MARs: Multi-view Attention Regularizations for Patch-based Feature Recognition of Space Terrain

Timothy Chase Jr, Karthik Dantu
University at Buffalo

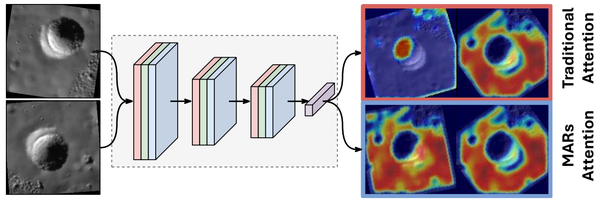
University at Buffalo
The State University of New York
DRONES Lab
Distributed Robotics and Networked Embedded Systems

ECCV MILANO 2024

Motivation

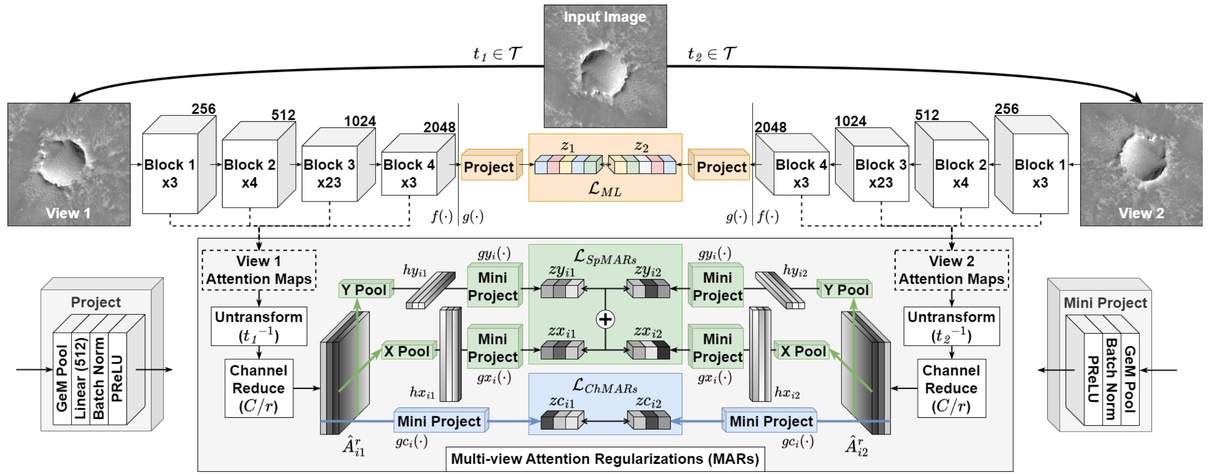
Background: Spacecraft track pre-gathered, hand-crafted patch features (landmarks) with template matching for Terrain Relative Navigation (TRN).

Problem: Object detection-style methods are being deployed for increased autonomy; description still an open problem. **Metric learning unable to handle challenging appearance change present in TRN.**



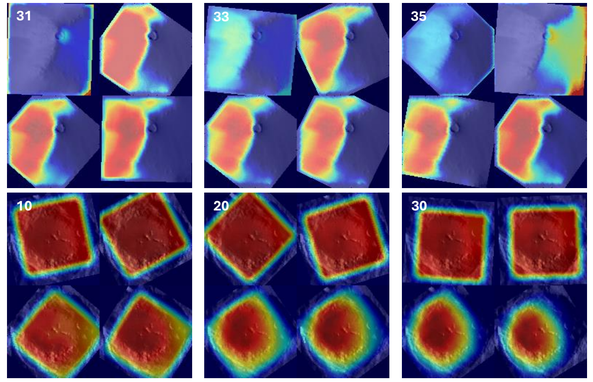
Contribution: We propose soft-similarity constraints to align the ***what*** and ***where*** of attention information between multi-view features, improving recognition.

Learning Multi-view Attention Similarity Constraints



Improved Training

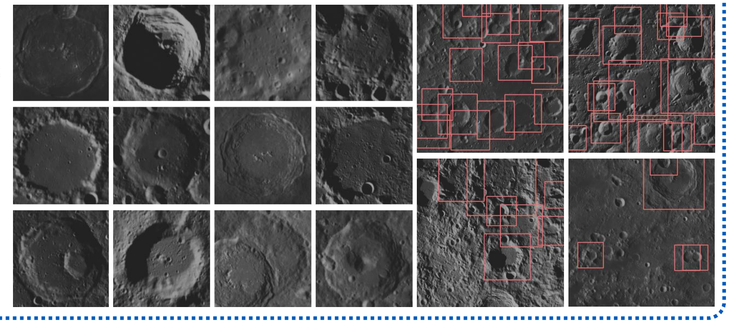
Attention learning with MARs (bottom pair) evolves similarly over time, **increasing training stability**…



… and promotes **faster, more uniform convergence** compared with RIC CA (top pair).

Luna-1: A New Dataset

5,067 **Moon Crater** landmarks coupled with 2,161 **emulated navigation frames** derived from real-world NASA data for TRN research.



Boosting Landmark Recognition Performance

conv2d SE: *conv2d* and Squeeze-Excitation (channel only) attention.
RIC CA: Rotation-invariant convolution and Coordinate (channel and spatial) Attention.
MARs: *RIC CA* with *MARs* multi-view attention alignment.

$\mathcal{L}_{ML}$	Incremental Recall@1														
	Earth Stadium			Mars Crater			Moon Crater			Moon Navigation			Moon Lost-in-Space		
	conv2d SE	RIC CA	MARs (Ours)	conv2d SE	RIC CA	MARs (Ours)	conv2d SE	RIC CA	MARs (Ours)	conv2d SE	RIC CA	MARs (Ours)	conv2d SE	RIC CA	MARs (Ours)
Circle	4.79	5.11	**5.13**	3.59	**60.81**	12.56	43.36	**60.09**	30.46	**58.07**	37.97	38.46	94.03	**96.68**	92.31
DR-MS	4.33	5.27	**54.32**	4.04	**48.48**	4.04	62.55	2.47	**64.57**	**37.69**	3.12	36.68	86.34	88.06	**90.05**
NTXent	**12.23**	7.75	8.59	4.49	**34.10**	34.01	69.85	77.06	**81.69**	48.25	32.00	**57.68**	94.83	83.16	**96.29**
PNP	4.64	**5.27**	4.66	3.14	**16.70**	5.14	10.24	15.54	**40.84**	14.34	23.34	**24.66**	61.41	**77.98**	75.46
Proxy Anchor	4.48	72.86	**78.06**	3.44	10.45	**71.10**	94.56	–	**94.78**	64.17	–	**66.31**	**97.21**	–	96.02
ProxyNCA++	4.33	4.64	**12.00**	4.04	**7.71**	4.79	56.03	**71.23**	70.27	**58.27**	53.92	35.87	**94.69**	93.24	91.38
Subcenter ArcFace	**4.49**	4.48	4.19	3.29	4.99	**38.69**	–	–	**20.45**	–	–	**40.63**	–	–	**81.17**
SupCon	4.19	5.80	**46.43**	4.65	**57.52**	49.19	16.73	79.42	**84.11**	17.92	**42.28**	37.50	89.39	90.32	**90.58**
Proxy Synthesis	4.33	**39.41**	22.57	4.03	4.97	**35.14**	**91.40**	64.71	17.47	**61.26**	60.53	32.67	**96.42**	93.77	36.87

MARs **raises recognition accuracy** for many metric learning losses on navigation-style multi-view description with challenging appearance change and leads to **new state-of-the-arts** across environments.

Visual Alignment