“华为杯”第七届中国研究生人工智能

创新大赛华为赛题

**赛题一**

1. 题目名称

问答场景中复杂问题的多层级拆解

1. 题目描述

在LLM智能体的RAG问答场景中，能否准确理解用户问题的真实目标，对回答结果的准确性影响很大。AI phone、AI PC等智能化产品，聚焦于在端侧构建智能体，为用户带来智能化体验的同时，实现隐私数据在端侧闭环。但端侧受限于计算资源、存储空间，只能用小尺寸的模型，因此很难准确理解用户问题的真实语义，在RAG场景中，这意味着难以识别出所需的关键词或检索条件，影响最终回答的效果。为解决以上痛点，本题目期望基于小尺寸（参数量不超过10B）的LLM，搭配其他MB级别参数量的模型或算法，完成用户问题的语义理解和回答。

本题目将重点解决以下问题：

（1）Text-to-SQL，识别问题中的关键实体。如，用户问“上周收到的账单”，该任务需要识别出检索条件“上周”和关键词“账单”，以便于在检索时，用“账单”做一次检索，再用“上周”这一时间要素过滤出最相关的内容。

（2）多轮问答中的上下文指代。如，用户问“上周收到的账单”，得到LLM生成的回答后；用户再问“上个月的呢？”，此时需能理解到，用户想问的可能是“上个月收到的账单”；若用户再问“这月多花了多少钱”，此时需能理解到，用户可能需要比较两轮提问中搜到的账单。

（3）基于用户反馈的动态优化。如，用户对于LLM生成的回答不满意，能够根据用户的反馈，给出调整后的回答。

1. 具体要求

（1）参赛方案需包含以下功能：

1. 问题意图理解：理解用户问题的真实意图，识别出关键信息，可能涉及实体识别、问题改写等；
2. 上下文内容分析：理解用户的连续提问意图，以及其中的上下文指代关系，可能涉及通过提示词工程合理编排上下文信息；
3. 反馈优化：当用户对答案提出质疑时，系统应能回溯拆解路径并优化；

（2）参赛者需给出算法模型的可解释性文档，系统阐述方案选型依据，并附上数据/图片等其他可证明观点的论据；

（3）可利用现有真实数据样本进行数据增强等方式来保证模型的泛化能力，或结合其他开源数据集完成；

（4）竞赛得分部分根据选手提交的方案在测试集（包括公开测试集和非公开测试集）上的运行结果给出；本次参赛的最终成绩会综合使用拆解完整度（40%）、答案准确率（30%）、交互修正效率（20%）、系统响应延时（10%）；

（5）赛题总分由竞赛得分（60%）和专家评分（40%）两部分组成。专家评分使用华为内部的问答测试数据进行指标评价和验收，考察算法的泛化性；

（6）专家评分由评委组对选手所提交的方案的新颖性、合理性等进行打分。因此，参赛选手还需要提交模型代码（用于非公开测试集评估）、模型使用说明文件（用于报告模型方案以及模型在公开测试集上的结果）。鼓励使用小模型。

4. 参考数据集

https://www.aicrowd.com/challenges/meta-comprehensive-rag-benchmark-kdd-cup-2024

1. 参考文献

[1]Decomposed Prompting: A Modular Approach for Solving Complex Tasks

[2]Buffer of Thoughts: Thought-Augmented Reasoning with Large Language Models

[3]A Survey on Employing Large Language Models for Text-to-SQL Tasks

6. 咨询专家及联系邮箱

孙飞扬 - sunfeiyang2@huawei.com

李有福 - [youfu.li@huawei.com](mailto:youfu.li@huawei.com)

7.赛题互动交流答疑社区链接：

[https://www.chaspark.com/#/races/competitions/1135661275977601024](https://www.chaspark.com/" \l "/races/competitions/1135661275977601024)

**赛题二**

1. 题目名称

海量多模态文档的精准语义检索与关联分析

1. 题目描述

在金融、法律、医疗等领域，每日产生海量多模态文档（文本、表格、图像、PDF等），传统基于关键词的检索方式难以满足复杂语义需求，且存在以下痛点：  
（1）语义鸿沟：用户查询意图与文档深层语义匹配度低；  
（2）多模态割裂：跨文本、表格、图像的联合检索能力不足；（3）效率瓶颈：文档库语义索引构建延迟高，难以满足端侧用户体验；  
（4）长尾问题：冷门专业术语或小样本场景下检索精度骤降。

