A Benchmark Dataset and Evaluation Methodology for Video Object Segmentation

Supplemental Material

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

1. Contents

In this document we include additional material related to the dataset and the evaluation.

- **Dataset**. In Figure 1 we visualize a representative frame of each sequence with ground-truth masks overlayed in red. We refer the reader to the accompanying video for a complete visualization of the dataset. A summary of the requirements detailed in Section 3 (*manuscript*) can be found in Table 1.
- **Attributes** In Table 3 we summarize the attributes assigned to each specific sequence. We refer the reader to Table 1 (*manuscript*) for a comprehensive description of the attributes. In Section 2 we provide an in-depth description of the statistical procedure we adopted to determine the attribute dependencies, visualized in Figure 2 (*right*, *manuscript*).
- **Evaluation and Running Times**. In Figure 2 we visualize the mean performance of all evaluated approaches, based the per-sequence region-similarity $\mathcal{J}$ and contour accuracy $\mathcal{F}$. The results are an estimator of the expected segmentation *difficulty* for a specific sequence. Sequences are sorted with respect to the estimated difficulty. In Tables 4, 5, 6 we summarize the per-sequence performance of each method, respectively in terms of region similarity $\mathcal{J}$, contour accuracy $\mathcal{F}$ and temporal stability $\mathcal{T}$. In Figures 3, 4, 5, 6. Estimated running times are given in Table 2.

2. Attributes Dependency - In Depth

We consider the presence or absence of each attribute in a video sequence to be represented as a binary random variable $X = (X_1, X_2, \ldots, X_d)$. The dependencies between the attributes can be modelled by a pairwise Markov random field (MRF) defined on the undirected graph $G = (V, E)$ where $V = \{1, \ldots, d\}$ is the set of vertices and E is the (unknown) set of edges. Each variable X_s is associated with a vertex $s \in V$. The pairwise MRF associated with G is the family of distributions which factorise as

$$\mathbb{P}_\theta(x) \propto \exp\left\{ \sum_{(s,t)\in E} \theta_{s,t}\, x_s x_t \right\}.$$

The absence of an edge between s and t means that X_s and X_t are *independent* conditioned on their respective Markov blankets[1]. In other words, given the state of the neighbours of s and t, knowing t gives us no information about s and *vice-versa*.

Equivalently, θ can be viewed as a $\binom{d}{2}$-dimensional vector which indexes all distinct pairs of vertices but is non-zero only when the vertex pair (s, t) belongs to the edge set E of the graph. Recovering E is equivalent to recovering the neighbourhood set $\mathcal{N}(r) := \{t \in V | (r, t) \in E\}$ for each $r \in V$. Estimating the neighbourhood set $\mathcal{N}(r)$ is equivalent to estimating the support (i.e. location of non-zero entries) of the $(d-1)$ dimensional sub-vector $\theta_{\setminus r} := \{\theta_{ru}, u \in V \setminus r\}$.

Dataset	SIZE	HD-Q	VARY	DENSE-GT	OBJ
OURS	✓	✓	✓	✓	✓
MoSeg [1]	✓				✓
BVSD [5]	✓	✓	✓		
SegTrack [6]				✓	✓
SegTrack v2 [2]			✓	✓	✓

Table 1: Summary of requirements fulfilled by datasets most relevant to video object segmentation. *From left*: large overall size of the dataset (SIZE), high-resolution videos (HD-Q), variety of content and challenges (VARY), pixel-accurate, per-frame ground-truth (DENSE-GT) and object presence (OBJ). A detailed overview of the requirements is described in Section 3. Our dataset is the only one meeting all requirements.

[1]For a MRF this consists of the neighbours of s and t, respectively

	Preprocessing			Unsupervised							Semi-Supervised				
	MCG	SF-LAB	SF-MOT	NLC	CVOS	TRC	MSG	KEY	SAL	FST	TSP	SEA	HVS	JMP	FCP
Time	1498.8s	85.3s	84.9s	880.3s	6190.4s	16612.3	2614.9s	55191.1s	149.4s	2319.2s	2870.3s	299.1s	634.1s	576.6s	2261.7s

Table 2: *Running times*. Estimated running times (in seconds) for each of the evaluated approaches. Due to the substantial processing power required to carry out this large scale evaluation, we used multiple machines and a cluster with thousands nodes and different CPU, therefore while computing times have been normalized to comparable processing power, they should be considered an aproximate estimate.

Following [4], given a collection of n observations $\mathcal{X}^n = \{x^{(1)}, \ldots, x^{(n)}\}$ of d-dimensional binary vectors $x^{(i)}$, the support of each $\theta_{\setminus r}$ can be estimated by solving the following minimization problem

$$\min_{\theta_{\setminus r} \in \mathbb{R}^{d-1}} -\frac{1}{n} \sum_{i=1}^{n} \log \mathbb{P}_\theta(x_r^{(i)} | x_{\setminus r}^{(i)}) + \lambda \|\theta_{\setminus r}\|. \qquad (1)$$

Since the random variables are binary, minimizing the penalised negative log likelihood above corresponds to solving ℓ_1 penalised logistic regression treating $x_{\setminus r}^{(i)} \in \mathbb{R}^{n \times (d-1)}$ as covariates and $x_r^{(i)}$ as the response.

The solution to (1) can be highly sensitive to the regularization strength λ which controls the sparsity of the solution. In order to determine the correct degree of sparsity we employ *stability selection* [3]. Briefly, this amounts to performing the above procedure on multiple $n/2$-sized subsamples of the data and computing the proportion of times each edge is selected. Setting an appropriate threshold on this selection probability allows us to control the number of wrongly estimated edges according to Theorem 1 in [3]. For example, for a threshold value of 0.6 and choosing a value of λ which on average selects neighbourhoods of size 4, the number of wrongly selected edges is at most 4 (out of $16^2 = 256$ possible edges).

References

[1] T. Brox and J. Malik. Object segmentation by long term analysis of point trajectories. In *ECCV*, 2010. 1

[2] F. Li, T. Kim, A. Humayun, D. Tsai, and J. M. Rehg. Video segmentation by tracking many figure-ground segments. In *ICCV*, 2013. 1

[3] N. Meinshausen and P. Bühlmann. Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4):417–473, 2010. 2

[4] P. Ravikumar, M. J. Wainwright, J. D. Lafferty, et al. High-dimensional ising model selection using l1-regularized logistic regression. *The Annals of Statistics*, 38(3):1287–1319, 2010. 2

[5] P. Sundberg, T. Brox, M. Maire, P. Arbelaez, and J. Malik. Occlusion boundary detection and figure/ground assignment from optical flow. In *CVPR*, 2011. 1

[6] D. Tsai, M. Flagg, and J. M. Rehg. Motion coherent tracking with multi-label MRF optimization. In *BMVC*, 2010. 1

