Video Object Segmentation Based on Pixel-level Annotated Dataset

EECS6893 Final Project Proposal

ch3467 Chong Hu EE
yl4238 Yanlin Liu EE

Content

- Introduction
- Data
- Methods
- System
- Schedule

Introduction

Video Object Segmentation

Goal: extracting foreground objects from video clips.

Application:

- video summarization/editing
- object tracking
- video action detection[3][4]
- autonomous driving
- etc…

Figure 1. separating foreground object(s) from the background region of a video[5]

Data

DAVIS 2016[2]

Densely **A**nnotated **VI**deo **S**egmentation

- 50 full HD video sequences
- pixel-accurate ground-truth data provided for every video frame
- Contain occlusions, fast-motion, non-linear deformation and motion-blur

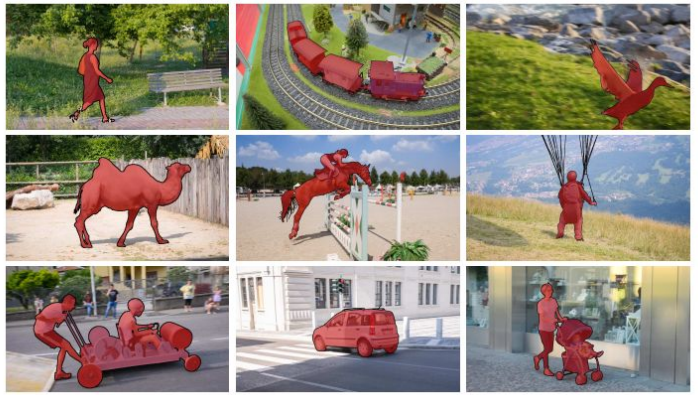


Figure 2. Sample images in DAVIS-2016 with annotation.[2]

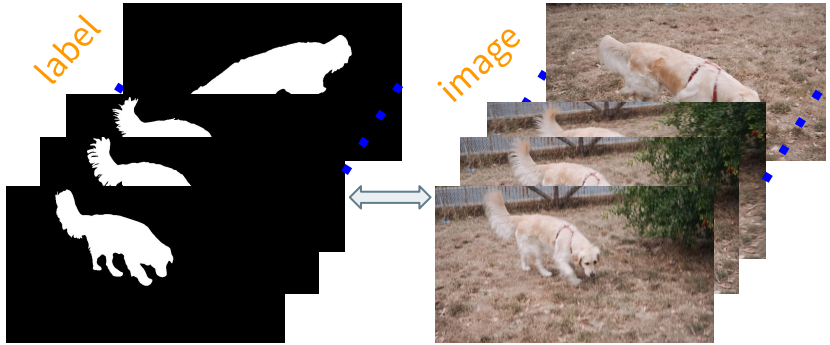


Figure 3. Image Sequence Data in DAVIS-2016.[2]

Methods

- Data argumentation:
  - Computational limit → resize image
  - Flip, Gaussian Noise, Brightness and Hue, etc.
- Model
  - Supervised vs. unsupervised → Spatiotemporal CNN[1]
- Evaluation: region similarity, contour accuracy and temporal instability (provided by dataset).

