A Benchmark Dataset and Evaluation Methodology for Video Object Segmentation

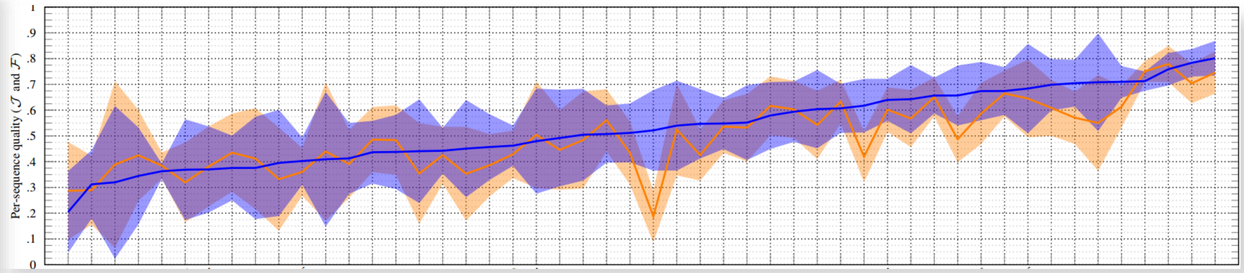
[1,2]F. Perazzi, [1]J. Pont-Tuset, [2]B. McWilliams, [1]L. Van Gool, [1,2]M. Gross, [2]A. Sorkine-Hornung

[1]ETH Zurich, [2]Disney Research

Introduction - DAVIS Densely Annotated VIdeo Segmentation

- New dataset and benchmark specific to the task of *video object segmentation.*
- **50 HD** video sequences with high-quality ground-truth - **14** state-of-the-art approaches evaluated.
- Analysis based on attributes that typically pose challenges to video segmentation.
- **Data and evaluation code available:** https://graphics.ethz.ch/~perazzif/davis/index.html

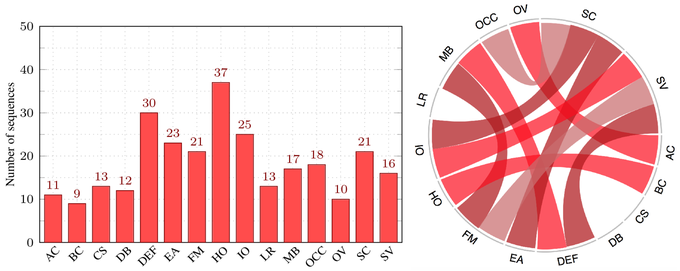
Per-sequence mean performance



Per-method spatio-temporal evaluation

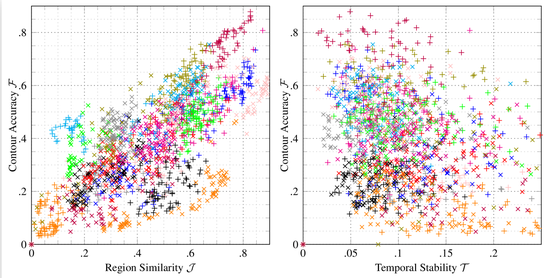
	Measure	MCG	SF-LAB	SF-MOT	NLC	CVOS	TRC	MSG	KEY	SAL	FST	TSP	SEA	HVS	JMP	FCP
		Preprocessing			Unsupervised							Semi-Supervised				
$\mathcal{J}$	Mean $\mathcal{M}\uparrow$	**0.724**	0.173	0.532	**0.641**	0.514	0.501	0.543	0.569	0.426	0.575	0.358	0.556	0.596	0.607	**0.631**
$\mathcal{J}$	Recall $\mathcal{O}\uparrow$	**0.912**	0.075	0.672	**0.731**	0.581	0.560	0.636	0.671	0.386	0.652	0.388	0.606	0.698	0.693	**0.778**
$\mathcal{J}$	Decay $\mathcal{D}\downarrow$	0.026	**-0.020**	0.050	0.086	0.127	0.050	**0.028**	0.075	0.084	0.044	0.385	0.355	0.197	0.372	**0.031**
$\mathcal{F}$	Mean $\mathcal{M}\uparrow$	**0.654**	0.218	0.452	**0.593**	0.490	0.478	0.525	0.503	0.383	0.536	0.346	0.533	0.576	**0.586**	0.546
$\mathcal{F}$	Recall $\mathcal{O}\uparrow$	**0.781**	0.052	0.440	**0.658**	0.578	0.519	0.613	0.534	0.264	0.579	0.329	0.559	**0.712**	0.656	0.604
$\mathcal{F}$	Decay $\mathcal{D}\downarrow$	0.046	**-0.016**	0.052	0.086	0.138	0.066	**0.057**	0.079	0.072	0.065	0.388	0.339	0.202	0.373	**0.039**
$\mathcal{T}$	Mean $\mathcal{M}\downarrow$	0.652	0.758	**0.637**	0.356	0.243	0.327	0.250	**0.190**	0.600	0.276	0.329	0.137	0.296	**0.131**	0.285

