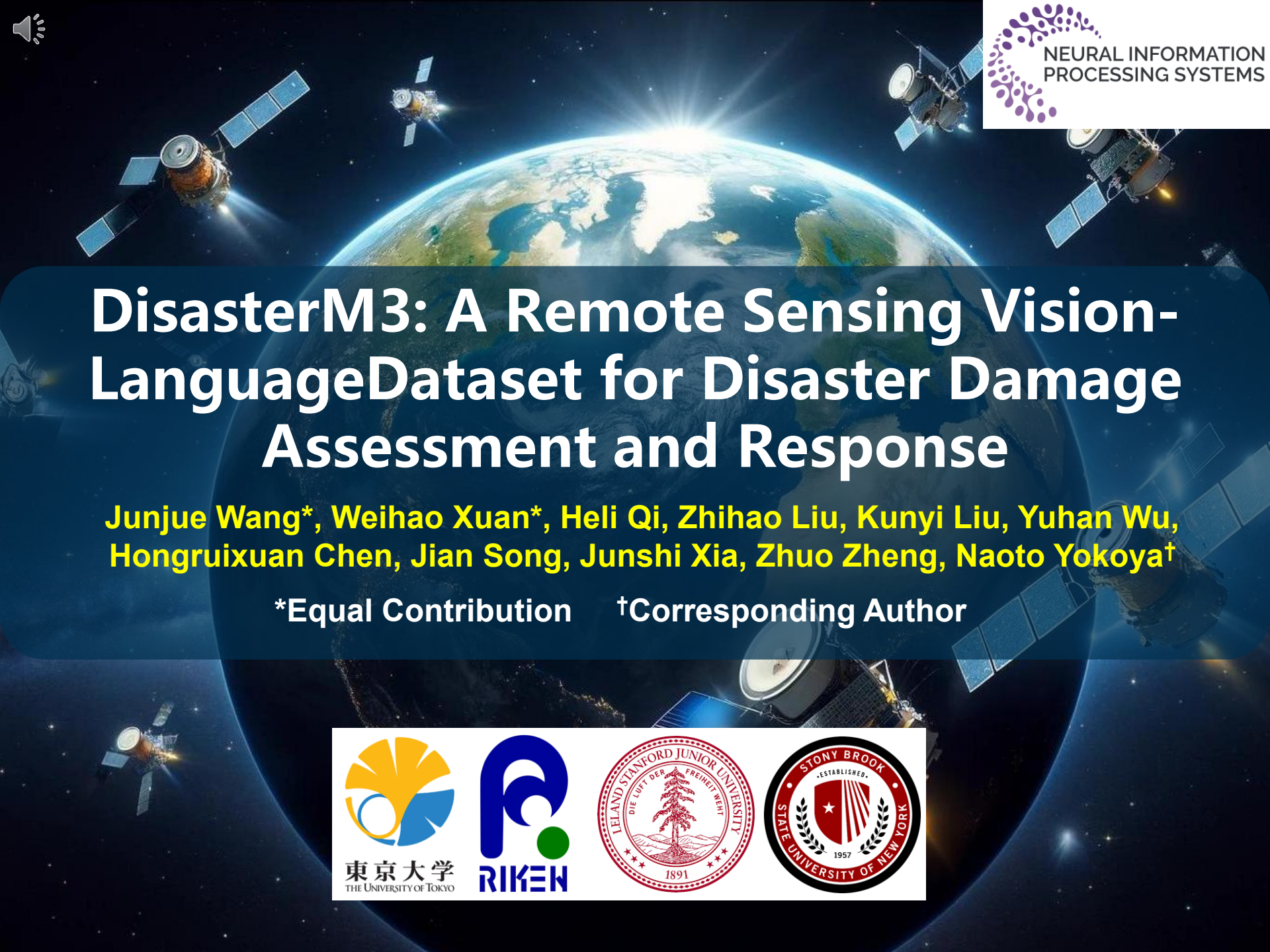


NEURAL INFORMATION PROCESSING SYSTEMS

DisasterM3: A Remote Sensing Vision-LanguageDataset for Disaster Damage Assessment and Response

