RVOS: End-to-End Recurrent Network for Video Object Segmentation

Universitat Oberta de Catalunya (UOC) — Barcelona Supercomputing Center (BSC) Centro Nacional de Supercomputación — Universitat Politècnica de Catalunya BARCELONATECH

Carles Ventura · Miriam Bellver · Andreu Girbau · Amaia Salvador · Ferran Marques · Xavier Giro

Summary

Multiple object video object segmentation is a challenging task, specially for the zero-shot case, when no object mask is given at the initial frame and the model has to find the objects to be segmented along the sequence. In our work, we propose a Recurrent network for multiple object Video Object Segmentation (RVOS) that is **fully end-to-end trainable**. Our model incorporates **recurrence** on two different domains: (i) the **spatial**, which allows to discover the different object instances within a frame, and (ii) the **temporal**, which allows to keep the coherence of the segmented objects along time.
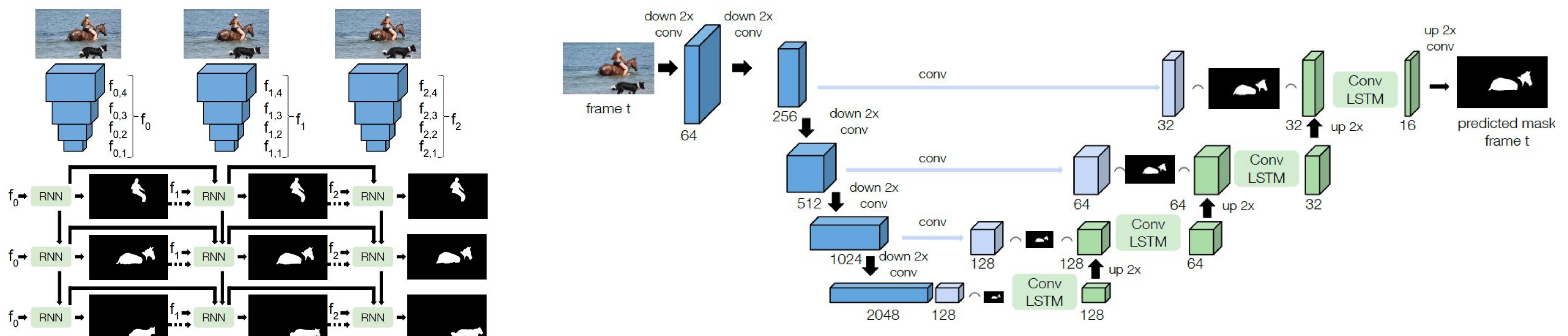
Previous work: Spatial Recurrence