核心挑战：

（1）多模态语义融合：实现文本、表格、图像等异构数据的统一语义表征；

（2）小样本学习：通过少量标注数据构建高泛化能力的语义模型；

（3）高效语义构建：在保证精度的前提下，将单个文档的语义索引构建耗时控制在百毫秒级；

（4）动态增量更新：支持新文档实时入库后的索引快速重建。

3. 具体要求

（1）模型设计：

1. 构建端到端语义检索模型；
2. 需支持跨模态检索（如用文本查询匹配相关表格/图像）；
3. 鼓励使用知识图谱、增量学习等技术实现领域知识语义增强。

（2）性能要求：

1. 模型参数量参考值≤500MB，支持GPU/CPU混合部署；
2. 设计分层索引策略（如粗筛+精排两阶段架构）。

（3）参赛者需给出算法模型的可解释性文档，系统阐述方案选型依据，并附上数据/图片等其他可证明观点的论据：

1. 需输出检索结果的语义关联图谱，标注文档片段与查询语句的匹配权重；
2. 提供bad case分析报告，说明模型在特定场景（如缺少领域知识）中的改进方向。

（3）可利用现有真实数据样本进行数据增强等方式来保证模型的泛化能力，或结合其他开源数据集完成；

（4）竞赛得分部分根据选手提交的方案在测试集（包括公开测试集和非公开测试集）上的运行结果给出；本次参赛的最终成绩会综合使用平均检索精度（40%）、跨模态检索成功率（25%）、构建时间（20%）、模型效率（15%）；

|  |  |  |
| --- | --- | --- |
| **指标** | **权重** | **说明** |
| 平均检索精度（mAP@50） | 40% | 前50条结果的精度均值 |
| 跨模态检索成功率 | 25% | 文本→表格/图像的匹配准确率 |
| 构建时间 | 20% | 单篇文档索引构建时间 |
| 模型效率 | 15% | 内存占用与计算吞吐量 |

（5）赛题总分由竞赛得分（60%）和专家评分（40%）两部分组成。专家评分使用华为内部的测试数据进行指标评价和验收，考察算法的泛化性；

（6）专家评分由评委组对选手所提交的方案的新颖性、合理性等进行打分。因此，参赛选手还需要提交模型代码（用于非公开测试集评估）、模型使用说明文件（用于报告模型方案以及模型在公开测试集上的结果），文件内容包括：

1. 语义表征方案（如Embedding生成策略）
2. 小样本场景下的优化方法（如Prompt Tuning）
3. 参考数据集

<https://ai.google.com/research/NaturalQuestions/dataset>

https://www.wikidata.org/wiki/Wikidata:Main\_Page

1. 参考文献

[1]Enhancing Distractor Generation for Multiple-Choice Questions with Retrieval Augmented Pretraining and Knowledge Graph Integration

[2]History-Aware Conversational Dense Retrieval

[3]CF-TCIR: A Compositor-Free Framework for Hierarchical Text-Conditioned Image Retrieval

6. 咨询专家及联系邮箱

孙飞扬 - sunfeiyang2@huawei.com

李有福 - [youfu.li@huawei.com](mailto:youfu.li@huawei.com)

7. 赛题互动交流答疑社区链接：

[https://www.chaspark.com/#/races/competitions/1135664052174835712](https://www.chaspark.com/" \l "/races/competitions/1135664052174835712)

**赛题三**

1. 题目名称

视频联合去噪去马赛克算法

1. 题目描述

视频去噪是计算机视觉的关键任务之一，相比图像去噪，除了恢复噪声退化的图像内容，还需要保持视频内容的时域一致性。近年来，深度神经网络在图像和视频恢复方面取得了巨大成功。然而通常计算成本很高，限制了它们在资源受限的移动设备上的应用。

此外，业界流行的去噪方法大都是从 sRGB 图像出发，进一步提升其质量。由于非线性 ISP 转换，在 sRGB 域中对退化进行建模非常复杂，导致效果受限。尽管存在这个已知问题，但只有少数方法可以直接处理 RAW 图像。

