

Object Segmentation Tracking from Generic Video Cues

Amirhossein Kardoost¹ Sabine Müller^{2,3}
Joachim Weickert⁴ Margret Keuper¹

¹Data and Web Science Group, University of Mannheim, Germany

²Fraunhofer ITWN Competence Center HPC, Kaiserslautern, Germany

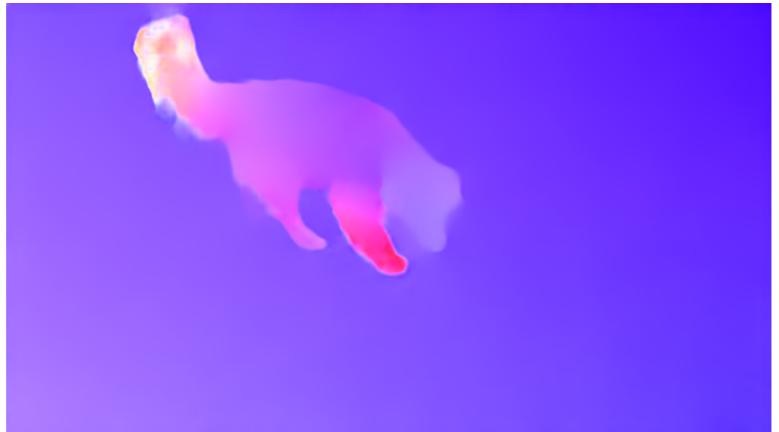
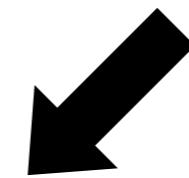
³Fraunhofer Center Machine Learning, Germany

⁴Mathematical Image Analysis Group, Saarland University, Germany





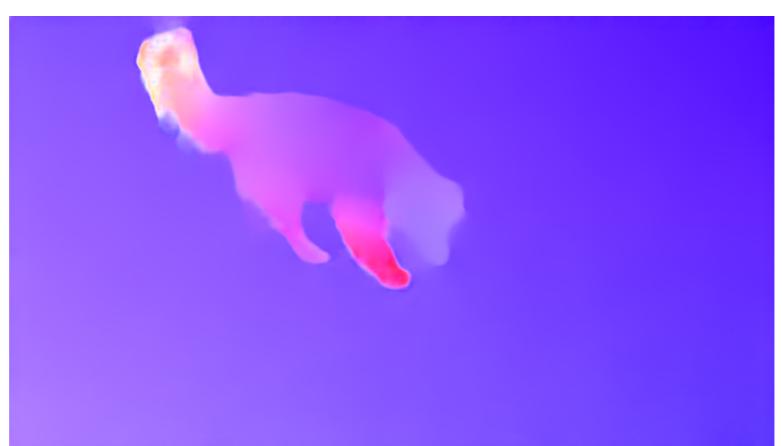
Optical Flow
Estimation



Forward Flow
(f)



Backward Flow
(b)



Forward Flow
(f)



Backward Flow
(b)



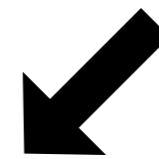
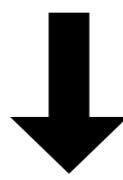
Segmentation at $frame_t$



Forward Flow
(f)

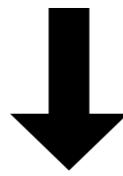


Backward Flow
(b)



Confident Label Propagation

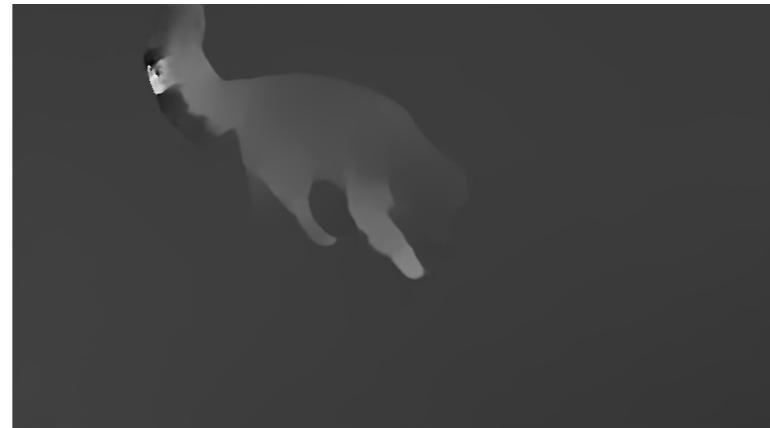
$$d(f, b, point_y) < \tau$$



Scribble Points



Scribble Points



Backward Flow
Direction



Backward Flow
Magnitude



Cost of Object 1 (*background*)

