ClayNet: Improved VOS Modeling

Increasing Video Object Segmentation Performance on Long-Term Obscuration

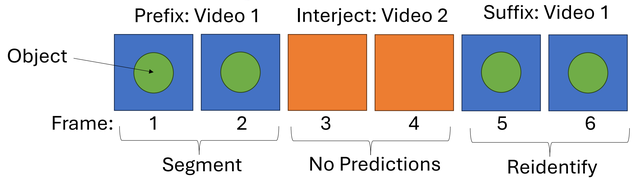
Clayton Bromley, Alex Moore (LLNL)

Video object segmentation (VOS) is a challenging task in the world of deep learning and computer vision and is essential for the proper perception and analysis of videos. Cutie, the current state-of-the-art model for VOS, performs very well on clean video. When artificially introducing long-term obscurations of the segmented object, Cutie's performance drops significantly. This project aims to design custom DAVIS datasets with artificial obscurations and propose a ClayNet model with improved performance.

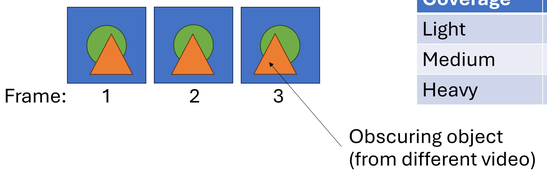
DATASETS

A few different datasets are designed that test the limitations of VOS on tasks that are very easy for humans but difficult for computer systems. These VOS benchmarks are more challenging for a new era of powerful AI. In all cases, the visible pixels, obscured pixels, and purple (all) pixels, are recorded for PVO analysis.

Method 1: Interjection

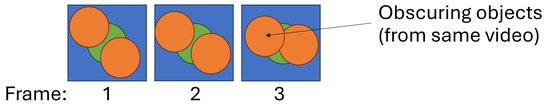


Method 2: Obscuration



Coverage	Minimum	Maximum
Light	0%	50%
Medium	25%	75%
Heavy	50%	100%

Method 3: N Object



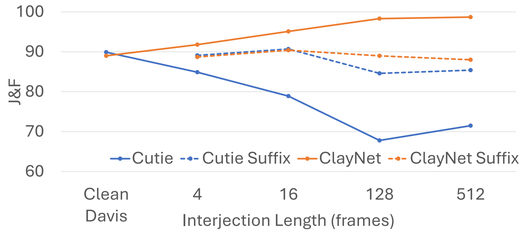


Figure #1: Cutie and ClayNet J&F Performance on Interjections

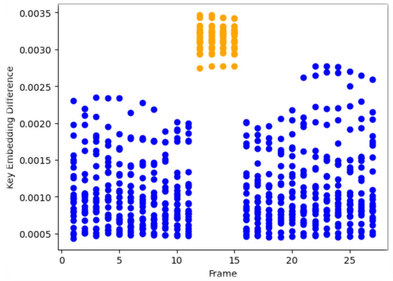
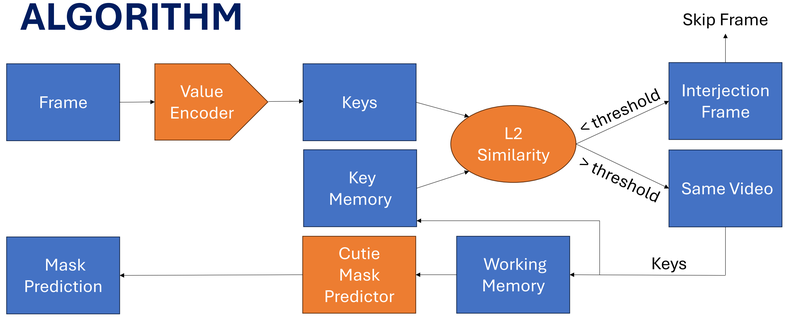


Figure #2: Key Embedding Differences for 25 Interjection Samples
Orange Frames are Interjected

Figure #3: Cutie and ClayNet Applied to 512 Frame Interjection

ALGORITHM



CONCLUSIONS

While Cutie maintains excellent performance on clean video data, simple obscuration augmentations cause this model to lose track of the segmented object. By introducing a frame-by-frame binary classifier in the algorithm with 99.6% accuracy, ClayNet can drastically improve performance on interjection datasets. Specifically, not writing bad frames to memory allows for suffix reidentification improvement.

CONTACT

bromley1@llnl.gov, moore278@llnl.gov

Cheng, Ho Kei, et al. "Putting the object back into video object segmentation." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024.

Pont-Tuset, Jordi, et al. "The 2017 davis challenge on video object segmentation." *arXiv preprint arXiv:1704.00675* (2017).

Lawrence Livermore National Laboratory

LLNL-POST-867770

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

National Nuclear Security Administration