Attributes distribution and correlation



BC Background Clutter. Back- and foreground have similar colors.
DEF Deformation. Object undergoes non-rigid deformations.
MB Motion Blur. Object has fuzzy boundaries.
FM Fast-Motion. The average, per-frame object motion computed.
LR Low Resolution. The ratio between object bounding-box area.
OCC Occlusion. Object becomes partially or fully occluded.
OV Out-of-view. Object is partially clipped by the image boundaries.
SV Scale-Variation. The area ratio among pairs of bounding boxes .
AC Appearance. Change. Noticeable appearance variation.
EA Edge Ambiguity. Unreliable edge detection.
CS Camera-Shake. Footage displays non-negligible vibrations.
HO Heterogeneous. Object. Object regions have distinct colors.
IO Interacting Objects. The target object is an ensemble of multiple, spatially-connected objects.
DB Dynamic Background. Background regions move or deform.
SC Shape Complexity. The object has complex boundaries.

Evaluation Metrics



Region Similarity
Intersection-over-union between the segmentation and the ground-truth masks.

Contour Accuracy
Precision and recall between the contour points the segmentation and the ground-truth.

Temporal Stability
Penalize jittery, unstable boundaries. Use the Dynamic Time Warping (DTW) to match the points that minimizes the Shape Context Descriptor distances between two segmentations at consecutive time frames.

Attribute-based aggregate performance

Attr	NLC	CVOS	TRC	MSG	KEY	SAL	FST	TSP	SEA	HVS	JMP	FCP
	Unsupervised							Semi-Supervised				
AC	0.54 *+0.13*	0.42 *+0.12*	0.37 *+0.17*	0.48 *+0.08*	0.42 *+0.19*	0.33 *+0.12*	**0.55** ***+0.04***	0.17 *+0.23*	0.46 *+0.12*	0.42 *+0.23*	**0.58** ***+0.03***	0.51 *+0.16*
DB	**0.53** ***+0.15***	0.37 *+0.18*	0.39 *+0.15*	0.43 *+0.15*	0.52 *+0.07*	0.35 *+0.10*	0.53 *+0.06*	0.40 *-0.06*	0.58 *-0.03*	0.60 *-0.01*	0.60 *+0.01*	**0.62** ***+0.01***
FM	**0.64** ***+0.00***	0.37 *+0.24*	0.41 *+0.16*	0.46 *+0.14*	0.50 *+0.12*	0.35 *+0.13*	0.50 *+0.12*	0.18 *+0.31*	0.40 *+0.28*	0.42 *+0.31*	0.50 *+0.18*	**0.55** ***+0.13***
MB	**0.61** ***+0.04***	0.36 *+0.23*	0.32 *+0.27*	0.35 *+0.29*	0.51 *+0.08*	0.33 *+0.15*	0.48 *+0.14*	0.15 *+0.32*	0.39 *+0.24*	0.44 *+0.24*	0.51 *+0.15*	**0.53** ***+0.15***
OCC	**0.70** ***-0.09***	0.43 *+0.13*	0.44 *+0.10*	0.48 *+0.10*	0.52 *+0.08*	0.44 *-0.02*	0.53 *+0.07*	0.27 *+0.14*	0.47 *+0.13*	0.53 *+0.11*	0.47 *+0.21*	**0.59** ***+0.07***

Related Works

- **MCG**: Multiscale combinatorial grouping for image segmentation and object proposal. *J. Pont-Tuset et al. TPAMI, 2016*
- **SF-***: Saliency filters: Contrast based filtering for salient region detection. *F. Perazzi et al. CVPR 2012*
- **NLC**: Video segmentation by non-local consensus voting. *A. Faktor and M. Irani BMVC 2014*
- **CVOS**: Causal video object segmentation from persistence of occlusions. *B. Taylor et al. CVPR 2015*
- **TRC**: Video segmentation by tracing discontinuities in a trajectory embedding. *K. Fragkiadaki et al. CVPR 2012*
- **MSG**: Object segmentation by long term analysis of point trajectories. *T. Brox and J. Malik ECCV 2010*
- **KEY**: Key-segments for video object segmentation. *Y. J. Lee et al. ICCV 2011*
- **SAL**: Saliency-Aware geodesic video object segmentation. *J. Shen et al. CVPR 2015*
- **FST**: Fast object segmentation in unconstrained video. *A. Papazoglou et al*. ICCV 2013
- **TSP**: A video representation using temporal superpixels. *J. Chang et al. CVPR 2013*
- **SEA**: Seamseg: Video object segmentation using patch seams. *S. A. Ramakanth and R. V. Babu CVPR 2014*
- **HVS**: Effi- cient hierarchical graph-based video segmentation. *M. Grundmann et al. CVPR 2010*
- **JMP**: Jumpcut: Non-successive mask transfer and interpolation for video cutout. *Q. Fan et al. SIGGRAPH ASIA 2015*
- **FCOP**: Fully connected object proposals for video segmentation. *F. Perazzi et al. ICCV 2015*