Hurricane Damage Detection using Computer Vision

Amrita Ramesh
Department of Electronics and Communication Engineering, PES University, India
amritaramesh08@gmail.com

Sanjana K R Prasad
Department of Electronics and Communication Engineering, PES University, India
sanjanaprasadkr@gmail.com

Siddhanth Srikanth
Department of Electronics and Communication Engineering, PES University, India
ssiddhanth06@gmail.com

Shikha Tripathi
Department of Electronics and Communication Engineering, PES University, India
shikha@pes.edu

ABSTRACT

Post hurricane damage assessment is a key requirement for first responders to effectively and accurately aid in disaster recovery. This paper focuses on the use of satellite imagery of buildings across an area, to identify four different degrees of damage - no damage, minor damage, major damage and destroyed. The main contribution of this work is to explore the state-of-the-art Vision Transformer (ViT) architecture, and analyse its efficacy in classifying the satellite image dataset into the above-mentioned classes. Various Convolutional Neural Networks (CNNs) have also been trained, and a comparative analysis of these architectures with the Vision Transformer has been carried out. The reported existing techniques have utilized image pre-processing and data augmentation using Generative Adversarial Networks. ViT successfully outperformed the CNNs, to give a higher accuracy.

CCS CONCEPTS

• **Computing methodologies**; • **Artificial intelligence**; • **Computer vision**; • **Computer vision problems**;

KEYWORDS

vision transformer, ResNet, InceptionNet, EfficientNet, Visual Geometry Group, generative adversarial network

ACM Reference Format:

Amrita Ramesh, Sanjana K R Prasad, Siddhanth Srikanth, and Shikha Tripathi. 2023. Hurricane Damage Detection using Computer Vision. In *2023 5th International Conference on Image, Video and Signal Processing (IVSP 2023), March 24–26, 2023, Singapore, Singapore.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3591156.3591174

1 INTRODUCTION

One of the most adverse effects of climate change comes in the form of natural disasters. Natural catastrophes are unpredictable and pose a risk to both life and property. One of nature's strongest storms, hurricanes can produce a variety of dangerous conditions, such as tornadoes, storm surges, heavy rain, and strong winds. Strong winds can damage buildings, bridges, and other exterior objects on land, causing loose material to become lethal flying projectiles. Damage from all hurricanes that were reported between 1980 and 2021 totals 1.1 trillion and the average cost of damage from these storms is 20.5 billion. Disaster aid is crucial to minimizing the harm caused by catastrophic events.

The first step in forming an effective post disaster rescue plan is damage detection. Various techniques such as image pre-processing [6, 10], image augmentation [8, 12], region of interest segmentation, feature extraction, classification and object detection using deep learning methods [3] such as convolutional neural networks [1, 2], recurrent neural networks, auto-encoders and deep neural networks have been used. It is also interesting to see the use of computer vision technology for hurricane damage detection.

The major contribution of this work is to compare the performance and accuracy of four pre-trained Convolutional Neural Networks [9] – VGG, EfficientNet, ResNet and InceptionNet with a Vision Transformer [4, 5, 7] in achieving classification of the degrees of damage post hurricanes into four classes – no damage, minor damage, major damage and destroyed.

The rest of the paper is organized as follows: Section 2 highlights the preprocessing techniques followed by Section 3 where implementation of GANs is discussed. Section 4 and 5 elaborated the CNN and GAN implementation. Section 5 discusses the Vision Transformers and Section 6 concludes the paper.

DATASET

We have use the xBD dataset [13] for the deep learning models, which consists of pairs of pre-disaster and post-disaster 1024×1024 satellite images. The train, validation, and test sets consist of 16470, 1833 and 1866 images respectively. The dataset provides 4-level damage labels, including no damage, minor damage, major damage and destroyed as shown in Figure 1.

2 PRE-PROCESSING

Preprocessing image datasets help highlight important features that may not be distinguishable to the naked eye. Additionally, it improves performance of models that take image inputs. We have used the following methods to pre-process the satellite images as shown in Figure 2:

Thresholding was performed on the images in order to separate object or foreground pixels, from background pixels. In our case, we used adaptive mean thresholding where the mean value of the

IVSP 2023, March 24–26, 2023, Singapore, Singapore

ACM ISBN 978-1-4503-9838-1/23/03…$15.00
https://doi.org/10.1145/3591156.3591174

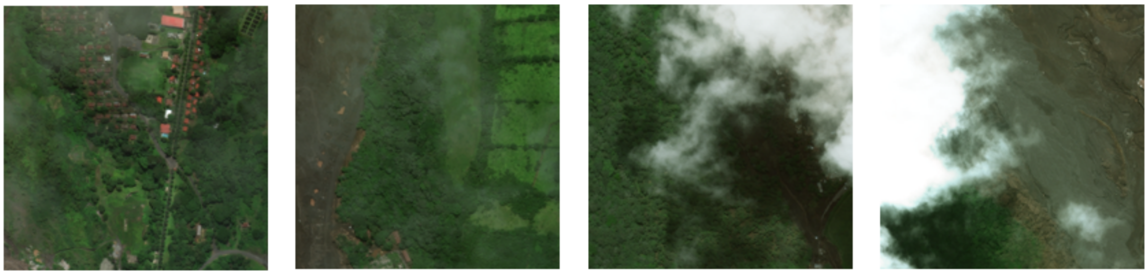

Figure 1: xBD dataset images of classes no damage, minor damage, major damage and destroyed respectively (https://arxiv.org/abs/1911.09296v1)

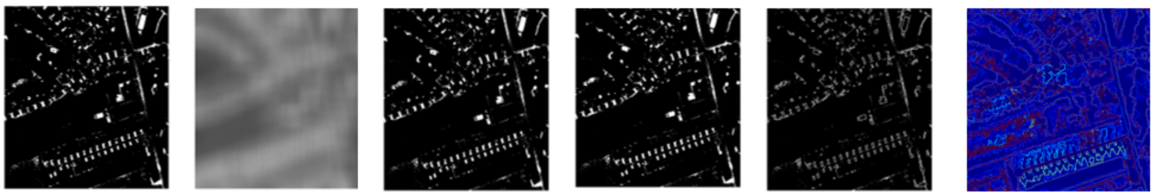

Figure 2: Processed images after steps thresholding, noise removal, opening, flood filling, boundary extraction and final processed output respectively.

pixel was considered as the threshold. This was followed by noise removal using Weiner filter to reduce Gaussian noise. Opening was used to relocate tiny items from the forefront of the image, which is typically seen as its brightest pixels, into the background. We used a structural element in the form of a disk of radius 4 pixels. Flood-filling was done to fill in gaps in images. This is useful in removing irrelevant artefacts from images. In order to gain information about the boundary of image features, boundary extraction was carried out by finding the difference of the original and dilated image.

3 GENERATIVE ADVERSARIAL NETWORK

Generative Adversarial Networks (GANs), are a type of generative modelling that uses deep learning methods like convolutional neural networks. It is a machine learning task that involves recognizing and understanding the patterns of chronicity in input images such that the model may be utilised to manufacture new images that could have been from the source dataset.

An interesting use of GANs for computer vision problems is in the form of data augmentation [11] to generate additional samples and help balance datasets. We opted for the deep convolutional generative adversarial network (DCGANs) [14] as shown in Figure 3 and Figure 4 for augmentation to improve the dataset size.

4 CONVOLUTIONAL NEURAL NETWORKS

Convolutional neural networks are one of the various types of network architectures for deep learning, and are commonly used to interpret visual imagery. A convolutional neural network could be made up of a large number of layers, and every layer is taught to identify different facets of an image. Every training image is run through a series of filters at different resolutions, and the resulting convolved image is used as the input for the layer after that. The filters start with very fundamental qualities like brightness and borders and then become more complex until they reach attributes that uniquely recognise the object. We have compared the outputs of the following pretrained models:

Visual Geometry Group

VGG or Visual Geometry Group is a deep convolutional neural network architecture with numerous layers. It is an innovative object-recognition model that supports up to 19 layers as shown in Figure 5.