Junjue Wang*, Weihao Xuan*, Heli Qi, Zhihao Liu, Kunyi Liu, Yuhan Wu, Hongruixuan Chen, Jian Song, Junshi Xia, Zhuo Zheng, Naoto Yokoya†

*Equal Contribution †Corresponding Author

東京大学 THE UNIVERSITY OF TOKYO

RIKEN

LELAND STANFORD JUNIOR UNIVERSITY · DIE LUFT DER FREIHEIT WEHT · 1891

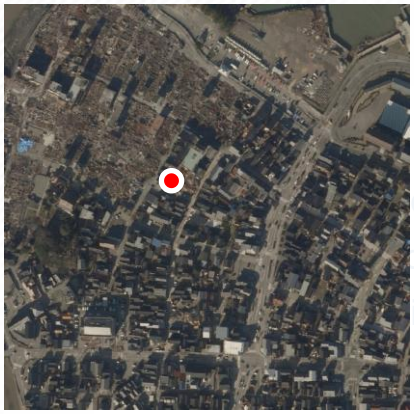
STONY BROOK · ESTABLISHED · 1957 · STATE UNIVERSITY OF NEW YORK

Motivation: Why we need DisasterM3?

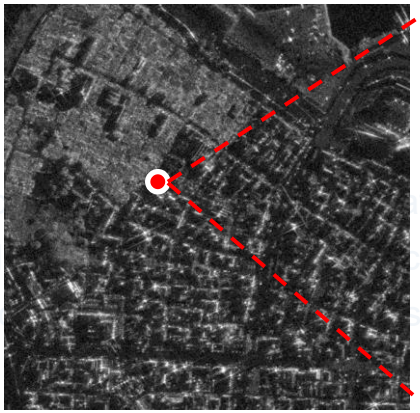
NEURAL INFORMATION PROCESSING SYSTEMS

When a disaster occurs, how to combine large vision-language model (VLM) and remote sensing data to achieve AI-based disaster response?

