Object Segmentation Tracking from Generic Video Cues

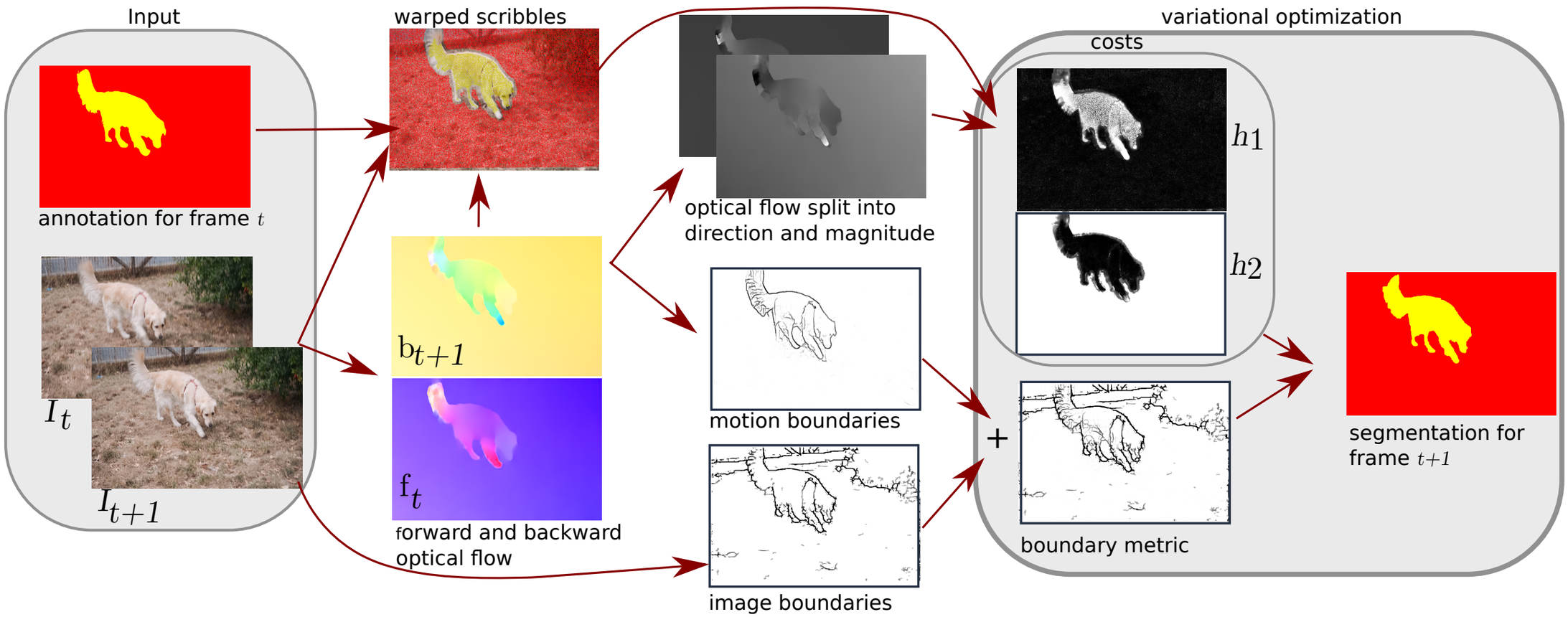
A. Kardoost[1] S. Müller[2,3] J. Weickert[4] M. Keuper[1]

[1]Data and Web Science Group, University of Mannheim, Germany
[2]Fraunhofer ITWN Competence Center HPC, Kaiserslautern, Germany
[3]Fraunhofer Center Machine Learning, Germany
[4]Mathematical Image Analysis Group, Saarland University, Germany

Our Contributions

- Light-weight formulation for video object segmentation
  - Using variational label propagation
- Segmentation of fine details and thin structures
- Retrieval of objects due to intermediate tracking mistakes
- Refine state-of-the-art CNN-based results