InceptionNet

InceptionNet is a deep neural network that carries out 1x1 convolutions. Its structure as shown in Figure 6 constitutes recurring components known as Inception modules.

ResNet

When the number of layers is raised, vanishing/exploding gradient is a common problem in convolutional neural networks. The ResNet design as shown in Figure 7 proposes the concept of Residual Blocks to overcome the vanishing gradient problem. In this network, we use the skip connections technique. The skip connection connects the activations of one layer to those of other layers by hopping over a few levels in between. This produces a residual block. ResNets are created by stacking these blocks.

EfficientNet

EfficientNet is a convolutional neural network model that utilises a compound coefficient to uniformly design and scale all dimensions of resolution as shown in Figure 8. Compared to earlier convolutional neural networks, EfficientNets reach substantially higher levels of accuracy and efficiency.

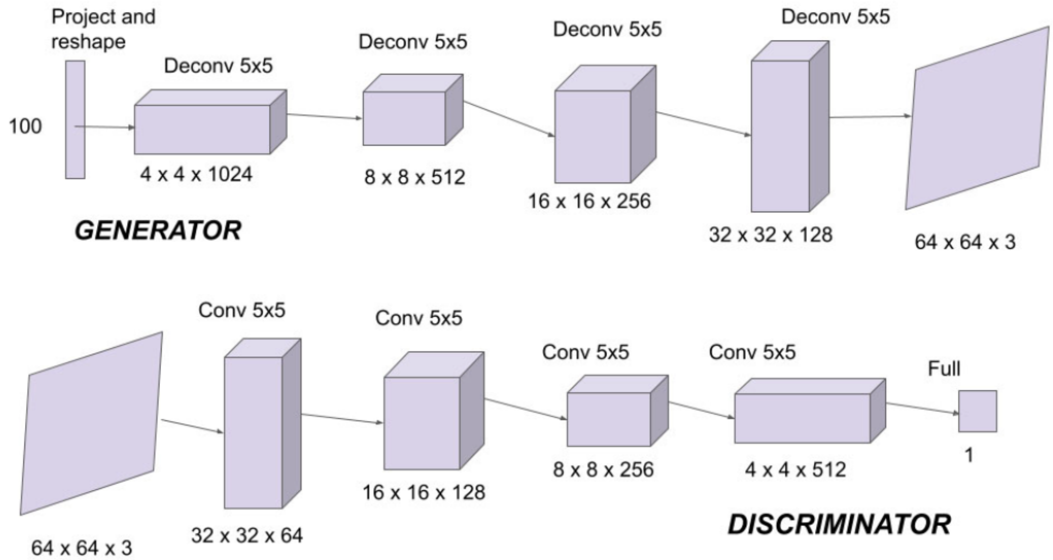

Figure 3: Architecture of deep convolutional generative adversarial networks. Via Various Generative Adversarial Networks Model for Synthetic Prohibitory Sign Image Generation. (https://www.mdpi.com/2076-3417/11/7/2913)

