文献阅读报告

文献阅读要求：

1. 请先明确报考方向，并根据报考方向从相应的论文列表中选择一篇论文进行精读。
2. 请根据自己的理解、用中文撰写阅读报告，避免直译原文。
3. 报告不少于1500字。

**论文列表：**

|  |  |
| --- | --- |
| 方向编号：1 | 系统软件（泛在操作系统、数联网系统软件等） |
| 1. FLASH: Heterogeneity-Aware Federated Learning at Scale. IEEE TMC 2024.  2. SoCFlow: Efficient and Scalable DNN Training on SoC-Clustered Edge Servers. ASPLOS 2024.  3. Battery-Aware Energy Optimization for Satellite Edge Computing. IEEE TSC 2024.  4. Adonis: Practical and Efficient Control Flow Recovery through OS-level Traces ACM TOSEM 2024.  5. FaaSLight: General Application-level Cold-start Latency Optimization for Function-as-a-Service in Serverless Computing. ACM TOSEM 2023. | |
| 方向编号：2 | 高可信软件（软件与系统安全、区块链与隐私计算等） |
| 1. REM:Resource-Efficient Mining for Blockchains, USENIX Security, 2017.  2. Secure Keyword Search and Data Sharing Mechanism for Cloud Computing, TDSC, 2021.  3. SegScope: Probing Fine-grained Interrupts via Architectural Footprints, HPCA 2024.  4. T-Counter: Trustworthy and Efficient CPU Resource Measurement Using SGX in the Cloud, TDSC 2023.  5. Text-to-Image Diffusion Models can be Easily Backdoored through Multimodal Data Poisoning, ACM MM 2023. | |
| 方向编号：3 | 领域智能软件（大数据机器学习、分布式智能运维等） |
| 1. Knowledge Graph-Based Behavior Denoising and Preference Learning for Sequential Recommendation, TKDE 2024.  2. Hilogx: noise-aware log-based anomaly detection with human feedback, VLDB Journal 2024.  3. Enhancing Job Recommendation through LLM-Based Generative Adversarial Networks, AAAI 2024.  4. MetaLog: Generalizable Cross-System Anomaly Detection from Logs with Meta-Learning, ICSE 2024.  5. Generalized Ambiguity Decomposition for Ranking Ensemble Learning, JMLR 2022. | |
| 方向编号：4 | 领域智能软件（多模态知识计算、程序分析与理解等） |
| 1. M3care: Learning with missing modalities in multimodal healthcare data. KDD 2022.  2. CoderUJB: An Executable and Unified Java Benchmark for Practical Programming Scenarios. ISSTA 2024.  3. KIEval: A Knowledge-grounded Interactive Evaluation Framework for Large Language Models. ACL 2024.  4. Vision Lanauge Pre-training by Contrastive Learning with Cross-Modal Similarity Regulation. ACL 2023.  5. Low-Resources Project-Specific Code Summarization. ASE 2022. | |
| 方向编号：5 | 领域智能软件 (智能计算与感知等) |
| 1. HEAL: Performance Troubleshooting Deep inside Data Center Hosts. SIGMETRICS 2024.  2. Pattern-Aware Transformer: Hierarchical Pattern Propagation in Sequential Medical Images, TMI 2024.  3. Placement Matters: Understanding the Effects of Device Placement for WiFi Sensing, UbiComp-IMWUT 2022.  4. Neural Parametric Mixtures for Path Guiding, SIGGRAPH 2023.  5. Incrementally Updateable Honey Password Vaults, USENIX Security 2021. | |

|  |  |
| --- | --- |
| 姓名：王文 | 本科学校：北京交通大学 |
| 邮箱：[wangwen.great@gmail.com](mailto:wangwen.great@gmail.com) | 联系电话：13808437449 |
| 报考方向（请打√）：  □1、系统软件（泛在操作系统、数联网系统软件等）  □2、高可信软件（软件与系统安全、区块链与隐私计算等）  □3、领域智能软件（大数据机器学习、分布式智能运维等）  □4、领域智能软件（多模态知识计算、程序分析与理解等）  ☑5、领域智能软件 (智能计算与感知等) | |
| 所选论文名称（须与本表中报考方向一致）：  Pattern-Aware Transformer: Hierarchical Pattern Propagation in Sequential Medical Images, TMI 2024. | |

1. 论文研究背景、动机与主要贡献
2. 论文研究背景：

在医学成像任务中，有效地挖掘序列图像之间的上下文信息并对其进行联合建模具有重要意义。这是因为序列图像相比单帧图像包含了更丰富的上下文信息，例如物体的运动模式和表示的一致性，为检测和分割复杂目标提供了有价值的线索。

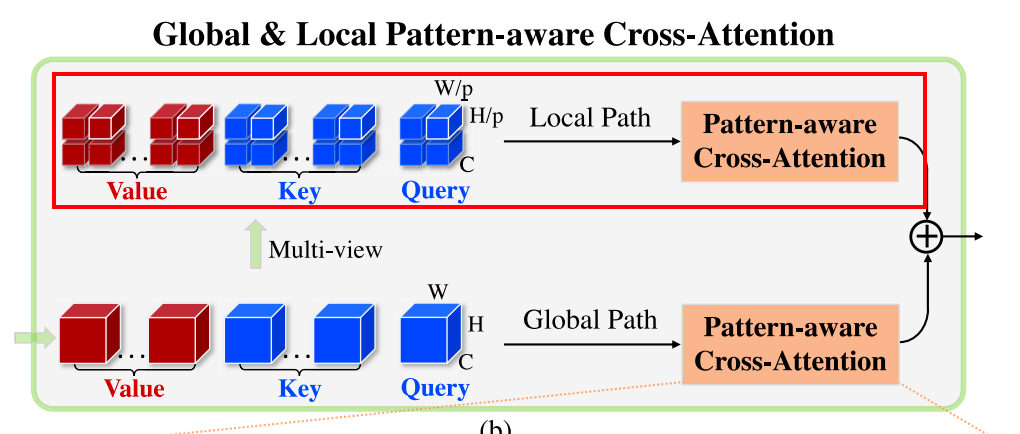
然而，目前处理医学成像任务的方法大多基于全局仿射变换假设，使用逐点token编码的注意力机制来获取序列图像之间的特征关联。这些方法存在两个主要缺点。首先，它们缺乏全局信息，例如图像结构中的位置信息，这在具有解剖空间先验的医学图像中尤为不利。其次，由于使用逐点token编码，这些方法将不同视觉模式（通道）的信息作为一个整体进行处理，忽视了不同通道特征的独特性。

1. 动机：

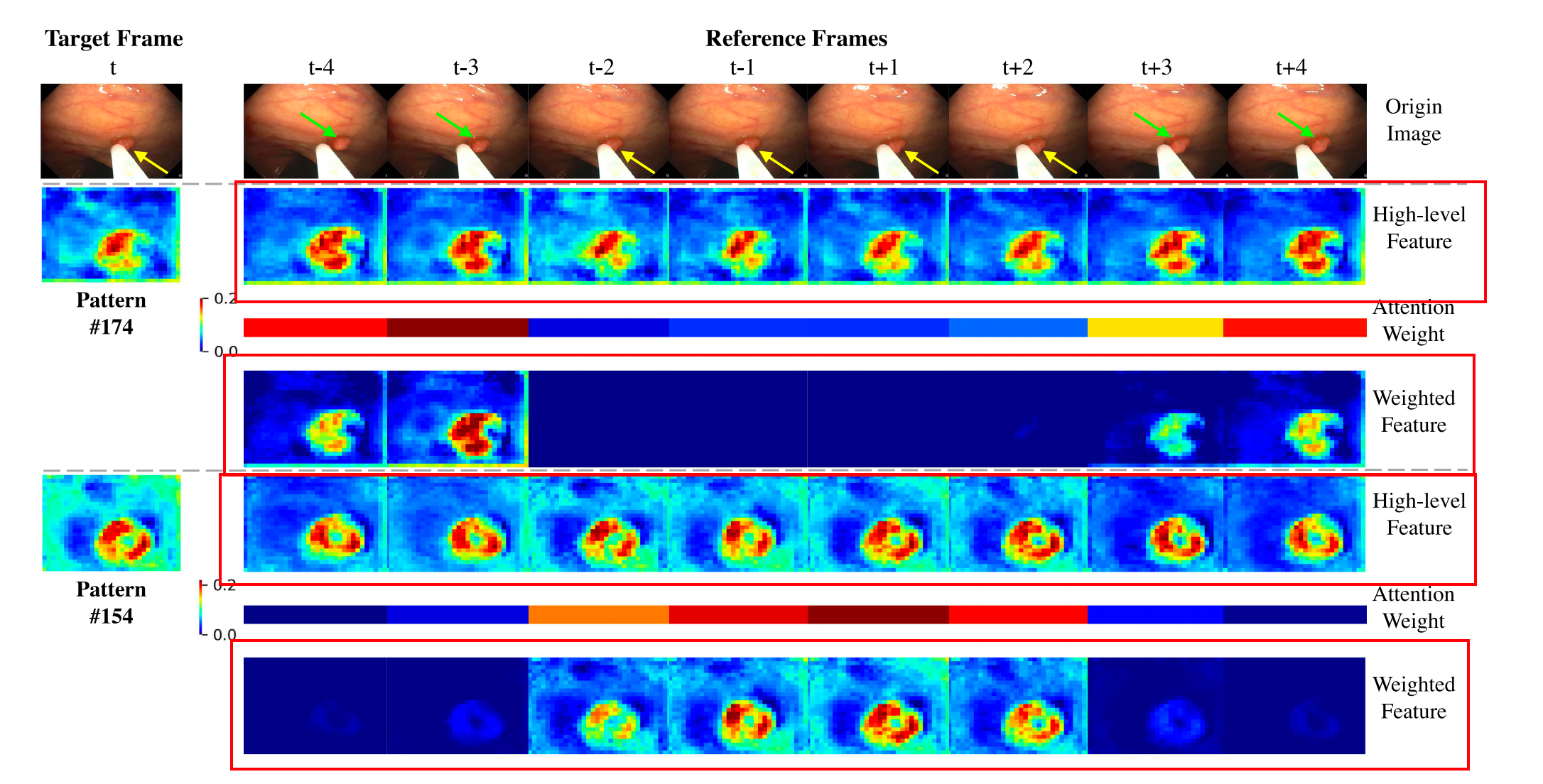
存在大量的前人工作，将CNN和Transformer结合在一起使用，由于CNN可以有效的提取图像局部特征，而Transformer又擅长捕捉长期依赖关系。因此作者在具体任务中也将CNN和Transformer进行结合，从而达到特征即有局部性又有全局性，此外逐点的token编码，缺乏全局的位置信息。

同时经过作者发现（文中图二），不同的视觉模式实际编码了不同的语义信息，因此作者认为应该将不同视觉模式进行独立的处理，从而确保在不同视觉模式中灵活地关注不同的参考帧，从而设计了Pattern-Aware Transformer。

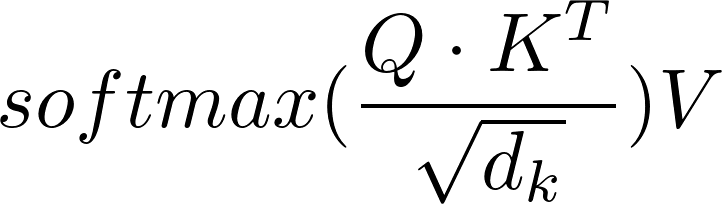
1. 主要贡献：
2. 在序列视觉任务领域中，**提出了一个全新的token编码角度**——pattern-aware tokenization，其与之前的逐点token编码不同，将不同的视觉模式分别单独进行处理，同时保留了局部和全局信息。
3. **提出了一个全新的Transformer——Pattern-Aware Transformer**，在Multi-head Self-Attention(MSA)模块其通过多头注意力机制以及位置编码，增强了目标图片内的信息交互同时保证了位置信息的不丢失；在PATrans模块通过**局部-全局双路径**的交叉注意力机制，实现了不同视觉模式下，目标图片和参考图片之间的交互和联系。
4. **封装了一个即插即用的PATrans模块**，可以在图片序列建模的不同下游任务中与主干网进行无缝结合，从而达到提高下游任务的准度，此篇论文主要展示了，其应用在医学的视频目标检测和医学3D语义分割。
5. 论文问题描述或定义
6. **缺乏全局信息**：逐点token编码策略在捕捉图像结构的全局位置信息方面存在不足，尤其在处理具有解剖空间先验的医学图像时，无法充分利用这些先验信息。
7. **忽视不同视觉模式的特性**：传统逐点token编码的方法将不同视觉模式（即不同通道的特征）视为整体进行处理，忽略了不同通道所代表的不同视觉模式及其独特的语义信息。
8. 论文提出的新思路、新理论、或新方法
9. 新思路：前人在处理序列图像建模时，通常采用逐点token编码的方法，但这种方法存在缺乏全局位置信息和忽视不同视觉模式特性的缺点。为此，本文提出了一种新思路，即在使用MSA提取目标特征的信息时，引入位置编码，同时在全局-局部PCA模块，将不同视觉模式特征的区域进行token编码，从而利用交叉注意力机制分别独立处理不同的视觉模式的特征。而且其中的Multi-view Pattern-aware Cross-Attention（MPCA）模块可以通过改变每个视觉模式的区域大小，也就是Patch的个数，例如下图中红框在进入PCA模块前将Q、K、V划分多个patch，分别对每个patch进行PCA操作最后再合并起来，与只有一个patch的全局PCA的结果进行结合，从而同时实现局部特征和全局特性的保留。



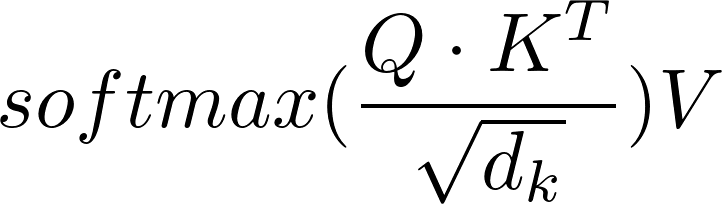
1. 新理论：通过实践发现，不同视觉模式下的特征信息，包含着不同的语义信息，再引入交叉注意力机制，从而可以使模型更加关注不同的参考帧，从下图红框中的高层特征权重以及注意力权重可视化可以看出，不同视觉模式所表示的语义信息以及注意力所关注的信息是不同的，从而单独处理不同视觉模式可以获得更多尺度的信息，提高下游任务的准确性。



1. 新方法：本文提出了全新的pattern-aware tokenization和Pattern-Aware Transformer（PATrans）。同时通过封装PATrans模块，实现了一个即插即用的组件，能够无缝集成到各种骨干网络中，适用于多种下游序列建模任务，例如视频目标检测以及3D语义分割。
2. 论文方法的理论分析或实验评估方法与效果
3. 理论分析：
   1. Pattern-Aware Transformer
      1. 多头自注意力模块

对于多头自注意力机制方面，首先会通过层归一化，然后在归一化后的结果加上位置编码得到/Users/wangwen/Library/Containers/com.kingsoft.wpsoffice.mac/Data/tmp/wpsoffice.aVySjkwpsoffice，再使用传统的逐点token化，将每个像素中所有通道作为一个token，从而得到矩阵wpsoffice，将其token化的结果作为Query、Key以及Value输入到多头自注意力机制中，从而使每个token都具有全局的信息。其中Query、Key以及Value的维度均为wpsoffice其中h为头的个数.通过多头自注意力机制的结果进行合并得到原维度wpsoffice，再与未进入层归一化的原始数据进行残差连接得到该模块的结果，将其作为Query输入到全局-局部PCA模块。

* + 1. 全局-局部PCA

将在多头自注意力模块得到的结果作为Query经过层归一化，与N个参考特征作为Key和Value进行全局和局部的PCA；而全局的PCA是指，将Query、Key以及Value每一层特征（不同的视觉模式特征）进行token化，分别得到/Users/wangwen/Library/Containers/com.kingsoft.wpsoffice.mac/Data/tmp/wpsoffice.ynuapvwpsoffice,再将所有特征图进行拉平成向量，开始做交叉注意力，此时Q来自目标特征，而K和V来自参考特征，从而获得目标特征与参考特征的全局关联；而局部的PCA，与全局的PCA的区别在于，会使用Multi-View的形式，将每个Query、Key和Value划分为个patch，然后对相应位置的patch分别做与全局PCA一样的处理，最后再将结果合并起来，从而得到目标特征与参考特征的局部关联。最后将全局PCA和局部PCA得到的结果进行合并，同样此处也会和前面一样做一个残差连接。