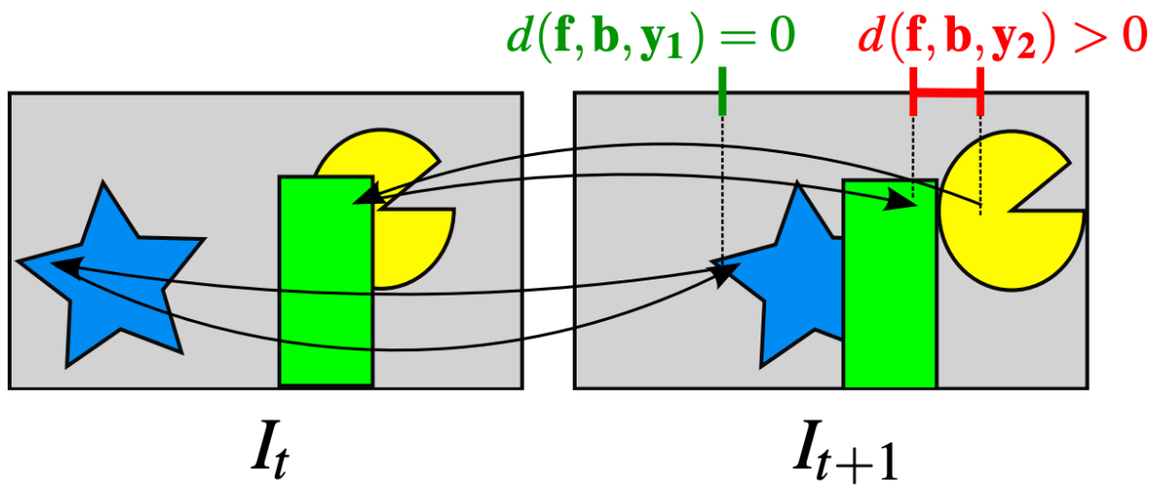
Proposed Model

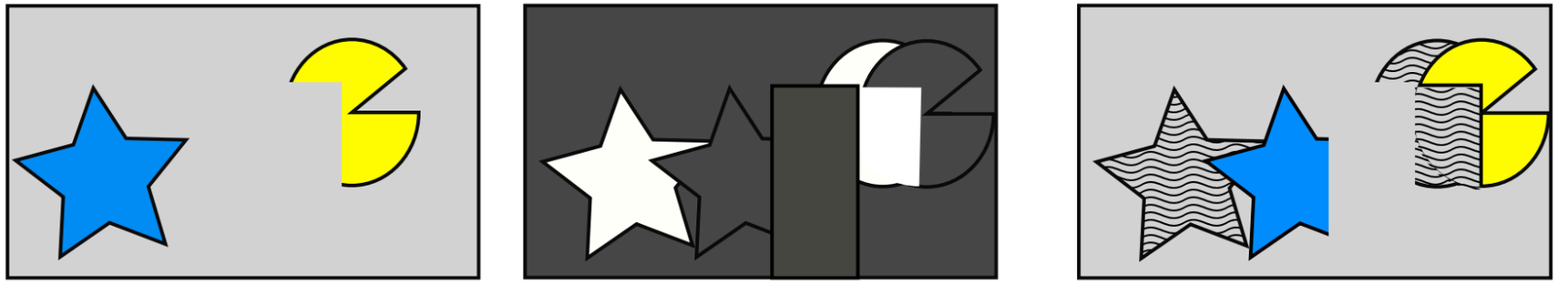


$$\min_{\Omega_1,\ldots,\Omega_n \subset \Omega} \frac{\lambda}{2} \sum_{i=1}^{n} \mathbf{Per}(\Omega_i; \Omega) + \sum_{i=1}^{n} \int_{\Omega_i} h_i(\mathbf{x})\, d\mathbf{x}$$

$$s.t. \quad \Omega = \cup_{i=1}^{n} \Omega_i, \quad \forall i \neq j \quad \Omega_i \cap \Omega_j = \varnothing .$$