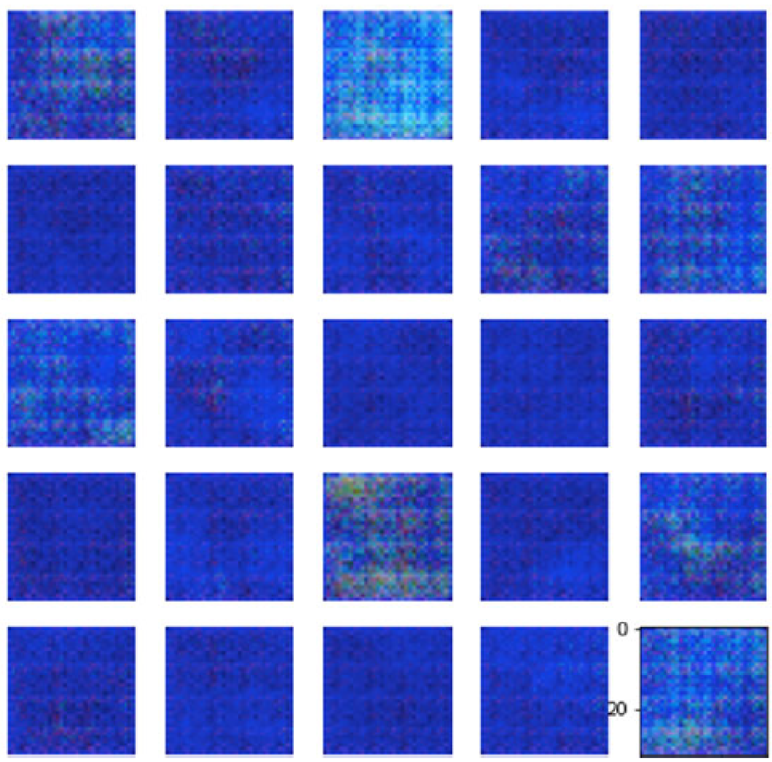

Figure 4: Synthetic images generated by the DCGAN

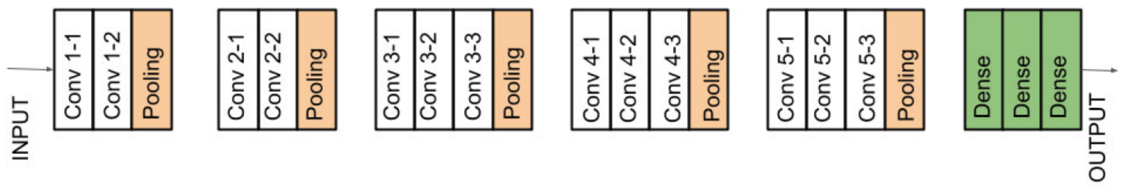

Figure 5: Architecture of VGG. Via GeeksforGeeks (https://www.geeksforgeeks.org/vgg-16-cnn-model/)

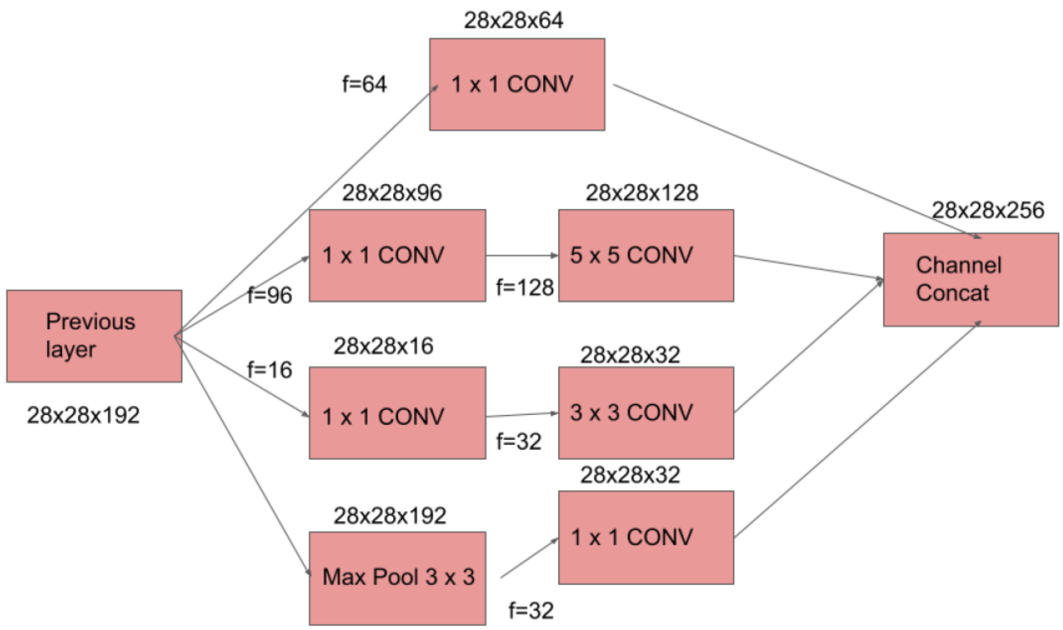

Figure 6: Architecture of InceptionNet. Via GeeksforGeeks (https://www.geeksforgeeks.org/ml-inception-network-v1/)