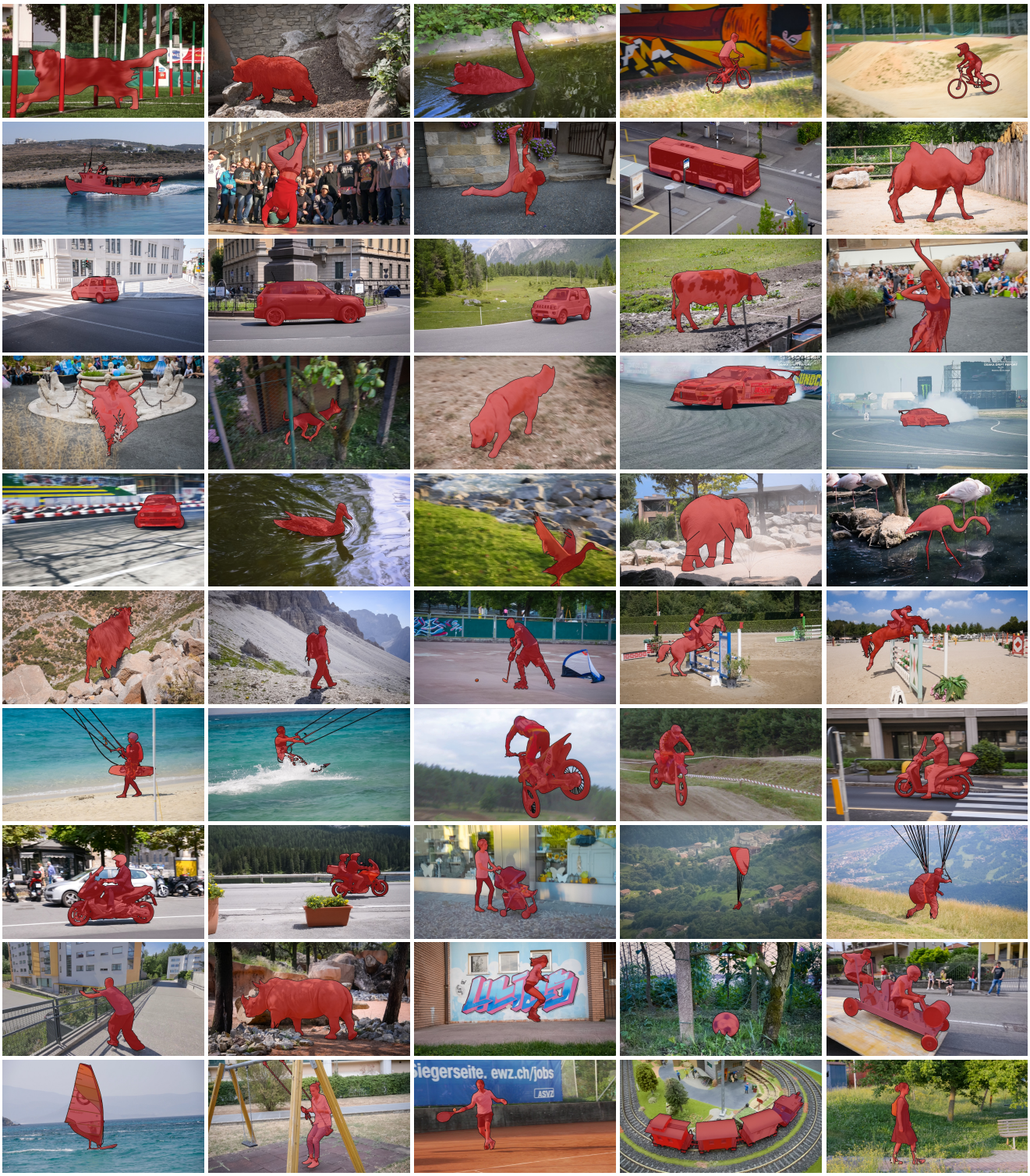

Figure 1: Sample sequences from our dataset, with ground truth segmentation masks overlayed. Please refer to the accompanying video for a complete visualization of the dataset.

Sequence	AC	BC	CS	DB	DEF	EA	FM	HO	IO	LR	MB	OCC	OV	SC	SV
bear					✓										
blackswan															
bmx-bumps			✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
bmx-trees		✓	✓		✓	✓	✓	✓	✓	✓	✓	✓		✓	✓
boat		✓		✓		✓								✓	
breakdance	✓			✓	✓		✓	✓			✓		✓		
breakdance-flare			✓		✓		✓	✓			✓				
bus						✓		✓				✓		✓	
camel			✓		✓				✓						
car-roundabout		✓													
car-shadow	✓	✓				✓				✓					
car-turn		✓													✓
cows			✓		✓			✓	✓			✓			
dance-jump				✓	✓	✓		✓			✓	✓		✓	
dance-twirl			✓		✓			✓	✓		✓		✓	✓	
dog			✓		✓	✓	✓				✓				
dog-agility	✓				✓	✓	✓	✓			✓	✓	✓		
drift-chicane	✓			✓		✓	✓	✓		✓					✓
drift-straight	✓		✓			✓	✓	✓		✓	✓		✓		✓
drift-turn	✓			✓			✓	✓	✓				✓		✓
elephant			✓	✓	✓	✓									
flamingo				✓	✓			✓	✓					✓	
goat		✓	✓		✓	✓									
hike					✓			✓		✓					
hockey					✓			✓	✓					✓	
horsejump-high					✓			✓	✓			✓		✓	
horsejump-low					✓	✓		✓	✓			✓		✓	
kite-surf				✓		✓		✓	✓		✓	✓		✓	✓
kite-walk				✓	✓			✓	✓			✓		✓	
libby					✓	✓		✓			✓	✓		✓	
lucia					✓			✓				✓			
mallard-fly	✓			✓	✓	✓	✓			✓	✓		✓		✓
mallard-water				✓		✓			✓	✓					
motocross-bumps	✓	✓					✓	✓	✓				✓		✓
motocross-jump	✓				✓	✓	✓	✓	✓		✓		✓	✓	✓
motorbike						✓	✓	✓	✓	✓		✓		✓	✓
paragliding								✓	✓	✓				✓	
paragliding-launch					✓	✓		✓	✓					✓	
parkour	✓				✓		✓	✓		✓		✓			✓
rhino		✓			✓							✓			
rollerblade			✓		✓		✓	✓		✓	✓				
scooter-black						✓		✓	✓						✓
scooter-gray		✓				✓	✓	✓	✓			✓		✓	
soapbox	✓				✓			✓	✓		✓				✓
soccerball							✓	✓		✓	✓	✓			
stroller			✓		✓		✓	✓	✓					✓	
surf			✓	✓			✓	✓	✓				✓		✓
swing					✓		✓	✓	✓			✓		✓	
tennis					✓		✓	✓	✓		✓				✓
train						✓		✓						✓	

Table 3: *List of attributes for each video in the dataset.*. Left to right: appearance changes (AC), background clutter (BC), camera shake (CS), dynamic background (DB), non-linear deformation (DEF), edge ambiguity (EA), fast-motion (FM), heterogeneus object (HO), interacting objects (IO), low resolution (LR), motion blur (MB), occlusions (OCC), out-of-view (OV), shape complexity (SC), scale variation (SV). See Table 1 in the paper for the description of each attribute.

