Scene Separation & Data Selection: Temporal Segmentation Algorithm for Real-time Video Stream Analysis

Yuelin Xin[1,2] Zihan Zhou[1] Yuxuan Xia[1]

[1]SWJTU-Leeds Joint School, CS
Southwest Jiaotong University

[2]School of Computing
University of Leeds

Spatio-Temporal Reasoning and Learning, 2022

Outline

Introduction

- The problem (Background & What we want to achieve)
- Our motivation (Why not neural networks?)

Remark

Scene separation(Temporal segmentation) is a problem in which we want to separate a video stream into different scenes. **A scene** is defined as a group of similar-looking frames that are temporally adjacent to each other.

The problem

- **Background**: real-time video stream interpretation, including video semantics / video accessibility / surveillance footage auto-interpretation, etc.
- **Difficulties**: algorithms do not see video as a continuous stream of images, but as discrete frames.

Figure 1: Video semantics.

The problem

- **The traditional approach**: 3D CNNs (CNN models with the additional temporal dimension)
- **What's missing**: hard to control when the video is very long or it is of indefinite length (like live streaming).

Example

It would be hard to pick up sudden moves in long videos because the longer the video, the worse the temporal resolution. (like a very tiny object in a very massive picture in 2D CNNs)

Our motivation

Why not neural networks?

- Neural networks are relatively slow, the inference time of a lot of NNs makes them difficult to be used in real-time video analysis.
- And the 2SDS algorithm is fully capable of handling simple scene separation tasks.

Algorithm	FPS (higher is better)
YOLOv5s	13
2SDS	100

Table 1: Comparison of inference speed under same hardware.[1]

[1] Apple M1 Pro (CPU)

Method: 2SDS

- Related work: SlowFast Networks architecture
- Our work: 2SDS architecture
- Scene separation
- Data selection

Related work

SlowFast Networks [Feichtenhofer *et al.*, 2019]

- **Slow pathway**:CNN with high spatial resolution(low FPS).
- **Fast pathway**:CNN with high temporal resolution(high FPS).

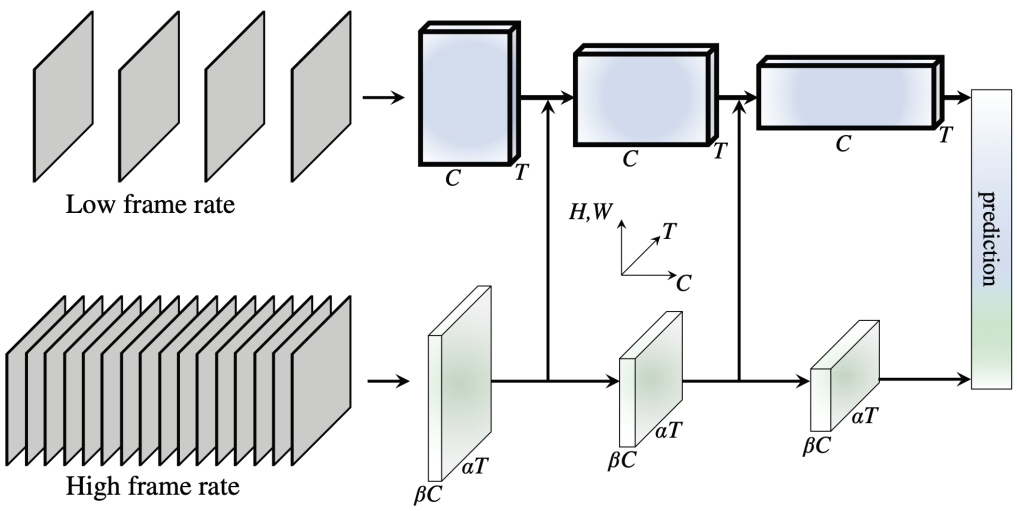

Figure 2: SlowFast Networks Architecture.

Our work

Similar with the SlowFast Networks architecture, but we replace the fast pathway with 2SDS.
This architecture has an even **finer temporal resolution** because we replaced the CNN with a faster algorithm.

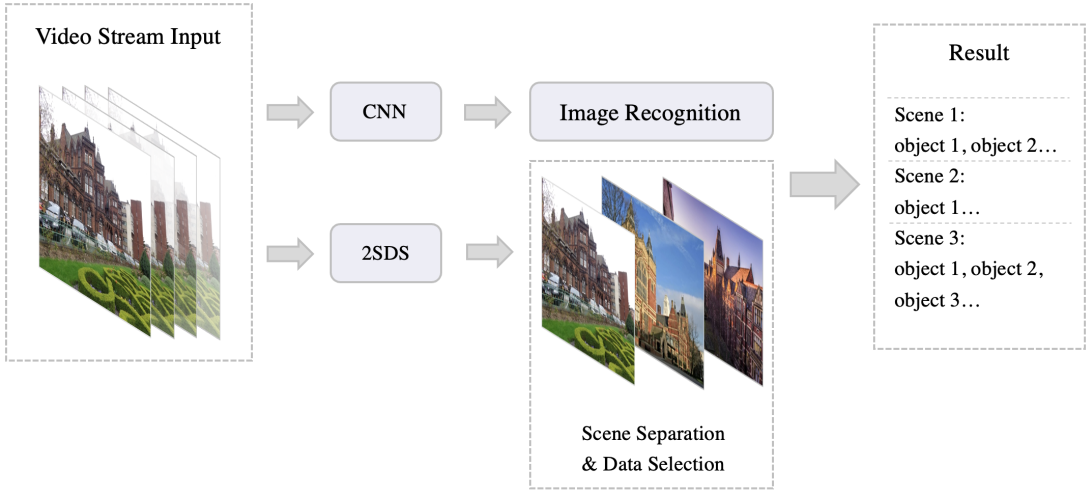

Figure 3: 2SDS Architecture.

2SDS: a two step method

- Step 1: **Scene separation**(Temporal segmentation)
- Step 2: **Data selection**

Miracle Factory

Scene separation

Data selection