Disaster occurs

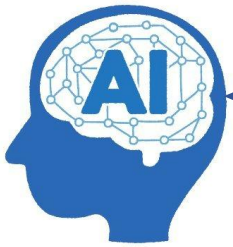


Optical image



SAR image

Disaster site

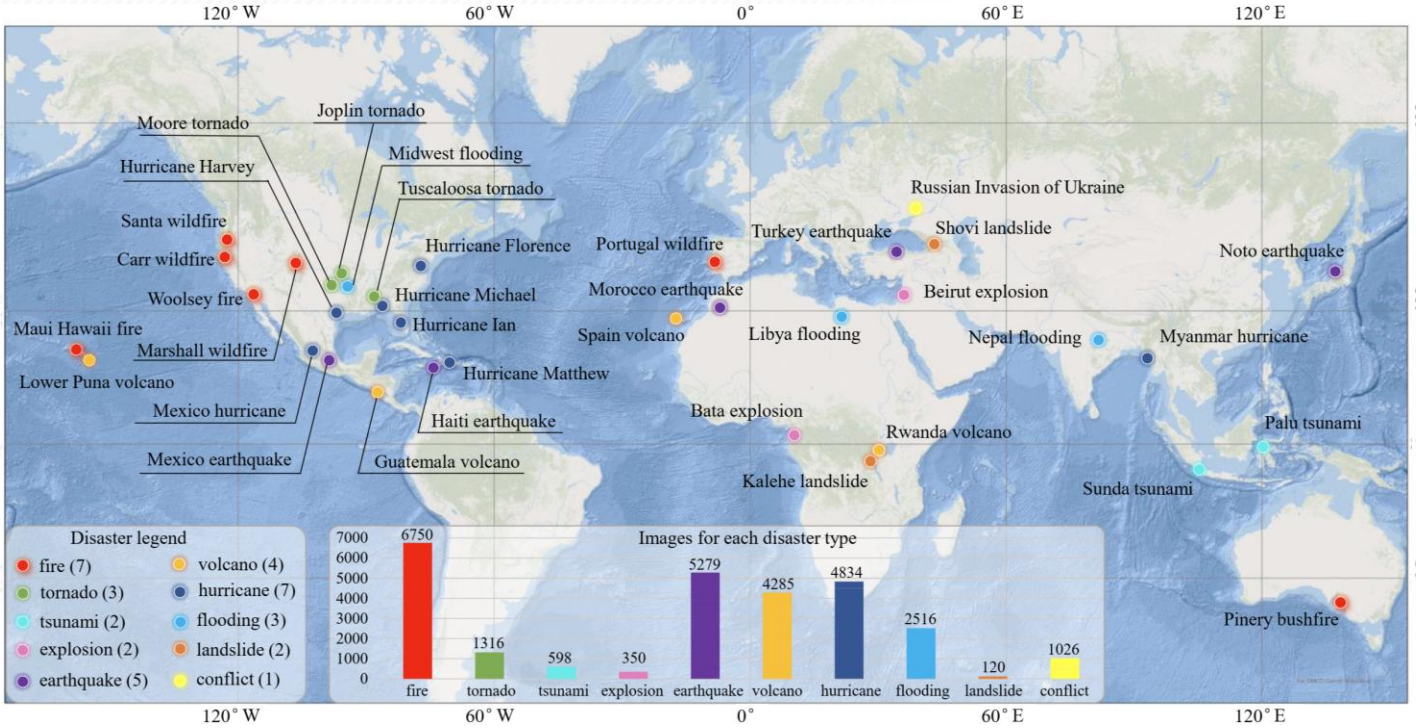


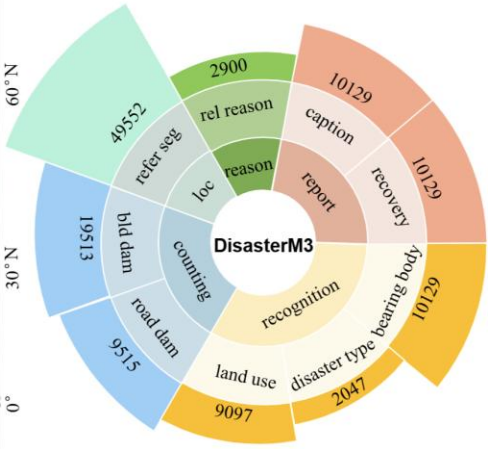
The Gap

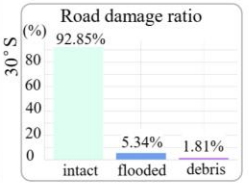
- No comprehensive vision-language disaster dataset exists
- Models lack disaster-specific corpus
- Complex damage patterns across geographies
- Extreme weather blocks optical sensors

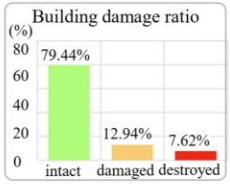
DisasterM3: Multi-Hazard, Multi-Sensor, Multi-Task

To address the gap, we propose the DiasterM3 dataset, includes nearly 27,000 bi-temporal image pairs from 36 major disaster events worldwide.









Scale & Coverage

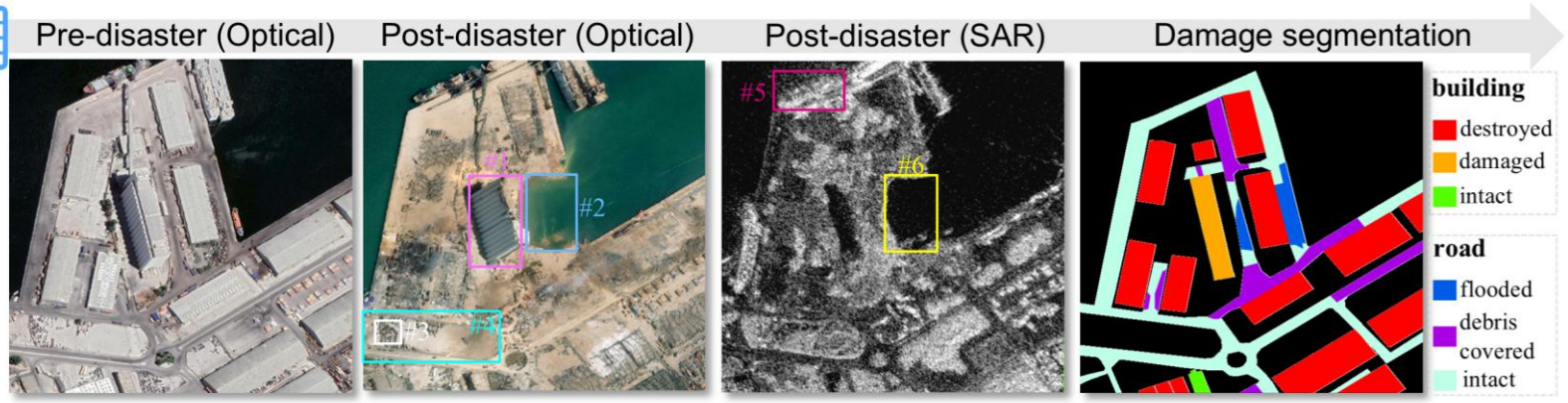
- 26,988 bi-temporal satellite images
- 123,010 instruction pairs
- 36 disasters across 5 continents