	Preprocessing			Unsupervised							Semi-Supervised				
Sequence	MCG	SF-LAB	SF-MOT	NLC	CVOS	TRC	MSG	KEY	SAL	FST	TSP	SEA	HVS	JMP	FCP
bear	**0.937**	0.126	0.556	**0.906**	0.864	0.873	0.851	0.891	0.657	0.898	0.778	0.912	**0.938**	0.929	0.906
blackswan	**0.871**	0.524	0.547	**0.874**	0.422	0.569	0.526	0.842	0.222	0.732	0.872	**0.933**	0.917	0.930	0.908
bmx-bumps	**0.490**	0.030	0.281	**0.635**	0.368	0.350	0.353	0.309	0.188	0.241	0.290	0.198	**0.428**	0.336	0.300
bmx-trees	**0.473**	0.021	0.468	**0.212**	0.121	0.162	0.188	0.193	0.194	0.180	0.095	0.113	0.178	0.229	**0.248**
boat	**0.619**	0.068	0.171	0.007	0.056	0.130	0.144	0.065	0.271	**0.361**	0.656	**0.793**	0.782	0.705	0.613
breakdance	**0.713**	0.204	0.649	**0.673**	0.183	0.114	0.237	0.549	0.422	0.467	0.056	0.329	0.550	0.478	**0.567**
breakdance-flare	**0.733**	0.206	0.629	**0.804**	0.317	0.245	0.157	0.559	0.476	0.616	0.040	0.131	0.499	0.430	**0.723**
bus	0.749	0.290	**0.797**	0.629	0.664	0.684	**0.885**	0.785	0.739	0.825	0.515	0.752	0.809	0.668	**0.832**
camel	**0.795**	0.015	0.620	0.768	**0.850**	0.778	0.756	0.579	0.320	0.562	0.654	0.649	**0.876**	0.640	0.734
car-roundabout	**0.786**	0.306	0.708	0.509	**0.871**	0.552	0.630	0.640	0.500	0.808	0.614	0.708	**0.777**	0.726	0.717
car-shadow	0.701	0.135	**0.765**	0.645	0.759	0.449	**0.880**	0.589	0.538	0.698	0.636	**0.775**	0.699	0.645	0.723
car-turn	**0.865**	0.071	0.593	0.833	0.820	0.805	0.621	0.806	0.611	**0.851**	0.323	**0.909**	0.810	0.834	0.724
cows	**0.811**	0.139	0.684	**0.883**	0.562	0.833	0.799	0.337	0.623	0.791	0.595	0.707	0.779	0.756	**0.812**
dance-jump	0.470	0.251	**0.582**	0.718	0.341	0.303	0.065	**0.748**	0.291	0.598	0.132	0.662	**0.680**	0.490	0.522
dance-twirl	**0.644**	0.093	0.624	0.347	0.452	0.366	0.366	0.380	0.372	**0.453**	0.099	0.117	0.318	0.444	**0.471**
dog	**0.621**	0.195	0.532	**0.809**	0.753	0.786	0.331	0.692	0.566	0.708	0.313	0.581	0.722	0.673	**0.774**
dog-agility	**0.663**	0.060	0.354	**0.652**	0.193	0.138	0.110	0.132	0.055	0.280	0.079	0.354	0.457	**0.699**	0.453
drift-chicane	**0.806**	0.046	0.396	0.324	0.313	0.722	**0.758**	0.188	0.244	0.667	0.018	0.119	0.331	0.243	**0.457**
drift-straight	**0.753**	0.171	0.427	0.473	0.344	0.431	0.575	0.194	0.268	**0.683**	0.197	0.513	0.295	0.618	**0.668**
drift-turn	**0.856**	0.161	0.359	0.154	0.615	0.412	**0.638**	0.255	0.349	0.533	0.162	0.667	0.276	**0.717**	0.606
elephant	**0.686**	0.099	0.640	0.518	0.494	0.760	0.689	0.675	0.510	**0.824**	0.666	0.553	0.742	**0.750**	0.655
flamingo	**0.850**	0.247	0.517	0.539	0.783	0.731	0.794	0.692	0.570	**0.817**	0.666	0.583	**0.811**	0.530	0.717
goat	**0.641**	0.057	0.138	0.010	0.074	**0.793**	0.736	0.705	0.257	0.554	0.444	0.535	0.580	**0.731**	0.677
hike	**0.900**	0.335	0.657	**0.918**	0.878	0.756	0.603	0.895	0.683	0.889	0.679	0.776	**0.877**	0.664	0.874
hockey	**0.775**	0.080	0.538	0.810	**0.817**	0.674	0.713	0.515	0.566	0.468	0.413	**0.714**	0.698	0.677	0.647
horsejump-high	**0.649**	0.263	0.596	**0.834**	0.830	0.364	0.734	0.370	0.568	0.578	0.236	0.637	**0.765**	0.586	0.676
horsejump-low	0.545	0.125	**0.618**	0.651	**0.743**	0.705	0.682	0.630	0.388	0.526	0.291	0.498	0.551	**0.663**	0.607
kite-surf	**0.654**	0.059	0.208	0.453	0.357	0.501	0.419	**0.585**	0.193	0.272	0.366	0.487	0.405	0.500	**0.577**
kite-walk	**0.736**	0.668	0.420	**0.813**	0.447	0.052	0.597	0.197	0.725	0.649	0.447	0.498	**0.765**	0.509	0.682
libby	**0.655**	0.097	0.443	**0.635**	0.169	0.073	0.050	0.611	0.470	0.507	0.070	0.226	**0.553**	0.295	0.316
lucia	**0.820**	0.119	0.760	**0.876**	0.840	0.669	0.417	0.847	0.706	0.644	0.377	0.626	0.776	**0.836**	0.801
mallard-fly	**0.799**	0.022	0.310	**0.617**	0.380	0.293	0.033	0.585	0.227	0.601	0.200	**0.557**	0.436	0.536	0.541
mallard-water	**0.755**	0.035	0.017	0.761	0.245	0.190	0.045	**0.785**	0.085	0.087	0.623	**0.865**	0.704	0.751	0.687
motocross-bumps	**0.827**	0.236	0.460	0.614	0.603	0.502	0.466	**0.689**	0.351	0.617	0.133	0.470	0.534	**0.761**	0.306
motocross-jump	**0.760**	0.204	0.428	0.251	0.245	0.338	**0.618**	0.288	0.491	0.602	0.123	0.386	0.099	**0.583**	0.511
motorbike	**0.688**	0.116	0.572	0.714	0.387	0.723	**0.737**	0.572	0.335	0.558	0.340	0.451	0.687	0.506	**0.713**
paragliding	**0.877**	0.140	0.743	0.880	0.890	0.816	**0.933**	0.861	0.568	0.725	0.735	0.863	0.907	**0.951**	0.866
paragliding-launch	**0.599**	0.255	0.501	**0.628**	0.591	0.555	0.513	0.559	0.539	0.506	0.301	0.577	0.537	**0.589**	0.571
parkour	**0.815**	0.288	0.491	**0.901**	0.146	0.345	0.295	0.410	0.392	0.458	0.070	0.121	0.240	**0.342**	0.322
rhino	**0.864**	0.253	0.615	0.682	0.520	0.846	**0.902**	0.675	0.685	0.776	0.694	0.736	**0.812**	0.716	0.794
rollerblade	0.554	0.004	**0.564**	**0.814**	0.406	0.566	0.801	0.510	0.141	0.318	0.098	0.138	0.461	**0.726**	0.449
scooter-black	**0.704**	0.125	0.613	0.162	**0.759**	0.435	0.579	0.502	0.348	0.522	0.378	**0.793**	0.624	0.626	0.504
scooter-gray	0.653	0.039	**0.703**	**0.586**	0.327	0.357	0.345	0.363	0.421	0.325	0.133	0.241	0.433	0.123	**0.483**
soapbox	**0.680**	0.163	0.579	0.634	**0.832**	0.294	0.672	0.757	0.332	0.410	0.247	**0.783**	0.684	0.759	0.449
soccerball	**0.856**	0.052	0.732	0.829	0.242	0.350	0.370	**0.878**	0.378	0.843	0.029	0.653	0.065	0.096	**0.820**
stroller	**0.600**	0.313	0.573	**0.850**	0.619	0.720	0.678	0.759	0.466	0.580	0.369	0.464	**0.662**	0.656	0.597
surf	**0.944**	0.521	0.653	0.775	0.273	0.464	0.770	**0.893**	0.312	0.475	0.814	0.821	0.759	**0.941**	0.843
swing	**0.709**	0.387	0.677	**0.851**	0.533	0.413	0.622	0.710	0.569	0.431	0.098	0.511	0.104	0.115	**0.648**
tennis	**0.714**	0.032	0.562	**0.871**	0.494	0.196	0.590	0.762	0.480	0.388	0.074	0.482	0.576	**0.765**	0.623
train	0.535	0.205	**0.548**	0.729	**0.903**	0.876	0.887	0.450	0.620	0.831	0.648	0.854	0.846	**0.873**	0.841
Mean	**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**

Table 4: *Results of region similarity ($\mathcal{J}$) for each video sequence in the dataset.* The best performing method of each category is highlighted in bold.