为了解决上述问题，我们提出了一个在算力受限约束下，基于RAW视频图像输入，联合去噪去马赛克的竞赛。目标是获得与GT具有最高像素保真度的输出视频（客观、主观度量），同时模型满足算力和参数量约束。

## 

1. 具体要求

（1）任务：给定一组RAW视频序列，参赛者设计和实现AI算法模型，推理输出去噪和去马赛克后的RGB视频。

（2）数据集：

1. 训练数据集：基于可获得的开源RGB视频，自行收集数据、数据增强，参考文献[1]的流程，进行数据退化，退化参数由华为提供，获得退化RAW视频序列
2. 验证数据集：用于参赛者验证的视频集，包含退化RAW视频序列和RGB GT视频序列，可以计算PSNR。

（3）算力和参数量约束：以4K（3840x2160）图像为例，单帧处理算力不超过1T FLOPS，参数量不超过1M。

（4）参赛者需撰写相关文档，说明所提出算法/方案的有效性和先进性，进行数据集收集方法、数据集增强、模型设计思路等内容的详细解释。

4. 评分标准

（1）打榜成绩（计50%分数）。基于华为提供的测试数据集（不公开）进行PSNR验收评分，取Top10进行人工评委审核，若存在并列成绩情况，可多取候选者。

（2）PSNR：

（3）人工评委审核（计50%分数）：人工审核主要基于输出视频进行主观异常问题识别，每个异常问题扣1分，扣完为止。异常问题至少包含：鬼影、闪烁、紫边、伪彩、伪细节等

（4）最终成绩为打榜成绩与人工评委审核之和。

5. 参考文献

[1] T. Brooks, B. Mildenhall, T. Xue, J. Chen, D. Sharlet, and J. T. Barron, “Unprocessing Images for Learned Raw Denoising,” in *2019 IEEE/CVF Conference on Computer Vision*