Multi-Hazard
10 disaster types worldwide

Multi-Sensor
Optical + SAR for all-weather

Multi-Task
9 disaster analysis tasks

Multi-level Task Taxonomy



Image

Rec.

Disaster type recognition
Q: What disaster has happened in this area? **A:** Explosion.

Disaster scene recognition
Q: What are land-use types in the pre-disaster scene? **A:** Harbor, industrial, and sea.

Disaster bearing bodies recognition
Q: What key structures are affected in this disaster? **A:** Building, road, and the harbor.

Counting

Damaged building counting
Q: How many intact buildings are in this disaster? **A:** 1
Q: How many major damaged buildings are in this disaster? **A:** 1
Q: How many total-destroyed buildings are in this disaster? **A:** 15

Damaged road area estimation
Q: What is the area ratio of the flooded roads? **A:** 0.59%
Q: What is the area ratio of the roads covered by debris? **A:** 1.99%
Q: What is the area ratio of the intact roads? **A:** 6.52%

Loc.&Reason.

Referring segmentation
Q: Segment the flooded road after the disaster?
A: Sure, the flooded roads are in ■
Q: Segment the road covered by debris?
A: Segmentation results are in ■
Q: Segment the total-destroyed building?
A: Totally destroyed buildings are in ■

Damaged object relational reasoning
Q: What's the relation between object#1 in pink box and object#2 in light blue box?
A: The damaged building#1 is located next to the left side of the subsidence area#2.
Q: What's the relation between object#3 in white box and object#4 in cyan box?
A: Many vehicles#3 are parked on the west side of the square#4, while building debris covers the east side.
Q: What's the relation between object#5 in pink red box and object#6 in yellow box?
A: The ship#5 is docked to the top left of the subsidence area#6.

Report

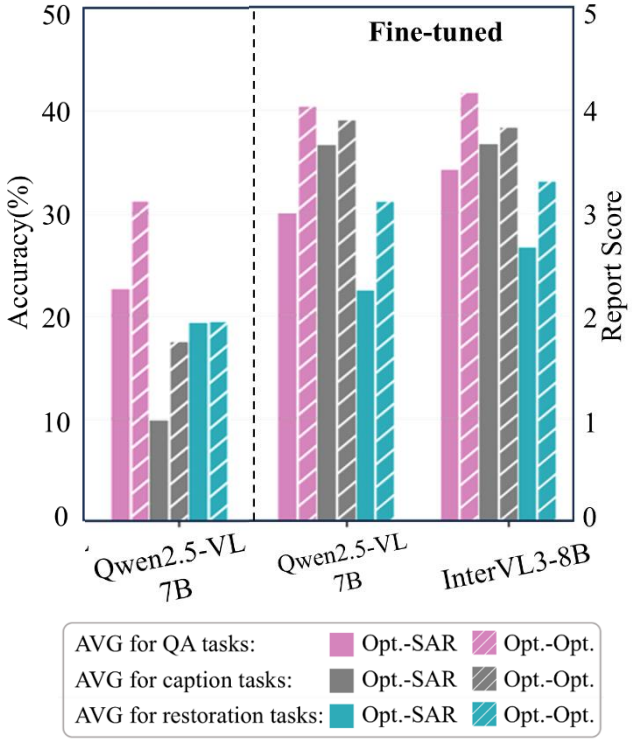
Disaster description
Building: Most of the buildings exhibit partial or complete collapse, evidenced by disrupted roofs and debris patterns. The building at the center of the image is totally destroyed, forming a crater. **Road:** Part of the road in the middle right of the image is buried by building rubble, while another section is flooded by seawater following the explosion. **Conclusion:** The explosion has caused severe damage to man-made structures, i.e., buildings and roads. Other features, such as water bodies, vegetation, and agricultural lands, show no visible impact.

