LEARNING VIDEO OBJECT SEGMENTATION FROM STATIC IMAGES

INTRODUCTION

Goal: Separating foreground objects from the background in a video given the 1st frame mask annotation.

- Guided CNN trained for video object segmentation using only static images.
- Handle different types of input annotations such as bounding boxes and segments while leveraging an arbitrary amount of annotated frames.
- Competitive results on three different datasets, independently from the type of input annotation.

Visit the project pages: https://graphics.ethz.ch/~perazzi/masktrack
https://www.mpi-inf.mpg.de/masktrack
Source code and trained models are available online.

LEARNING FROM STATIC IMAGES

Video Dataset

- DAVIS 2016 [Perazzi et al. 16]: 50* videos, 4K images
- SegTrack-v2 [Li et al. 13]: 14 videos

- lack of large-scale diverse video data for training

Image Dataset

- MSRA10K [Cheng et al. 14]: 10K images

EVALUATION

Competitive results despite using the same model and parameters across all videos.

SUPERVISION EFFORT

REFERENCES

ObjFlow: Video Segmentation via Object Flow, Yi-Hsuan Tsai et al., CVPR 2017

NLC: Video segmentation by non-local consensus voting, A. Faktor and M. Irani, BMVC 2014

BVS: Bilateral Space Video Segmentation, N. Maerki et al. CVPR 2016

FCP: Fully connected object proposals for video segmentation, F. Perazzi et al. ICCV 2015

TRS: Track and segment: An iterative unsupervised approach for video object proposals, Xiao et al. CVPR 2016


Z15: Semantic object segmentation via detection in weakly labeled video, Zhang et al. CVPR 2015

Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs, Chen et al. ICLR 2015