

# 基于场景草图的细粒度视频检索

## Fine-Grained Video Retrieval With Scene Sketches

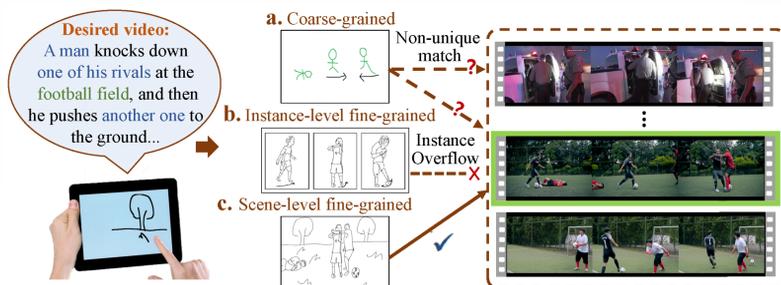
Ran Zuo, Xiaoming Deng\*, Keqi Chen, Zhengming Zhang, Yu-Kun Lai, Fang Liu, Cuixia Ma\*, Hao Wang, Yong-Jin Liu\*, Hongan Wang

IEEE Transactions on Image Processing  
(TIP 2023, Volume: 32, Page(s): 3136-3149, CCF-A)

联系人: 马翠霞, 邮箱: cuixia@iscas.ac.cn

### Introduction

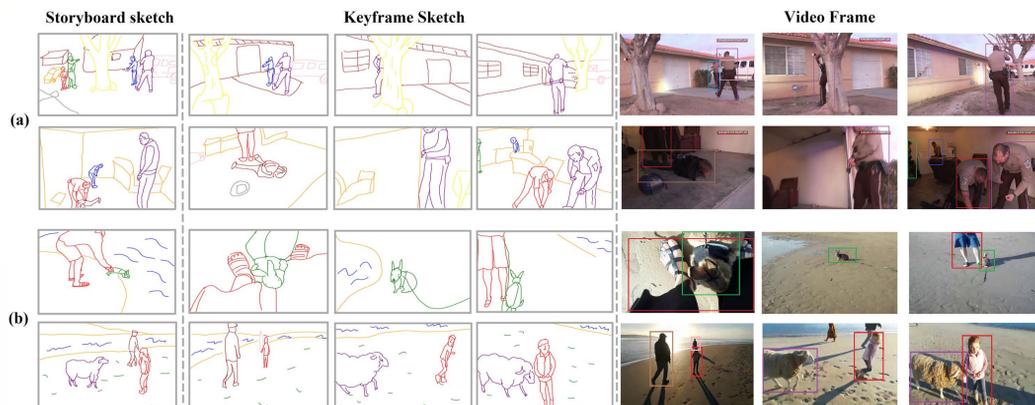
- 基于草图的视频检索通过草图直观表达特定视频场景的外观细节, 进而检索对应场景内容的视频, 目前缺少包含多个物体复杂场景的细粒度场景级草图检索视频工作
- 论文针对基于场景草图的细粒度视频检索问题, 通过单张故事板草图描述视频场景(包括背景元素以及出现在不同视频帧中多个物体的外观、大小和姿势等), 检索对应场景内容的视频片段



### Datasets & Methods

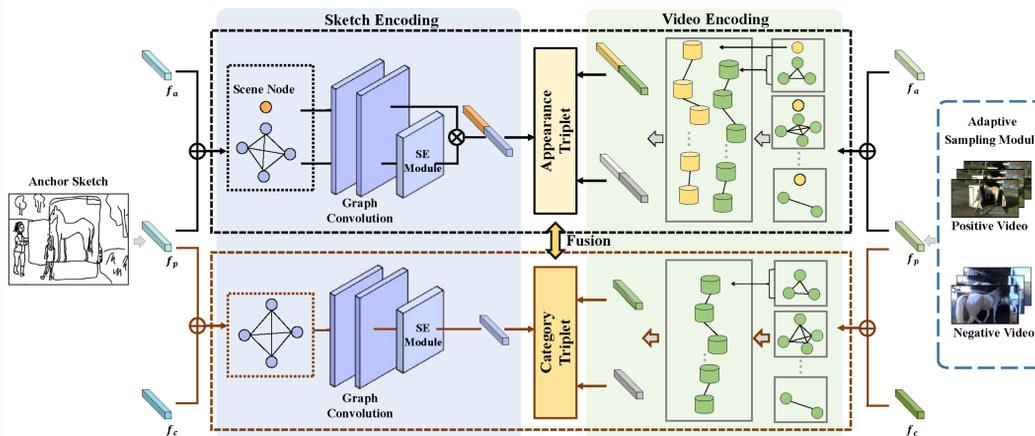
#### Dataset

- 构建细粒度场景级草图-视频数据集 SketchVideo, 每个视频片段对应单张故事板草图概括视频场景内容和多张关键帧草图描述对应的视频帧
- 草图按照笔画进行实例级标注, 并与特定视频帧构建实例对应, 为细粒度场景级草图-视频的研究提供数据基础



#### Method

- 提出基于空间图卷积神经网络和时空图卷积神经网络的检索框架 SQ-GCN, 为草图和视频构建以实例为节点的空间图和时空图
- 构建外观和类别分支模型对图结构节点初始化, 分别经过图卷积运算进行特征更新
- 采用自适应视频采样策略获取视频的关键帧以提高视频编码的效率

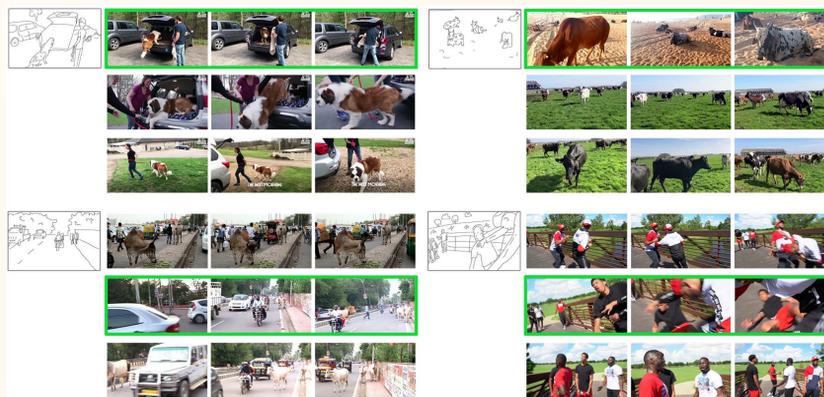


### Results & Application

#### Comparison with SoTA

- 在 SketchVideo 数据集上进行草图检索视频实验, 实验结果表明, 本文的 SQ-GCN 检索精度超过其它基于草图的视频检索方法

Method	mAP	Acc.@1	Acc.@5	Acc.@10
FG-SBVR[8]	23.25	14.02	32.23	42.99
Scene Sketcher[27]	24.14	17.76	31.78	37.78
Baseline	35.22	22.68	49.19	63.90
SQ-GCN	<b>66.74</b>	<b>52.78</b>	<b>83.98</b>	<b>93.80</b>



#### Fine-grained Scene-level Retrieval Capability

- 输入包含相似场景内容的草图, 模型仍可以细粒度地检索到对应视频



#### Retrieval Prototype System

- 构建草图检索视频原型系统, 支持草图绘制和草图合成两种场景草图构建方式