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



Cost of Object 2 (*foreground*)

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



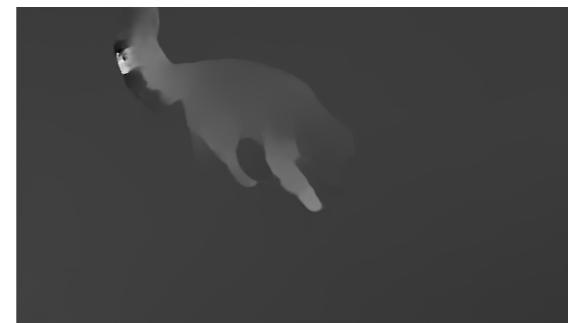
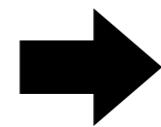
Cost of Object 1 (*background*)



Cost of Object 2 (*foreground*)

$$\min_{\Omega_1, \dots, \Omega_n \subset \Omega} \frac{\lambda}{2} \sum_{i=1}^n \mathbf{Per}(\Omega_i; \Omega) + \sum_{i=1}^n \int_{\Omega_i} h_i(\mathbf{x}) d\mathbf{x}$$
$$s.t. \quad \Omega = \cup_{i=1}^n \Omega_i, \quad \forall i \neq j \quad \Omega_i \cap \Omega_j = \emptyset.$$

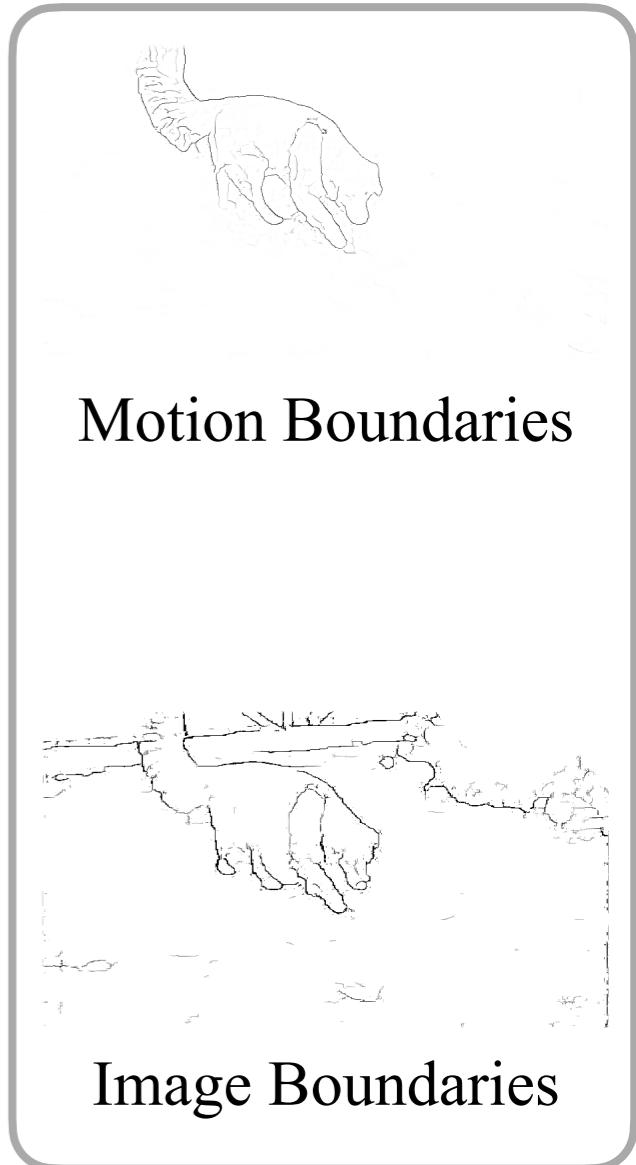
$$h_i(\mathbf{x}) = -\log \frac{1}{|\mathcal{S}_i|} \int_{\mathcal{S}_i} G_{\rho_i} \cdot G_{\sigma} d\mathbf{x}_{S_i} \quad \rightarrow \quad \text{Data Cost Term}$$



Backward Flow
Direction



Backward Flow
Magnitude



Boundary Metric

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

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



Cost of Object 1 (*background*)



Cost of Object 2 (*foreground*)