* + 1. 前向传播

将全局-局部PCA得到的结果，通过两层的全连接层，同时激活函数使用GELU；还有一个残差连接

1. 实验评估方法与效果
   1. 视频目标检测
      1. 数据集：CVCVideoClinicDB、ASU-Mayo Clinic Colonoscopy Video；同时训练集和测试集与SOTA方法的设置相同
      2. 测试指标：精准率、召回率以及F1分数
      3. 优化器：随机梯度下降
      4. 损失函数：分类头使用Focal Loss，回归头使用L1范式
      5. 方法：使用ResNet50作为主干网络，FCOS作为检测器，对ResNet50得到的多尺度特征输入到本文的PATrans架构中
      6. 效果：在两个数据集上F1分数均超过了SOTA，而且更加具有稳定性，相比FCOS单帧检测器，虽然其精准率高，但是其召回率非常低，也就是容易出现漏检的情况，但是其他的多帧检测器方法，则出现漏检的情况较少，这也另一方面证明了在图片序列建模时，图片序列之间有着大量的上下文信息可以帮助下游任务的完成
   2. 3D语义分割
      1. 数据集：Lung Tumor、Nasopharynx Tumor、Kidney Tumor；前两个数据集的训练集和测试集分割的方式采用的MICCAI 2019挑战的规则；而第三个数据集，则是采用SOTA方法中的规则。
      2. 测试指标：Dice相似系数
      3. 优化器：Adam
      4. 损失函数：交叉熵损失和tversky损失，其中tversky损失是为了惩罚那些难以分割的样本
      5. 方法：使用3D-UNet作为主干网络，在UNet每个skip connection时插入PATrans；训练时把病变面积最大的作为第一个切片和Mask，同时每个当前切片的参考图像只有两个，分别为第一个切片（即病变面积最大的）和当前切片的前一个切片，而且由于**初期**模型的不稳定，防止因错误mask信息的传递，导致模型无法正常训练，因此将所有预测的mask都替换成真实的mask；在推理过程中，默认的参考图像数量为10；
      6. 效果：作者主要对比了两类方法，一个是基于二维的CNN，另一个是与本文类似的基于Transformer；同时作为先进行了无输入mask的自动化实验，发现Dice基本在50%到64%之间，证明了该任务的困难而目前无法达到临床要求；然后作者有进行了加入mask的实验，发现PATrans的Dice超过其他方法，尤其对基于二维CNN的方法Mem3D；同时在一些challenging的样例中有着较好的鲁棒性，而AOP方法（基于Transformer）虽然分割的准确高，但是其出现了假阳性的情况，将一些不属于的地方进行了分割。
   3. 消融实验
      1. 对于全局-局部双路径的必要性；作者通过是否使用局部或全局路径的方式，证明了不论是全局路径还是局部路径对于性能的提高都是至关重要的
      2. 对于多头自注意力机制中的头的数量，作者进行实验发现当头为8时，模型的效果最好
      3. 对于MPCA模块patch的数量，作者发现对大尺度的低层次特征使用更多的patch数量，而对小尺度的高层次特征使用较少的patch数量，会使模型达到最佳，从而也验证了模型对形状大小的鲁棒性。
      4. 作者通过对比不同参考图像数量的实验，发现都可以超过基线，因此模型对N大小不是很敏感。
      5. 作者对比了传统逐点的token化和本文的pattern-aware 的token化，发现不论是局部上还是全局上，pattern-aware的token化的方法都会比逐点token化的方式好。
2. 总结

（包括分析论文的优缺点，本文带来的启示、后续研究思路等）

1. 优点：
   1. 本论文的创新点好，从目前方法缺点的角度分析，提出了全新的tokenization的方式以及设计了PATrans的架构；能够更好地保留全局和局部的信息。
   2. 模型泛化性好，在视频目标检测任务中，不易漏检同时保证了定位的准确性；在3D语义分割任务中，不论是在对于大的腹部肿瘤还是小的胸部肿瘤都可以有较好的分割效果，同时也没有出现假阳性的现象。
   3. 设计了一个即插即用的PATrans模块，其可以很好的与现有的主干网进行缝合。
   4. 性能优越，其在实验中证明了其模型的性能均稳定已超过SOTA方法。
2. 缺点
   1. 模型存在大量超参需要调整，例如MPCA模块中pacth的个数、MSA模块注意力头的数量以及参考图像的数量。
   2. 实验中涉及的数据集太少，缺少对该模型zero-shot或者few-shot性能的验证，从而无法判断模型对数据集的依赖性。
   3. 在3D语义分割任务中，文章没有对**提供一个MASK会比没有提供MASK有显著提高性能**这一下现象有解释。
3. 启示
   1. 对于图片序列建模时，挖掘序列的上下文信息，能够显著提高检测和分割的准确性。
   2. 针对不同的视觉模式特征，需要单独处理，从而获得更多尺度的信息。
4. 后续研究思路
   1. **集成模型**，利用PATrans模块提出一个统一的模型框架，可以统一训练多个任务的数据集，并且在对应的任务的性能都有较好的表现。
   2. 本文仅利用了视觉信息，未来的工作中我们计划**加入其他模态的信息**。目前，大量研究表明，多模态数据的性能通常优于单模态数据，有望进一步提升模型的效果。