	✓	×	0.5651	0.6870
	CPR-CLIP+	✓	✓	✓	0.7076	0.7602
Video Swin[86]	CPR-CLIP	✓	✓	✓	0.6685	0.7567
		×	✓	✓	0.6351	0.7328
		✓	×	✓	0.4525	0.5910
		✓	✓	×	0.5591	0.7470
	CPR-CLIP+	✓	✓	✓	0.7439	0.7924



目 录

6.1 时序行为分析算法-研究现状

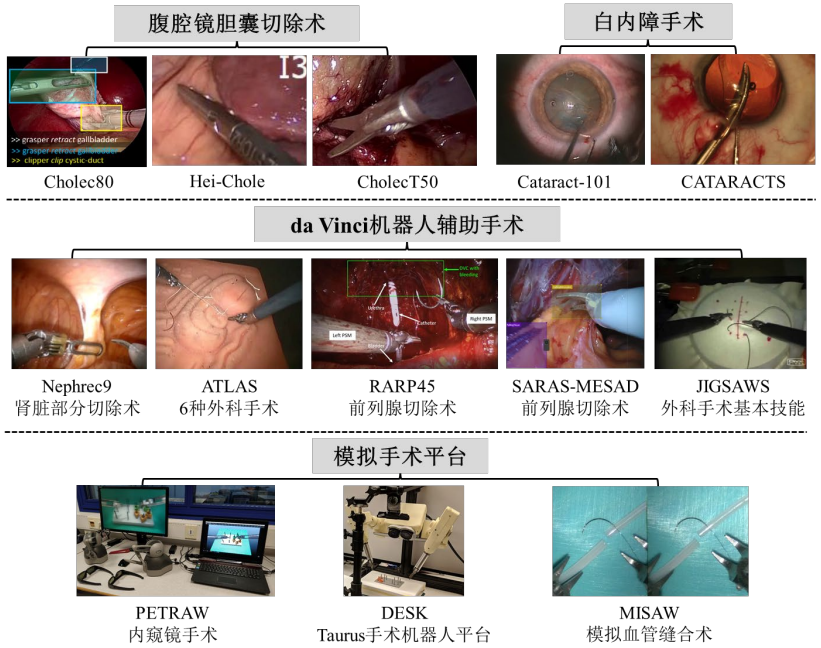
目前时序行为分析研究面临着以下问题：

- **数据集复杂度低**：数据集场景相对独立，无统一的行为标签表示体系、时序行为种类划分粒度粗、**未提供错误行为案例**；
- **时序分割算法精度差**：现有算法通常**对全序列进行无差别建模**，忽略特征信息在时间维度上的差异性；
- **不支持行为合规性评估**：现有算法只支持时序行为分割，**无法支持时序行为纠错与分析功能**。

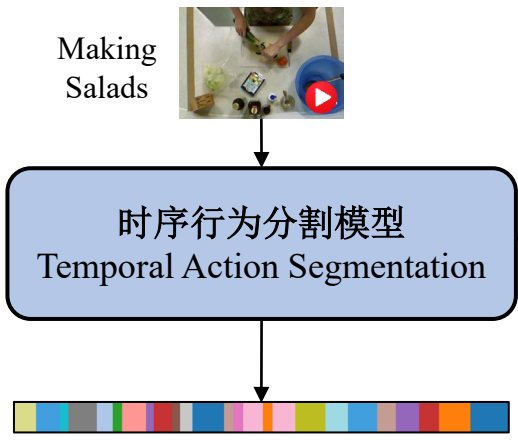
时序行为分割任务 / TAS：对持续时间较长的视频进行逐帧行为分类，医疗场景中又称为手术流程识别任务；

时序行为分析任务 / TAA：在时序分割任务的基础上，完成行为的合规性判断。

现有手术流程识别数据集



时序行为分割任务形式



现有时序行为分割算法分类

基于RNN的模型	TCN+LSTM
基于时序卷积的模型	MS-TCN, MS-TCN++
基于图卷积网络的模型	ASFormer, ViViT
基于Transformer的模型	ASFormer, CETNet, DiffAct

6.2 时序医疗行为知识图谱构建

- **研究现状**：学界中现有的医疗知识图谱关注于医疗知识的表示，**并不支持对医疗行为这种具有时序特性的知识进行表示**；
- **构建目的**：通过构建**时序医疗行为知识图谱**，为后续时序医疗行为分析研究提供**细粒度的行为标签表示框架**；
- **构建素材**：本文以《中国医学生临床技能操作指南（第二版）》教材为依据进行医疗行为知识图谱的构建。此教材囊括了 60 类常见临床技能操作，细类临床操作数目达 73 种。

中文医疗领域开源知识图谱

知识图谱名称	知识图谱主题	实体数量	关系数量	提出单位
DiseaseKG[264]	基于 cnSchma 常见疾病信息知识图谱	44,656	312,159	OpenKG
DiaKG[265]	糖尿病知识图谱数据集	22,050	6890	妙健康、阿里云
COVID-19[266]	基于 COVID-19 论文集的学术知识图谱	80 万	120 万	云南省高校数据科学与智能计算重点实验室
中文症状库[267]	包含症状实体和症状相关三元组数据集	135,485	617,499	华东理工大学

时序医疗行为的三个特性

- 类别有限：教材对每个行为均有结构化的介绍
- 层次化：医疗行为可划分为系列子行为
- 有序性：部分医疗子行为是严格有序的

《中国医学生临床技能操作指南》目录

6.2 时序医疗行为知识图谱构建

- **医疗行为流程知识图谱**：包含教材中对**每种医疗行为的操作顺序流程**规定；
- **医疗行为文本知识图谱**：对临床技能操作指南教材中的**大量知识性和说明性文本**进行知识图谱构建；
- **知识图谱统计信息**：3,630条医疗行为相关语句；5,052个命名实体；2,087个三元组。

基于正则匹配表达式的医疗行为流程知识提取

正则匹配表达式

```
result1 = re.match('[一二三四五六七八九十]、[\u4e00-\u9fa5]+\n' , row)
result2 = re.match('[0-9]+[.].*\n', row)
result3 = re.match('\([0-9]+\).*\n', row)
result4 = re.match('[1-9]\).*\n', row)
```

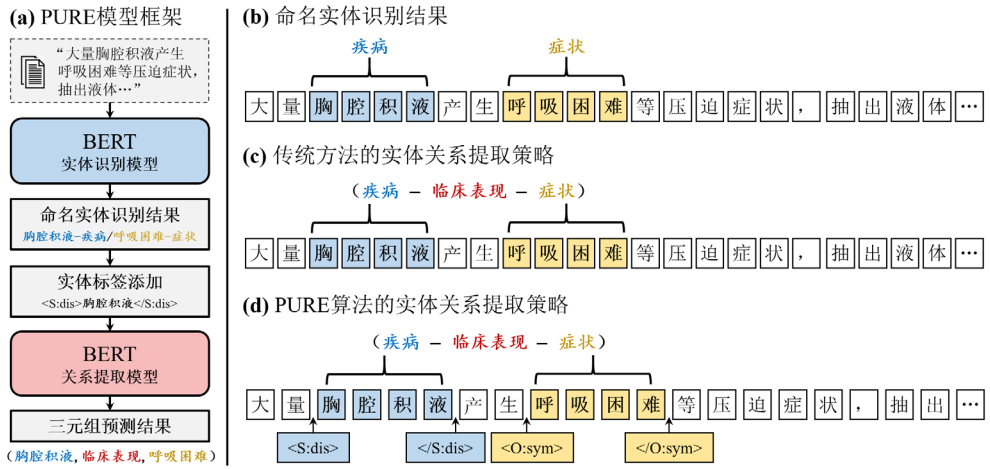
TXT文本

```
一、目的
1. 诊断作用：抽取少量胸腔内液体标本检测，以明确胸腔积液病因。
2. 治疗作用：抽出胸腔内液体，促进肺复张；胸膜腔内给药，达到治疗作用。

二、适应证
1. 胸腔积液需要明确诊断。
2. 大量胸腔积液产生呼吸困难等压迫症状，抽出液体促进肺复张，缓解症状。
3. 胸膜腔内给药。

三、禁忌证
对有凝血功能障碍或重症血小板减少者应慎用，必要时可补充一定量的凝血因子或血小板，使血液的出凝血功能得到部分纠正后，再行胸腔穿刺。

四、操作前准备
1. 患者准备
（1）测量生命体征（心率、血压、呼吸）。
（2）向患者解释胸腔穿刺的目的操作过程、可能的风险确认患者无穿刺禁忌、无利多卡因过敏。
（3）告知需要配合的事项（操作过程中避免剧烈咳嗽，保持体位，如有头晕、心悸、气促等不适及时报告）。
（4）签署知情同意书。
2. 材料准备
（1）胸腔穿刺包：内含弯盘2个、尾部连接乳胶管的16号和18号胸腔穿刺针各1根、中弯止血钳4把、孔巾1块、巾钳2把、棉球10个、纱布2块、小消毒杯2个、标本留置小瓶5个。
（2）消毒用品：2.5%碘酊和75%酒精，或0.5%碘伏。
（3）麻醉药物：2%利多卡因5ml。
（4）其他：5ml和50ml注射器各1个、500ml标本容器2个、胶布1卷、1000ml量筒或量杯1个、有靠背的座椅1个、抢救车1个、无菌手套2副。
3. 操作者准备
（1）两人操作。
（2）操作者洗手，戴帽子、口罩和无菌手套；助手协助患者体位摆放，观察穿刺过程中患者情况等。
（3）了解患者病情穿刺目的胸片情况。
（4）掌握胸腔穿刺操作相关知识、并发症的诊断与处理。

五、操作步骤
1. 体位：再次确认病变位于左侧还是右侧。常规取直立坐位，上身略前倾，必要时双前臂合抱或将前胸靠在床头桌上，以使肋间隙能够充分暴露（图1-1）。卧床患者可以采取仰卧高坡卧位，患侧略向健侧转，便于显露穿刺部位。
2. 穿刺点选择
（1）穿刺点选择：穿刺点主要是根据患者胸液的范围而定，常选择腋前线第5肋间，腋中线第6~7肋间，腋后线第7~8肋间，肩胛下角线第7~8肋间。穿刺点应避开局部皮肤感染灶。
（2）标记穿刺点：确定后要标记穿刺点。
（3）叩诊寻找：一般通过叩诊结合X线胸片确定穿刺部位，必要时可通过超声检查来进一步确定穿刺点及穿刺深度，甚至在B超引导下完成穿刺。
3. 消毒铺单
（1）准备：术者戴好无菌手套，在两个消毒小杯内分别放入数个棉球，助手协助，分别倒入少量2.5%碘
```

