Sim2Real-Fire: A Multi-modal Simulation Dataset for Forecast and Backtracking of Real-world Forest Fire

Yanzhi Li Keqiu Li Guohui Li Zumin Wang Chanqing Ji Lubo Wang Die Zuo Qing Guo Feng Zhang Manyu Wang* Di Lin*

NEURAL INFORMATION PROCESSING SYSTEMS

Agency for Science, Technology and Research SINGAPORE

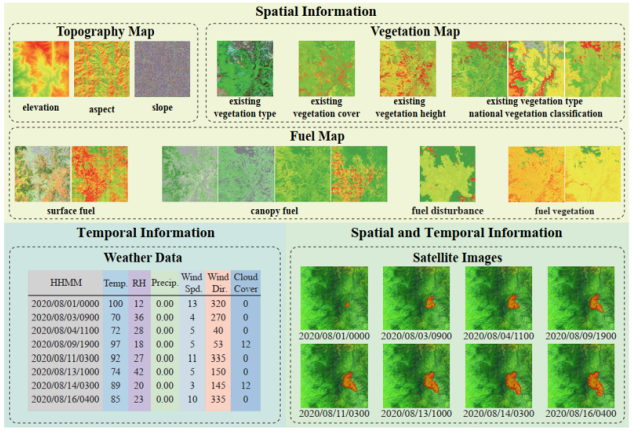
Introduction

The latest research on wildfire forecast and backtracking has adopted AI models, which require a large amount of data from wildfire scenarios to capture fire spread patterns. We explore using cost-effective simulated wildfire scenarios to train AI models and apply them to the analysis of real-world wildfire. This solution requires AI models to minimize the Sim2Real gap, a brand-new topic in the fire spread analysis research community.

To investigate the possibility of minimizing the Sim2Real gap, we collect the **Sim2Real-Fire dataset** that contains 1M simulated scenarios with multi-modal environmental information for training AI models. We prepare 1K real-world wildfire scenarios for testing the AI models. We also propose **a deep transformer, S2R-FireTr,** which excels in considering the multimodal environmental information for forecasting and backtracking the wildfire. S2R-FireTr surpasses state-of-the-art methods in real-world wildfire scenarios.

Sim2Real-Fire dataset

The Sim2Real-Fire dataset contains wildfire simulation and real-world data. The set includes 1M and 1K wildfire scenarios.



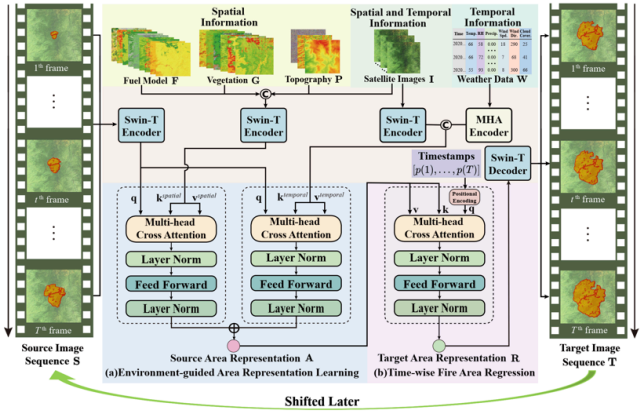
HHMM	Temp.	RH	Precip.	Wind Spd.	Wind Dir.	Cloud Cover
2020/08/01/0000	100	12	0.00	13	320	0
2020/08/03/0900	70	36	0.00	4	270	0
2020/08/04/1100	72	28	0.00	5	40	0
2020/08/09/1900	97	18	0.00	5	53	12
2020/08/11/0300	92	27	0.00	11	335	0
2020/08/13/1000	74	42	0.00	5	150	0
2020/08/14/0300	89	20	0.00	3	145	12
2020/08/16/0400	85	23	0.00	10	335	0

Topography, vegetation, fuel, weather, and the satellite data in the Sim2Real-Fire dataset.

Name	Scenarios	Countries	Areas	Tasks	Period	Spatial Resolution	Temporal Resolution	Sim/Real	Modalities
GABAM	10K	Worldwide	149,000,000	Fire Behavior	1990-2021	30m	1year	Real Only	1
Fire Atlas	13M	Worldwide	149,000,000	Fire Behavior	2003-2016	500m	1day	Real Only	1
GlobFire	100M	Worldwide	149,000,000	Fire Behavior	2001-2017	500m	1day	Real Only	1
WildfireDB	17M	USA	9,834,000	Spread Forecast	2012-2017	375m	1day	Real Only	4
SeasFire Cube	20K	Worldwide	149,000,000	Burned Area Forecast	2001-2021	27km	8days	Real Only	4
Next Day Wildfire	18K	USA	9,834,000	Spread Forecast	2012-2020	1km	1day	Real Only	4
WildfireSpreadTS	607	USA	9,834,000	Spread Forecast Spread Backtrack	2018-2021	375m	1day	Real Only	4
PT-FireSprd	80	Portugal	92,150	Fire Behavior Spread Forecast Danger Forecast Burned Area Forecast	2015-2021	1m-4km	30mins-14hours30mins	Real Only	1
Mesogeos	25K	Mediterranean	9,000,000	Danger Forecast Burned Area Forecast	2006-2022	1km	1day	Real Only	4
MODIS Thermal Anomaly	40K	Worldwide	149,000,000	Danger Forecast Spread Forecast	2000-2024	1km	1day	Real Only	3
VIIRS Thermal Anomaly	40K	Worldwide	149,000,000	Danger Forecast Spread Forecast	2012-2024	375m	12hours	Real Only	3
NOAA HMS Fire	1K	North America	24,710,000	Danger Forecast	2003-2024	2km	1day	Real Only	3
NOAA HMS Smoke	1K	North America	24,710,000	Danger Forecast	2005-2024	2km	1day	Real Only	1
GOES Wildfire	1K	Western Hemisphere	61,000,000	Danger Forecast Spread Forecast Burned Area Forecast	2017-2024	2km	5mins	Real Only	4
NIFC Wildfire Perimeters	20K	USA	9,834,000	Spread Forecast Burned Area Forecast	2000-2024	2km	5mins	Real Only	1
Sim2Real-Fire	**1M**	**Worldwide**	**20,000,000**	**Spread Forecast Spread Backtrack**	**2013-2023**	**30m**	**1hour**	**Sim&Real**	**5**

Comparison with the related datasets for wildfire analysis.

S2R-FireTr



S2R-FireTr predicts future wildfire areas from source fire areas. (a) Source fire areas and environmental data are input into the transformer to compute source area representation A. (b) A and the target timestamp are used to compute target area representation R for future predictions. Interchanging source and target areas allows wildfire backtracking.

Acknowledgement

The Key Science and Technology Program of the Ministry of Emergency Management of the People's Republic of China (2024EMST010102) fully supported this work.