SHLS: Superfeatures Learned from Still Images for Self-supervised VOS

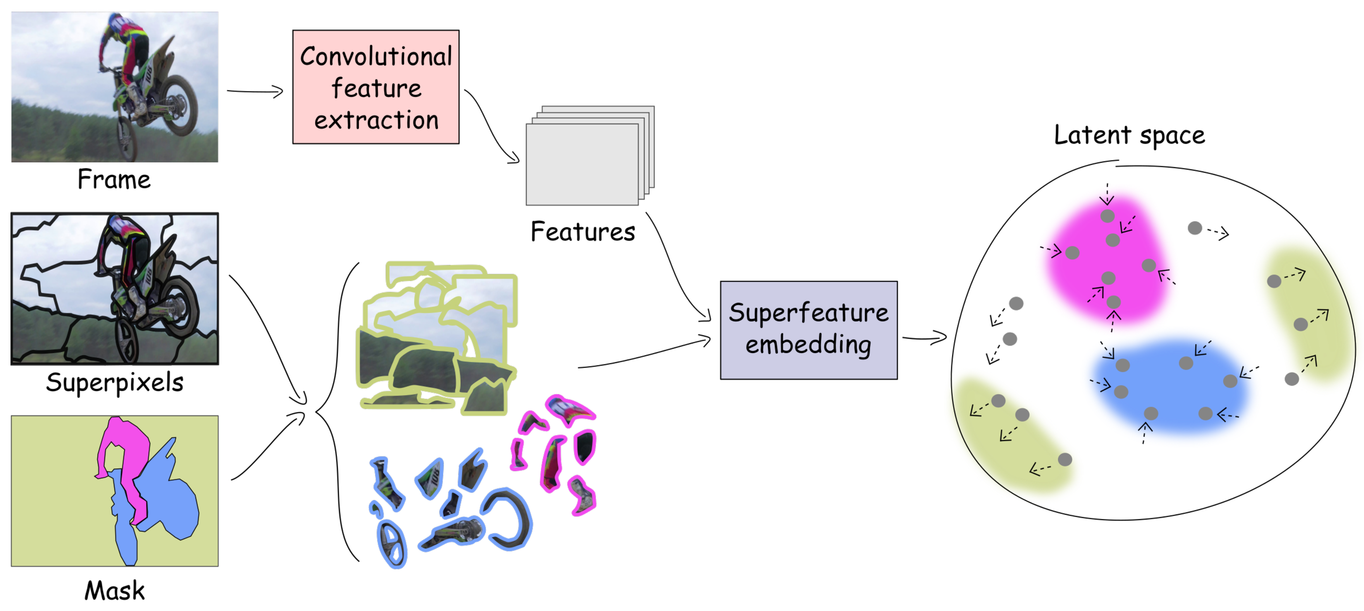
Marcelo Mendonça, Jefferson Fontinele, Luciano Oliveira