Boundary Metric

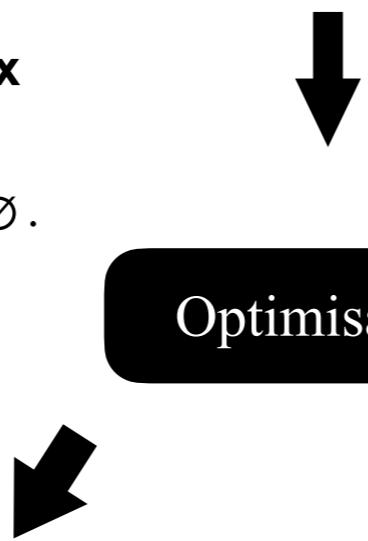
$$\min_{\Omega_1, \dots, \Omega_n \subset \Omega} \frac{\lambda}{2} \sum_{i=1}^n \mathbf{Per}(\Omega_i; \Omega) + \sum_{i=1}^n \int_{\Omega_i} h_i(\mathbf{x}) d\mathbf{x}$$

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

Optimisation



Segmentation at *frame*_{t+1}



...

Lost Object Retrieval

Without
Lost Object Retrieval



With
Lost Object Retrieval



Quantitative Results

Results on Train and Validation sets of DAVIS 2016

Measure	train						
	F(%)			J(%)			
	M	R	D	M	R	D	
CNN	MSK [5]	76.1	88.9	9.8	80.7	93.9	8.8
	VPN [31]	77.0	94.3	13.1	78.3	95.4	7.2
	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
	BVS [38]	70.1	83.7	25.1	70.9	82.7	24.1
	FCP [39]	58.3	67.6	7.2	66.2	82.0	6.5
	JMP [40]	62.3	73.2	36.5	63.2	73.7	35.8
	<i>ours</i>	73.5	88.6	13.8	77.7	90.2	12.5

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

Results on Validation and Test sets of DAVIS 2017

Method	$F(\%)$		$J(\%)$	
	val	test	val	test
<i>ours</i>	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] + <i>ours</i>	71.4	62.2	65.2	53.7
CINM [28]	74.0	70.5	67.2	64.5
[28] + <i>ours</i>	74.1	70.6	67.4	64.7

Results on SegTrack v2

Sequence/Object	SPT+CSI [19]	MSK [5]	JOTS [44]	SPT [19]	<i>ours</i>
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