Recurrent Semantic Instance Segmentation (RSIS)
- Model based on spatial recurrence for instance segmentation in images
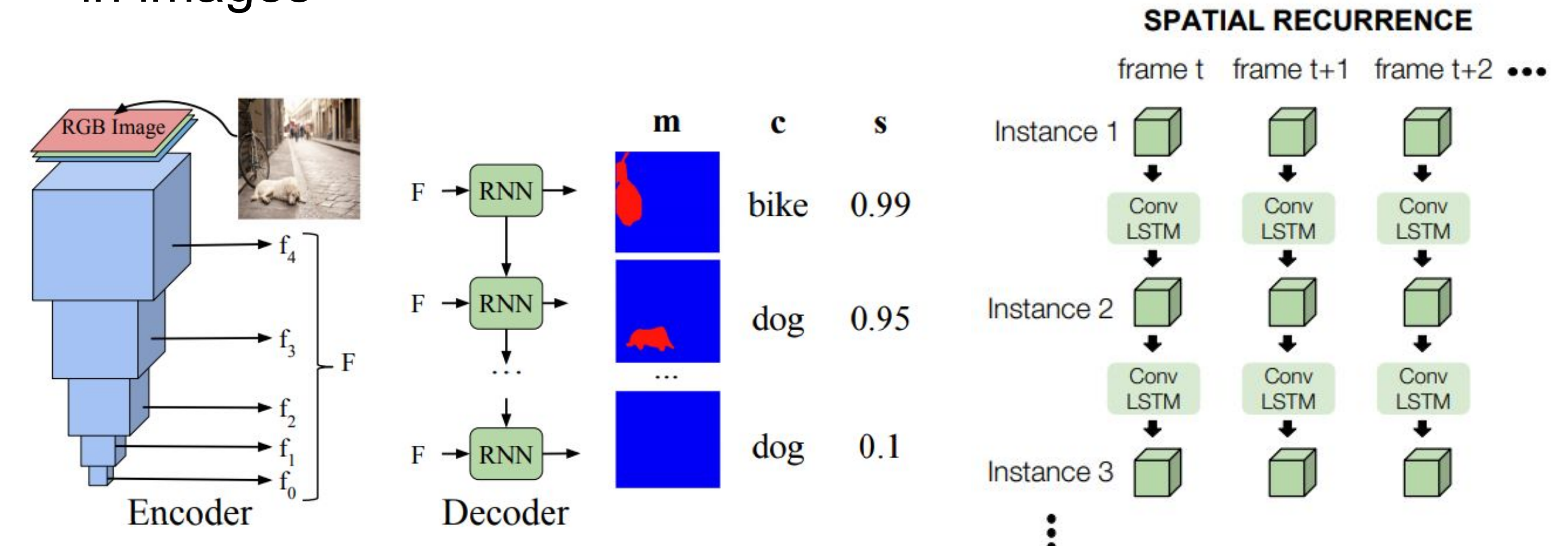
Proposed Model: Spatio-temporal Recurrence

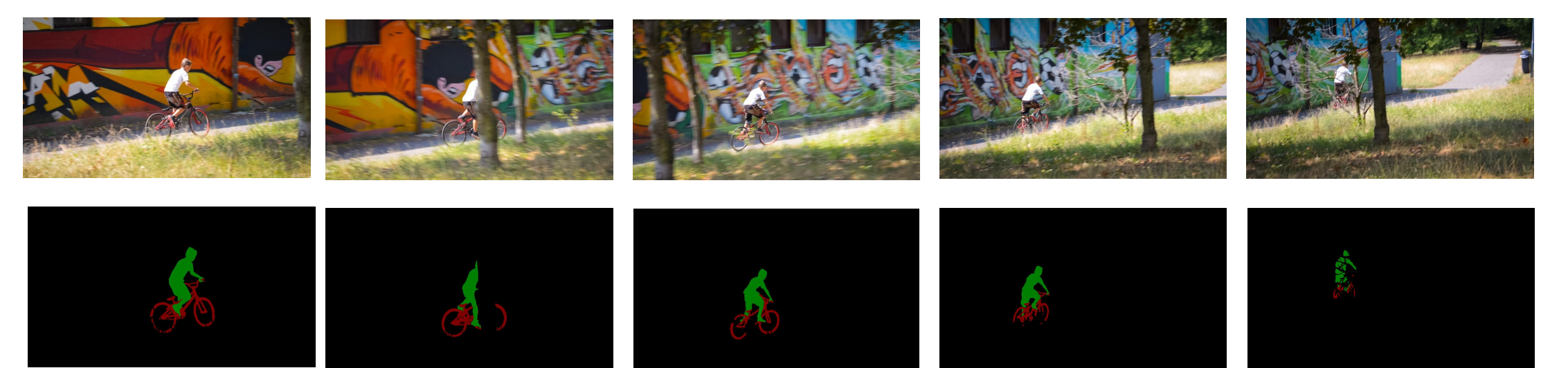We propose to extend RSIS (spatial recurrent model for images) to RVOS (spatio-temporal recurrent model for videos):