6. 咨询专家及邮箱

moyong@huawei.com

7.赛题互动交流答疑社区链接：

https://www.chaspark.com/#/races/competitions/1135664058728919040

**赛题四**

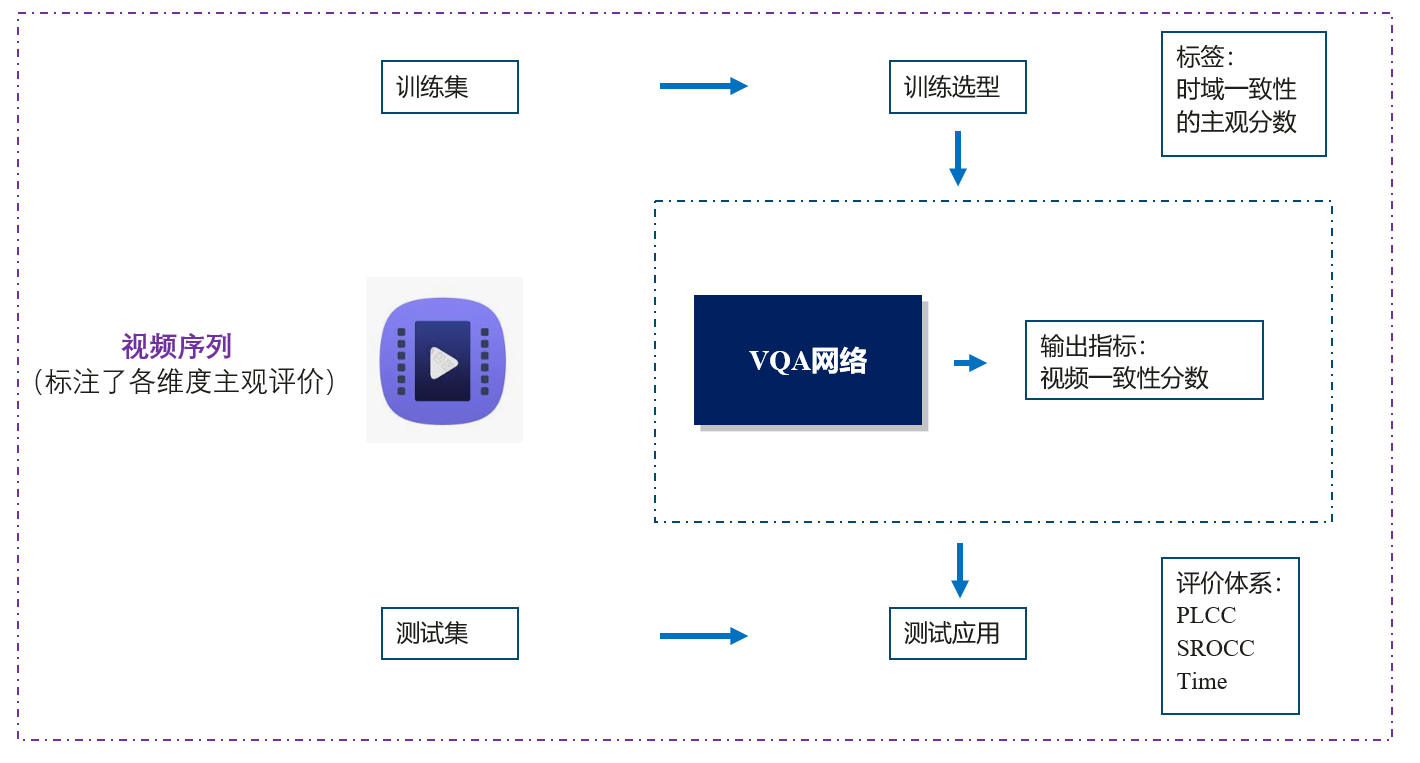
1. 题目名称

视频质量评估算法设计

1. 赛题背景

随着5G网络、超高清视频和流媒体技术的快速发展，视频已成为信息传递的核心载体。然而，在视频采集、压缩、传输及播放过程中，常因器件性能、网络波动、编解码失真等因素导致画质受损，例如卡顿、模糊、伪影、色彩失真等问题，直接影响用户体验。如何精准评估视频质量并针对性优化，成为提升视频服务竞争力的关键技术挑战。传统客观评价指标（如PSNR、SSIM）虽计算高效，但难以贴合人类主观感知；基于深度学习的方法虽在特定场景表现优异，但其泛化能力受限于训练数据分布，面对复杂真实场景时鲁棒性不足。为突破现有技术瓶颈，本赛题聚焦**复杂场景下的视频质量评价（VQA）算法研发**，要求参赛者构建能够评价视频时域一致性的评估模型。赛题将提供存在不同程度一致性问题的视频及其主观评价作为验证集，参赛方案的评价体系包括主客观一致性和计算效率。本次竞赛，旨在推动视频质量评价技术向更智能、更实用的方向发展，为手机录像、视频监控、XR等领域提供可靠的算法支持。

1. 题目描述



赛题示意框图

1. 赛题描述：

（1）赛题说明：选手接收出题方提供的视频集用于做VQA算法代码训练选型及各维度打分，最终PLCC及SROCC结果逼近1且耗时短为效果最优。正式比赛选手提供训练选型好的VQA算法，出题方提供比测试视频集（不对选手开放），评测选手算法的PLCC、SROCC和方案耗时。

（2）相关领域：视频处理，视觉评价，人工智能（AI）。

（3）数学根技术：多模态智能体表达问题。3.具体要求

5. 评分标准

判题指标

其中，分别表示每个场景客观和主观评价中的排序位置

判题指标

其中，分别是第i幅图像主观分数和客观分数

6. 咨询专家及邮箱

[moyong@huawei.com](mailto:moyong@huawei.com)

7. 赛题互动交流答疑社区链接：

[https://www.chaspark.com/#/races/competitions/1135667292120866816](https://www.chaspark.com/" \l "/races/competitions/1135667292120866816)

# 赛题五

1. 题目名称

低功耗设备上的实时视频超分辨率重建

1. 赛题背景

随着移动互联网和物联网的快速发展，越来越多的视频内容通过手机、平板、智能摄像头等低功耗设备进行采集、传输和播放。然而，受限于设备的硬件性能和能耗限制，这些设备采集的视频往往分辨率较低，在大屏幕上播放或进行细节分析时，图像模糊、信息丢失的问题严重影响用户体验和应用效果。因此，研究在低功耗设备上实现实时视频超分辨率重建技术具有重要的现实意义和应用价值。