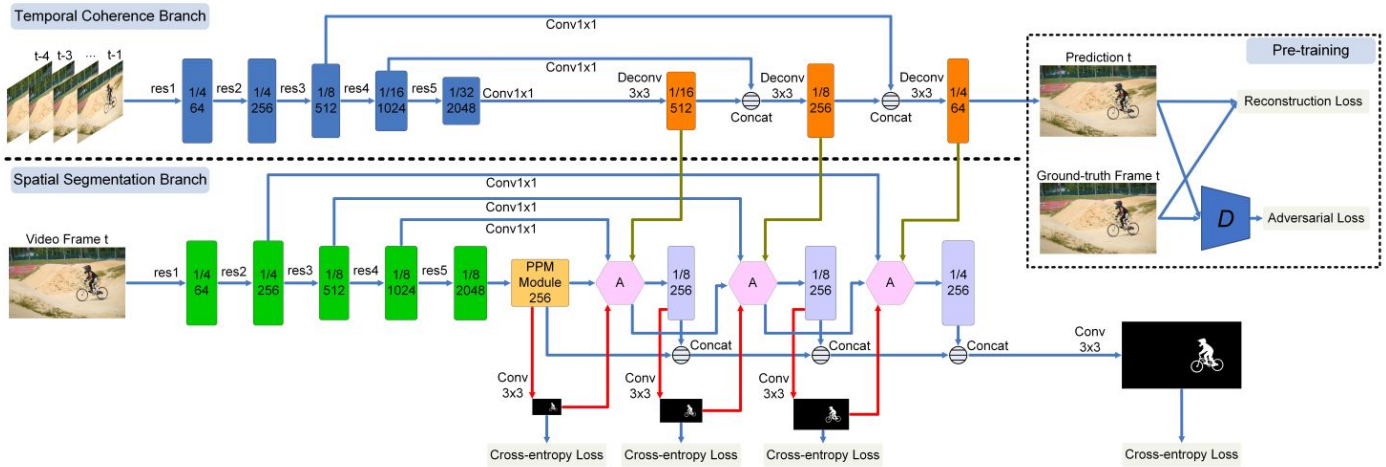


Figure 4. Network Structure Overview for STCNN.[1]

System

- Aims:
  - To support previewing uploaded video with segmented foreground object.
- Expected outcome:
  - Separate foreground objects with larger than 80% overlapping with ground-truth on average
  - Provide API for video website and simple web front-end for demo

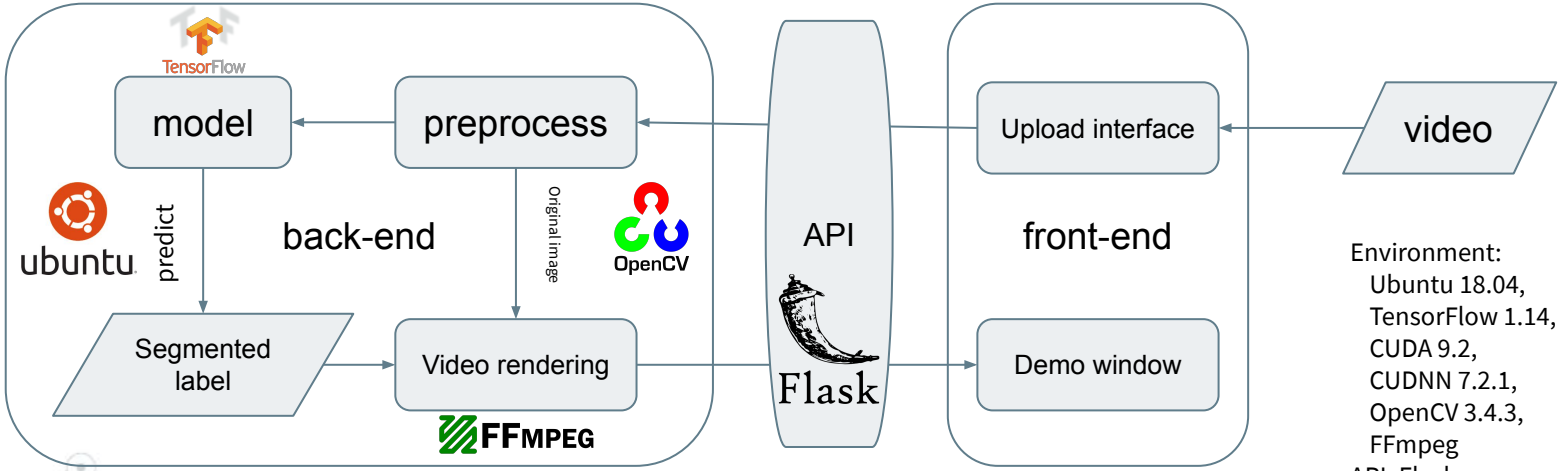
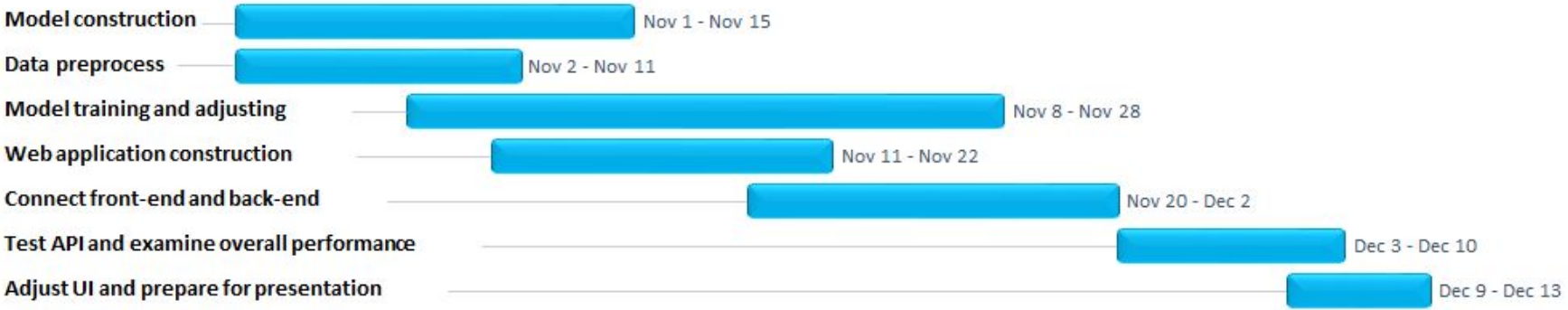