	Preprocessing			Unsupervised							Semi-Supervised				
Sequence	MCG	SF-LAB	SF-MOT	NLC	CVOS	TRC	MSG	KEY	SAL	FST	TSP	SEA	HVS	JMP	FCP
bear	**0.934**	0.180	0.451	0.850	0.845	0.832	0.781	0.775	0.495	**0.860**	0.635	0.899	**0.905**	0.904	0.845
blackswan	**0.873**	0.491	0.500	**0.820**	0.695	0.654	0.700	0.787	0.430	0.736	0.857	**0.957**	0.910	0.945	0.905
bmx-bumps	**0.597**	0.106	0.283	**0.734**	0.409	0.325	0.410	0.453	0.313	0.349	0.338	0.254	**0.525**	0.397	0.340
bmx-trees	**0.605**	0.145	0.557	0.330	0.118	0.189	0.263	**0.366**	0.206	0.348	0.138	0.125	0.282	0.309	**0.324**
boat	**0.536**	0.289	0.135	0.036	0.108	0.403	**0.485**	0.000	0.264	0.197	0.682	0.764	**0.807**	0.607	0.460
breakdance	**0.670**	0.265	0.635	**0.661**	0.191	0.121	0.231	0.463	0.300	0.411	0.070	0.389	0.473	**0.511**	0.473
breakdance-flare	**0.781**	0.215	0.626	**0.808**	0.335	0.301	0.230	0.585	0.512	0.694	0.116	0.167	0.625	0.523	**0.738**
bus	0.497	0.296	**0.505**	0.406	0.535	0.542	**0.657**	0.635	0.570	0.584	0.477	**0.724**	0.682	0.604	0.539
camel	**0.728**	0.085	0.498	0.719	**0.873**	0.698	0.629	0.437	0.432	0.590	0.529	0.614	**0.871**	0.711	0.617
car-roundabout	**0.512**	0.246	0.491	0.250	**0.678**	0.451	0.602	0.362	0.301	0.625	0.435	**0.710**	0.551	0.619	0.478
car-shadow	0.588	0.233	**0.628**	0.546	0.617	0.474	**0.858**	0.459	0.441	0.540	0.513	**0.755**	0.594	0.625	0.642
car-turn	**0.760**	0.113	0.431	0.634	0.703	**0.741**	0.677	0.632	0.485	0.731	0.379	**0.883**	0.605	0.742	0.614
cows	**0.736**	0.168	0.554	**0.807**	0.499	0.721	0.621	0.293	0.499	0.681	0.544	0.677	0.632	**0.700**	0.667
dance-jump	0.372	0.204	**0.497**	0.567	0.282	0.272	0.038	**0.569**	0.262	0.462	0.186	0.567	**0.571**	0.526	0.418
dance-twirl	**0.604**	0.121	0.589	0.365	0.444	0.376	0.325	0.317	0.301	**0.471**	0.128	0.213	0.516	**0.520**	0.427
dog	**0.582**	0.140	0.495	0.707	**0.761**	0.695	0.304	0.633	0.418	0.659	0.295	0.543	0.635	0.596	**0.672**
dog-agility	**0.517**	0.153	0.288	**0.551**	0.262	0.122	0.076	0.095	0.102	0.265	0.083	0.410	0.446	**0.654**	0.315
drift-chicane	**0.906**	0.154	0.364	0.312	0.397	0.823	**0.886**	0.192	0.206	0.731	0.033	0.159	**0.547**	0.338	0.477
drift-straight	**0.593**	0.152	0.312	0.385	0.330	0.408	**0.509**	0.053	0.167	0.470	0.213	**0.500**	0.266	0.473	0.479
drift-turn	**0.700**	0.207	0.199	0.185	**0.480**	0.310	0.459	0.018	0.231	0.442	0.217	0.512	0.216	**0.631**	0.488
elephant	**0.506**	0.080	0.446	0.251	0.359	0.546	0.505	0.324	0.231	**0.569**	0.523	0.399	**0.579**	0.542	0.430
flamingo	**0.876**	0.253	0.622	0.610	**0.806**	0.663	0.776	0.589	0.621	0.763	0.544	0.563	**0.790**	0.650	0.641
goat	**0.526**	0.171	0.225	0.133	0.241	**0.724**	0.657	0.552	0.187	0.400	0.404	0.470	0.546	**0.617**	0.576
hike	**0.926**	0.312	0.540	**0.943**	0.922	0.804	0.702	0.925	0.691	0.918	0.675	0.796	0.878	0.744	**0.912**
hockey	**0.742**	0.198	0.430	**0.808**	0.789	0.651	0.761	0.560	0.559	0.584	0.579	0.721	**0.778**	0.726	0.612
horsejump-high	**0.703**	0.326	0.561	**0.881**	0.841	0.405	0.748	0.392	0.613	0.621	0.343	0.655	**0.807**	0.653	0.699
horsejump-low	**0.548**	0.152	0.516	0.659	**0.709**	0.672	0.637	0.533	0.419	0.490	0.356	0.548	0.572	**0.696**	0.533
kite-surf	**0.447**	0.221	0.286	0.448	0.241	0.422	**0.521**	0.504	0.368	0.346	0.268	0.285	**0.375**	0.309	0.362
kite-walk	0.486	**0.524**	0.285	**0.662**	0.438	0.014	0.577	0.128	0.526	0.561	0.435	0.355	**0.624**	0.359	0.411
libby	**0.733**	0.244	0.581	**0.748**	0.185	0.086	0.118	0.730	0.529	0.718	0.091	0.209	**0.641**	0.365	0.389
lucia	**0.743**	0.233	0.736	**0.872**	0.801	0.663	0.491	0.819	0.691	0.568	0.453	0.542	0.782	**0.870**	0.708
mallard-fly	**0.824**	0.071	0.338	**0.660**	0.391	0.332	0.019	0.631	0.293	0.633	0.235	**0.607**	0.441	0.579	0.539
mallard-water	**0.701**	0.115	0.034	0.692	0.254	0.225	0.000	**0.733**	0.115	0.079	0.585	**0.886**	0.646	0.755	0.557
motocross-bumps	**0.710**	0.242	0.338	0.560	0.567	0.497	0.466	**0.674**	0.300	0.610	0.184	0.520	0.548	**0.743**	0.302
motocross-jump	**0.568**	0.274	0.290	0.303	0.186	0.307	0.393	0.237	0.388	**0.453**	0.116	0.404	0.137	**0.539**	0.386
motorbike	**0.642**	0.247	0.264	0.571	0.380	0.541	0.594	**0.726**	0.391	0.585	0.406	0.481	**0.823**	0.578	0.632
paragliding	0.742	0.133	**0.798**	0.744	0.744	0.724	**0.909**	0.681	0.541	0.675	0.634	0.744	0.857	**0.907**	0.727
paragliding-launch	**0.200**	0.188	0.171	0.243	0.182	0.157	0.196	**0.253**	0.169	0.185	0.122	0.180	**0.206**	0.176	0.183
parkour	**0.811**	0.304	0.553	**0.916**	0.158	0.421	0.401	0.374	0.359	0.478	0.094	0.278	0.323	**0.418**	0.292
rhino	**0.767**	0.193	0.472	0.431	0.469	0.739	**0.826**	0.429	0.487	0.634	0.499	0.658	**0.658**	0.653	0.647
rollerblade	0.577	0.114	**0.586**	**0.868**	0.475	0.687	0.822	0.351	0.211	0.411	0.143	0.155	0.552	**0.759**	0.576
scooter-black	**0.548**	0.177	0.451	0.228	0.557	0.304	**0.565**	0.420	0.257	0.395	0.411	**0.722**	0.574	0.529	0.363
scooter-gray	**0.494**	0.126	0.434	**0.467**	0.212	0.266	0.272	0.367	0.333	0.321	0.122	0.275	**0.545**	0.123	0.437
soapbox	**0.647**	0.190	0.453	0.658	**0.754**	0.389	0.633	0.719	0.307	0.355	0.336	**0.750**	0.690	0.677	0.423
soccerball	**0.898**	0.043	0.754	0.855	0.262	0.377	0.401	**0.924**	0.355	0.900	0.059	0.697	0.074	0.127	**0.836**
stroller	**0.635**	0.426	0.488	**0.874**	0.606	0.691	0.662	0.751	0.417	0.558	0.404	0.525	0.708	**0.718**	0.581
surf	**0.881**	0.524	0.372	0.673	0.515	0.637	0.804	**0.820**	0.395	0.445	0.641	0.732	0.652	**0.872**	0.713
swing	**0.599**	0.373	0.571	**0.778**	0.493	0.417	0.611	0.614	0.502	0.491	0.087	0.409	0.091	0.109	**0.538**
tennis	**0.751**	0.151	0.450	**0.927**	0.547	0.301	0.670	0.818	0.530	0.567	0.114	0.537	0.579	**0.818**	0.652
train	0.403	0.347	**0.504**	0.521	**0.831**	0.766	0.770	0.464	0.440	0.660	0.589	0.713	0.688	**0.770**	0.736
Mean	**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

Table 5: *Results of boundary precision ($\mathcal{F}$) for each video sequence in the dataset.* The best performing method of each category is highlighted in bold.

	Preprocessing			Unsupervised							Semi-Supervised				
Sequence	MCG	SF-LAB	SF-MOT	NLC	CVOS	TRC	MSG	KEY	SAL	FST	TSP	SEA	HVS	JMP	FCP
bear	**0.222**	0.633	0.840	0.151	**0.059**	0.272	0.156	0.068	0.448	0.227	0.077	**0.047**	0.086	0.051	0.114
blackswan	**0.309**	0.332	0.322	0.110	0.058	0.219	0.145	**0.049**	0.660	0.225	0.049	0.032	0.060	**0.029**	0.064
boat	0.813	**0.437**	0.971	0.559	1.159	0.350	0.163	**0.015**	0.382	0.177	0.067	**0.055**	0.125	0.062	0.136
bus	0.741	0.757	**0.397**	0.178	0.146	0.194	0.154	**0.143**	0.369	0.270	0.293	**0.109**	0.306	0.193	0.156
camel	0.593	0.438	**0.341**	0.232	**0.123**	0.172	0.129	0.138	0.380	0.161	0.084	**0.055**	0.117	0.062	0.212
car-roundabout	0.545	1.178	**0.428**	0.352	**0.064**	0.382	0.291	0.161	0.536	0.242	0.158	**0.071**	0.255	0.078	0.283
car-shadow	1.162	1.270	**0.227**	0.361	**0.180**	0.452	0.206	0.313	0.793	0.353	**0.206**	0.230	0.351	0.274	0.339
car-turn	**0.274**	0.714	0.477	0.235	0.118	0.202	0.204	**0.108**	0.566	0.214	0.796	0.117	0.135	**0.065**	0.256
cows	0.494	1.022	**0.457**	0.147	**0.133**	0.148	0.196	0.412	0.511	0.282	0.179	**0.044**	0.164	0.055	0.163
dance-jump	1.083	**0.630**	0.870	0.316	0.459	0.576	**0.110**	0.214	0.586	0.242	0.272	0.286	0.324	**0.173**	0.506
drift-straight	0.969	1.170	**0.775**	0.599	0.900	0.638	0.543	**0.292**	0.950	0.482	0.826	0.396	0.823	**0.317**	0.597
drift-turn	**0.391**	0.987	0.984	0.850	0.334	0.475	0.402	**0.150**	1.002	0.258	0.633	0.168	0.703	**0.128**	0.328
elephant	1.103	0.960	**0.557**	0.315	0.118	0.236	0.236	**0.085**	0.426	0.139	0.097	0.076	0.213	**0.075**	0.404
flamingo	0.523	**0.319**	0.769	0.138	0.173	0.215	0.382	**0.113**	0.486	0.175	0.118	**0.069**	0.133	0.089	0.182
hike	0.331	0.873	**0.282**	0.158	0.125	0.230	0.251	**0.117**	0.412	0.247	0.141	0.122	0.120	**0.092**	0.164
hockey	0.519	1.052	**0.378**	0.227	**0.159**	0.228	0.211	0.162	0.377	0.276	0.403	0.103	0.258	**0.102**	0.228
kite-surf	**0.420**	0.464	0.888	0.944	0.248	0.432	0.507	**0.233**	0.568	0.404	0.278	0.125	0.497	**0.117**	0.212
kite-walk	0.409	**0.322**	0.829	0.221	0.127	**0.002**	0.328	0.366	0.356	0.301	0.241	0.173	0.185	**0.154**	0.166
mallard-water	0.815	**0.579**	1.579	0.242	0.394	0.641	**0.000**	0.184	1.070	0.230	0.287	**0.123**	0.295	0.219	0.317
motocross-bumps	**0.606**	0.756	0.919	0.541	**0.327**	0.566	0.481	0.344	0.903	0.329	0.628	0.289	0.767	**0.211**	0.486
paragliding-launch	0.498	0.598	**0.324**	0.259	0.273	0.347	0.331	**0.213**	0.602	0.703	0.660	0.208	0.316	**0.180**	0.329
rhino	**0.308**	0.396	0.700	0.188	0.064	0.153	0.093	**0.056**	0.390	0.138	0.066	**0.037**	0.093	0.037	0.151
scooter-black	1.020	0.910	**0.419**	0.760	**0.320**	0.577	0.364	0.558	0.790	0.475	0.960	**0.216**	0.599	0.282	0.423
soapbox	0.983	0.613	**0.504**	0.390	**0.154**	0.413	0.214	0.160	0.613	0.158	0.763	**0.112**	0.314	0.126	0.379
stroller	1.195	0.968	**0.864**	0.205	**0.116**	0.235	0.366	0.127	0.546	0.184	0.134	**0.130**	0.363	0.189	0.376
surf	**0.234**	1.338	0.728	0.364	0.169	0.375	0.223	**0.086**	1.093	0.398	0.207	0.238	0.291	**0.127**	0.276
train	1.038	0.739	**0.377**	0.576	**0.056**	0.110	0.070	0.270	0.396	0.159	0.249	0.069	0.106	**0.047**	0.448
Mean	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

Table 6: *Results of temporal stability ($\mathcal{T}$) for each video sequence in the dataset*. The best performing method of each category is highlighted in bold. Please note that this measure is only computed on those sequences without occlusions and strong deformations.