Disaster restoration advice
Immediate: Immediate recovery actions should focus on clearing debris from the affected buildings and roads to restore connectivity and ensure safe access for emergency responders and reconstruction efforts. Stabilize the partially damaged buildings to prevent further collapse, while initiating sediment removal from the water body to mitigate environmental contamination and support operational recovery of the port.
Long-term: Long-term recovery prioritizes reconstructing sturdy warehouses and roads with explosion-proof and earthquake-proof features, meanwhile implementing strict safety training for staff. Consider future risks, and fortifying the shoreline to prevent erosion is recommended.

DisasterM3: Multi-Hazard, Multi-Sensor, Multi-Task

- **Larger VLMs achieve higher performances.**
- **Remote sensing VLMs still struggle with disaster tasks.**
- **Fine-tuned models improve comprehensively.**

Method	Accuracy (%)							Descrip.	Advice
	AVG	DSR	DTR	BBR	BDC	DRE	ORR		
Random Guess	-	-	20	-	20	20	20	-	-
LLaVA-1.5-7B	12.1	4.2	-	-	-	-	20.0	-	-
Kimi-Instruct	25.6	28.9	66.3	4.0	20.4	15.0	18.9	1.69	2.67
Kimi-Think	26.7	27.0	51.6	7.4	24.4	25.4	24.4	1.61	2.61
InternVL3-8B	31.3	39.6	53.5	4.0	30.3	24.1	36.2	1.96	2.75
Qwen2.5-VL-3B	26.2	30.8	56.1	5.7	29.9	21.2	13.8	1.00	2.15
Qwen2.5-VL-7B	31.2	28.3	66.6	4.7	34.2	29.3	23.9	1.75	1.95
Qwen2.5-VL-32B	35.3	36.7	54.7	11.6	33.2	**30.9**	44.8	1.55	2.96
Qwen2.5-VL-72B	40.5	47.0	74.8	6.8	**34.8**	28.9	50.8	2.01	2.92
GeoChat-7B	10.7	6.1	-	-	-	-	15.3	-	-
TeoChat-7B	23.0	6.9	64.9	2.0	22.5	22.3	18.2	1.77	1.95
EarthDial-4B	22.9	10.6	58.1	3.2	30.2	20.8	14.5	1.53	2.42
GPT-4o	39.3	49.4	**80.5**	10.6	24.2	21.4	49.8	2.27	3.19
GPT4.1	**42.3**	**52.4**	79.6	7.2	25.5	25.0	**64.0**	2.57	3.14
Fine-tuned on DisasterM3 Instruct set									
Qwen2.5-VL-7B	40.4	37.7	83.6	21.5	34.3	29.4	36.2	**3.90**	3.11
Improve↑	**9.2**	**9.4**	**17.0**	**16.8**	**0.1**	**0.1**	**12.3**	**2.15**	**1.26**
InternVL3-8B	41.7	42.6	79.3	**23.9**	29.1	24.9	50.6	3.83	**3.31**
Improve↑	**10.4**	**3.0**	**25.8**	**19.9**	**-1.2**	**0.8**	**14.4**	**1.87**	**0.56**



Enabling AI-Assisted Disaster Response

NEURAL INFORMATION PROCESSING SYSTEMS

Immediate Impact

- First comprehensive disaster VLM benchmark
- Enables rapid damage assessment at scale
- All-weather capability via multi-sensor

Next Steps

- Multi-resolution generalization
- Enhanced sensor diversity
- Living benchmark with new disasters

https://github.com/Junjue-Wang/DisasterM3

Paper

Github

Data

Homepage