Optical Flow Consistency





annotation for I_t | flow inconsistencies | propagated labels

Quantitative Results

Train set of DAVIS-2016

	Measure	$F(\%)$			$J(\%)$		
		M	R	D	M	R	D
	train						
CNN	MSK [5]	76.1	88.9	9.8	**80.7**	93.9	8.8
CNN	VPN [31]	**77.0**	**94.3**	13.1	78.3	**95.4**	7.2
CNN	CTN [37]	72.8	88.3	14.7	76.9	90.0	13.5
non-CNN	OFL [25]	70.9	83.1	21.9	73.2	83.0	20.2
non-CNN	BVS [38]	70.1	83.7	25.1	70.9	82.7	24.1
non-CNN	FCP [39]	58.3	67.6	**7.2**	66.2	82.0	**6.5**
non-CNN	JMP [40]	62.3	73.2	36.5	63.2	73.7	35.8
non-CNN	*ours*	**73.5**	**88.6**	13.8	**77.7**	**90.2**	12.5

Validation set of DAVIS-2016

	Measure	$F(\%)$			$J(\%)$		
		M	R	D	M	R	D
	validation						
CNN	OSVOS-S [29]	87.5	**95.9**	8.2	85.6	96.8	5.5
CNN	[29] + *ours*	87.6	**95.9**	8.1	**86.0**	**96.9**	5.6
CNN	CINM [28]	85.0	92.1	14.7	83.4	94.9	12.3
CNN	[28] + *ours*	**87.7**	93.0	14.3	84.2	95.6	12.1
CNN	MSK [5]	75.4	87.1	9.0	79.7	93.1	8.9
CNN	VPN [31]	65.5	69.0	14.4	70.2	82.3	12.4
CNN	SIAMMASK [41]	67.8	79.8	2.1	71.7	86.8	3.0
CNN	CTN [37]	69.3	79.6	12.9	73.5	87.4	15.6
CNN	PLM [42]	62.5	73.2	14.7	70.2	86.3	11.2
non-CNN	OFL [25]	63.4	70.4	27.2	68.0	75.6	26.4
non-CNN	BVS [38]	58.8	67.9	21.3	60.0	66.9	28.9
non-CNN	FCP [39]	49.2	49.5	**−1.1**	58.4	71.5	**−2.0**
non-CNN	JMP [40]	53.1	54.2	38.4	57.0	62.6	39.4
non-CNN	*ours*	**68.4**	**78.4**	17.8	**71.6**	**81.0**	16.8

Train and Test sets of DAVIS-2017

	$F(\%)$		$J(\%)$	
Method	val	test	val	test
ours	56.5	44.0	54.5	41.5
OSMN [36]	57.1	44.9	52.5	37.7
FAVOS [43]	61.8	44.2	54.6	42.9
OSVOS-S [29]	71.3	62.1	64.7	52.9
[29] + *ours*	71.4	62.2	65.2	53.7
CINM [28]	74.0	70.5	67.2	64.5
[28] + *ours*	**74.1**	**70.6**	**67.4**	**64.7**

SegTrack v2

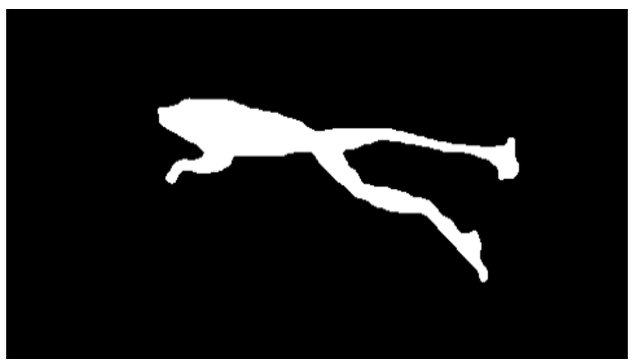
Sequence/Object	SPT+CSI [19]	MSK [5]	JOTS [44]	SPT [19]	*ours*
Unsupervised	✓	×	×	✓	✓
Online	×	✓	✓	✓	✓
Mean per object	65.9	67.4	71.8	62.7	65.7
Mean per seq.	71.2	-	72.2	68.0	68.6

Qualitative Results








Video Frame **Ground-truth** **Ours**