BMVC 2023

Project page

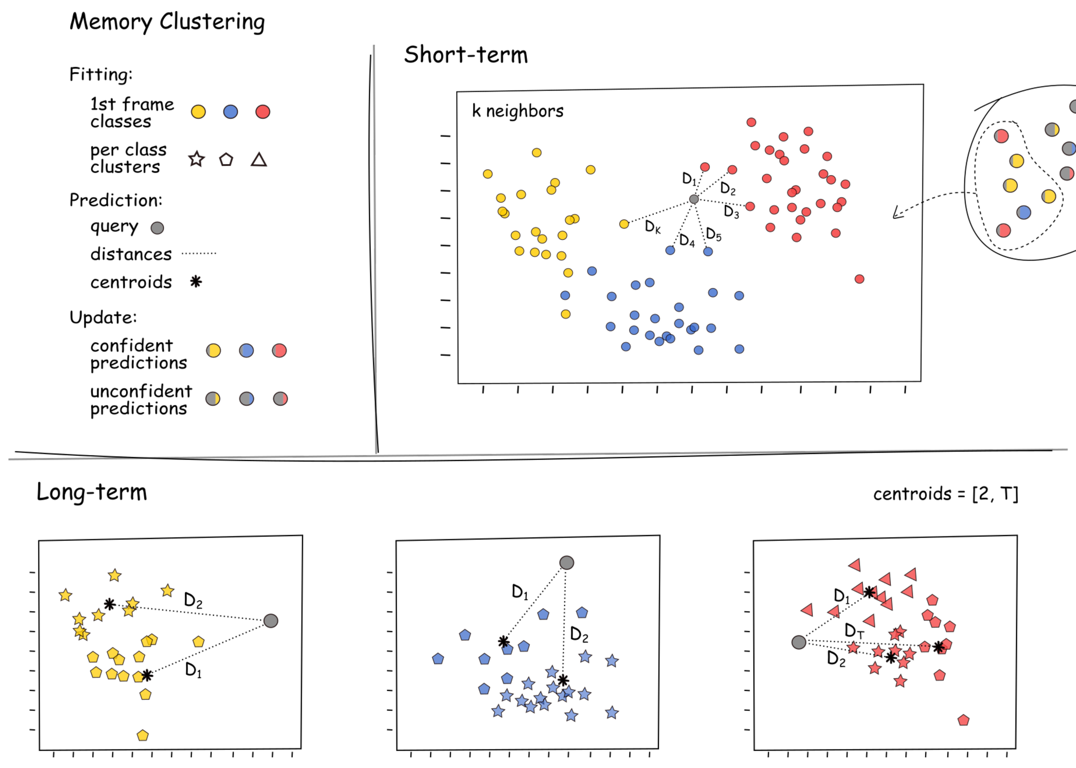
Superfeatures in a Highly Compressed Latent Space

A novel approach that embeds convolutional features into the corresponding superpixel areas through metric learning. The resulting ultra-compact image representations enable us to learn video object segmentation (VOS) from a small dataset of unlabeled still images.



Memory Clustering

Our memory clustering mechanism provides short- and long-term information by measuring similarity distances among superfeatures in the latent space.

