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.

Combining superpixels and features in superfeatures

The features inside a superpixel are averaged, for each channel, yielding $N_xC_1$ and $N_xC_4$ 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.

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.

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.

Acknowledgements

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

References


[github.com/IvisionLab/SHLS]