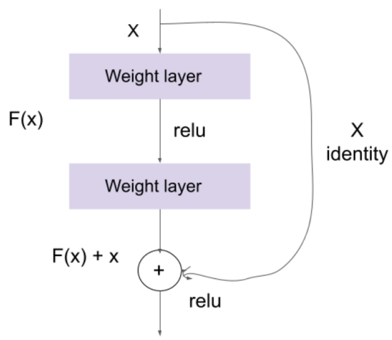

Figure 7: Architecture of ResNet. Via GeeksforGeeks (https://www.geeksforgeeks.org/residual-networks-resnet-deep-learning/)

5 VISION TRANSFORMERS

The vision transformer, or ViT as shown in Figure 9 is a methodology for classifying images. Patches of an image are processed using an architecture akin to the old to a transformer. A vector sequence is formed by segmenting the input images into patches of fixed-size, embedding each one linearly, linking its positional embedding, and finally feeding the resulting vectors to a standard transformer encoder. The standard classification process involves inserting a second learnable “classification token" into the sequence.

Vision transformer has a relatively smaller inductive bias than CNNs, leading to a greater reliance on model normalisation, which may be problematic when training on smaller datasets. However, the dependency is diminished if a substantial dataset can be utilised.

While analysing images, if only pixel values are considered, it may mistake noisy or undesirable pixels for important features, which is a possibility while using CNN. In ViT however, the images are divided into visual tokens. The ViT analyses data effectively by breaking a picture up into fixed-size patches and adding positional embedding as an input to the transformer encoder. This aids in better results when the quality of the images is not sufficient.

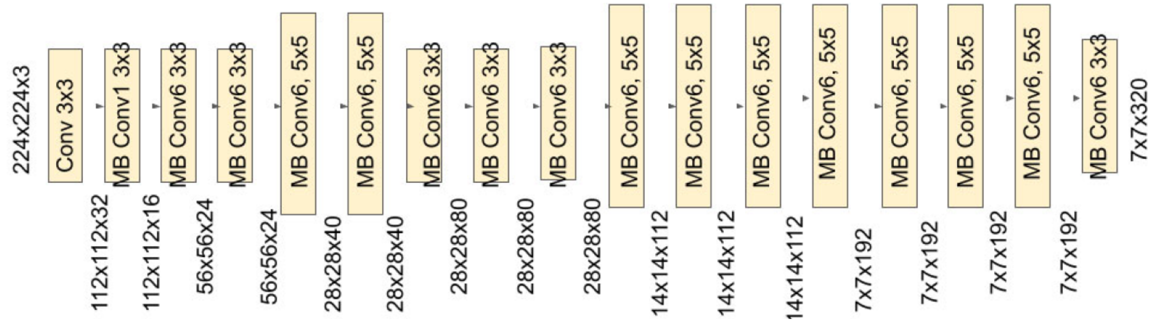

Figure 8: Architecture of EfficientNet. Via Google AI Blog (https://ai.googleblog.com/2019/05/efficientnet-improving-accuracy-and.html)