1. 题目描述

参赛队伍需要设计并实现一种适用于低功耗设备（如麒麟SoC等）的实时视频超分辨率重建算法和系统。该系统应能够实时处理输入的低分辨率视频流（如4K），将其分辨率提升至更高水平（如8K），并在保证视频流畅度（帧率不低于60fps）的前提下，尽可能提高重建视频的图像质量，使重建后的视频在主观视觉效果和客观评价指标（如 PSNR、SSIM 、时域一致性等）上达到较好的表现。

1. 低功耗硬件限制：低功耗设备的计算资源（如 CPU、GPU 性能）和内存有限，如何设计轻量级的神经网络模型，使其在有限的硬件资源下高效运行，是需要解决的关键问题。
2. 实时性要求：视频是连续的图像序列，需要在极短的时间内完成每一帧的超分辨率处理，以保证视频的流畅播放。这对算法的计算速度和处理效率提出了极高的要求。
3. 图像质量保持：在根据时域、空域信息提升分辨率的过程中，要避免引入过多的噪声、伪影、时域闪烁等问题，保持图像的细节和纹理信息，提高重建视频的视觉质量和清晰度。
4. 评分标准
5. 数据提供：主办方将提供一定数量的低分辨率视频数据集，包括不同场景（如室内、室外、人物、风景等）和不同分辨率的视频片段，用于算法训练和测试。参赛队伍也可以自行收集额外的相关数据进行补充训练，但需确保数据来源合法合规。
6. 评估指标：采用客观评价指标（如 PSNR、SSIM）和主观评价指标（通过多人主观打分）相结合的方式对参赛作品进行评估。客观评价指标用于量化衡量重建视频与原始高分辨率视频之间的相似度和图像质量提升程度；主观评价指标则从人的视觉感受出发，评估重建视频的视觉效果和观看体验。
7. 咨询专家及邮箱

[fanwengui@huawei.com](mailto:fanwengui@huawei.com)

1. 赛题互动交流答疑社区链接：

[https://www.chaspark.com/#/races/competitions/1135702868594470912](https://www.chaspark.com/" \l "/races/competitions/1135702868594470912)

赛题六

1. 题目名称

大模型优化NPU算子代码

1. 题目描述

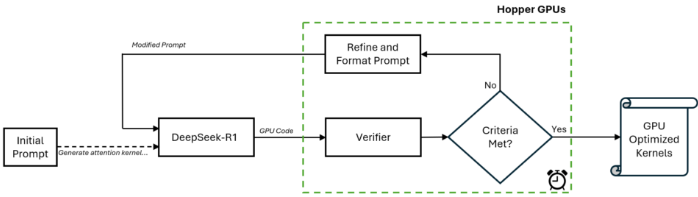
AI辅助代码生成不仅是一个前沿领域，更在快速成为一个不可阻挡的趋势。一个很让人震撼的数据，在大模型普及后，stack overflow的流量下降了50%有余。借助人工智能技术的高速发展，AI辅助代码生成作为一项具有巨大潜力的技术，有望彻底改变人们的软件开发模式，提高开发效率，降低成本，提升软件质量。

AI辅助代码生成的原理与大模型对话类似，通过对大量代码数据的学习与分析，实现开发者意图、自然语言、与代码之间的高精度映射；AI辅助代码生成工具能够精确解析开发者的编码需求，结合优秀的代码规范，并将其准确地转化为程序代码。

随着国产芯片的快速崛起，相关开源代码数据量快速增加，但高质量代码总量仍相对较少；主流辅助代码工具生成国产芯片代码，尤其生成NPU处理器Kernel代码的能力相对于通用代码仍有差距。

1. 具体要求

随着 AI 模型能力拓展，以及DeepSeek-R1等推理模型的落地，推理时间缩放 (inference-time scaling) 技术正让 AI能解决更复杂问题。

NPU有不同于GPGPU的硬件架构和指令集设计，其内核有专门适配AI加速计算单元和存储结构，计算效率极高。同时，NPU kernel代码也需采用不同于GPGPU的算法和优化技术，基于NPU特点设计特有的并行计算、数据复用等算法，提升计算速度、降低能耗。使其在处理大规模神经网络任务时，能快速完成复杂计算任务。对硬件资源精细化管理，合理分配计算资源，确保内核高效运行，减少资源冲突和浪费。