	Unsupervised							Semi-Supervised				
Attr	NLC	CVOS	TRC	MSG	KEY	SAL	FST	TSP	SEA	HVS	JMP	FCOP
LR	**0.65 *-0.03***	0.40 *+0.14*	0.45 *+0.05*	0.48 *+0.06*	0.53 *+0.05*	0.32 *+0.13*	0.52 *+0.06*	0.29 *+0.08*	0.46 *+0.11*	0.46 *+0.16*	0.50 *+0.13*	**0.58 *+0.06***
SV	**0.52 *+0.16***	0.42 *+0.12*	0.42 *+0.11*	0.50 *+0.05*	0.49 *+0.10*	0.33 *+0.14*	0.49 *+0.11*	0.23 *+0.18*	0.48 *+0.09*	0.44 *+0.21*	**0.57 *+0.04***	0.51 *+0.16*
SC	**0.59 *+0.06***	0.49 *+0.02*	0.46 *+0.06*	0.52 *+0.01*	0.50 *+0.11*	0.45 *-0.05*	0.51 *+0.09*	0.32 *+0.05*	0.50 *+0.09*	0.56 *+0.05*	0.51 *+0.14*	**0.58 *+0.07***
FM	**0.62 *+0.01***	0.36 *+0.25*	0.40 *+0.16*	0.44 *+0.15*	0.49 *+0.12*	0.34 *+0.13*	0.49 *+0.13*	0.17 *+0.31*	0.39 *+0.28*	0.40 *+0.31*	0.49 *+0.18*	**0.54 *+0.13***
CS	**0.59 *+0.06***	0.41 *+0.12*	0.54 *-0.06*	0.53 *-0.00*	0.51 *+0.06*	0.36 *+0.09*	0.53 *+0.05*	0.34 *+0.01*	0.42 *+0.17*	0.54 *+0.06*	0.60 *+0.00*	**0.60 *+0.03***
IO	**0.61 *+0.03***	0.54 *-0.06*	0.46 *+0.05*	0.57 *-0.07*	0.53 *+0.05*	0.41 *+0.02*	0.48 *+0.16*	0.34 *+0.03*	0.53 *+0.03*	0.55 *+0.06*	0.58 *+0.04*	**0.59 *+0.07***
DB	**0.52 *+0.14***	0.37 *+0.18*	0.38 *+0.14*	0.43 *+0.14*	0.51 *+0.07*	0.34 *+0.10*	0.52 *+0.06*	0.39 *-0.06*	0.57 *-0.03*	0.59 *-0.01*	0.59 *+0.01*	**0.61 *+0.01***
MB	**0.59 *+0.05***	0.35 *+0.23*	0.31 *+0.28*	0.33 *+0.30*	0.50 *+0.08*	0.32 *+0.15*	0.47 *+0.15*	0.14 *+0.32*	0.39 *+0.24*	0.42 *+0.25*	0.50 *+0.15*	**0.52 *+0.15***
DEF	**0.67 *-0.11***	0.51 *-0.01*	0.48 *+0.01*	0.51 *+0.05*	0.57 *-0.01*	0.45 *-0.08*	0.56 *+0.01*	0.31 *+0.10*	0.49 *+0.14*	0.58 *+0.00*	0.58 *+0.03*	**0.60 *+0.04***
OCC	**0.68 *-0.08***	0.42 *+0.13*	0.42 *+0.11*	0.46 *+0.11*	0.51 *+0.08*	0.43 *-0.02*	0.52 *+0.08*	0.26 *+0.14*	0.46 *+0.13*	0.51 *+0.12*	0.46 *+0.21*	**0.58 *+0.07***
HO	**0.64 *-0.04***	0.49 *+0.05*	0.45 *+0.14*	0.53 *+0.01*	0.53 *+0.12*	0.42 *-0.01*	0.54 *+0.10*	0.27 *+0.29*	0.49 *+0.23*	0.53 *+0.21*	0.55 *+0.17*	**0.59 *+0.13***
EA	0.50 *+0.24*	0.40 *+0.19*	0.45 *+0.07*	0.45 *+0.16*	0.48 *+0.15*	0.35 *+0.12*	**0.51 *+0.10***	0.31 *+0.06*	0.51 *+0.07*	0.53 *+0.10*	0.54 *+0.09*	**0.57 *+0.10***
OV	**0.49 *+0.17***	0.34 *+0.20*	0.32 *+0.21*	0.39 *+0.17*	0.42 *+0.18*	0.30 *+0.15*	0.48 *+0.10*	0.20 *+0.19*	0.43 *+0.14*	0.39 *+0.24*	**0.59 *+0.00***	0.51 *+0.13*
AC	0.53 *+0.13*	0.41 *+0.12*	0.36 *+0.17*	0.47 *+0.08*	0.41 *+0.19*	0.33 *+0.12*	**0.54 *+0.04***	0.17 *+0.23*	0.45 *+0.12*	0.41 *+0.22*	**0.57 *+0.04***	0.50 *+0.16*
BC	0.45 *+0.22*	0.45 *+0.06*	0.50 *-0.02*	0.54 *-0.01*	0.52 *+0.05*	0.42 *+0.00*	**0.57 *-0.00***	0.41 *-0.07*	0.58 *-0.04*	**0.61 *-0.04***	0.60 *-0.00*	0.58 *+0.05*

Table 7: Attribute-based aggregate performance. For each method, the respective left column corresponds to the average region similarity $\mathcal{J}$ over all sequences with that specific attribute (e.g., AC), while the right column indicates the performance gain (or loss) for that method for the remaining sequences without that respective attribute. The best performing method of each category is highlighted in bold.

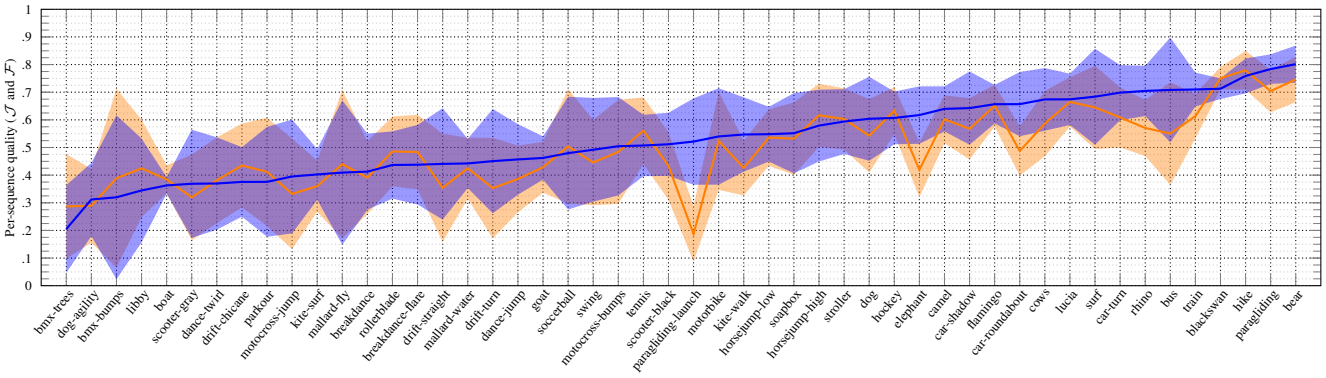

Figure 2: **Per-sequence mean performance**: Mean and variance of region Jaccard $\mathcal{J}$ (blue) and boundary F measure $\mathcal{F}$ (orange). Sequences are sorted by *difficulty*, *i.e.* mean performance of $\mathcal{J}$ over all techniques.

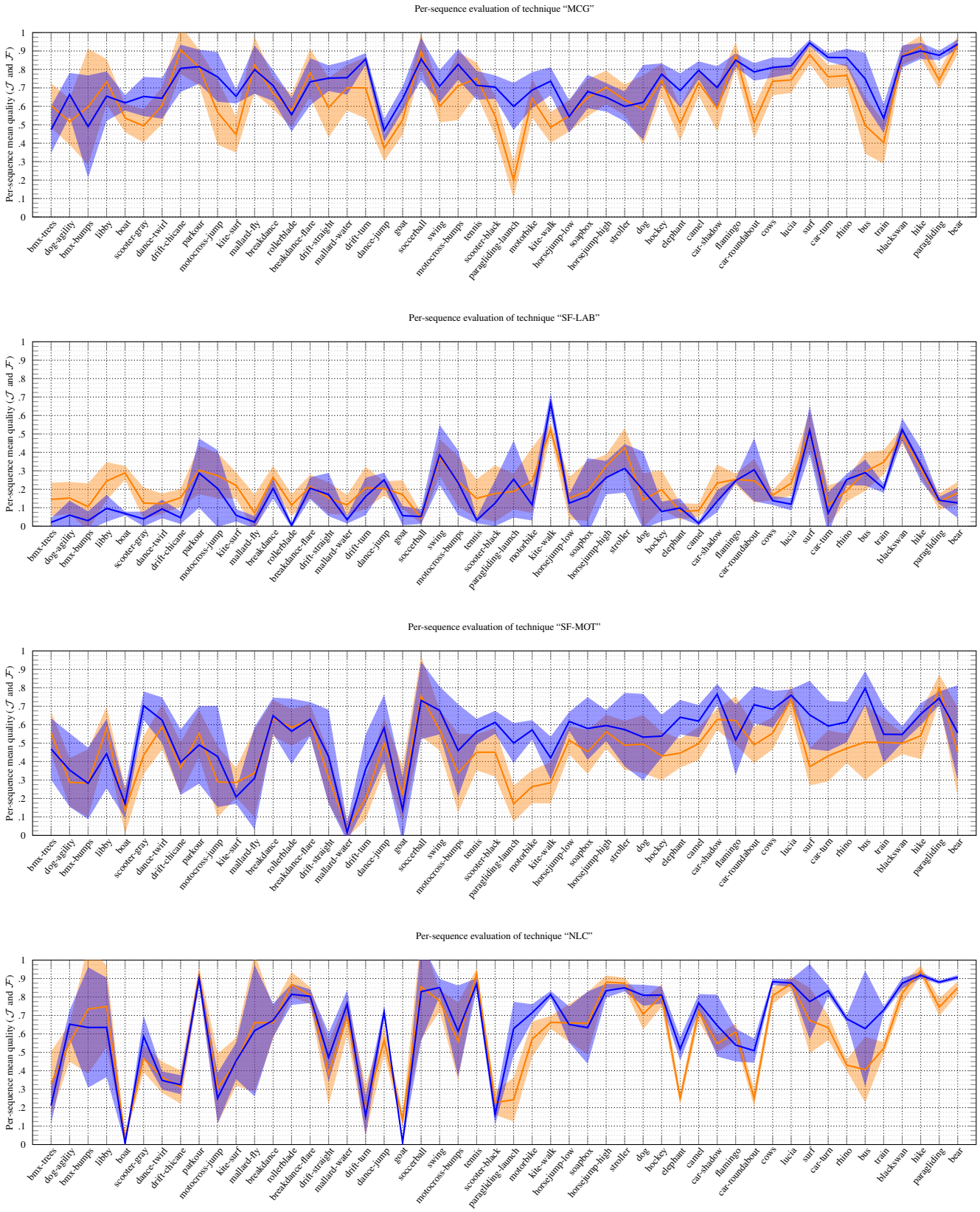

Figure 3: **Per-sequence per-method evaluation**: Mean and variance of region Jaccard $\mathcal{J}$ (blue) and boundary F measure $\mathcal{F}$ (orange)

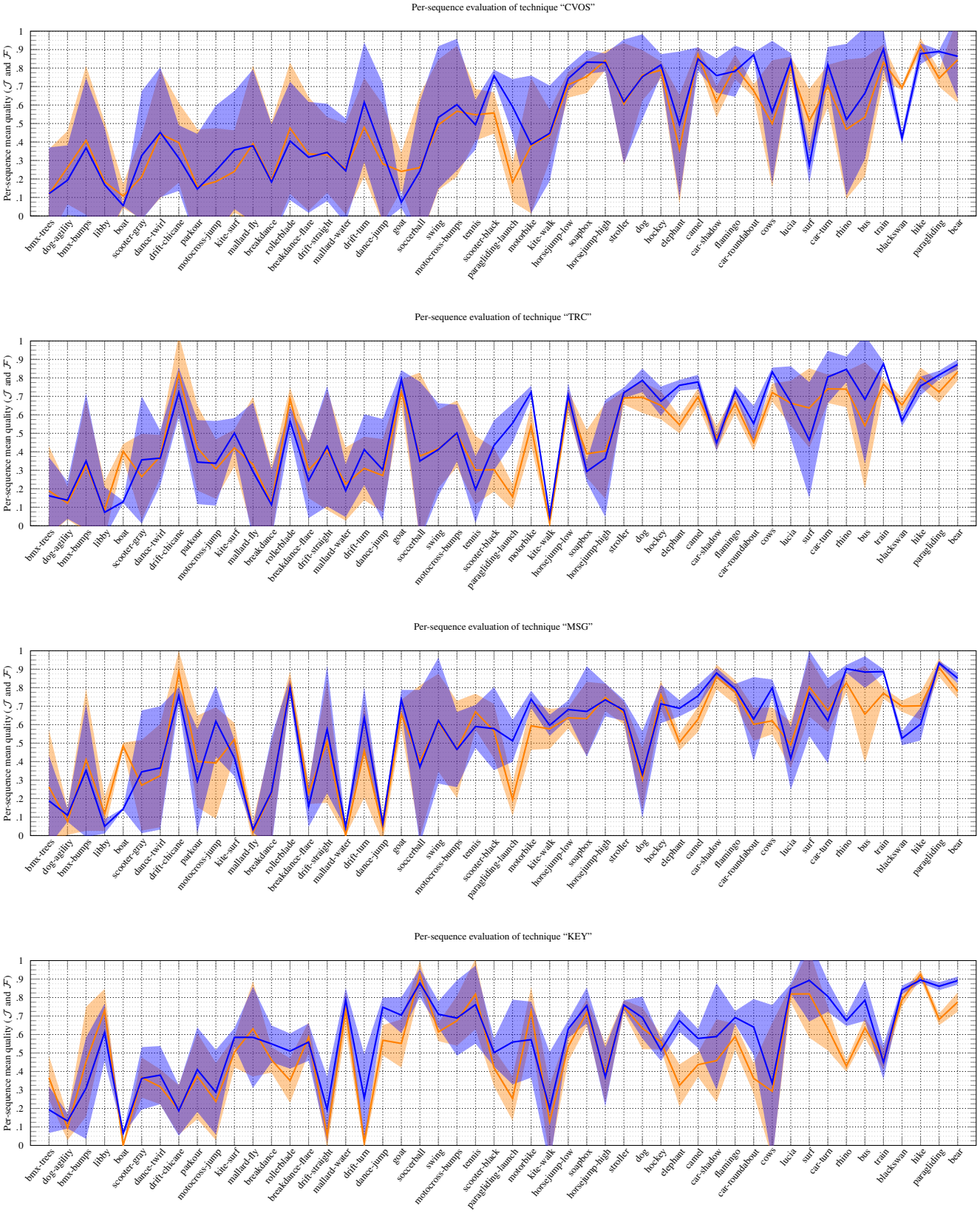

Figure 4: **Per-sequence per-method evaluation**: Mean and variance of region Jaccard $\mathcal{J}$ (blue) and boundary F measure $\mathcal{F}$ (orange)

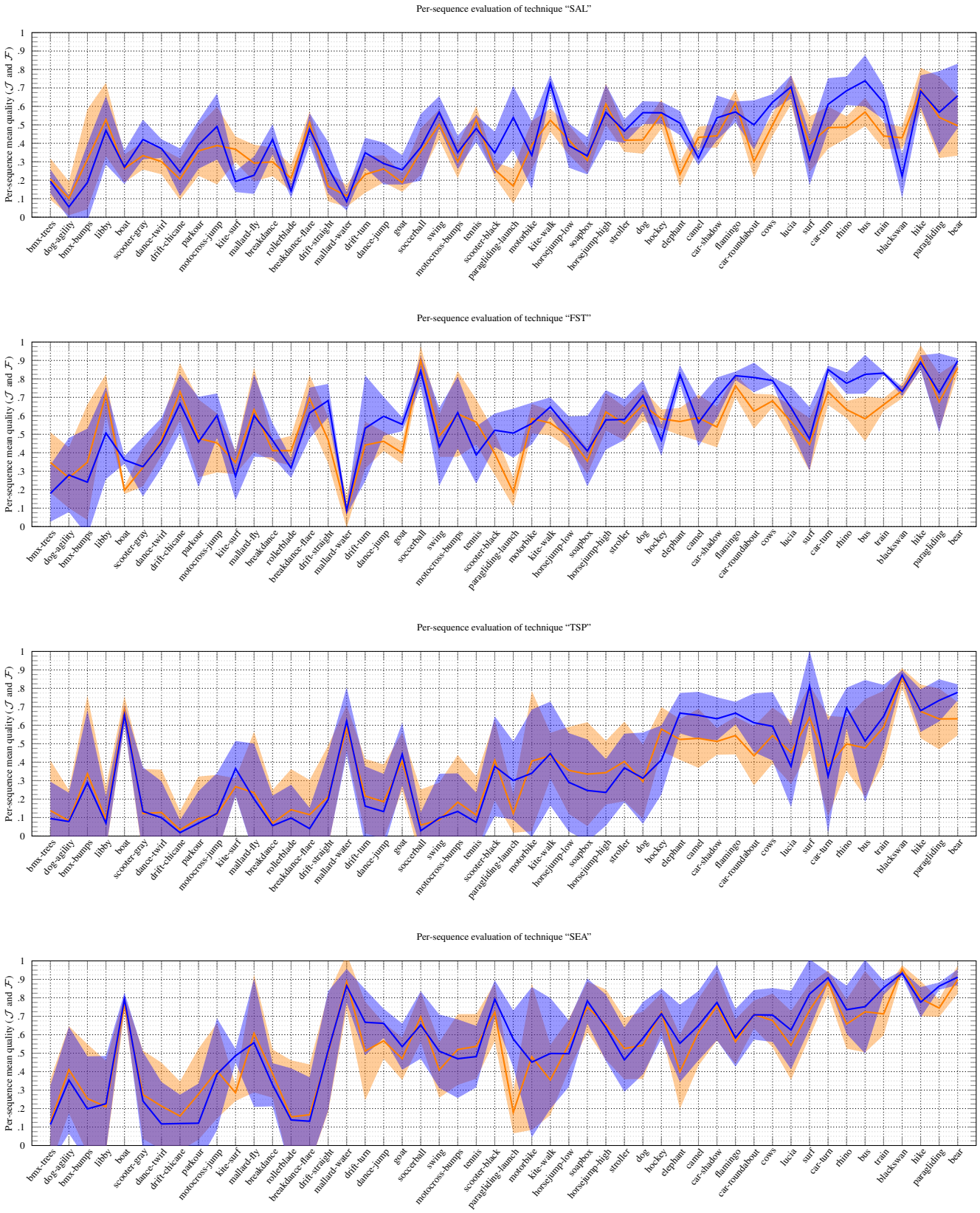

Figure 5: **Per-sequence per-method evaluation**: Mean and variance of region Jaccard $\mathcal{J}$ (blue) and boundary F measure $\mathcal{F}$ (orange)

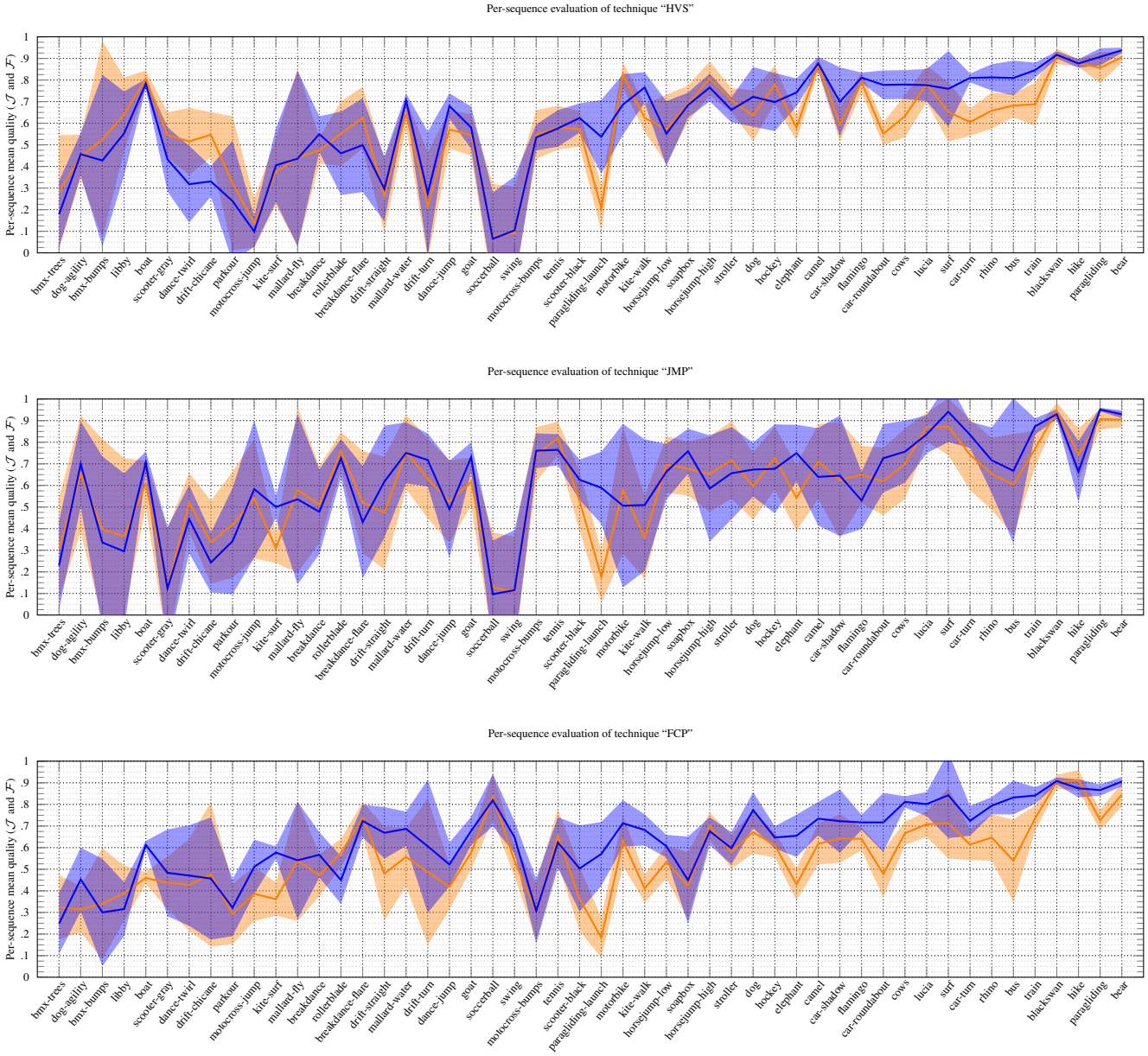

Figure 6: **Per-sequence per-method evaluation**: Mean and variance of region Jaccard $\mathcal{J}$ (blue) and boundary F measure $\mathcal{F}$ (orange)