Experimental Results

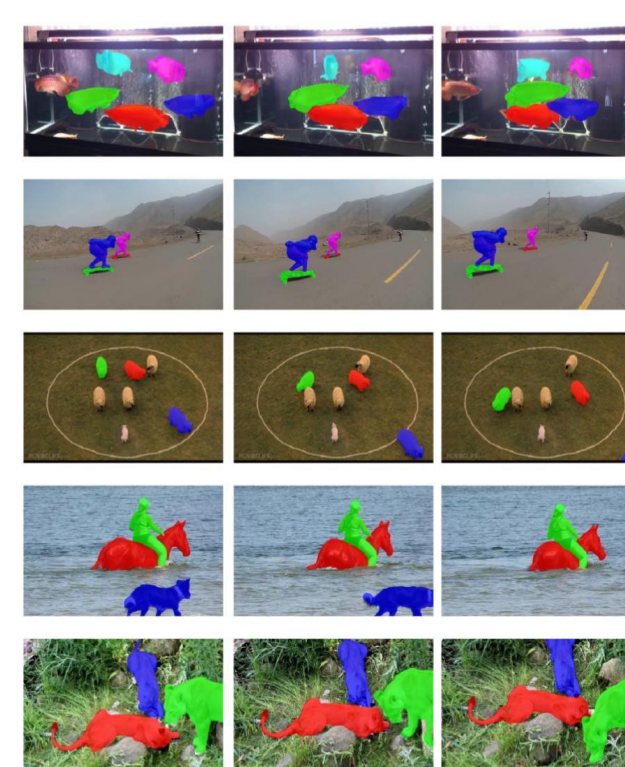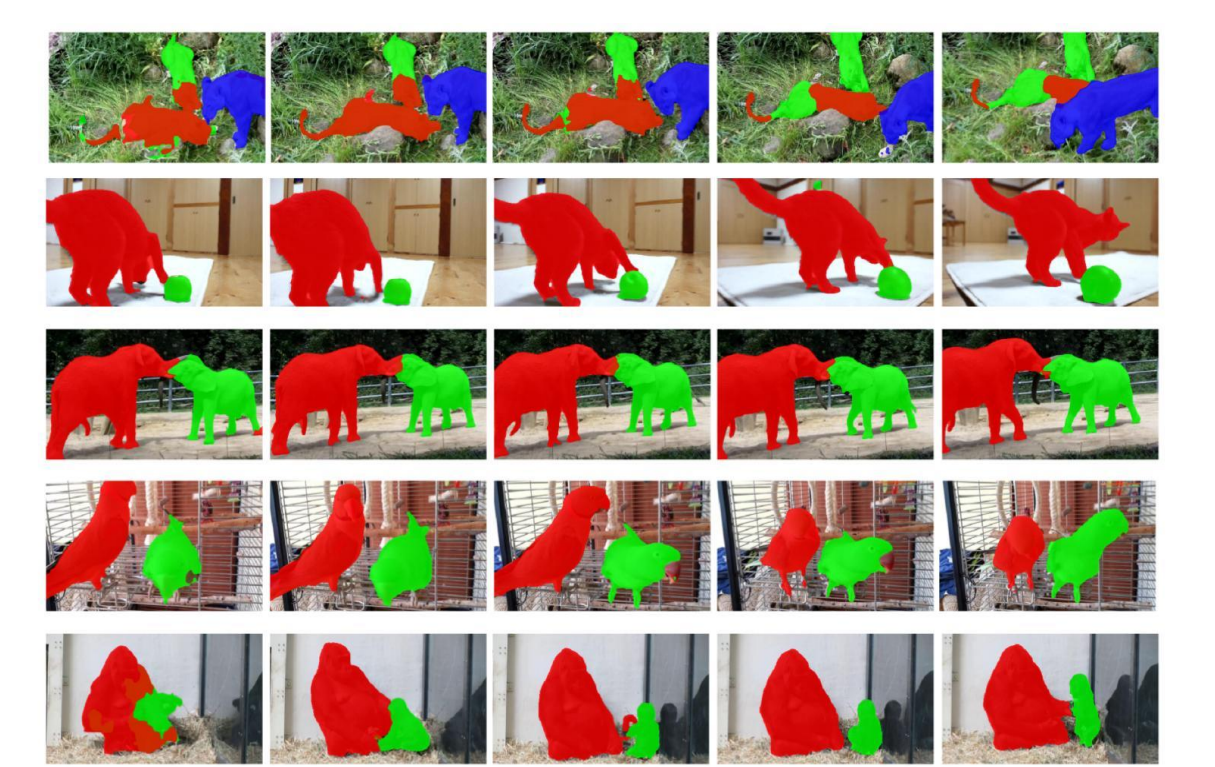### ONE-SHOT VIDEO OBJECT SEGMENTATION