NVIDIA的研究院使用DeepSeek-R1模型在Kernel Bench下进行评测，不但成功生成了功能正确的GPU基础算子和高阶算子(准确率：98%)，而且性能提升了1.1~2.1x(相比基线算子)。

具体参考https://developer.nvidia.com/blog/automating-gpu-kernel-generation-with-deepseek-r1-and-inference-time-scaling/

1. 得分规则：

（1）参赛选手基于昇腾NPU kernel公开代码，自主设计初始提示词，实现大模型自动编写昇腾NPU Kernel算子。

（2）根据3个指标进行综合评分，分别为生成准确率，生成代码性能，以及大模型迭代耗时，与算子复杂度。

* 1. 算子正确生成(输入输出结果正确)得10分。
  2. 生成代码性能每超越公开代码仓每个百分点得10分。
  3. 大模型迭代耗时(等效8卡A100算力)，每使用秒扣0.1分。
  4. 基础算子：卷积、矩阵乘、层归一化等，加10分。
  5. 高阶算子：conv、bias、ReLU等，加20分。
  6. 可以使用[1]中的代码，对大模型进行微调。

1. [昇腾Gitee仓](https://gitee.com/ascend)
2. [昇腾社区官网](https://www.hiascend.com/)
3. 咨询专家及联系邮箱：

[qiuyangtao@hisilicon.com](mailto:qiuyangtao@hisilicon.com)

[fengyuan5@huawei.com](mailto:Fengyuan5@huawei.com)

1. 赛题互动交流答疑社区链接：

[https://www.chaspark.com/#/races/competitions/1135704143654498304](https://www.chaspark.com/" \l "/races/competitions/1135704143654498304)

**赛题七**

1．题目名称

多主体视频生成技术

2．题目描述

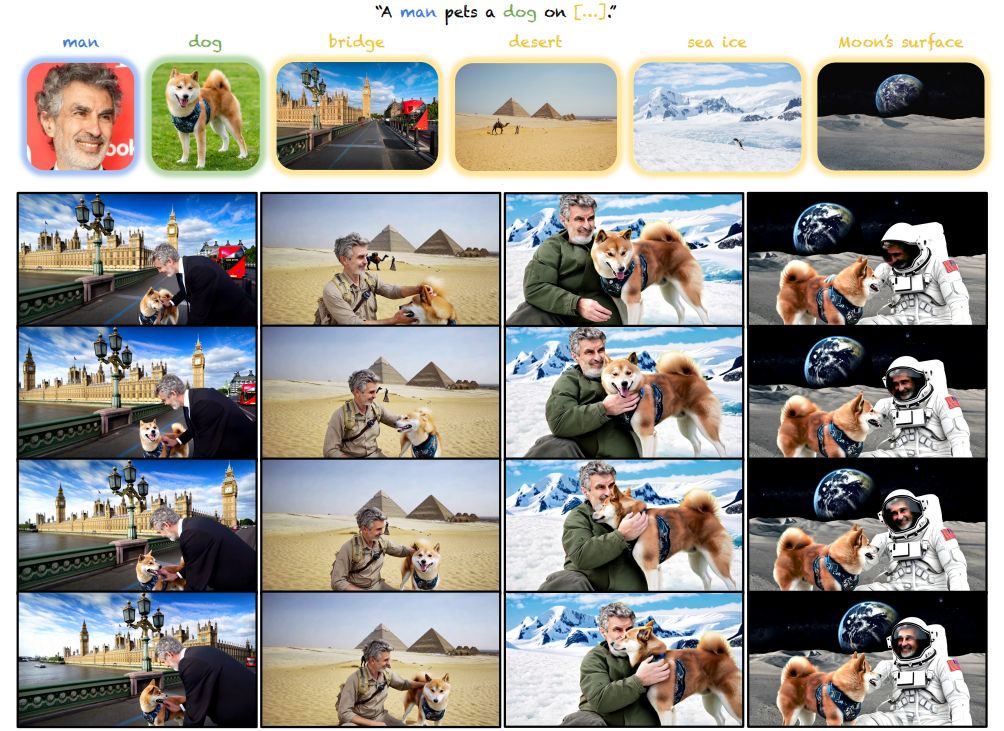
随着在短视频爆发、元宇宙再次崛起，人们对视频内容的需求正从“单一叙事”转向“多角色互动”的沉浸式体验。然而，传统视频生成领域长期存在角色失真、场景割裂等痛点。如何确保视频中不同主体（人物、动物、物品等）在各种动作和互动中都能够保持一致，并且自然地融合在一起，一直存在较多挑战。为了解决这些痛点和挑战，多主体视频生成技术成当下研究热点，多主体视频生成技术要解决主要问题：

（1）主体一致性问题：不同镜头下或视角下，确保各主体特征保持一致，避免出现主体“变脸”或消失；

（2）主体交互问题：多个主体之间动作交互自然和谐，能准确处理主体间的空间位置关系，确保多个主体交互动作和空间位置关系符合现实世界的物理规律。

（3）场景融合问题：当多个主体出现在同一场景中时，确保场景的光影效果、色彩风格、背景元素等保持一致，避免场景不融合、不自然的问题。

多主体视频生成示意图



3.具体要求

（1）参赛者需要基于开源数据和开源视频生成基模（万相、混元等），完成算法原型设计和实现；

（2）参赛者需要在开源测试集上进行效果验证，并给出对比结果报告；

（3）参赛者需给出算法的可解释性文档，系统阐述算法关键模块，并附上数据/图片等可证明观点的论据；

（4）评测

1. 客观评测

① 文图相似度：文本和每帧图像CLIP特征的平均COS相似度。

其中，*CLIP*为*Vit-L*模块，*Tfea*为文本*CLIP*特征，*Im*为单帧图像*CLIP*特征，*t*为视频中第*t*帧，*L*为视频帧数。*Ts*为图文*CLIP*特征平均*COS*相似度，范围[0,1]，越大表明相似性越好。



图1. 文图相似度对比示意图

② 主体相似度：参考主体和生成主体*DINO*特征的平均*COS*相似度。

其中,为第*i*个参考主体的DINO特征， *(i)*为第*i*个生成主体的*DINO*特征，*N*为主体个数。为参考主体和生成主题*DINO*特征的平均*COS*相似度，范围[0,1]，越大表明相似度越好。



图2. 主体相似度对比示意图

③ 人脸相似度：参考人脸和生成人脸的*ArcFace*人脸特征的平均*COS*相似度。

其中,为第*i*个参考人脸的*ArcFace*特征， (*i*)为第*i*个生成人脸的ArcFace特征，*N*为人脸个数。为参考人脸和生成人脸的平均*COS*相似度，范围[0,1]，越大表明相似度越好。



图3. 人像相似度对比示意图

达标标准：文图相似度、主体相似度以及人像相似度，分别满足：*Ts≥ 0.5*，*Ss*≥0.7， *Fs*≥0.7。

2. 主观评测：人工评委会准备100组测试样本，召集100人的众测团队，验证参赛者的算法效果，评估维度：（1）整体MOS分，（2）成片率（交互合理性、主体特征保持等维度）。

达标标准：整体MOS≥3.5（满分5分，很差：1，差：2，一般：3， 好：4， 很好：5 ），成片率≥85%。

4．参考评估数据集

https://github.com/snap-research/MSRVTT-Personalization

5．参考文献

[1]Multi-subject Open-set Personalization in Video Generation

[2]ConceptMaster multi-concept video customization on diffusion transformer models without test-time tuning

[3] <https://snap-research.github.io/open-set-video-personalization/>

6．咨询专家及联系邮箱

李安平–lianping@huawei.com

7.赛题互动交流社区链接：

[https://www.chaspark.com/#/races/competitions/1135707174513094656](https://www.chaspark.com/" \l "/races/competitions/1135707174513094656)

**赛题八**

1．题目名称

影视剧说话人识别

2．题目描述

一部长篇影视剧通常包含大量台词。识别出每句台词的说话人，对于理解影视剧剧情具有重要辅助作用。然而当前尚不存在能够结合视频、文本、语音的说话人识别方法，通常只能通过单一模态（如语音输入）来判断说话人身份，在影视剧数据中的识别精度难以提升。算法面临的主要挑战包括：