Binary-Object Segmentation

DAVIS 2016



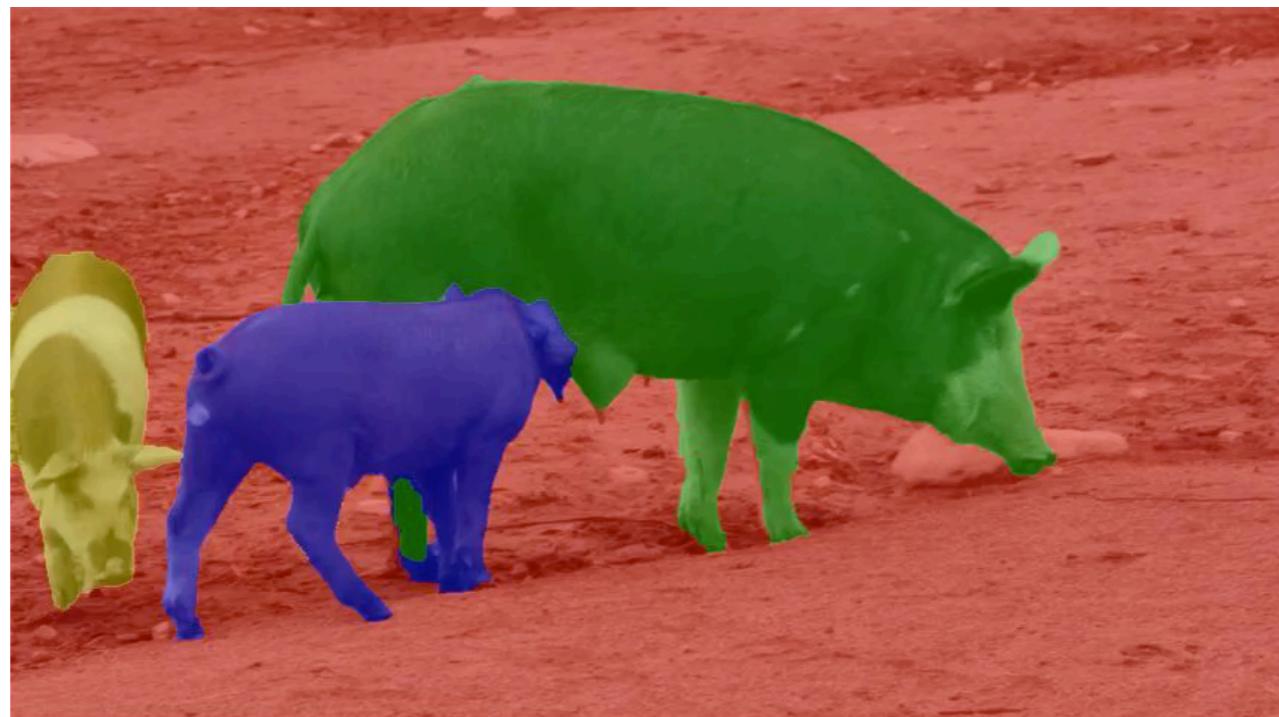
Binary-Object Segmentation

SegTrack v2

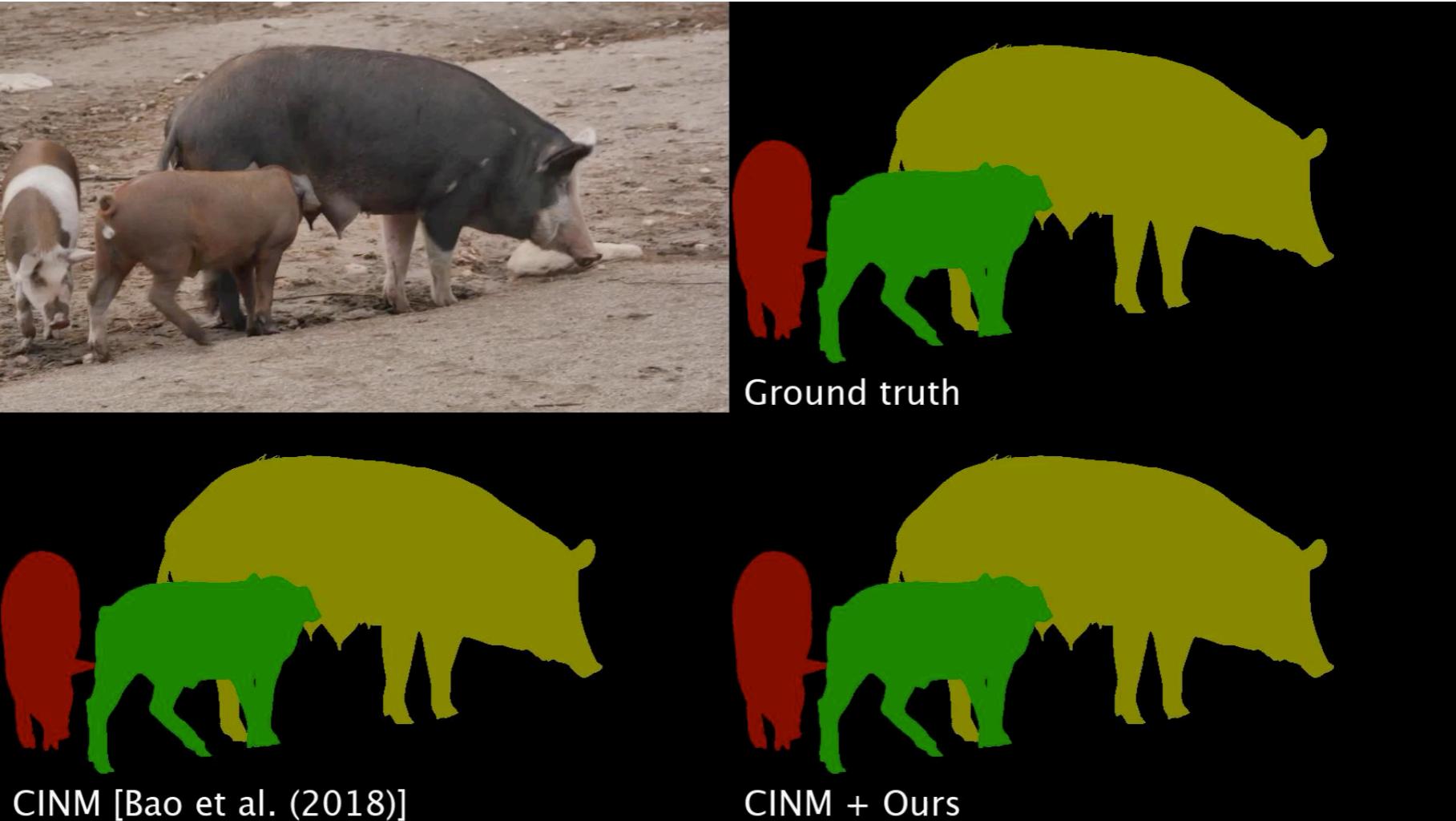


Multi-Object Segmentation

DAVIS 2017



Post-Processing CNN-Based Results



Thank You!