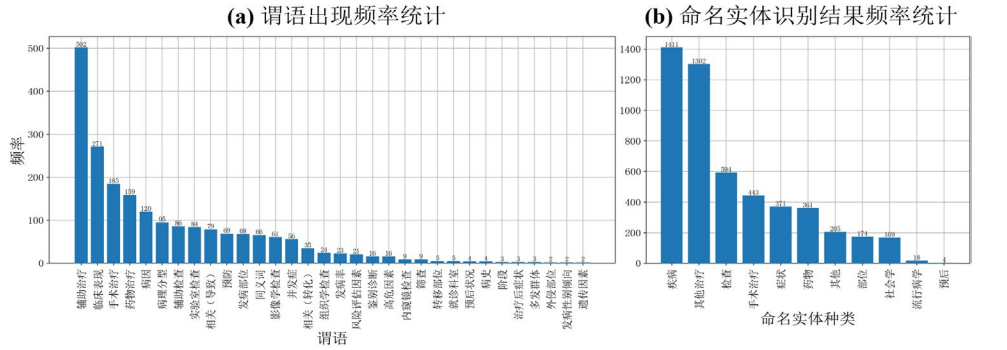
JSON结构化文本

```
{
    "目的": [
        "1. 诊断作用：抽取少量胸腔内液体标本检测，以明确胸腔积液病因。",
        "2. 治疗作用：抽出胸腔内液体，促进肺复张；胸膜腔内给药，达到治疗作用。"
    ],
    "适应证": [
        "1. 胸腔积液需要明确诊断。",
        "2. 大量胸腔积液产生呼吸困难等压迫症状，抽出液体促进肺复张，缓解症状。",
        "3. 胸膜腔内给药。"
    ],
    "禁忌证": [
        "对有凝血功能障碍或重症血小板减少者应慎用，必要时可补充一定量的凝血因子或血小板，使血液
    ],
    "操作前准备": {
        "1. 患者准备": [
            "（1）测量生命体征（心率、血压、呼吸）。",
            "（2）向患者解释胸腔穿刺的目的操作过程、可能的风险确认患者无穿刺禁忌、无利多卡因过
            "（3）告知需要配合的事项（操作过程中避免剧烈咳嗽，保持体位，如有头晕、心悸、气促等
            "（4）签署知情同意书。"
        ],
        "2. 材料准备": [
            "（1）胸腔穿刺包：内含弯盘2个、尾部连接乳胶管的16号和18号胸腔穿刺针各1根、中弯止血
            "（2）消毒用品：2.5%碘酊和75%酒精，或0.5%碘伏。",
            "（3）麻醉药物：2%利多卡因5ml。",
            "（4）其他：5ml和50ml注射器各1个、500ml标本容器2个、胶布1卷、1000ml量筒或量杯1个
        ],
        "3. 操作者准备": [
            "（1）两人操作。",
            "（2）操作者洗手，戴帽子、口罩和无菌手套；助手协助患者体位摆放，观察穿刺过程中患者
            "（3）了解患者病情穿刺目的胸片情况。",
            "（4）掌握胸腔穿刺操作相关知识、并发症的诊断与处理。"
        ]
    },
    "操作步骤": {
        "1. 体位": [
            "再次确认病变位于左侧还是右侧。常规取直立坐位，上身略前倾，必要时双前臂合抱或将前
        ],
        "2. 穿刺点选择": [
            "（1）穿刺点选择：穿刺点主要是根据患者胸液的范围而定，常选择腋前线第5肋间，腋中线第
            "（2）标记穿刺点：确定后要标记穿刺点。",
            "（3）叩诊寻找：一般通过叩诊结合X线胸片确定穿刺部位，必要时可通过超声检查来进一步
        ],
        "3. 消毒铺单": [
            "（1）准备：术者戴好无菌手套，在两个消毒小杯内分别放入数个棉球，助手协助，分别倒入
            "（2）消毒：用2.5%碘酊以穿刺点为中心，向周边环形扩展消毒至少15cm；以75%酒精脱碘
            "（3）铺巾：无菌孔巾中心对准穿刺点，上方以胶布或巾钳固定于患者衣服上。"
```

基于PURE模型的医疗行为知识提取



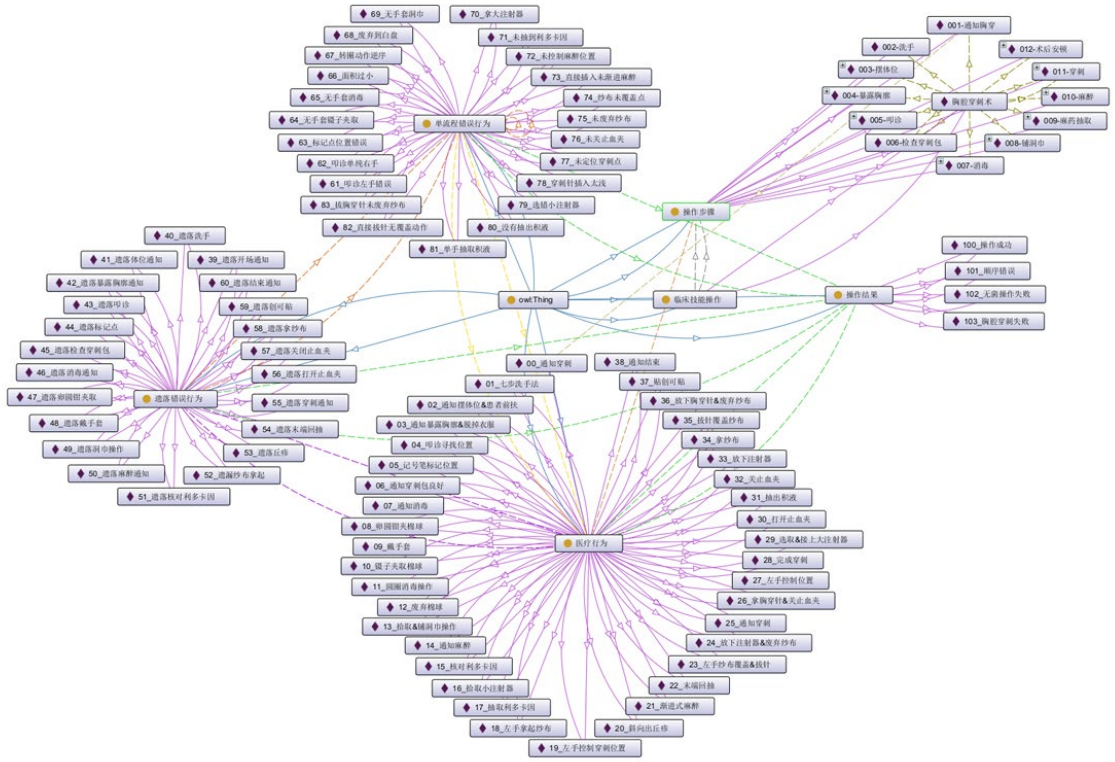
时序医疗行为知识图谱统计数据



6.2 时序医疗行为知识图谱构建

- **研究对象**：四大穿刺术之一的**胸腔穿刺术**，具备一定的时序复杂度；
- **细化方法**：在教材划分的9个流程基础上，细化为**12个流程**和**39个子行为**；
- **错误行为探究**：**22种遗漏错误行为**、**23种单流程错误行为**；
- **行为因果关系**：流程遗漏或错误行为会导致后续出现关联错误。

胸腔穿刺术时序行为知识图谱



胸腔穿刺术时序行为细化表

序号	流程划分	细分项	遗漏错误行为	单流程错误行为类
1	通知胸穿	0_通知穿刺	39_遗漏通知	
2	洗手	1_七步洗手法	40_遗漏洗手	
3	摆体位	2_通知摆体位&前扶	41_遗漏通知	
4	暴露胸廓	3_通知暴露胸廓&脱衣	42_遗漏通知	
5	叩诊	4_叩诊寻找位置	43_遗漏叩诊	61_叩诊左手错误 62_叩诊单纯右手
		5_记号笔标记位置	44_遗漏标记点	63_标记点位置错误
6	检查穿刺包	6_通知穿刺包良好	45_遗漏检查	
7	消毒	7_通知消毒	46_遗漏通知	
		8_卵圆钳夹棉球	47_遗漏卵圆钳夹取	
		9_戴手套	48_遗漏手套	
		10_镊子夹取棉球		64_无手套镊子夹取
		11_螺旋消毒操作		65_无手套消毒 66_面积过小 67_转圈动作逆序
		12_废弃棉球		68_废弃到白盘
8	铺洞巾	13_拾取&铺洞巾操作	49_遗漏洞巾操作	69_无手套洞巾
9	核对+利多卡因抽取	14_通知麻醉	50_遗漏通知	
		15_核对利多卡因	51_遗漏核对	
		16_拾取小注射器		70_拿大注射器
		17_抽取利多卡因		71_未抽到
10	麻醉	18_麻醉前拿起纱布	52_遗漏纱布拿起	
		19_左手控制穿刺位置		72_未控制麻醉位置
		20_斜向出丘疹	53_遗漏丘疹	
		21_渐进式麻醉		73_未渐进麻醉
		22_未端回抽	54_遗漏回抽	
		23_纱布覆盖&拔针		74_纱布未覆盖点
		24_废弃纱布		75_未废弃纱布
11	穿刺	25_通知穿刺	55_遗漏通知	
		26_胸穿针&关止血夹		76_未关止血夹
		27_左手控制位置		77_未定位穿刺点
		28_完成穿刺		78_插入太浅
		29_接大注射器		79_选错注射器
		30_打开止血夹	56_遗漏打开止血夹	
		31_抽出积液		80_没有抽出积液 81_单手抽取
		32_关止血夹	57_遗漏关闭止血夹	
		33_放下注射器		
		34_拿纱布	58_遗漏纱布	
		35_拔针覆盖纱布		82_直接拔针
		36_废弃纱布		83_未废弃纱布
12	术后	37_贴创可贴	59_遗漏创可贴步骤	
		38_通知结束	60_遗漏结束通知	

6.2 时序医疗行为知识图谱构建

- **研究对象**：四大穿刺术之一的**胸腔穿刺术**，具备一定的时序复杂度；
- **细化方法**：在教材划分的9个流程基础上，细化为**12个流程**和**39个子行为**；
- **错误行为探究**：**22种遗漏错误行为**、**23种单流程错误行为**；
- **行为因果关系**：流程遗漏或错误行为会导致后续出现关联错误。

胸腔穿刺术时序行为知识图谱

行为因果关系

流程遗落导致后续不合格类	48_遗落手套	⇒	64_无手套镊子夹取
	48_遗落手套	⇒	65_无手套消毒
	48_遗落手套	⇒	69_无手套洞巾
	52_遗漏纱布拿起	⇒	74_纱布未覆盖点
	52_遗漏纱布拿起	⇒	75_未废弃纱布
	56_遗落打开止血夹	⇒	80_没有抽出积液
	58_遗落纱布	⇒	82_直接拔针无覆盖动作
	58_遗落纱布	⇒	83_未废弃纱布
流程错误导致后续不合格类	63_标记点位置错误	⇒	80_没有抽出积液
	78_插入太浅&穿刺失败	⇒	80_没有抽出积液
	76_未关止血夹	⇒	80_没有抽出积液

