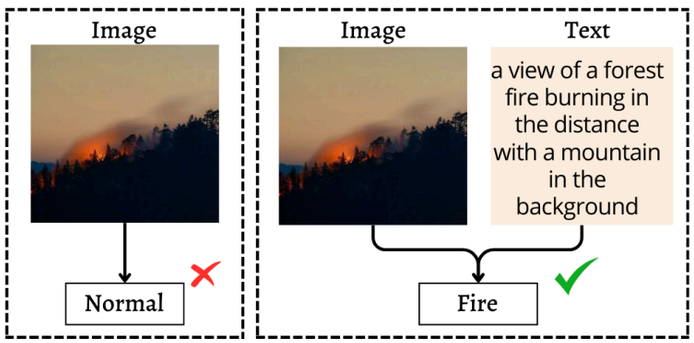
Synergizing Vision and Language in Remote Sensing: A Multimodal Approach for Enhanced Disaster Classification in Emergency Response Systems

Shubham Gupta, Nandini Saini, Suman Kundu, Chiranjoy Chattopadhyay*, Debasis Das

Indian Institute of Technology Jodhpur India, *FLAME University India

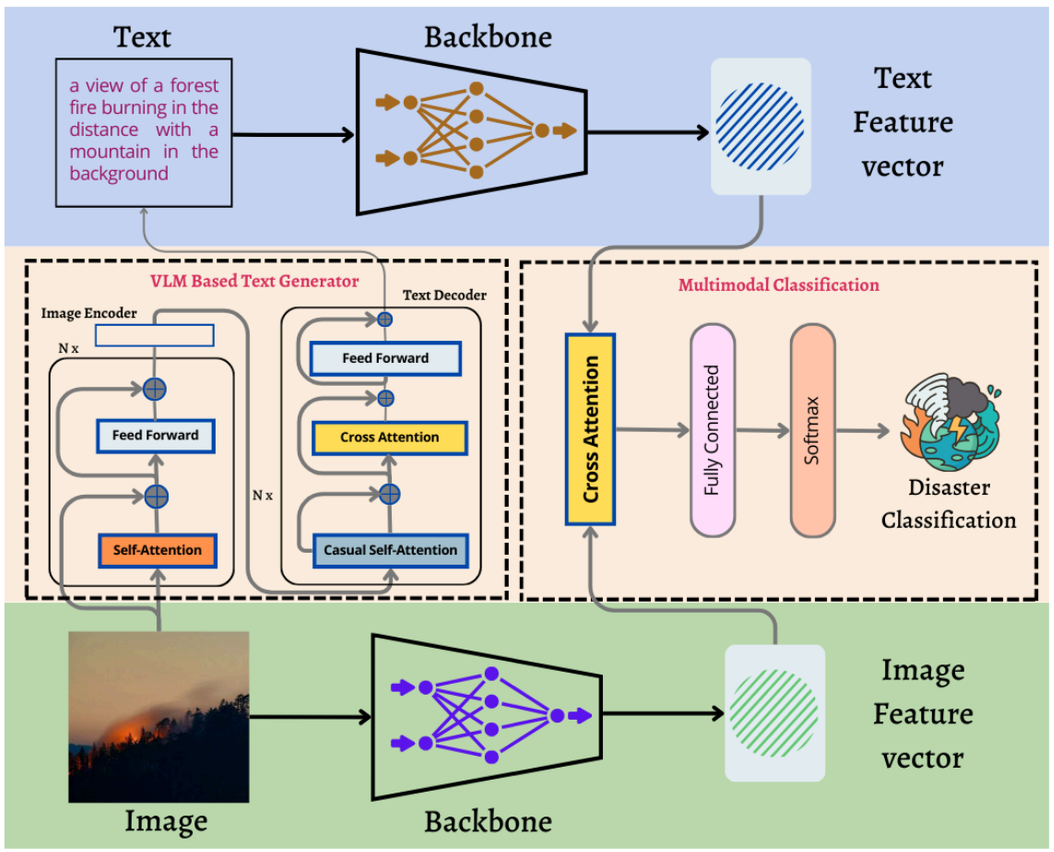
IEEE IGARSS 2024 July 7-12 ATHENS GREECE

Introduction



- Image-only model fails to recognize the disaster when visuals are ambiguous, complex scenes.
- Text can help to specify objects present in remote sensing images
- Lack of method to fill the semantic gap between individual modalities due to inconsistent encoding methods.

Framework



Results

Model	Modality	Overall Accuracy
VGG16	I	91.9
ResNet50	I	90.2
Xception	I	95.3
MobileNet V1	I	95.0
MobileNet V2	I	95.2
MobileNet V3	I	95.3
SqueezeNet	I	91.5
ShuffleNet	I	91.1
FireNet	I	90.5
E^2AlertNet	I	96.0
Ours/ image	T	95.6
Ours/ text	I	96.1
Ours	I+T	**98.0**

- Experiments are conducted on a publicly available AIDER emergency response dataset.
- AIDER consists four classes Collapsed Building, Fire, Flooded Areas, and Traffic Incident.
- Proposed model achieves a new state-of-the-art performance, surpassing the benchmark results of existing models.

Class Label	Precision	Recall	F1-score	OA
Collapsed Building	0.98	0.96	0.97	
Fire	0.98	0.98	0.98	
Flooded Areas	0.98	1.00	0.99	98.04
Traffic Incident	0.96	0.96	0.96	

Contact

{gupta.37, saini.9, suman, debasis}@iitj.ac.in
*chiranjoy.chattopadhyay@flame.edu.in

Scan Me