Figure 5. Simple overview diagram of system

Environment:
Ubuntu 18.04,
TensorFlow 1.14,
CUDA 9.2,
CUDNN 7.2.1,
OpenCV 3.4.3,
FFmpeg
API: Flask
Front-end: video js

Schedule


Proposal
Nov 1
Midpoint
Nov 22
Final presetation
Dec 13
2019
Nov
Dec
2019
Today


Model construction
Nov 1 - Nov 15
Data preprocess
Nov 2 - Nov 11
Model training and adjusting
Nov 8 - Nov 28
Web application construction
Nov 11 - Nov 22
Connect front-end and back-end
Nov 20 - Dec 2
Test API and examine overall performance
Dec 3 - Dec 10
Adjust UI and prepare for presentation
Dec 9 - Dec 13

Reference

[1] Kai Xu, Longyin Wen, Guorong Li, Liefeng Bo, and Qingming Huang. Spatiotemporal cnn for video object segmentation. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2019.

[2] Federico Perazzi, Jordi Pont-Tuset, Brian McWilliams, Luc Van Gool, Markus Gross, and Alexander Sorkine-Hornung. A benchmark dataset and evaluation methodology for video object segmentation. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[3] Chunhui Gu, Chen Sun, Sudheendra Vijayanarasimhan, Caroline Pantofaru, David A. Ross, George Toderici, Yeqing Li, Susanna Ricco, Rahul Sukthankar, Cordelia Schmid, and Jitendra Malik. AVA: A video dataset of spatio-temporally localized atomic visual actions. CoRR, abs/1705.08421, 2017

[4] Xitong Yang, Xiaodong Yang, Ming-Yu Liu, Fanyi Xiao, Larry S. Davis, and Jan Kautz. Step: Spatio-temporal progressive learning for video action detection. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2019.

[5] Jordi Pont-Tuset, Federico Perazzi, Sergi Caelles, Pablo Arbel´aez, Alexander Sorkine-Hornung, and Luc Van Gool. The 2017 davis challenge on video object segmentation. arXiv:1704.00675, 2017

[6] Fanyi Xiao and Yong Jae Lee. Track and segment: An iterative unsupervised approach for video object proposals. In CVPR, 2016. 1, 2

[7] Nicolas Marki, Federico Perazzi, Oliver Wang, and Alexander Sorkine-Hornung. Bilateral space video segmentation. In CVPR, pages 743–751, 2016. 6, 8

[8] Anestis Papazoglou and Vittorio Ferrari. Fast object segmentation in unconstraint video. In ICCV, pages 1–6, 2013. 2

[9] Yi-Hsuan Tsai, Ming-Hsuan Yang, and Michael J. Black. Video segmentation via object flow. In CVPR, pages 3899–3908, 2016. 1, 2, 6, 8

[10] Longyin Wen, Dawei Du, Zhen Lei, Stan Z. Li, and Ming-Hsuan Yang. JOTS: joint online tracking and segmentation. In CVPR, pages 2226–2234, 2015. 1, 2

[11] Siyang Li, Bryan Seybold, Alexey Vorobyov, Xuejing Lei, and C.-C. Jay Kuo. Unsupervised video object segmentation with motion-based bilateral networks. In ECCV, pages 215– 231, 2018. 1

Thanks!