YOUTUBE-VOS

	OL	J_{seen}	J_{unseen}	F_{seen}	F_{unseen}	Inference time (s/frame)
OSVOS [3]	✓	59.8	**54.2**	60.5	**60.7**	10
MaskTrack [20]	✓	59.9	45.0	59.5	47.9	12
OnAVOS [30]	✓	**60.1**	46.6	**62.7**	51.4	13
S2S w/o OL [33]	✗	**66.7**	48.2	65.5	50.3	0.160
OSMN [34]	✗	60.0	40.6	60.1	44.0	0.065
RVOS-Mask-ST+	✗	63.6	45.5	**67.2**	51.0	**0.044**

DAVIS 2017

	OL	J	F
OSVOS [3]	✓	47.0	54.8
OnAVOS [30]	✓	49.9	55.7
OSVOS-S [17]	✓	52.9	62.1
CINM [2]	✓	**64.5**	**70.5**
OSMN [34]	✗	37.7	44.9
FAVOS [4]	✗	42.9	44.2
RVOS-Mask-ST+ (pre)	✗	46.4	50.6
RVOS-Mask-ST+ (ft)	✗	**48.0**	**52.6**

ZERO-SHOT VIDEO OBJECT SEGMENTATION

YouTube-VOS zero-shot

	J_{seen}	J_{unseen}	F_{seen}	F_{unseen}
RVOS-S	40.8	19.9	43.9	23.2
RVOS-T	37.1	20.2	38.7	21.6
RVOS-ST	**44.7**	**21.2**	**45.0**	**23.9**
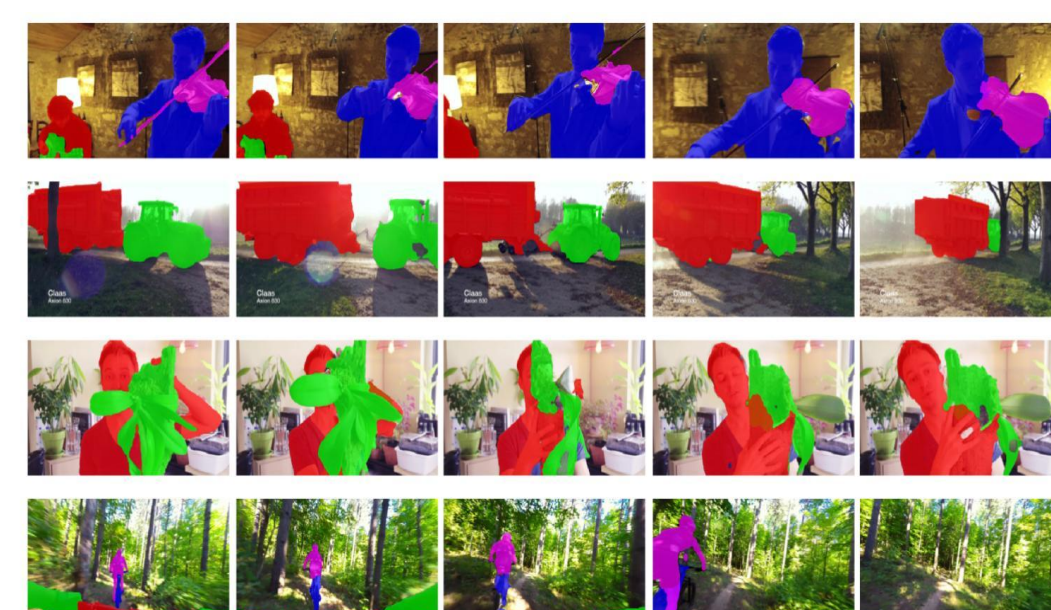
DAVIS-2017 zero-shot

	J	F
RVOS-ST (pre)	21.7	27.3
RVOS-ST (ft)	**23.0**	**29.9**

Acknowledgments: