Big Data System to Support Natural Disaster Analysis

PRESENTER: SHUXIANG ZHU

Natural Disaster

A natural disaster is a major adverse event resulting from natural processes of the Earth; examples are floods, hurricanes, tornadoes, volcanic eruptions, earthquakes, tsunamis, storms, and other geologic processes. A natural disaster can cause loss of life or damage property, and typically leaves some economic damage in its wake.

Damage Caused by Natural Disaster

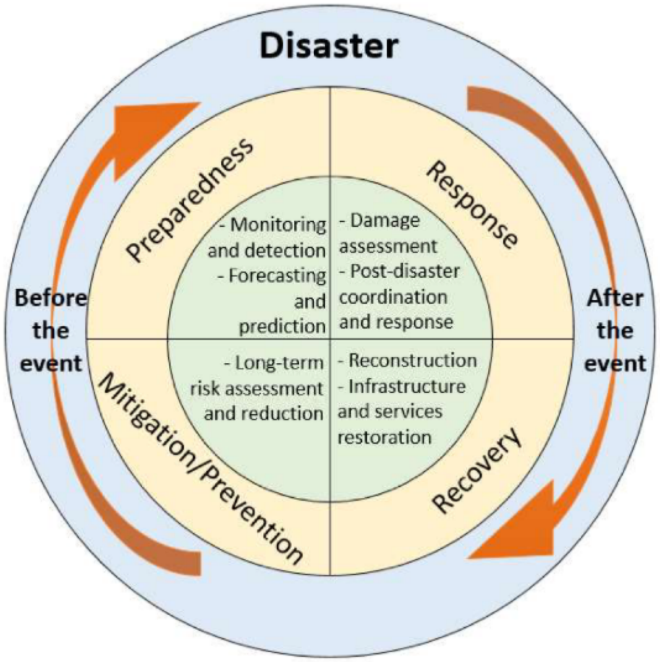
- According to The Centre for Research on the Epidemiology of Disasters (CRED), in 2018 there were 315 natural disaster events recorded with 11,804 deaths, over 68 million people affected, and US$131.7 billion in economic losses across the world.
- Hurricane Maria occurred in 2017 was one of the deadliest Atlantic hurricane which caused 5,740 deaths and $91.61 billion losses in the U.S.
- Camp Fire, the deadliest and most destructive wildfire in California, caused at least 85 civilian fatalities and destroyed 18,804 structures. The total damage was estimated at $16.5 billion

Disaster Management

Emergency management is the organization and management of the resources and responsibilities for dealing with all disasters.

It is a continuous process involving four phases: mitigation/prevention, preparedness, response, and recovery.

1. Preparedness: Detection, monitoring, and prediction
2. Response: Damage assessment, emergency response
3. Recovery: Reconstruction
4. Mitigation: Long-term risk assessment and reduction



Satellite Imagery

- The ability to gather different types of information such as optical and infrared images over wide swaths of the ocean or in remote areas of land
- With the increasing number of satellites constantly sensing our planet, the amount, timeliness, and capability of satellite imagery monitoring natural disasters have improved substantially
- The availability of high-resolution satellite imagery can provide accurate and detailed information for natural disaster analysis

Traditional Disaster Analysis Methods

- Remote sensor based detection and classification of many kinds of disasters have been proved to be useful, however
  - Some of them use visual identification of images for classification, and experts rely primarily on judgment and experience to identify and locate disasters.
  - Some based on the comparison of the changes in the affected area with the satellite images both before and after the disaster, which are more complicated and difficult for real-time analysis.
- Numerous methods using high-resolution satellite images have been raised for building damage assessment.
  - Using methods based on post-disaster images, buildings with smaller damage levels cannot be classified very well.
  - In most studies, the building footprint was provided in the dataset that researchers can focus on how to identify the damage, but not practical in real life.
- Most of these systems focus on one specific natural disaster, which is not robust enough and cannot be expanded for other disaster analysis.

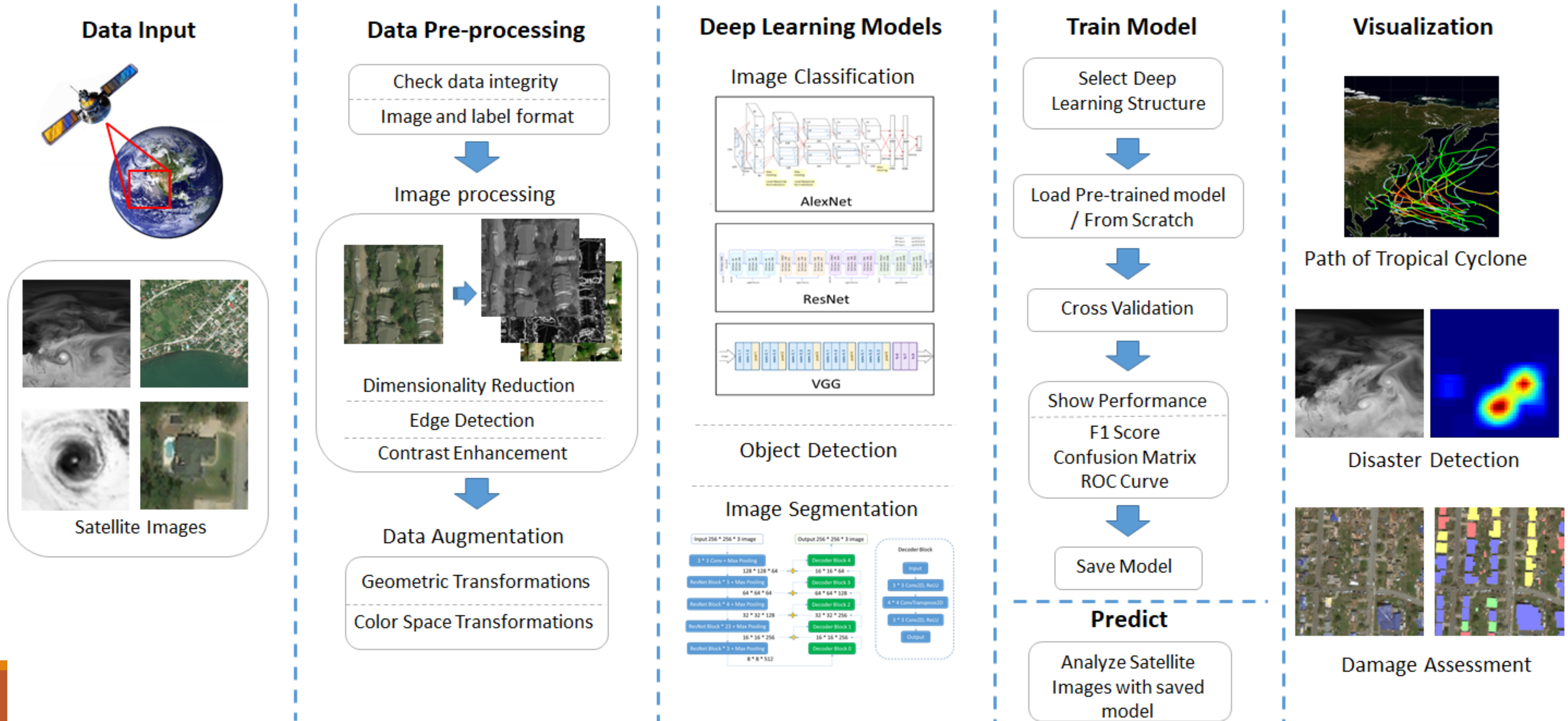
Motivation

- Most of the information in a satellite image is redundant and finding useful information in satellite imagery manually is always time-consuming
- Dealing with different parts of disaster management using multiple systems is not convenient and efficient
- Seeking a scalable and generic automated system that can extract features from different kind of satellite images to support meteorologists collecting valid information and making efforts during the whole process of disaster management.

How to design a robust system?

- Availability to receive different kinds of satellite images
- Customized on-demand image processing
- Multiple functions provided for different disaster analysis requirement
- Near-instantaneous result map visualization

System Overview

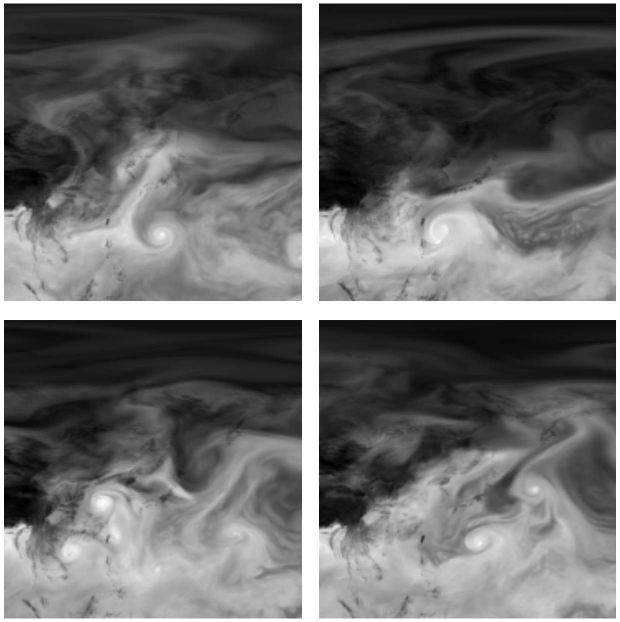


Data Input

- Optical, infrared satellite images
- Colorful images, gray-scale images
- Different image size
- Different data format

Dataset discription

- Global Forecast System (GFS) data
  - Size: 180 * 180
  - Format: grb2
  - Characteristic: GFS data is produced 4 times per day at 00, 06, 12, 18 UTC time
- International Best Track Archive for Climate Stewardship (IBTrACS) data
  - Providing best-track data for tropical cyclones



Examples of GFS data

Dataset discription

- xBD dataset
  - The xBD dataset is released by a challenge of xView2[1]
  - Contains more than 18,000 high-resolution satellite images of different disasters
  - Size: 1024 * 1024
  - Pre-disaster and post-disaster images of the same location
  - Polygons for all 632,228 buildings in the images

Score	Label	Visual Description of the Structure
0	No damage	Undisturbed. No sign of water, structural damage, shingle damage, or burn marks.
1	Minor damage	Building partially burnt, water surrounding the structure, volcanic flow nearby, roof elements missing, or visible cracks.
2	Major damage	Partial wall or roof collapse, encroaching volcanic flow, or the structure is surrounded by water or mud.
3	Destroyed	Structure is scorched, completely collapsed, partially or completely covered with water or mud, or no longer present.

Hurricane-matthew Mexico-earthquake Midwest-flooding Palu-tsunami

Examples of xBD dataset

[1] Ritwik Gupta, Richard Hosfelt, Sandra Sajeev, xbd: A dataset for assessing building damage from satellite imagery. arXiv preprint arXiv:1911.09296, 2019.
https://xview2.org/dataset

System Overview

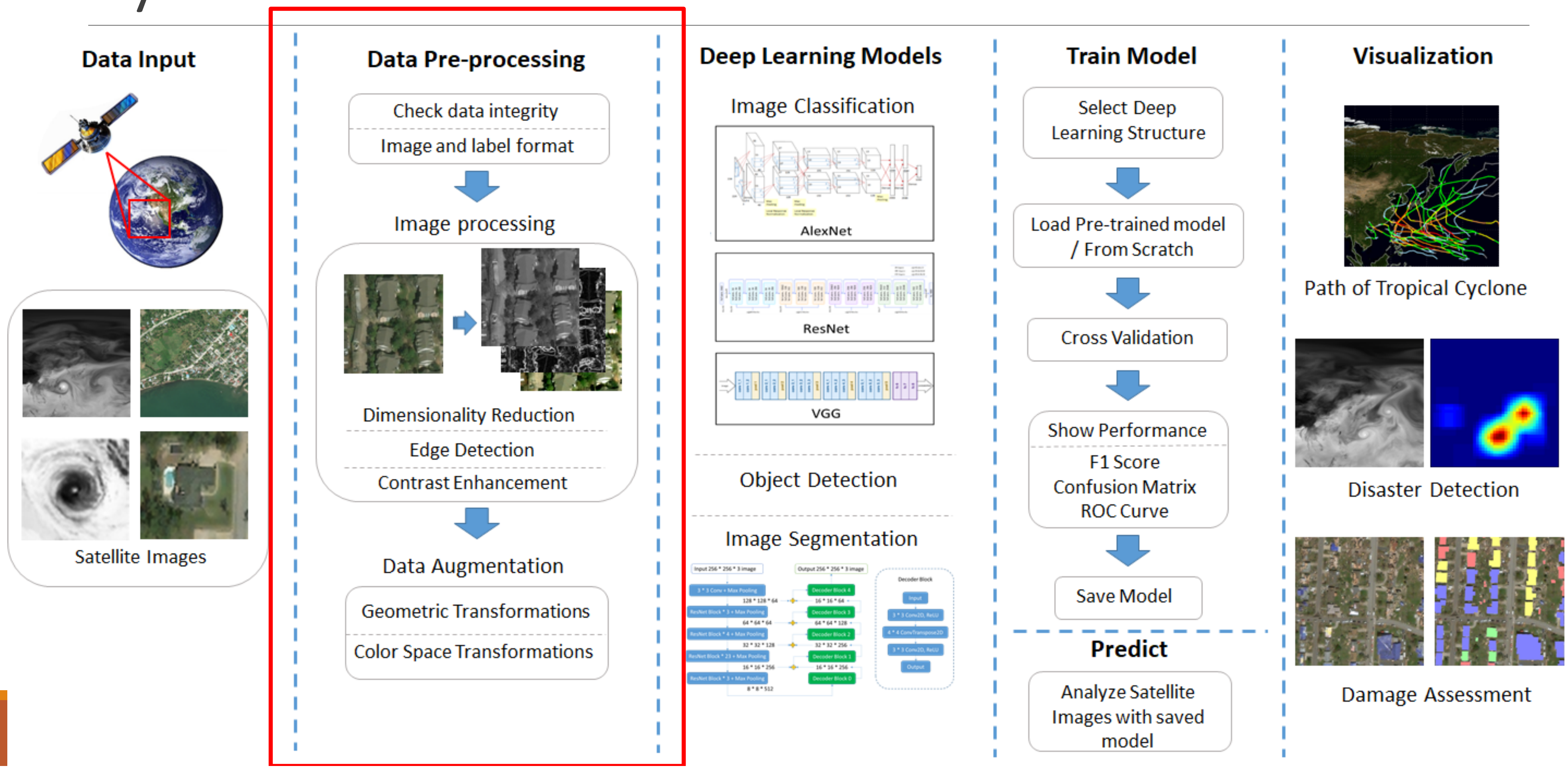


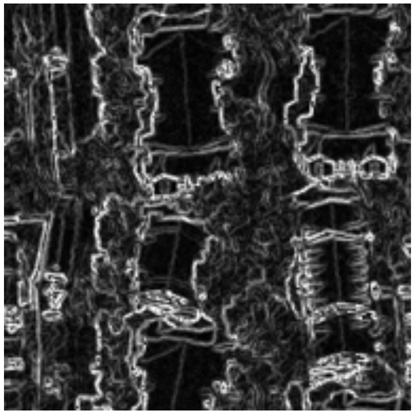
Image Processing Methods

- Mean Normalization and Standardization
  - Mean Normalization: rescale the input data into a certain range to maintain a similar data distribution
  - Standardization: rescales the data to have a mean of 0 and a standard deviation of 1
- Dimensionality Reduction
  - Reduce the number of redundant features, e.g. RGB image to Gray-scale image
- Contrast Enhancement
  - A process that changes the image value distribution to cover a wide range, which can optimize the quality of an image
- Edge Detection
  - E.g. Sobel kernel, Canny kernel

Image Processing Methods

Original RGB Image

Dimensionality Reduction



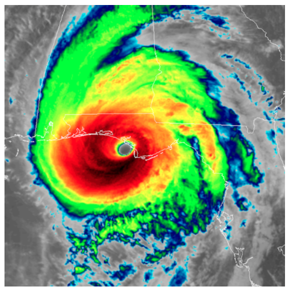
Edge Detection

Contrast Enhancement

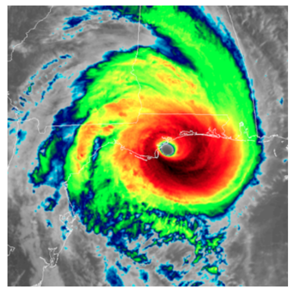
Data Augmentation

- To get a good performance and avoid overfitting, deep learning models are heavily reliant on big data
- When limited data is provided for training, data augmentation is used to enhance the size and quality of the training dataset.
- Oversampling the minority data in the classification task to solve the data imbalance problem
- Two ways of data augmentation:
  - Geometric Transformations: flipping, rotating, cropping, scaling, translation
  - Color Space Transformations: generate more data in different lighting conditions

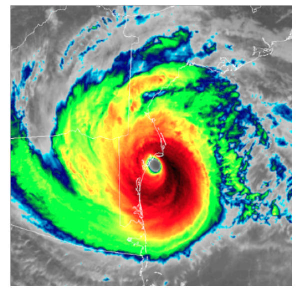
Data Augmentation



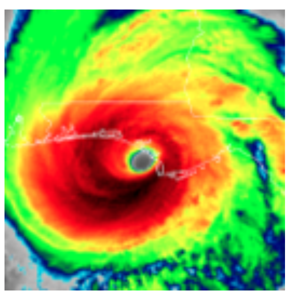
Original RGB Image



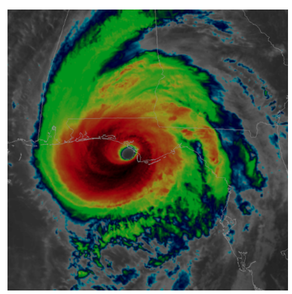
Flip

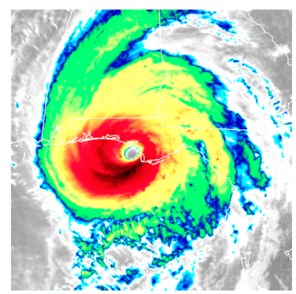


Rotate



Scale and Crop

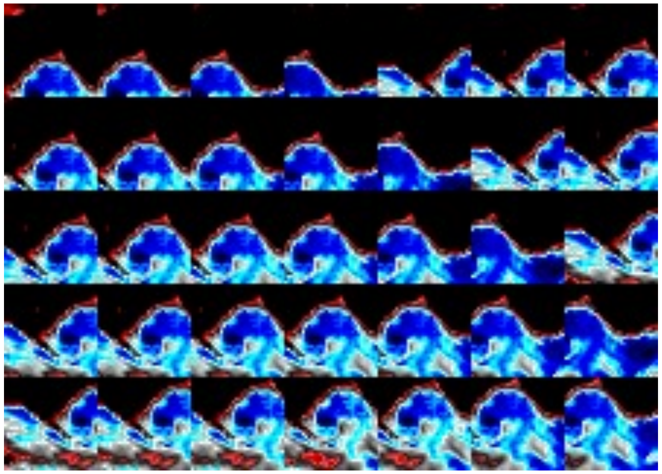




Color Space Transformation

Application of Sliding Window

- Sliding windows play an integral role in object detection and classification.
- Used to localize exactly where in an image an object resides.
- Each image can generate W * H samples with sliding window



W = H = (R - A) / S

where R = length of region

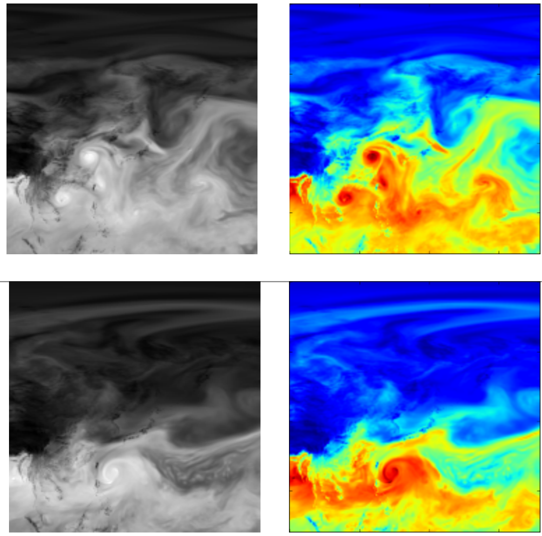
A = length of area

S = Interval points between area

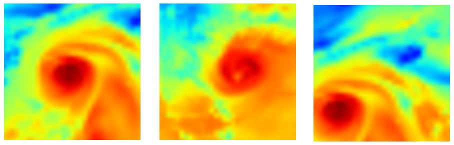
Sliding window example

Data Pre-processing

- Color Enhancement
  - Apply color enhancement to allow easy identification of features and areas of interest
- Sliding Window
  - Generate training samples: positive samples as regions of data containing a tropical cyclone
  - Sliding window size: 32 * 32



Color Enhancement



Positive samples

Negative samples

Image Pre-processing
Building Segmentation

- Segmentation task was based on the pre-disaster images
  - The structure of the buildings is more complete before the disaster
  - The surrounding area of the buildings is more regular
- Enhanced the sharpness and contrast of the images
- Divided the original masks into two parts: the boundary masks and the masks which overlapping boundaries are removed.
- Each image was randomly rotated and cropped into 256 * 256 images

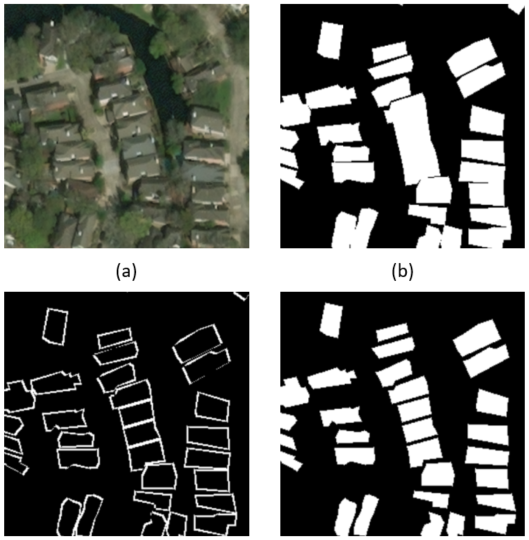
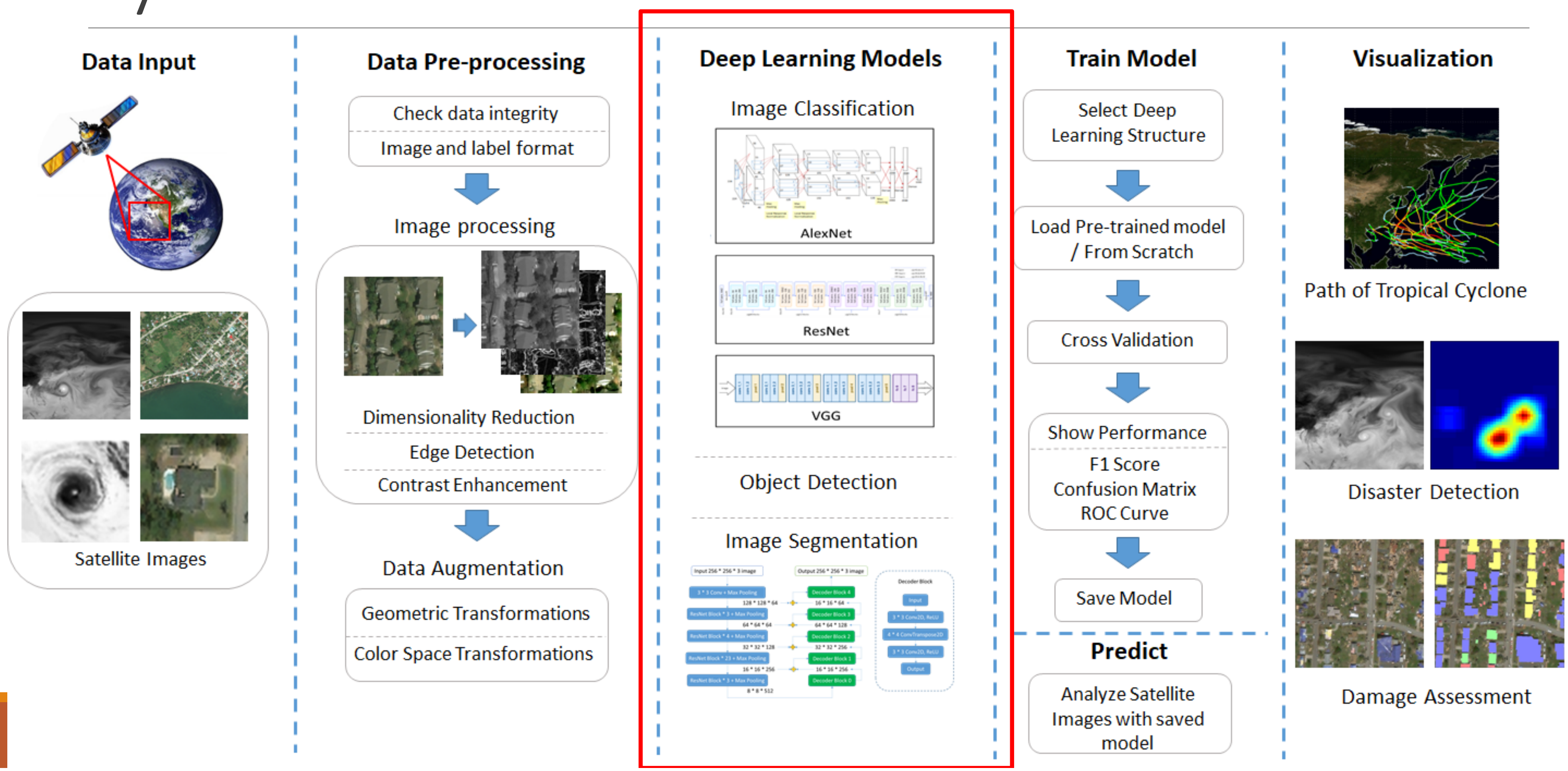

(a) (b) (c) (d)

Image Pre-processing
Damage Classification

- The building images are cropped according to the segmented polygons and resized to 64 * 64
- Applied different crop scheme on the buildings to obtain different ratio of surrounding area
  - Building damage level is classified by both the structure damage and the surrounding area damage
- The input data consists of 8 channels:
  - Pre-disaster and post-disaster color image pair with more surrounding area
  - Pre-disaster and post-disaster gray-scale image pair focusing on building structure
- Data augmentation: using random rotation and flipping to solve the data imbalance problem

(a) (b) (c) (d)

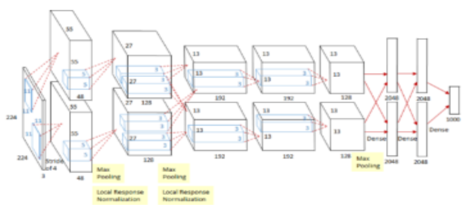
System Overview



Deep Learning Models

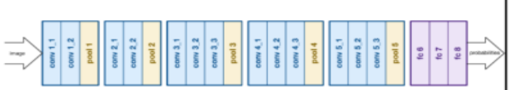
- Image Classification
  - Distinguish an object in the image or the scene of the image into a pre-defined category
- Image Segmentation
  - The process of partitioning an image into multiple objects.
- Object Detection
  - Identifies and localizes the object in an image
  - With the help of sliding window, use image classification model to identify target

Image Classification Models



AlexNet

ResNet

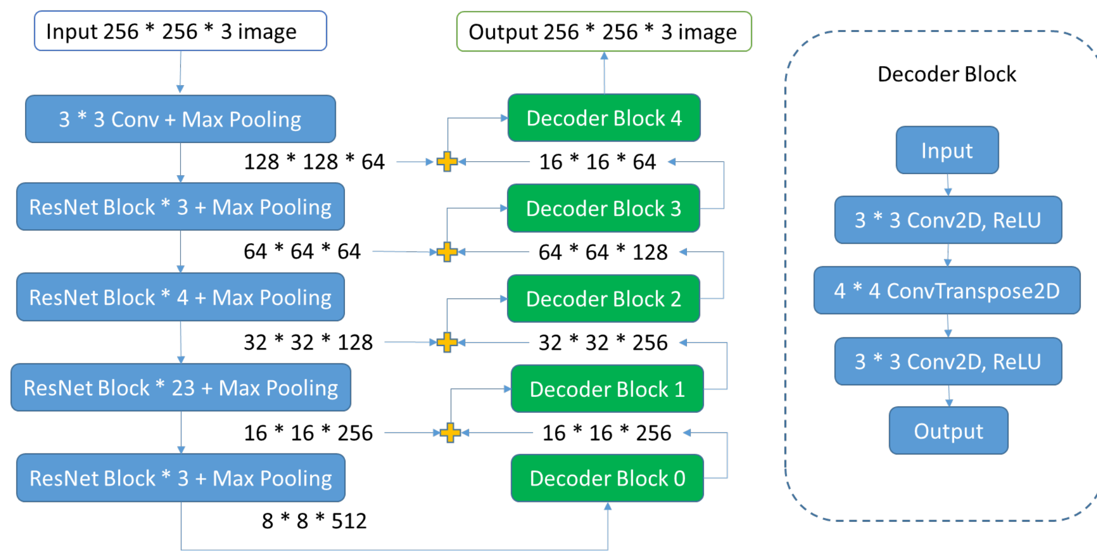


VGG

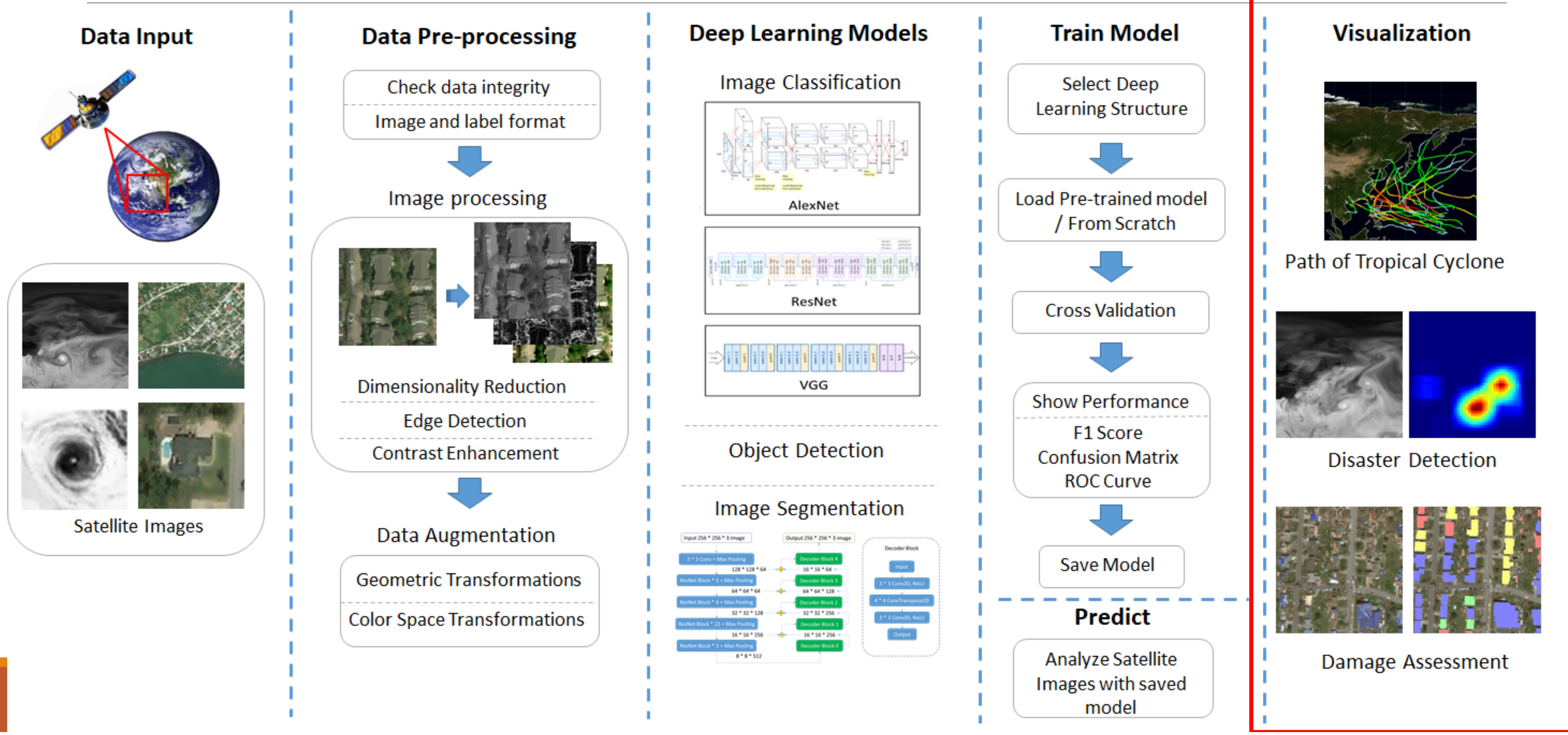
ConvNet configuration							
CNN		Alexnet		VGG		ResNet	
conv	$5 \times 5 \times 3$ - 64	conv	$3 \times 3 \times 3$ - 64	conv	$3 \times 3 \times 3$ - 64	conv	$3 \times 3 \times 3$ - 64
norm		norm		pool	2×2	conv	$3 \times 3 \times 3$ - 64
pool	3×3	pool	2×2	norm		pool	2×2
conv	$5 \times 5 \times 64$ - 64	conv	$3 \times 3 \times 64$ - 192	conv	$3 \times 3 \times 64$ - 64	conv	$3 \times 3 \times 3$ - 128
norm		norm		norm		conv	$3 \times 3 \times 3$ - 128
pool	3×3	pool	2×2	pool	2×2	pool	2×2
fc	384	conv	$3 \times 3 \times 192$ - 384	conv	$3 \times 3 \times 64$ - 128	conv	$3 \times 3 \times 3$ - 256
fc	192	norm		pool	2×2	conv	$3 \times 3 \times 3$ - 256
softmax		conv	$3 \times 3 \times 384$ - 256	norm		pool	2×2
		norm		conv	$3 \times 3 \times 64$ - 128	conv	$3 \times 3 \times 3$ - 512
		conv	$3 \times 3 \times 256$ - 256	norm		conv	$3 \times 3 \times 3$ - 512
		norn		pool	2×2	pool	2×2
		pool	2×2	fc	768	conv	$3 \times 3 \times 3$ - 512
		fc	4096	fc	384	conv	$3 \times 3 \times 3$ - 512
		fc	4096	softmax		pool	2×2
		softmax				fc	4096
						fc	4096
						fc	1000
						softmax	

Image Segmentation Model

- Consists of an encoder, a decoder and skipped bridges that connect the features extracted by the encoder to the decoder.
- The input of a decoder block is a combination of the previous decoder blocks' output and the output of an encoder block
- Use ResNet as the encoder

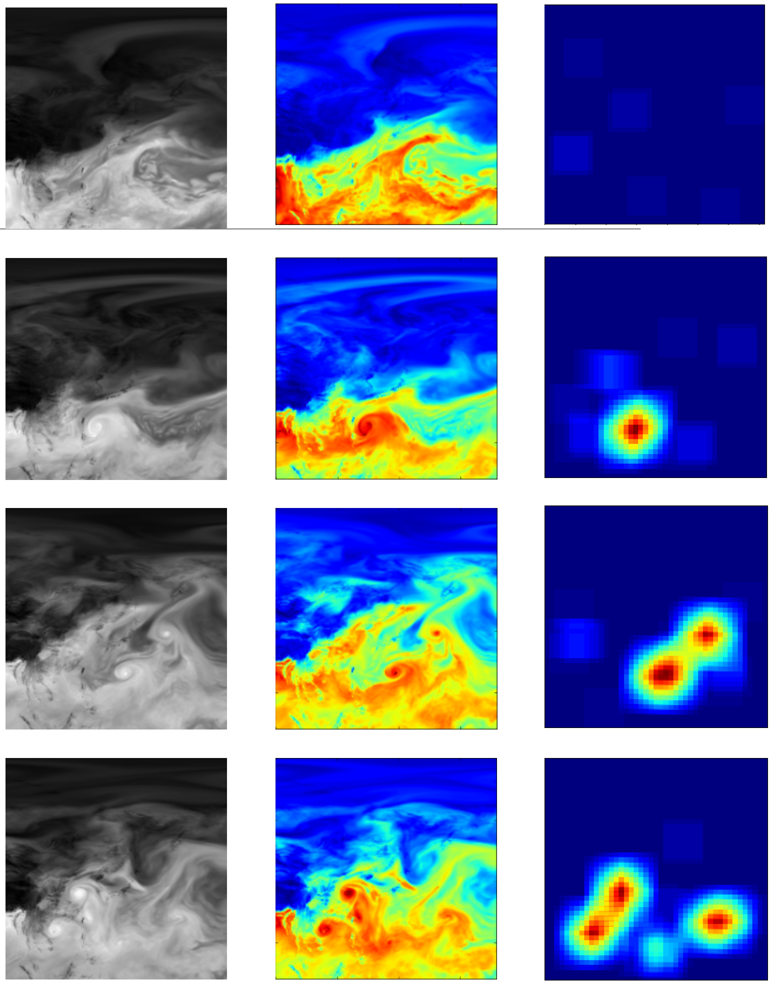


System Overview



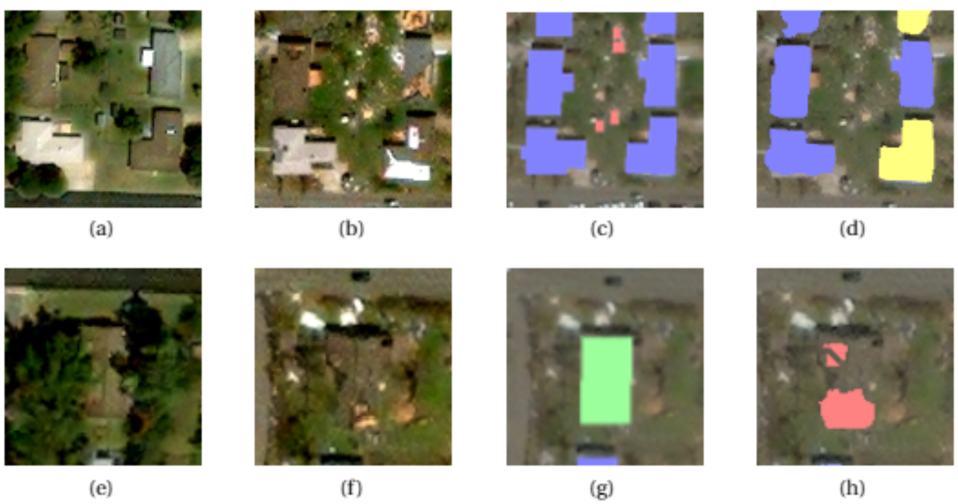
Visualization

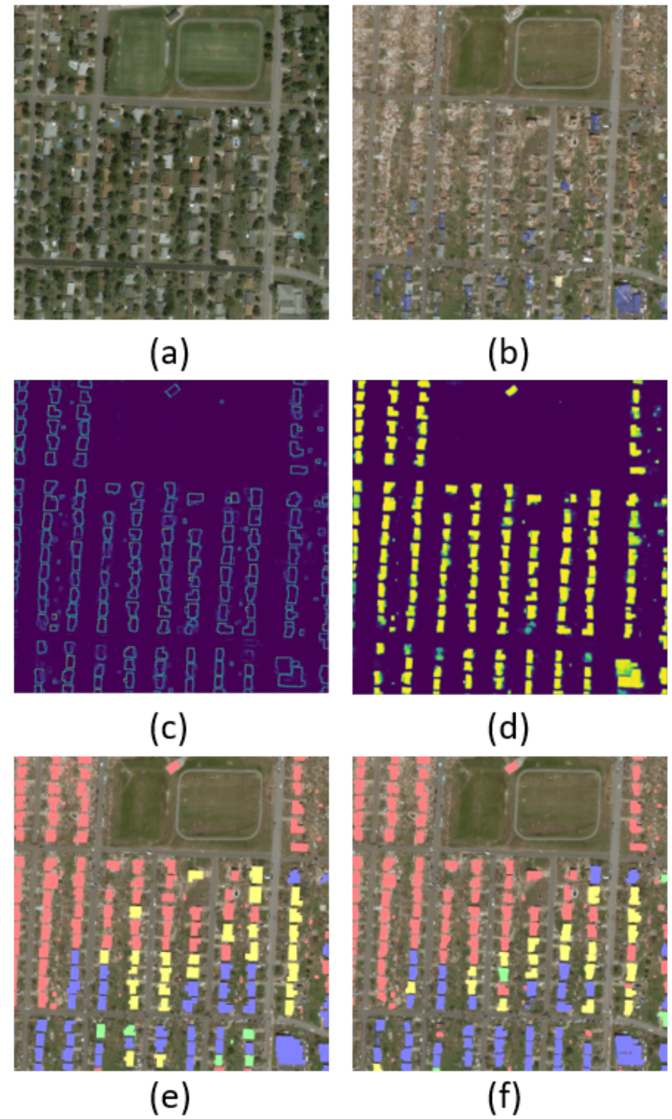
- ➢The deep learning model produces multiple predictions at each spatial location.
- ➢Pixel-wise probability of tropical cyclone detection is calculated
- ➢The system does a good job of identifying and localizing the tropical cyclone.
- ➢The big data system is capable of identifying the area of interest in a given region.



Visualization

- Green, blue, yellow and red represents no-damage, minor-damage, major-damage and destroyed respectively.
- The overall performance is great.
- From the visualized damage level images, experts can easily assess the damage caused by a disaster and organize efficient emergency response as soon as possible.



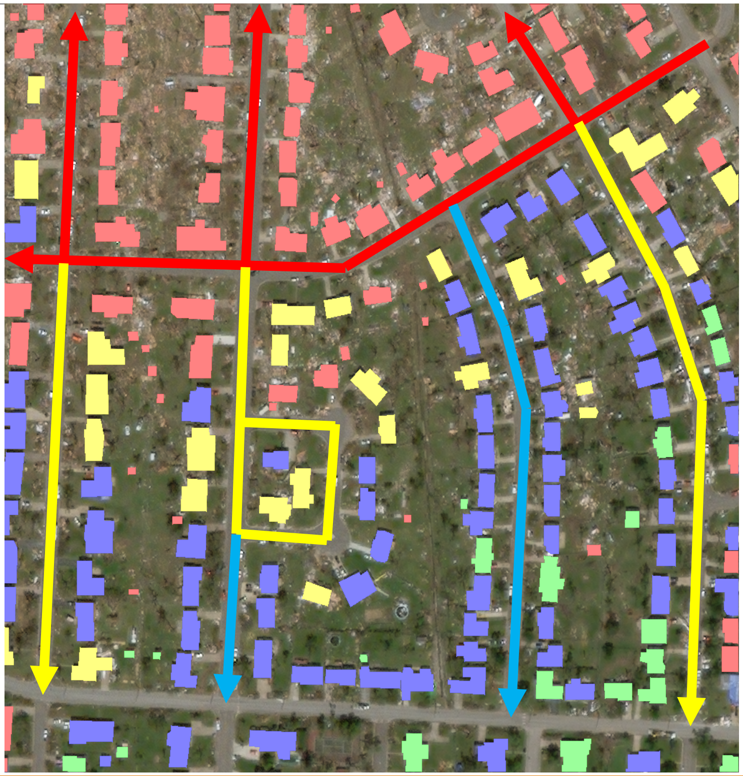


Path planning for Rescue and Recovery

The selection of suitable rescue/evacuation paths is very important to reduce the loss of disasters.

An inspection of damage severity through colorful damage maps can pinpoint major-damaged and destroyed buildings where victims may be trapped.

After disasters happen, the visualized building damage map provides clear information about the areas with different damage levels, which can intuitively support decision making during the recovery phase.



Conclusion

- Natural disasters can cause enormous impacts on daily life and industrial production. In time analysis of disaster based on satellite images can be an efficient approach to reduce effects. The big data system we designed is an ideal solution to support disaster analysis.
- The system we built is composed of a data pre-processing part, three aspects of deep neural network models of which purposes consist of image classification, image segmentation and object detection, a structure to train and use the models, and a visualization part.
- Two test cases – Tropical cyclone detection and building damage classification, illustrated that this big data system is capable of producing real-time detection probability maps, and segmentation masks, which has the ability to support the experts to analyze the natural disasters and help communities that demand robust and resilient responses for rapid recovery.