胸腔穿刺术时序行为细化表

序号	流程划分	细分项	遗漏错误行为	单流程错误行为类
1	通知胸穿	0_通知穿刺	39_遗漏通知	
2	洗手	1_七步洗手法	40_遗漏洗手	
3	摆体位	2_通知摆体位&前扶	41_遗漏通知	
4	暴露胸廓	3_通知暴露胸廓&脱衣	42_遗漏通知	
5	叩诊	4_叩诊寻找位置	43_遗漏叩诊	61_叩诊左手错误 62_叩诊单纯右手
		5_记号笔标记位置	44_遗漏标记点	63_标记点位置错误
6	检查穿刺包	6_通知穿刺包良好	45_遗漏检查	
7	消毒	7_通知消毒	46_遗漏通知	
		8_卵圆钳夹棉球	47_遗漏卵圆钳夹取	
		9_戴手套	48_遗漏手套	
		10_镊子夹取棉球		64_无手套镊子夹取
		11_螺旋消毒操作		65_无手套消毒 66_面积过小 67_转圈动作逆序
		12_废弃棉球		68_废弃到白盘
8	铺洞巾	13_拾取&铺洞巾操作	49_遗漏洞巾操作	69_无手套洞巾
9	核对+利多卡因抽取	14_通知麻醉	50_遗漏通知	
		15_核对利多卡因	51_遗漏核对	
		16_拾取小注射器		70_拿大注射器
		17_抽取利多卡因		71_未抽到
10	麻醉	18_麻醉前拿起纱布	52_遗漏纱布拿起	
		19_左手控制穿刺位置		72_未控制麻醉位置
		20_斜向出丘疹	53_遗漏丘疹	
		21_渐进式麻醉		73_未渐进麻醉
		22_末端回抽	54_遗漏回抽	
		23_纱布覆盖&拔针		74_纱布未覆盖点
		24_废弃纱布		75_未废弃纱布
11	穿刺	25_通知穿刺	55_遗漏通知	
		26_胸穿针&关止血夹		76_未关止血夹
		27_左手控制位置		77_未定位穿刺点
		28_完成穿刺		78_插入太浅
		29_接大注射器		79_选错注射器
		30_打开止血夹	56_遗漏打开止血夹	
		31_抽出积液		80_没有抽出积液 81_单手抽取
		32_关止血夹	57_遗漏关闭止血夹	
		33_放下注射器		
		34_拿纱布	58_遗漏纱布	
		35_拔针覆盖纱布		82_直接拔针
		36_废弃纱布		83_未废弃纱布
12	术后	37_贴创可贴	59_遗漏创可贴步骤	
		38_通知结束	60_遗漏结束通知	

6.3 ThoSet 数据集构建

ThoSet： Thoracocentesis Dataset

- **第一视角行为采集平台**：本文搭建了一套基于第一视角的**胸腔穿刺行为采集平台**，操作台上的所有医疗器械摆放与真实考核场景保持一致；
- **时序医疗行为知识图谱**：所有行为均按照**细化后的知识图谱**进行划分，并完成正确行为和错误行为的案例采集；
- **支持任务**：时序行为分割任务、行为合规性评估任务（遗漏行为&错误行为检测）

ThoSet数据集采集平台搭建

胸腔穿刺数据集采集平台

第一人称视角摄像头

摄像头佩戴方式

胸腔穿刺完整流程展示

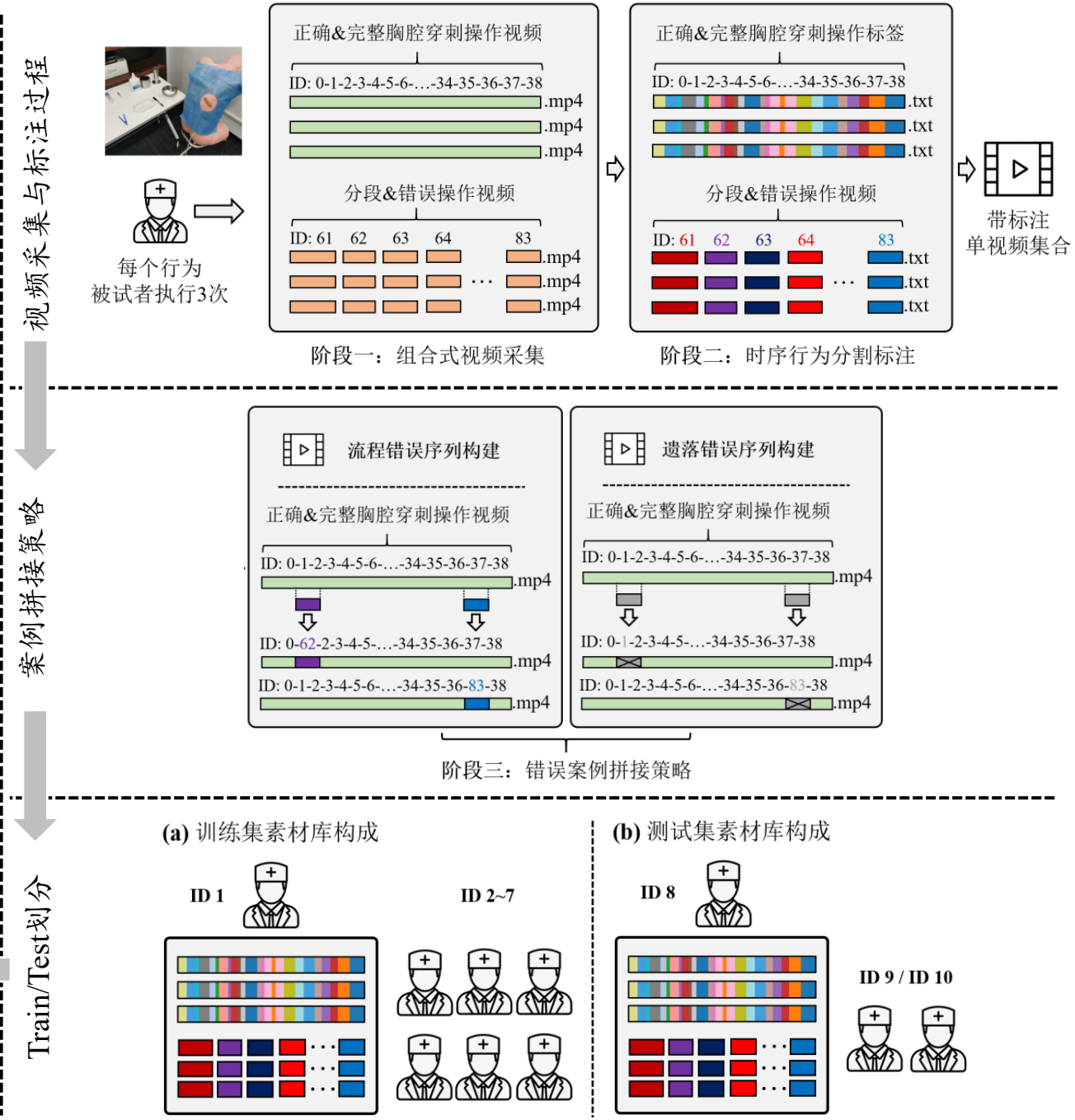
6.3 ThoSet 数据集构建

ThoSet数据集构建流程

- **阶段1-原始视频采集**：对**正确完整的胸穿操作**和**单独错误行为**视频进行采集；
- **阶段2-时序行为分割标注**：完成所有视频的时序行为标注，并依据标签连贯性进行视频分割，**构建视频素材库**。共包含3,720条单类别视频；
- **阶段3-错误案例拼接**：依据特定错误在素材库中进行视频采样，**构造错误行为序列**，满足因果条件约束；
- **ThoSet统计信息**：共包含**2,800条操作序列**，总时长近6h；通过“素材库+视频 ID 索引表”进行逻辑存储。

表5-9. ThoSet 数据集训练集与测试集统计信息

统计信息	训练集	Avg. Len.	Avg. Err.	测试集	Avg. Len.	Avg. Err.
人数	7	—	—	3	—	—
#正确行为序列	280	248.04s	0	120	243.09s	0
#较少错误序列	560	242.69s	2.08	240	233.16s	2.06
#中等数量错误序列	560	232.02s	4.50	240	225.60s	4.07
#较多数量错误序列	560	215.53s	8.39	240	211.42s	8.24
序列总数	1960	233.00s	—	840	227.02s	—

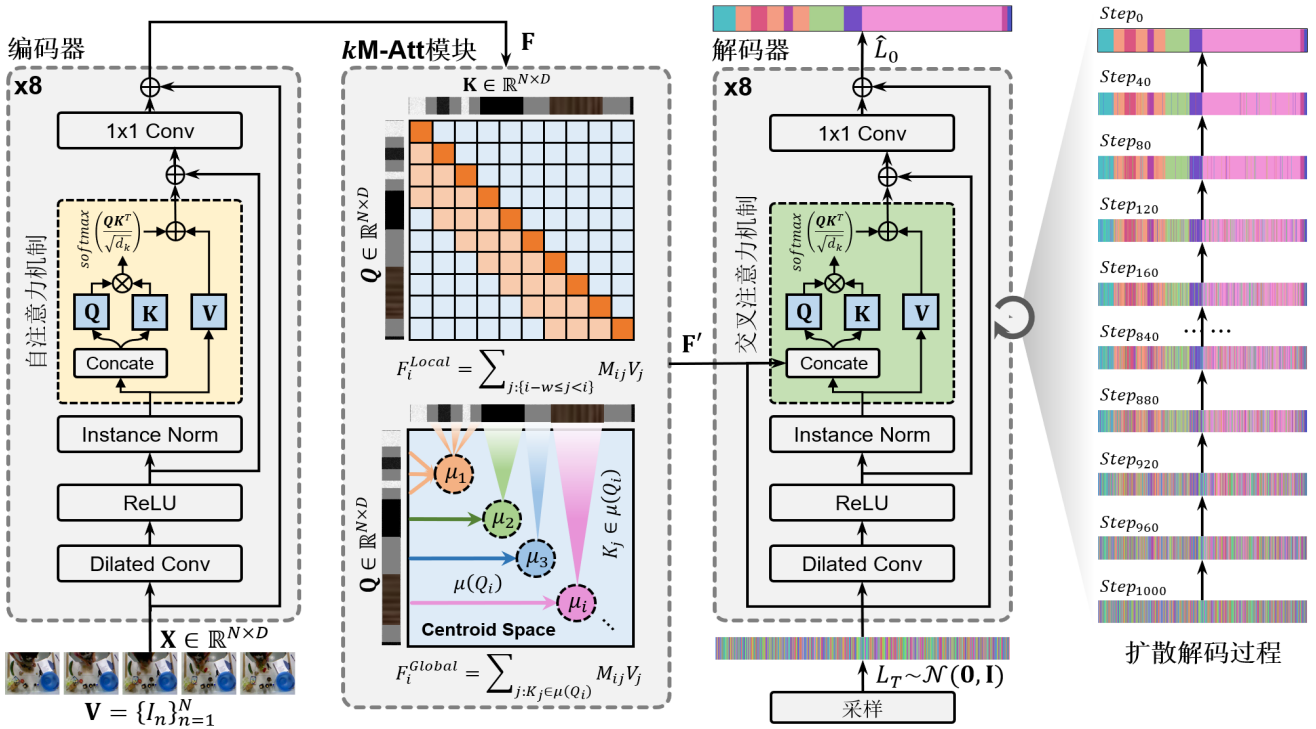


6.4 基于时序聚类注意力机制的扩散时序分割算法

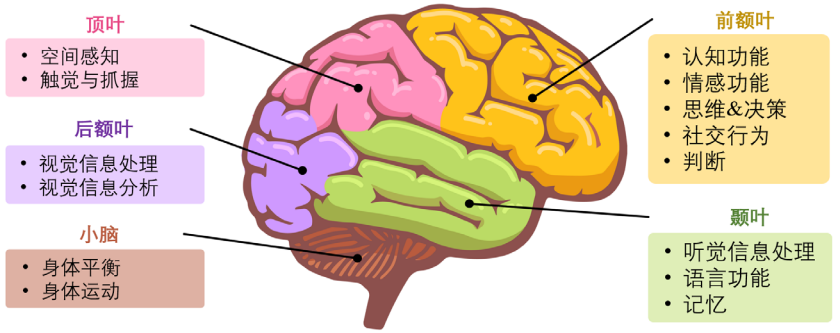
- **动机**：传统的基于Transformer的时序分割算法往往对全序列进行**无差别建模**，忽略特征信息在**时间维度上的差异性**；
- **启发**：受**人类大脑的功能分区结构**启发，本文提出了基于时序聚类注意力机制的***k*M-Att模块**，完成时序特征增强功能；

$$\mathbf{F} = \mathcal{F}_{Encoder}(\mathbf{X}) \quad \Rightarrow \quad \mathbf{F}' = \mathcal{F}_{kMAtt}(\mathbf{F}) \quad \Rightarrow \quad \hat{L}_0 = \mathcal{F}_{Decoder}(\mathbf{F}', L_T, T)$$