（1）背景声音干扰：影视剧通常有背景噪声、背景音乐等干扰；

（2）多说话人交叠：在争吵等场景，不同角色的声音容易混杂在一起；

（3）多模态理解：通常需要结合其他模态，如视频中说话者口型、上下文剧情等，才能准确判断说话人；

（4）开放集合识别：可能存在大量无名配角或者路人角色，其说话声音不在初始语音库内。

3. 具体要求

（1）对于1部长篇影视剧，给定原始视频数据、通过OCR得到的主要演员表、通过OCR得到的逐句台词及时间戳、通过人脸识别得到的主要角色在剧中逐帧出现时间，需要输出每句台词的说话人：对于演员表中存在的角色，直接给出其角色名字；对于演员表中不存在的角色，给出“其他”；对于不同的“其他”角色，最好能够给出区分这些角色的信息，如不同描述；

（2）参赛者需给出算法模型的可解释性文档，系统阐述模型的哪些关键设计对于说话人识别有重要帮助；

（3）可利用额外数据来加强模型的泛化能力，或结合其他开源数据集完成；

**（4）注意：在识别过程中，可以适当加入人工校验和标注，使得结果更加准确；但是人工工作量应当保持在合理范围内：例如不能通过逐句人工标注以达到接近100%的准确率；在最终报告中，必须给出全流程的详细说明；**

（5）参赛者的算法模型需给出逐句说话人ID识别结果，参考指标：识别精度；竞赛得分部分根据选手提交的方案在测试集（包括公开测试集和非公开测试集）上的运行结果给出；

（6）本次参赛的最终成绩会综合使用的算法先进程度、对于调参的敏感性、人工工作量大小、模型的检测精度等方面进行考量，其中精度指标最重要；

（7）赛题总分由竞赛得分（40%）和专家评分（60%）两部分组成。专家评分使用华为内部的缺陷测试数据进行指标评价和验收，考察算法的泛化性；

（8）专家评分由评委组对选手所提交的方案的新颖性、合理性等进行打分。因此，参赛选手还需要提交模型代码（用于非公开测试集评估）、模型使用说明文件（用于报告模型方案以及模型在公开测试集上的结果）。

（9）参考案例：

**被处理视频：唐山大地震.mp4**

**输出json文件格式结果：**

**{**

**"roles":[ "李元妮", "大姑", "方达", "方登", "四川母亲","王德清","奶奶", "方登（幼年）","老牛",**

**"方达（幼年）", "王志国妻","董桂兰","小河","方大强", "赵主任", "杨志", "杨志同学"],**

**"results" :{**

**"1": {**

**"start\_time": "00:02:39.333",**

**"end\_time": "00:02:40.875",**

**"text": "爸 快看快看",**

**"role": "方登（幼年）"**

**},**

**"2": {**

**"start\_time": "00:02:40.917",**

**"end\_time": "00:02:43.125",**

**"text": "蜻蜓 蜻蜓",**

**"role": "方登（幼年）"**

**},**

**…,**

**"226": {**

**"start\_time": "00:22:45.000",**

**"end\_time": "00:22:46.833",**

**"text": "我们上那边救人去啦",**

**"role": "其他：救援者（区分信息）"**

**},**

**...**

**}**

**}**

**竞赛得分评分规则：**

**1）角色判断正确台词句数/总台词句数（非演职表人物判断出其他即为正确），满分100**

**2）对于不同的“其他”角色，给出区分这些角色的信息，额外加10分**

**可选影视剧列表（任选一部提交前15集结果即可）：**

**《人世间》、《狂飙》、《甄嬛传》、《开端》、《三体》**

4．参考数据集

https://github.com/clovaai/voxceleb\_trainer

5．参考文献和链接

[1]Chung et al., VoxCeleb2: Deep Speaker Recognition, in INTERSPEECH, 2018.

[2]Jung et al., ESPnet-SPK: full pipeline speaker embedding toolkit with reproducible recipes, self-supervised front-ends, and off-the-shelf models, in arXiv preprint, 2024.

[3]Yakovlov et al., Reshape Dimensions Network for Speaker Recognition, in INTERSPEECH, 2024.

6．咨询专家及联系邮箱

谢凌曦–xielingxi@huawei.com

霍馨月-huoxinyue2@huawei.com

1. 赛题互动交流社区链接：

https://www.chaspark.com/#/races/competitions/1135708655057883136