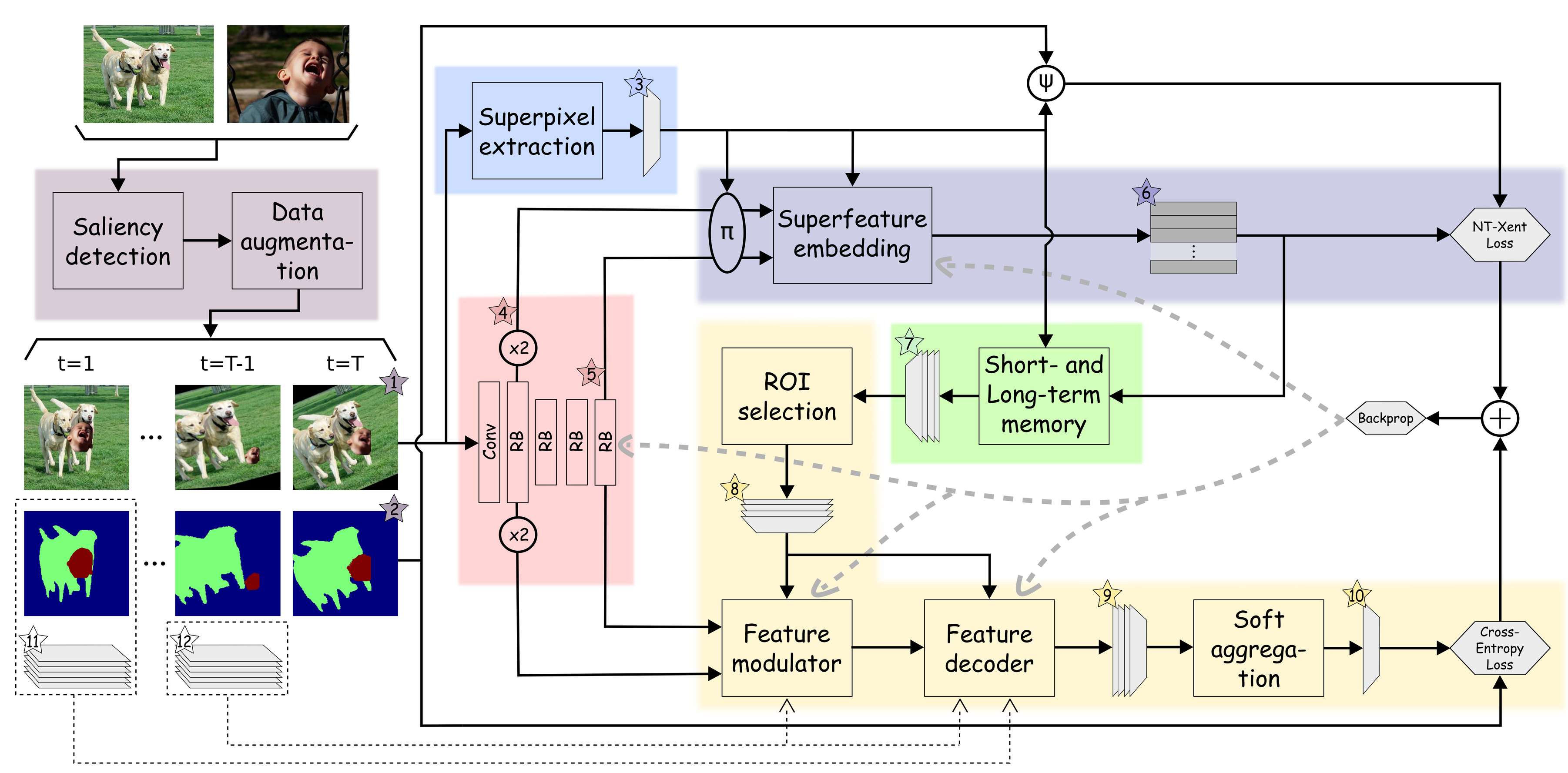


Short-term: is based on k-NN searches and responds quickly to immediate changes in the objects during short intervals.

Long-term: computes distances from the query superfeatures to the centroids of class-specific clusters.

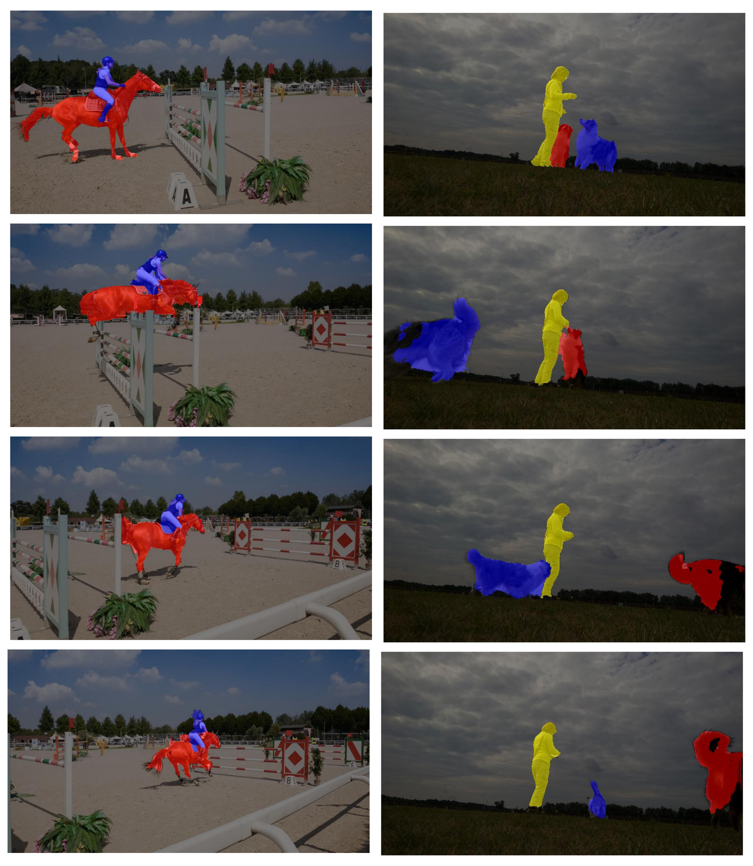
Single-object Segmentation

SHLS Training



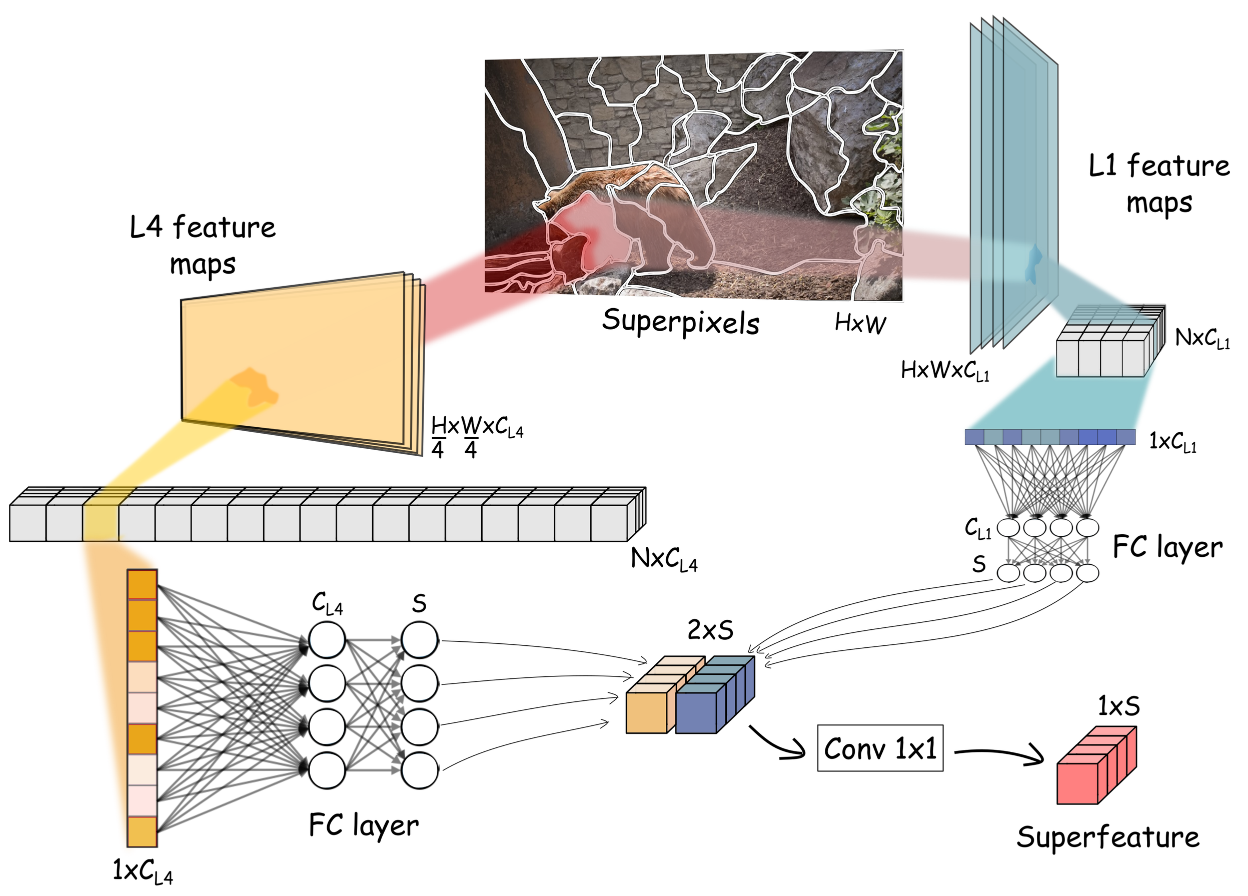
Offline phase: frame-mask sequences are synthesized using saliency detection and data augmentation; superpixels are extracted.
Online phase: Convolutional features are shared between two branches, one for superfeature generation with a contrastive NT-Xent loss and the other for segmentation refinement with a cross-entropy loss. A memory clustering module is used to store and retrieve information from past frames.

Multi-object Segmentation



Combining superpixels and features in superfeatures

The features inside a superpixel are averaged, for each channel, yielding NxC_{L1} and NxC_{L4} vectors. These vectors are fed into fully-connected layers, resulting in a 2xS vector, which is passed through a 1x1 convolution to generate the superfeature.

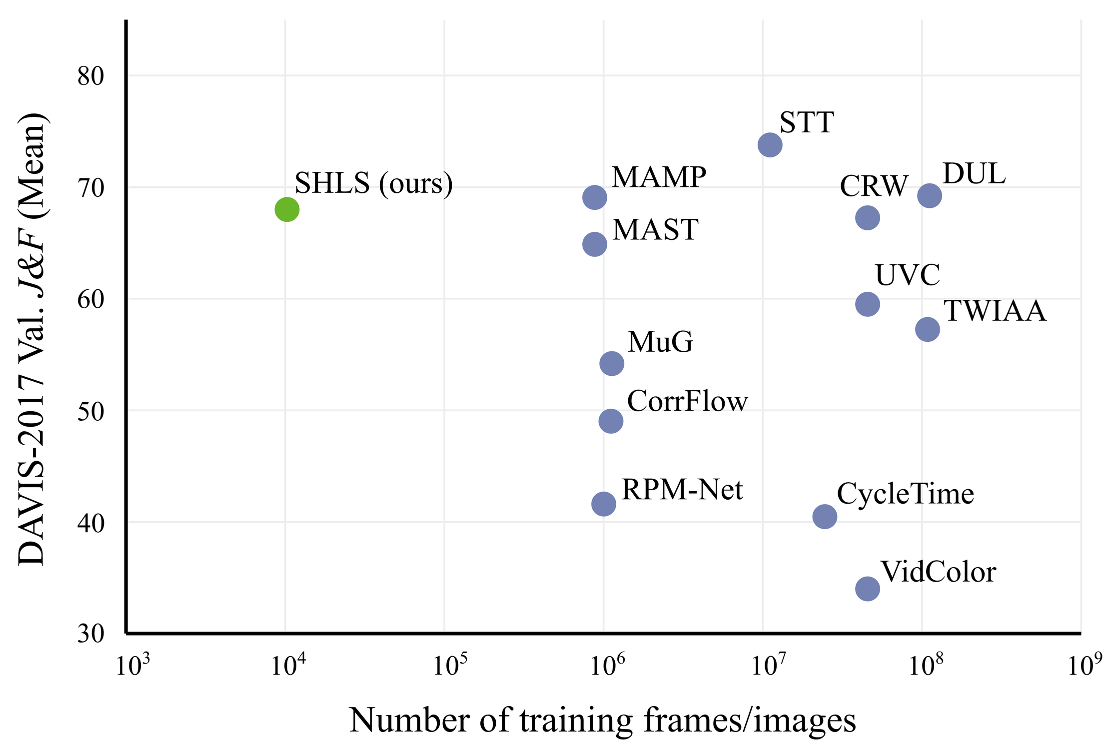


Takeaways

- A superfeature model that provides highly compressed superpixel-based representations.
- A memory clustering approach for retrieving information from past frames efficiently.
- A fully self-supervised VOS method trained with only 10k still images.

Quantitative Results

Benchmark on DAVIS-2017 validation set. SHLS is trained with at least 10^2 orders of magnitude fewer images than other self-supervised methods.



Acknowledgements

This work has been supported by the following Brazilian research agencies: FAPESB, CAPES and CNPq.

References

[1] MENDONÇA, M.; OLIVEIRA, L. ISEC: Iterative over-segmentation via edge clustering. IMAVIS, 2018.
[2] FONTINELE, J.; MENDONÇA, M.; RUIZ, M.; PAPA, J.; OLIVEIRA, L. Faster α-expansion via dynamic programming and image partitioning. IJCNN, 2020.

UFBA Universidade Federal da Bahia
IVISION INTELLIGENT VISION RESEARCH LAB
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA Bahia
UFMA

github.com/IvisionLab/SHLS