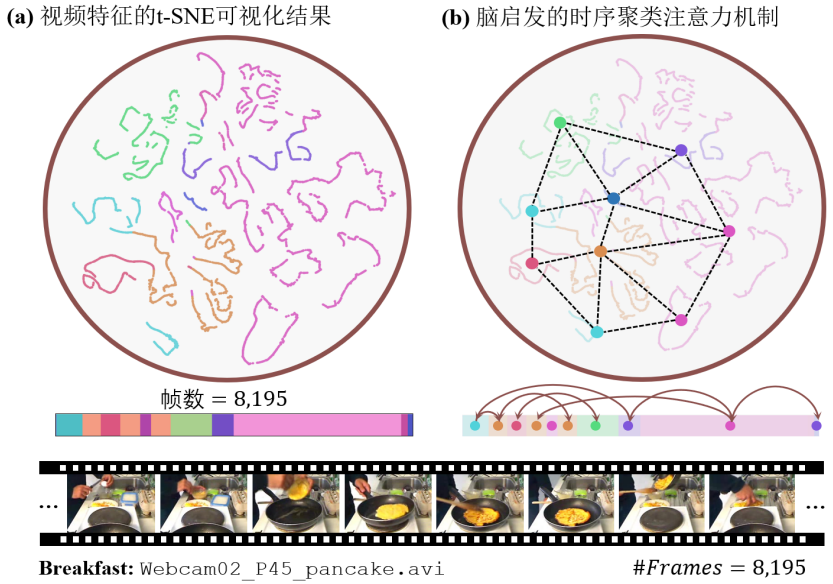
本文所提出的扩散时序行为分割框架示意图



人类大脑功能分区示意图



特征 t-SNE 可视化结果&
时序聚类注意力机制原理

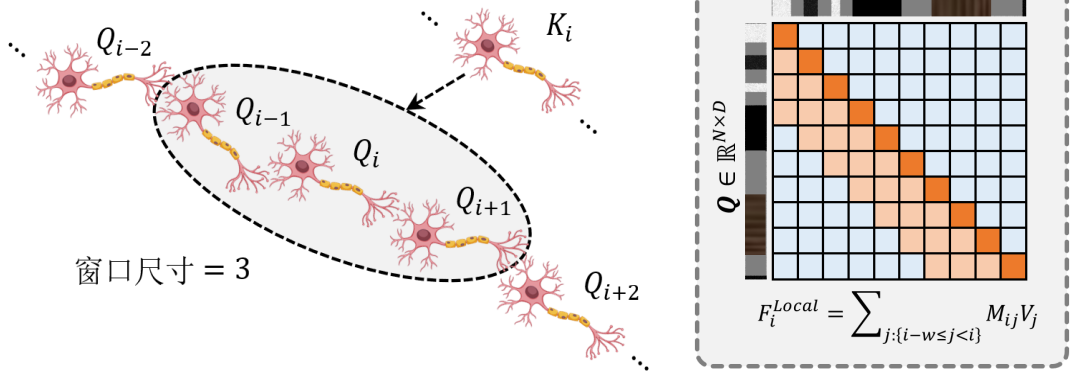


6.4 基于时序聚类注意力机制的*k*M-Att模块

- **物理临近交互（Physical Proximity Interactions）**：从局部来看，物理临近交互完成各**神经元之间的信息传递**；特征增强模块使用划定窗口的自注意力机制完成模拟；

- **逻辑分区交互（Logical Partitioning Interactions）**：从全局来看，逻辑分区交互完成各神经元**在功能区域内的信息传递**；本文通过基于时序*k*-means聚类注意力机制特征增强方法，完成不同时序分区信息中的特征交互。

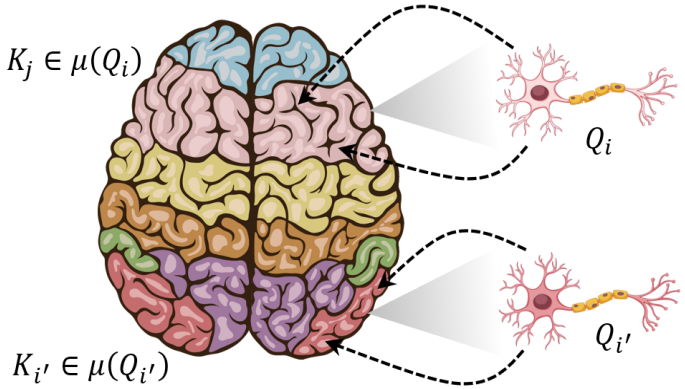
(a) 物理临近交互与局部注意力机制

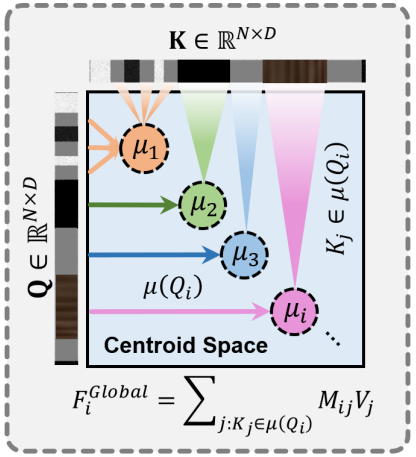


注意力图计算方法：$\mathbf{M} = \mathrm{softmax}\left(\frac{\mathbf{QK}^T}{\sqrt{d_k}}\right)$

窗口注意力机制：$F_i^{Local} = \sum_{j:\{i-w\le j<i\}} M_{ij}V_j$

(b) 逻辑分区交互与*k*-means聚类注意力机制





中心特征集合：$\boldsymbol{\mu} = \{\mu_1, \mu_2, \cdots, \mu_k\}, \mu_k \in \mathbb{R}^D$

k-means注意力机制：$F_i^{Global} = \sum_{j:\{\mu(K_j)=\mu(Q_i)\}} M_{ij}V_j$

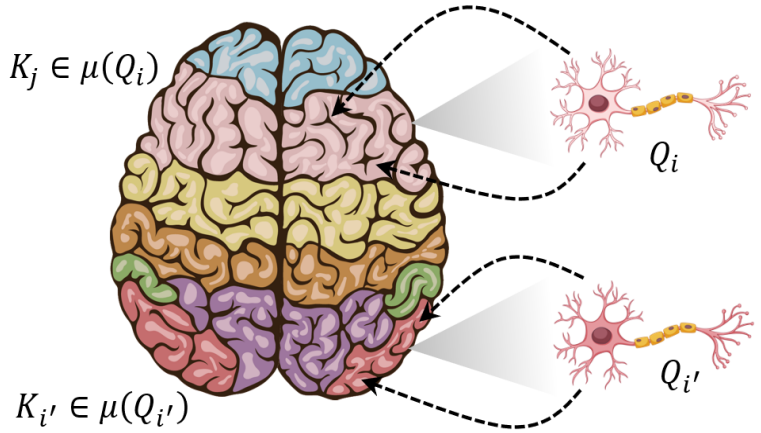
聚类中心更新：$\mu_k \leftarrow \lambda\mu_k + \frac{1-\lambda}{2}\sum_{i:\{\mu(Q_i)=\mu_k\}} Q_i + \frac{1-\lambda}{2}\sum_{j:\{\mu(K_j)=\mu_k\}} K_j$

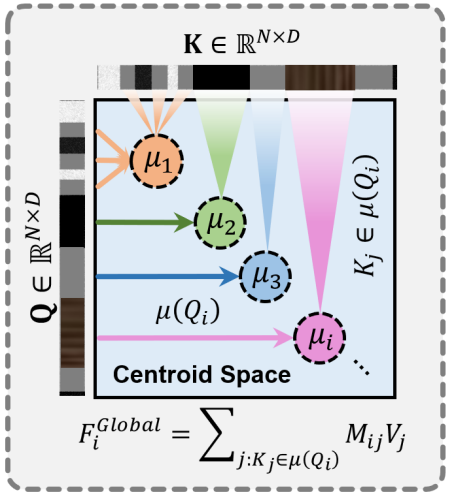
特征聚合：$\mathbf{F} = \mathcal{F}_{Linear}\left(\mathbf{F}^{Local} \oplus \mathbf{F}^{Global}\right)$

6.4 基于时序聚类注意力机制的*k*M-Att模块

➢ 时序*k*-means聚类注意力机制实现流程

(b) 逻辑分区交互与*k*-means聚类注意力机制





算法 1：时序 *k*-means 聚类注意力机制实现流程

Input：视频特征 $\mathbf{F} \in \mathbb{R}^{N\times D}$

Parameter：中心点集合 $\boldsymbol{\mu} \in \mathbb{R}^{k\times D}$，衰减参数 $\lambda = 0.999$

Output：视频特征 $\mathbf{F}^{Global} \in \mathbb{R}^{N\times D}$

// 线性层映射&归一化层

$\mathbf{Q} \leftarrow \mathbf{F}\mathbf{W}_Q, \mathbf{K} \leftarrow \mathbf{F}\mathbf{W}_K, \mathbf{V} \leftarrow \mathbf{F}\mathbf{W}_V$

$\mathbf{Q} \leftarrow \mathrm{LayerNorm}(\mathbf{Q}), \mathbf{K} \leftarrow \mathrm{LayerNorm}(\mathbf{K})$

// *k*-means 聚类注意力机制

$\widehat{\mathbf{Q}} \leftarrow \boldsymbol{\mu}\mathbf{Q}^T, \widehat{\mathbf{K}} \leftarrow \boldsymbol{\mu}\mathbf{K}^T$

$Q_{idx} \leftarrow \mathrm{Sort}[\text{Top-k}(\widehat{\mathbf{Q}}), N/k]$

$K_{idx} \leftarrow \mathrm{Sort}[\text{Top-k}(\widehat{\mathbf{K}}), N/k]$

// 收集选中特征

$\mathbf{Q}' \leftarrow \mathbf{Q}[Q_{idx}], \mathbf{K}' \leftarrow \mathbf{K}[K_{idx}], \mathbf{V}' \leftarrow \mathbf{V}[K_{idx}]$

// 生成注意力图

$\mathbf{M} \leftarrow \mathrm{softmax}(\mathbf{Q}'\mathbf{K}'^T/d_{K'})$

$\mathbf{V}' \leftarrow \mathbf{M}\mathbf{V}'$

$\mathbf{F} \leftarrow \mathbf{V}'[K_{idx}]$

// 更新中心点

$\mathbf{Q}_m \leftarrow \text{One-hot}[\mathrm{argmax}(\widehat{\mathbf{Q}}))]$

$\mathbf{K}_m \leftarrow \text{One-hot}[\mathrm{argmax}(\widehat{\mathbf{K}}))]$

$\boldsymbol{\mu} \leftarrow \lambda\boldsymbol{\mu} + (1-\lambda)\mathbf{Q}_m\widehat{\mathbf{Q}}/2 + (1-\lambda)\mathbf{K}_m\widehat{\mathbf{K}}/2$

return $\mathbf{F}^{Global}$

6.5 损失函数

本模型在训练过程中使用**三种损失函数**：$\mathcal{L} = \mathcal{L}_{CE} + \mathcal{L}_{Smooth} + \mathcal{L}_{Align}$

- **交叉熵损失**：对**帧级别的预测差异性**进行度量；

$$\mathcal{L}_{CE} = \sum_{n=1}^{N-1}\sum_{m=1}^{M} - L_0[n,m] \cdot \log \hat{L}_0[n,m]$$

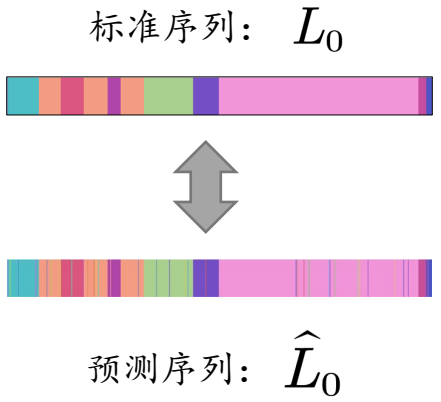
- **时序平滑损失**：对**标签的局部相似性**进行度量，从而降低网络预测结果中的标签切换频率；

$$\mathcal{L}_{Smooth} = \frac{1}{(N-1)M}\sum_{n=1}^{N-1}\sum_{m=1}^{M}\left(\log \hat{L}_0[n,m] - \log \hat{L}_0[n+1,m]\right)^2$$

- **边界对齐损失**：度量模型对**行为切换位置**预测的偏差；

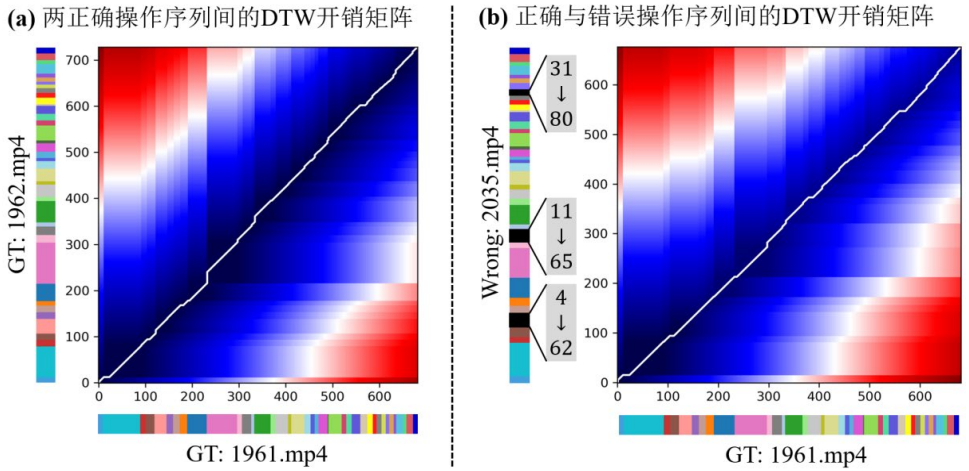
$$\text{边界序列生成：} B_n = \begin{cases} 0, & L_0[i] = L_0[i+1] \\ 1, & L_0[i] \neq L_0[i+1] \end{cases}$$

$$\text{边界对齐损失：} \mathcal{L}_{Align} = \frac{1}{N-1}\sum_{n=1}^{N-1}\Big[-\bar{B}_n \log\left(1 - \hat{L}_0[n] \cdot \hat{L}_0[n+1]\right) - \left(1-\bar{B}_n\right)\log\left(\hat{L}_0[n] \cdot \hat{L}_0[n+1]\right)\Big]$$

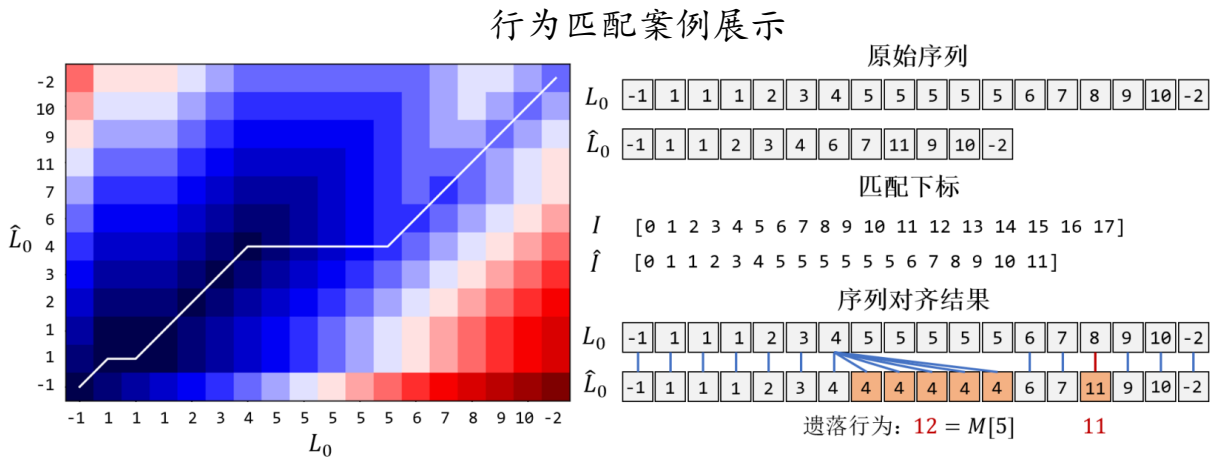


6.6 时序行为合规性检测算法

- **DTW 算法关联匹配环节**：使用动态规整算法完成**预测序列与标准序列的对齐**；



- **行为合规性检测环节**：根据序列对齐结果生成**遗漏行为**与**错误行为**检测结果；



算法 2：时序行为合规性检测算法

Input：预测序列 $\hat{L}_0$，标准序列 L_0，丢失行为映射表 M

Output：错误行为集合：$[S_{\text{Loss}}, S_{\text{Err}}]$

// DTW 算法匹配，获取两个下标集合

$(\hat{I}, I) \leftarrow \text{DTW}(\hat{L}_0, L_0)$, $\text{Len}(\hat{I}) = \text{Len}(I)$

// 序列遍历&差异检测

for i **in** $\text{Len}(I)$:

  if $\hat{L}_0[\hat{I}_i] \neq L_0[I_i]$:

    $S_{\text{Dig}}.add(i)$ // 获取差异位置集合

// 遗落行为&错误行为检测

for i **in** S:

  if $\hat{L}_0[\hat{I}_i] \in S_{\text{Correct}}$:

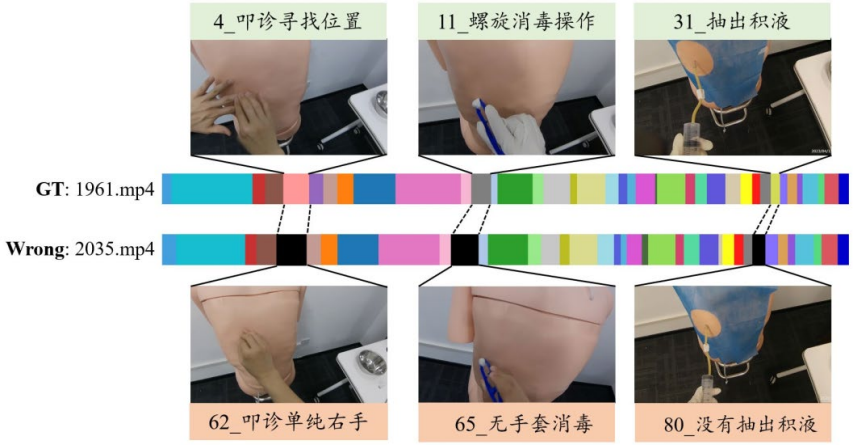
    $L_{\text{Loss}} \leftarrow M[L_0[I_i]]$

    $S_{\text{Loss}}.add(L_{\text{Loss}})$ // 遗落行为

  else :

    $S_{\text{Err}}.add(\hat{L}_0[\hat{I}_i])$ // 错误行为

return $[S_{\text{Loss}}, S_{\text{Err}}]$



6.7 实验结果-时序分割性能对比

- **公开时序行为分割数据集**：Breakfast、50Salads、GTEA、ThoSet；
- **与SOTA方法对比结果**：结果显示本文所提出的*k*M-Att模型能够在公开数据集和ThoSet数据集中均能取得比现有算法更精准的分割性能；

表5-11. *k*M-Att 方法与现有算法性能对比

模型	Breakfast				50Salads				GTEA			
	F1@{10,25,50}	Edit	Acc	Avg	F1@{10,25,50}	Edit	Acc	Avg	F1@{10,25,50}	Edit	Acc	Avg
MS-TCN++[147]	64.1/58.6/45.9	65.6	67.6	60.4	80.7/78.5/70.1	74.3	83.7	77.5	88.8/85.7/76.0	83.5	80.1	82.8
SSTDA[273]	75.0/69.1/55.2	73.7	70.2	68.6	83.0/81.5/73.8	75.8	83.2	79.5	90.0/89.1/78.0	86.2	79.8	84.6
GTRM[150]	57.5/54.0/43.3	58.7	65.0	55.7	75.4/72.8/63.9	67.5	82.6	72.4	-/-/-	-	-	-
BCN[274]	68.7/65.5/55.0	66.2	70.4	65.2	82.3/81.3/74.0	74.3	84.4	79.3	88.5/87.1/77.3	84.4	79.8	83.4
MTDA[275]	74.2/68.6/56.5	73.6	71.0	68.8	82.0/80.1/72.5	75.2	83.2	78.6	90.5/88.4/76.2	85.8	80.0	84.2
C2F-TCN[276]	72.2/68.7/57.6	69.6	76.0	68.8	84.3/81.8/72.6	76.4	84.9	80.0	90.3/88.8/77.7	86.4	80.8	84.8
G2L[149]	74.9/69.0/55.2	73.3	70.7	68.6	80.3/78.0/69.8	73.4	82.2	76.7	89.9/87.3/75.8	84.6	78.5	83.2
HASR[277]	74.7/69.5/57.0	71.9	69.4	68.5	86.6/85.7/78.5	81.0	83.9	83.1	90.9/88.6/76.4	87.5	78.7	84.4
ASRF[278]	74.3/68.9/56.1	72.4	67.6	67.9	84.9/83.5/77.3	79.3	84.5	81.9	89.4/87.8/79.8	83.7	77.3	83.6
ASFormer[152]	76.0/70.6/57.4	75.0	73.5	70.5	85.1/83.4/76.0	79.6	85.6	81.9	90.1/88.8/79.2	84.6	79.7	84.5
UARL[167]	65.2/59.4/47.4	66.2	67.8	61.2	85.3/83.5/77.8	78.2	84.1	81.8	92.7/91.5/82.8	88.1	79.6	86.9
DPRN[279]	75.6/70.5/57.6	75.1	71.7	70.1	87.8/86.3/79.4	82.0	87.2	84.5	92.9/92.0/82.9	90.9	**82.0**	88.1
TUT[154]	76.2/71.9/60.0	73.7	76.0	71.6	89.3/88.3/81.7	84.0	87.2	86.1	89.0/86.4/73.3	84.1	76.1	81.8
CETNet[155]	79.3/74.3/61.9	77.8	74.9	73.6	87.6/86.5/80.1	81.7	86.9	84.6	91.8/91.2/81.3	87.9	80.3	86.5
SEDT[280]	-/-/-	-	-	-	89.9/88.7/81.1	84.7	86.5	86.2	93.7/**92.4**/84.0	91.3	81.3	**88.5**
TCTr[153]	76.6/71.1/58.5	76.1	77.5	72.0	87.5/86.1/80.2	83.4	86.6	84.8	91.3/90.1/80.0	87.9	81.1	86.1
DTL[281]	78.8/74.5/62.9	77.7	75.8	73.9	87.1/85.7/78.5	80.5	86.9	83.7	-/-/-	-	-	-
UVAST[271]	76.9/71.5/58.0	77.1	69.7	70.6	89.1/87.6/81.7	83.9	87.4	85.9	92.7/91.3/81.0	**92.1**	80.2	87.5
BrPrompt[157]	-/-/-	-	-	-	89.2/87.8/81.3	83.8	88.1	86.0	**94.1**/92.0/83.0	91.6	81.2	88.4
TST[282]	77.5/72.3/59.5	76.7	73.7	71.9	87.9/86.6/80.5	82.7	86.6	84.9	91.4/90.2/82.1	86.6	80.3	86.1
FAMMSDTN[156]	78.5/72.9/60.2	77.5	74.8	72.8	86.2/84.4/77.9	79.9	86.4	83.0	91.6/90.9/80.9	88.3	80.7	86.5
DiffAct[159]	80.3/75.9/64.6	78.4	76.4	75.1	90.1/89.2/83.7	85.0	88.9	87.4	92.5/91.5/84.7	89.6	80.3	87.7
Ours	**81.1/76.8/65.4**	**79.3**	**76.9**	**75.9**	**91.8/90.7/85.7**	**87.6**	**89.9**	**89.1**	93.3/92.3/**85.7**	89.9	81.5	**88.5**

表5-12. *k*M-Att 方法与 DiffAct 模型的性能对比

数据集	模型	F1@{10,25,50}	Edit	Acc	Avg
Breakfast	DiffAct[159]	80.3/75.9/64.6	78.4	76.4	75.1
	Ours	**81.1/76.8/65.4**	**79.3**	**76.9**	**75.9**
	Δ	+0.8/+0.9/+0.9	+0.9	+0.5	+0.8
50Salads	DiffAct[159]	90.1/89.2/83.7	85.0	88.9	87.4
	Ours	**91.8/90.7/85.7**	**87.6**	**89.9**	**89.1**
	Δ	+1.7/+0.9/+2.0	+2.6	+1.0	+1.7
GTEA	DiffAct[159]	92.5/91.5/84.7	89.6	80.3	87.7
	Ours	**93.3/92.3/85.7**	**89.9**	**81.5**	**88.5**
	Δ	+0.8/+0.8/+1.0	+0.3	+1.2	+0.8

表5-15. ThoSet 数据集上的模型性能对比实验结果

模型	F1@{10,25,50}	Edit	Acc	Avg
ASFormer[152]	93.20 / 92.78 / 91.35	90.94	92.54	92.16
Ours	95.11 / 94.75 / 93.56	94.13	92.67	94.04
Δ	+1.91 / +1.97 / +2.21	+3.19	+0.13	+1.88
DiffAct[159]	94.79 / 94.51 / 93.29	93.83	91.66	93.62
Ours	95.11 / 94.75 / 93.56	94.13	92.67	94.04
Δ	+0.32 / +0.24 / +0.27	+0.30	+1.01	+0.42

6.7 实验结果-时序分割性能对比

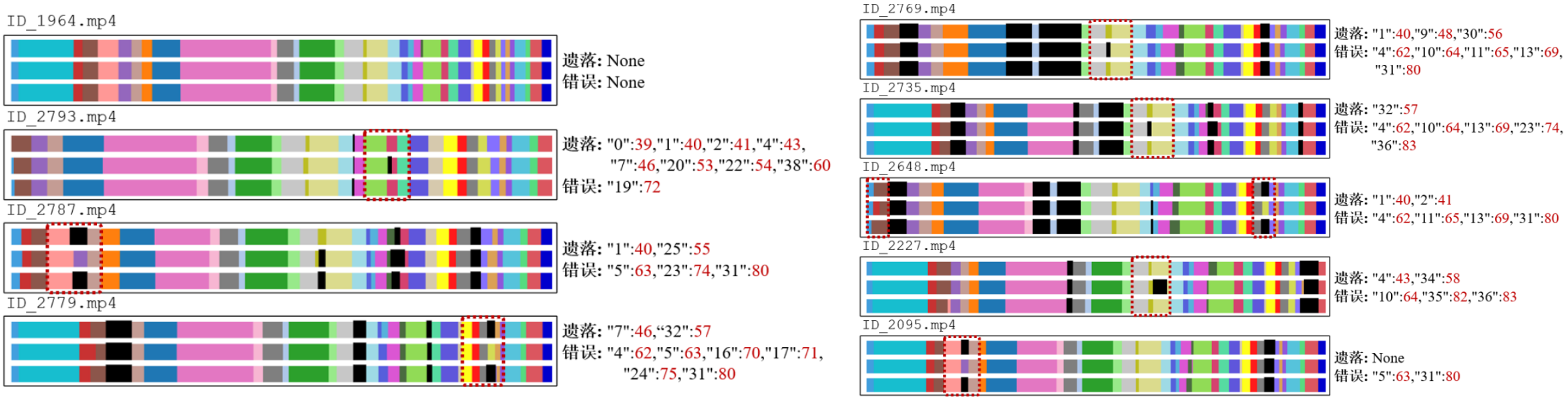
- **ThoSet数据集时序分割&合规性检测结果**

- **与DiffAct模型对比**：本文提出的*k*M-Att模型能够生成**更精准的时序分割结果**；
- **行为合规性检测结果**：相较于Transformer模型，本文所提出的方法能够生成更准确的**遗漏行为**和**错误行为**识别结果。

表5-17. ThoSet 数据集上的行为合规性检测结果

模型	错误行为识别		遗漏行为识别	
	mAP	**mmit mAP**	**mAP**	**mmit mAP**
Transformer[172]	88.52	92.95	90.47	94.09
Ours	89.23	93.64	92.61	95.86

图5-25. ThoSet数据集部分序列可视化结果



6.7 实验结果-消融实验

- **测试数据集**：Breakfast、50Salads、GTEA、ThoSet；
- **注意力机制消融结果**：*k*-means 聚类注意力机制与局部注意力机制**具有明显的互补特性**；
- **非锁步跳跃去噪步数影响**：不同的中心数量会造成性能的波动，但并不显著，**k=64可取得最优分割结果。**

表5-18. 公开数据集中 *k*-means 聚类注意力与局部注意力机制消融实验

数据集	模型	F1@{10,25,50}	Edit	Acc	Avg
Breakfast	*w/o k*-means Att	80.23 / 75.87 / 64.11	78.62	76.79	75.12
	w/o Local Att	80.83 / 76.38 / 65.16	79.02	76.26	75.53
	All	**81.06 / 76.80 / 65.36**	**79.28**	**76.91**	**75.88**
50Salads	*w/o k*-means Att	91.25 / 90.52 / 85.66	86.12	89.63	88.64
	w/o Local Att	91.61 / 90.61 / **85.97**	87.01	89.41	88.92
	All	**91.79 / 90.67** / 85.71	**87.57**	**89.86**	**89.12**
GTEA	*w/o k*-means Att	93.19 / 92.06 / 85.29	90.48	81.56	88.52
	w/o Local Att	93.17 / 91.31 / 84.18	**90.89**	**81.89**	88.29
	All	**93.28 / 92.33 / 85.70**	89.87	81.48	**88.53**

表5-19. ThoSet 中 *k*-means 聚类注意力与局部注意力机制消融实验

数据集	模型	F1@{10,25,50}	Edit	Acc	Avg
ThoSet	*w/o k*-means Att	94.92 / 94.49 / 92.91	93.70	91.88	93.58
	w/o Local Att	94.01 / 93.64 / 92.37	92.92	91.75	92.94
	All	**95.00 / 94.57 / 93.42**	**93.89**	**92.82**	**93.94**

表5-20. 公开数据集中 *k*M-Att 模块的聚类数量影响

数据集	#*k*	F1@{10,25,50}	Edit	Acc	Avg
Breakfast	128	**81.09** / 76.54 / 65.60	79.13	76.84	75.84
	64	81.06 / **76.80** / **65.63**	79.28	**76.91**	**75.88**
	32	80.94 / 76.49 / 65.22	**79.37**	76.71	76.75
	16	80.86 / 76.52 / 65.38	79.19	76.75	75.74
50Salads	128	91.53 / **91.12** / 85.68	86.08	89.25	88.73
	64	**91.79** / 90.67 / **85.71**	**87.57**	**89.86**	**89.12**
	32	90.89 / 90.19 / 84.97	85.86	89.58	88.30
	16	91.26 / 90.16 / 84.58	86.38	89.36	88.35
GTEA	128	91.36 / 90.42 / 84.41	88.11	81.27	87.11
	64	**93.28 / 92.33 / 85.70**	89.87	**81.48**	**88.53**
	32	92.59 / 91.85 / 83.95	89.41	81.24	87.81
	16	93.05 / 92.11 / 84.79	**90.30**	80.51	88.15

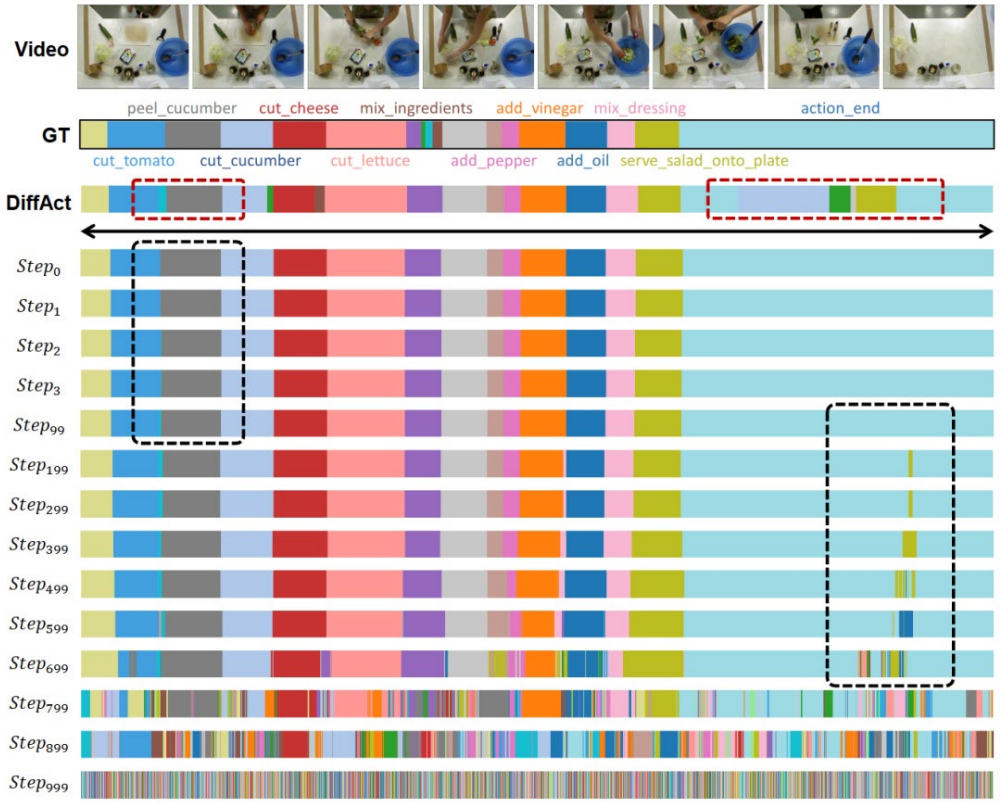
表5-21. ThoSet 中 *k*M-Att 模块的聚类数量影响

数据集	#*k*	F1@{10,25,50}	Edit	Acc	Avg
ThoSet	128	94.74 / 94.36 / 93.23	93.56	92.58	93.70
	64	**95.11 / 94.75 / 93.56**	**94.13**	**92.67**	**94.04**
	32	94.94 / 94.74 / 93.48	94.07	91.84	93.81
	16	94.93 / 94.60 / 93.26	93.66	92.51	93.79

6.7 实验结果-可视化结果

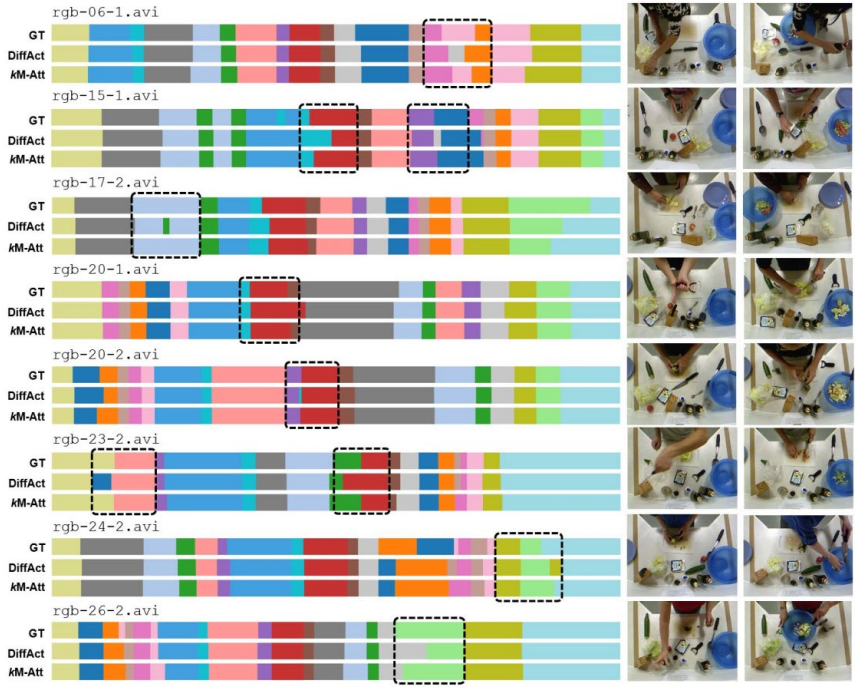
- **可视化数据集**：Breakfast、50Salads、GTEA；
- **非锁步跳跃去噪机制**：随着迭代次数的增加，**分割结果逐渐精准**；
- **时序分割结果可视化**：相较于SOTA方法DiffAct，本文所提出的模型**能够生成更加精准的时序分割结果**。

50Salads数据集部分序列可视化结果

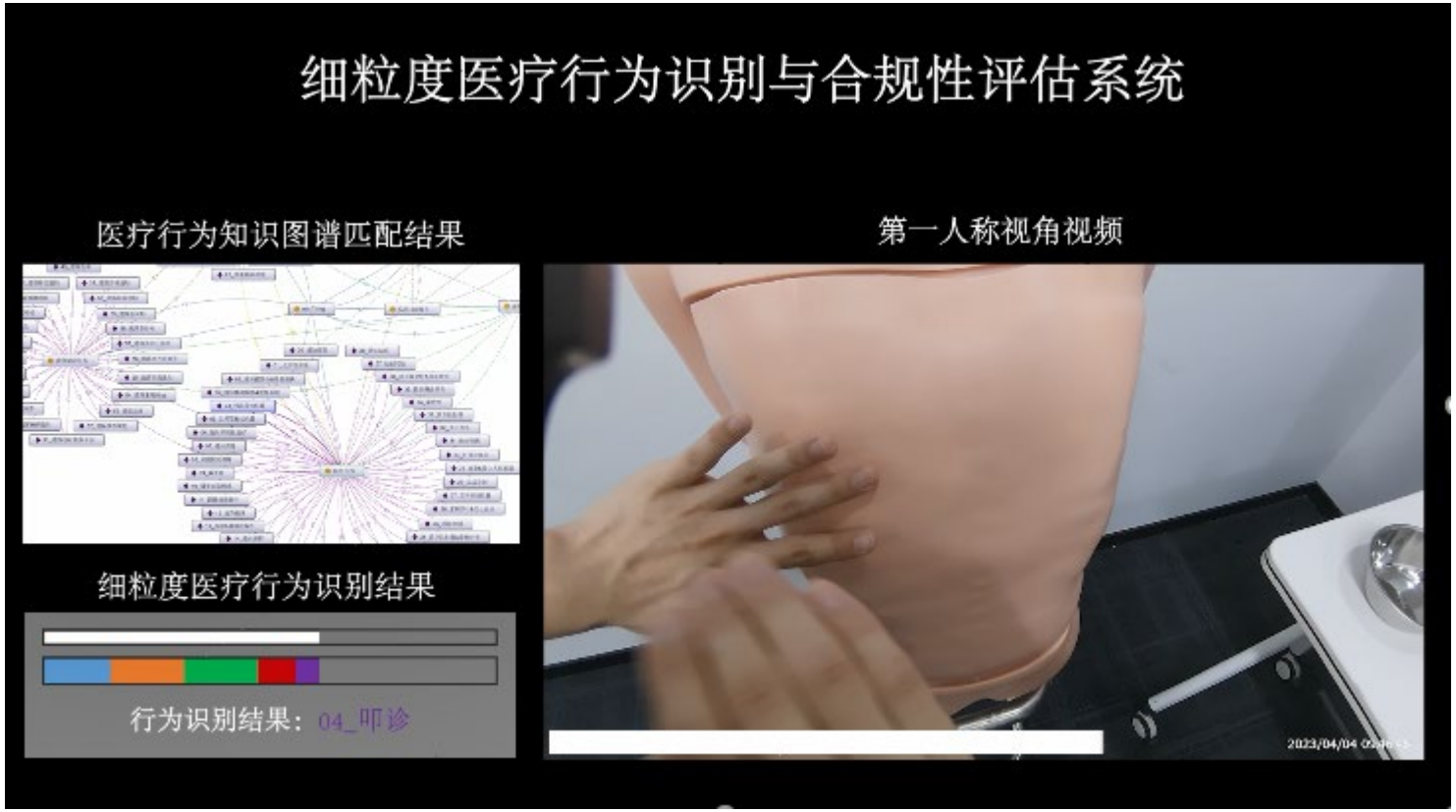


Breakfast 数据集部分序列可视化结果

50Salads数据集部分序列可视化结果



6.8 系统效果展示



胸腔穿刺术时序行为细化表

序号	流程划分	细分项	遗漏错误行为	单流程错误行为类
1	通知胸穿	0_通知穿刺	39_遗漏通知	
2	洗手	1_七步洗手法	40_遗漏洗手	
3	摆体位	2_通知摆体位&前扶	41_遗漏通知	
4	暴露胸廓	3_通知暴露胸廓&脱衣	42_遗漏通知	
5	叩诊	4_叩诊寻找位置	43_遗漏叩诊	61_叩诊左手错误 62_叩诊单纯右手
		5_记号笔标记位置	44_遗漏标记点	63_标记点位置错误
6	检查穿刺包	6_通知穿刺包良好	45_遗漏检查	
7	消毒	7_通知消毒	46_遗漏通知	
		8_卵圆钳夹棉球	47_遗漏卵圆钳夹取	
		9_戴手套	48_遗漏手套	
		10_镊子夹取棉球		64_无手套镊子夹取
		11_螺旋消毒操作		65_无手套消毒 66_面积过小 67_转圈动作逆序
		12_废弃棉球		68_废弃到白盘
8	铺洞巾	13_拾取&铺洞巾操作	49_遗漏洞巾操作	69_无手套洞巾
9	核对+利多卡因抽取	14_通知麻醉	50_遗漏通知	
		15_核对利多卡因	51_遗漏核对	
		16_拾取小注射器		70_拿大注射器
		17_抽取利多卡因		71_未抽到
10	麻醉	18_麻醉前拿起纱布	52_遗漏纱布拿起	
		19_左手控制穿刺位置		72_未控制麻醉位置
		20_斜向出丘疹	53_遗漏丘疹	
		21_渐进式麻醉		73_未渐进麻醉
		22_末端回抽	54_遗漏回抽	
		23_纱布覆盖&拔针		74_纱布未覆盖点
		24_废弃纱布		75_未废弃纱布
11	穿刺	25_通知穿刺	55_遗漏通知	
		26_胸穿针&关止血夹		76_未关止血夹
		27_左手控制位置		77_未定位穿刺点
		28_完成穿刺		78_插入太浅
		29_接大注射器		79_选错注射器
		30_打开止血夹	56_遗漏打开止血夹	
		31_抽出积液		80_没有抽出积液 81_单手抽取
		32_关止血夹	57_遗漏关闭止血夹	
		33_放下注射器		
		34_拿纱布	58_遗漏纱布	
		35_拔针覆盖纱布		82_直接拔针
		36_废弃纱布		83_未废弃纱布
12	术后	37_贴创可贴	59_遗漏创可贴步骤	
		38_通知结束	60_遗漏结束通知	

目 录

7.1 研究内容总结

细粒度医疗行为识别与技能评估技术研究

研究内容

- **第二章：**在**医疗技能评估任务**中设计了**高效且有效的管道自注意力特征增强策略**，提出TSA-Net框架。
- **第三章：**提出了细粒度**复合错误行为识别的任务范式**，在构建的CPR-Coach数据集基础上提出了**ImagineNet算法**。
- **第四章：**将**多模态预训练框架和提示词工程**引入到医疗技能评估领域，提出了**CPR-CLIP框架**，并开展了随机对照试验。
- **第五章：**创建了首个**时序医疗行为知识图谱**，并基于此构建了细粒度时序行为分析**数据集ThoSet**，在扩散时序行为分割模型中提出了特征增强**模块kM-Att**，并提出了**行为合规性检测算法**。

2个数据集贡献

- **CPR-Coach：**心肺复苏场景中复合错误行为识别数据集
- **ThoSet：**胸腔穿刺术中细粒度时序医疗行为分析数据集

1个知识图谱贡献

- 依据教材构建学界内首个时序医疗行为指示图谱

4个医疗技能评估算法贡献

- **TSA-Net：**基于管道自注意力机制的行为质量评估算法
- **ImagineNet：**基于特征组合机制的复合错误行为识别算法
- **CPR-CLIP：**基于多模态预训练的复合错误识别算法
- ***k*M-Att：**基于时序聚类注意力机制的扩散时序行为分析算法

1个随机对照实验贡献

- 在CPR按压错误行为识别任务中设计随机对照实验，对CPR-CLIP框架对辅助技能评估的有效性进行了验证。

7.2 本文创新点总结

- 本文首次将单目标跟踪器引入到行为质量评估任务中，提出了**基于时空管道自注意力机制的TSA模块和TSA-Net框架**。TSA模块能够依据跟踪框结果对视频特征进行选择性增强。在医疗与体育场景中的多个技能评估数据集上的实验结果证实，TSA-Net以**更少的计算开销达到了更优良的行为质量评估性能。**

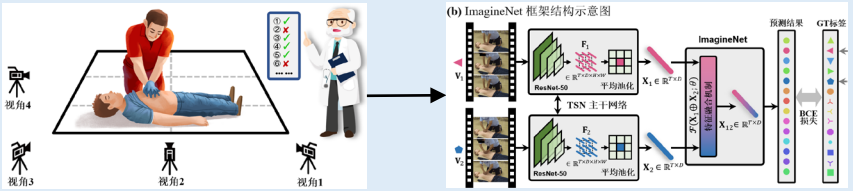
- 本文首次提出了复合错误行为识别任务范式，并将CPR胸外按压行为设定为研究对象，**构建了首个支持细粒度错误辨识任务的数据集CPR-Coach**。针对“单类训练，多类测试”的监督信息受限条件，本文提出了**基于特征组合训练机制的ImagineNet框架**，并通过充分的实验证实了此框架的有效性。

- 本文首次**将多模态预训练框架和提示词工程方法引入到复合错误行为识别任务中**，并提出了支持语言检索与批量评估功能的多模态医疗技能评估框架CPR-CLIP，充分的**模型性能对比和随机对照试验结果证实了该框架的有效性。**

- 本文首次提出了一套完整的时序医疗行为分析方法，在设计并构建**时序医疗行为知识图谱**的基础上，以胸腔穿刺术为研究对象构建了**同时支持时序行为分割与分析任务、具有高细粒度行为标签的ThoSet数据集**。提出了**基于时序聚类注意力机制的kM-Att模块**并将其应用在扩散时序行为分割模型中。基于高质量时序行为分割结果，基于DTW算法设计了医疗行为合规性评估算法。

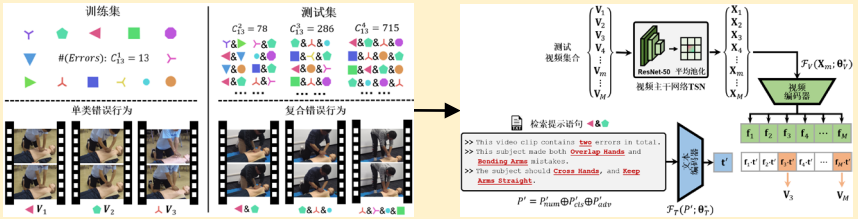
基于管道自注意力机制的行为质量评估算法TSA-Net

基于特征组合机制的CPR复合错误行为识别算法ImagineNet



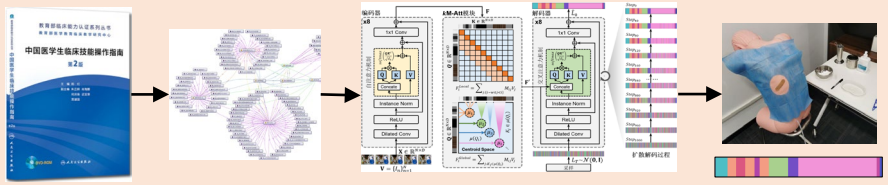
CPR-Coach数据集　　ImagineNet算法

基于多模态预训练机制的复合错误行为识别算法CPR-CLIP



复合错误识别任务　　对比预训练框架

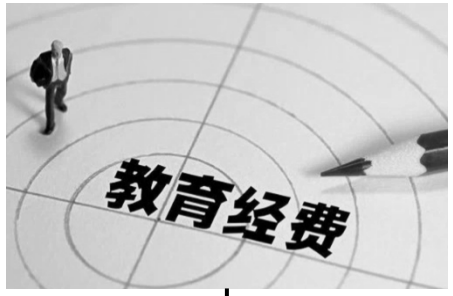
基于时序聚类注意力机制的扩散时序行为分析算法*k*M-Att



时序医疗行为KG　　*k*M-Att扩散分割算法　　时序分割结果

7.2 本文研究价值

医护比指标	2020~2025目标
每千人口执业医师数	2.9➔3.2
每千人口注册护士数	3.34➔3.8
医护比	1:1.15➔1:1.20
卫生人员比	1:1.48➔1:1.62
执业医师缺口数量	43.29万人
护士缺口数量	66.37万人



庞大的医护数量缺口

较高的培养质量要求

居高不下的培养成本

细粒度医疗行为识别与技能评估技术研究

- 基于管道自注意力机制的行为质量评估算法TSA-Net
- 基于特征组合机制的复合错误行为识别算法ImagineNet
- 基于多模态预训练的复合错误识别算法CPR-CLIP
- 基于时序聚类注意力机制的扩散时序行为分析算法*k*M-Att

算法

数据集

任务范式

知识图谱

本研究有望为医疗技能培训与考核效率的提升、医务人员数量缺口的快速补充和医疗服务系统压力的缓解做出一定贡献。

7.3 研究展望

数据集构建方面

- **数据集体量与丰富度问题**：医疗技能评估数据集的体量和丰富度会对技能评估算法的性能和实用性产生重要影响。
- **医疗行为数据的自动生成**：通过虚拟现实技术实现医疗行为数据的自动生成，可以大幅降低数据集的构建成本。

医疗技能评估算法设计

- **基于多模态信息的医疗技能评估模型**：充分利用多个模态之间信息的互补性，在提升模型性能同时改善系统易用性。
- **超长操作序列的高效建模**：参照NLP任务中处理超长序列信息的经验，设计出高效且有效的长序列建模算法。
- **小样本学习在医疗技能评估模型中的应用**：引入小样本学习策略实现对现有医疗行为数据的充分利用。
- **无监督与半监督学习**：在海量的操作视频库中进行智能化负例挖掘，实现错误行为标签空间和数据集的自动构建。

评估系统的落地应用

- **系统的人机交互能力改善**：充分利用多模态学习领域中的先进框架，进一步改善技能评估系统的人机交互能力。
- **模型推理性能的提升与优化**：在医疗技能评估模型的落地应用过程中，模型的推理加速会为系统的响应速度提供保障。

攻读博士学位期间取得的学术成果

- **已发表论文：**

[1] **Shunli Wang**, Dingkang Yang, Peng Zhai, Qing Yu, Tao Suo, Zhan Sun, Ka Li, Lihua Zhang*. A Survey of Video-based Action Quality Assessment[C]. *In Proceedings of the International Conference on Networking Systems of AI* (INSAI 2021, EI会议综述) , 对应正文第一章

[2] **Shunli Wang**, Dingkang Yang, Peng Zhai, Chixiao Chen, Lihua Zhang*. TSA-Net: Tube Self-Attention Network for Action Quality Assessment[C]. *In Proceedings of the ACM International Conference on Multimedia* (ACM MM 2021, CCF-A), 对应正文第二章

[3] **Shunli Wang**, Shuaibing Wang, Dingkang Yang, Mingcheng Li, Haopeng Kuang, Xiao Zhao, Liuzhen Su, Peng Zhai, Lihua Zhang*. CPR-Coach: Recognizing Composite Error Actions based on Single-class Training[C]. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (CVPR 2024, CCF-A), 对应正文第三章

[4] **Shunli Wang**, Dingkang Yang, Peng Zhai, Lihua Zhang*. CPR-CLIP: Multimodal Pre-training for Composite Error Recognition in CPR Training[J]. *IEEE Signal Processing Letters* (IEEE SPL 2023, SCI二区, IF=3.9), 对应正文第四章

[5] **Shunli Wang**, Shuaibing Wang, Bo Jiao, Dingkang Yang, Liuzhen Su, Peng Zhai, Chixiao Chen, Lihua Zhang*. CA-SpaceNet: Counterfactual Analysis for 6D Pose Estimation in Space[C]. *In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems* (IROS 2022, CCF-C)

- **竞赛获奖与Demo收录：**

[1] ACM-MM 2021 3rd VRU Challenge 国际视频关系理解挑战赛 二等奖

[2] MICCAI 2022 P2ILF Challenge 国际医学图像分割团队挑战赛 排名6/55

[3] 2021年“华为杯”第18届中国研究生数学建模竞赛, 三等奖

[4] ICCV 2023 Demo: Composite Error Action Recognition System for Cardio-pulmonary Resuscitation (CPR), 第一作者

[5] ICCV 2023 Demo: CA-SpaceNet: Counterfactual Analysis for 6D Pose Estimation in Space, 第一作者

- **投稿中论文：**

[1] **Shunli Wang**, Shuaibing Wang, Dingkang Yang, Yaxin Xu, Mingcheng Li, Ziyun Qian, Haopeng Kuang, Xiao Zhao, Peng Zhai, Lihua Zhang*. Sequential k-Means Clustering Attention for Diffusion Video Action Segmentation[C]. *In Proceedings of the Conference on Neural Information Processing Systems* (NeurIPS 2024, CCF-A) , 对应正文第五章

攻读博士学位期间取得的学术成果

- **实审中发明专利：**
  [1] 一种基于深度学习的运动员行为质量评估方法, 202111193385.4
  [2] 一种太空目标6D位姿估计系统, 202211159308.1
  [3] 一种基于CNN和迁移学习的心脏病发作检测系统, 202210948571.2
  [4] 一种基于情境感知的多模态情感识别方法和系统, 202111080047.x
  [5] 一种用于陪伴机器人的多模态情感识别方法和系统, 202111079583.8
  [6] 一种基于增强现实的儿科气管插管培训系统及方法, 202211511923.4

- **软件著作权：**
  [1] 基于机器学习的表情识别系统v1.0, 2020SR1661701, 排名第二
  [2] 基于深度学习的行为识别系统v1.0, 2020SR1661702, 排名第二

- **参与科研项目：**
  [1] 科技创新2030—“新一代人工智能”重大项目, 标准化儿童患者模型关键技术与应用 (项目编号: 2021ZD0113500)
  [2] 上海市人工智能科技重大专项, 人工智能前沿基础理论与关键技术 (项目编号: 2021SHZDZX0103)

感谢张立华教授的辛勤指导
感谢各位专家在百忙之中出席答辩会
请各位专家批评指正！

细粒度医疗行为识别与技能评估技术研究

Research of Fine-grained Medical Action Recognition and Skill Assessment Technologies

答辩人：王顺利 (19级本科直博生)

专　业：计算机应用技术

导　师：张立华 教授

日　期：2024年5月25日