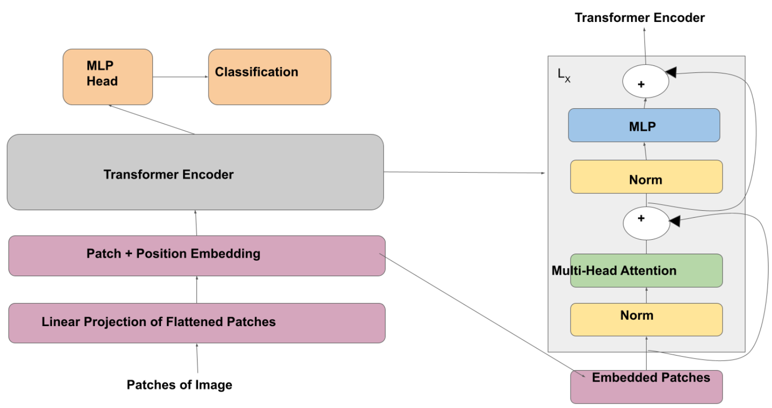

Figure 9: Architecture of Vision Transformer. Via Medium (https://medium.com/analytics-vidhya/understanding-the-vision-transformer-and-counting-its-parameters-988a4ea2b8f3)

The following steps describe the working of a vision transformer: First, patches are created from an image, and these are flattened. These flattened patches are transformer into linear embeddings of lower dimension. Each position is assigned a unique representation and the sequence is fed into a transformer encoder as input. The model is pretrained using labels from images (fully supervised on a substantially big dataset) and the dataset is calibrated for the image classification step.

IMPLEMENTATION

Our implementation began with procuring the satellite image dataset and performing pre-processing techniques mentioned in Section 2 on the images. This was followed by the process of data augmentation with the generative adversarial network. We settled on deep convolutional generative adversarial network as the best fit for our requirements. Our objective was to understand how ViTs compare to Convolutional Neural Networks (CNNs) in terms of their performance and classification abilities. Our focus was on both accuracy and efficiency.

Visual Geometry Group can accurately perform image recognition using small convolutional filters with a deep network. The selection can also be owed to its requirement of a smaller amount of training samples per time, faster training speed as well as generally higher accuracy. InceptionNet covers bigger areas at a time while still maintaining fine resolution to not lose small details and information. It also has feature extractors that assist the network in performing better. This results in good accuracy as well as efficiency. ResNet18 was selected due to its favorable balance between classification accuracy and training time. ViT utilizes more global information than ResNet at the lower layers, resulting in distinctive features. Furthermore, skip connections play a more crucial role in ViT performance compared to ResNets. EfficientNetV2 has proven to be a game-changer in the field of image classification as it not only improves accuracy but also reduces training time and latency. However, when using large image resolutions, the training process becomes slower, particularly in the early layers of the network architecture where depthwise convolutional layers take longer to train.

Following the previous steps, we trained the dataset which consisted of a combination of the original dataset with the synthetically generated samples on the four pre-trained convolutional neural network models. The vision transformer was similarly trained on the dataset. The metric employed was accuracy which measures the model's performance by determining the number of correct

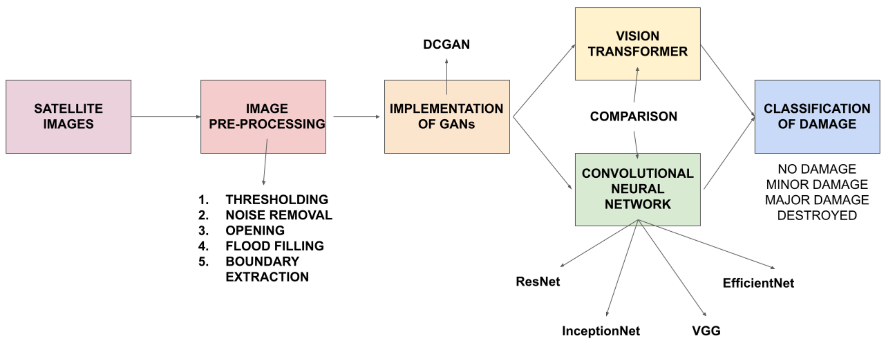

Figure 10: Block Diagram.

Table 1: Accuracy Values of CNN and vision transformer

Model	VGG	InceptionNet	ResNet	EfficientNet	ViT
Accuracy	71.02 %	71.02 %	71.02 %	71.02 %	88.49 %

predictions made compared to the total number of predictions. Each model was trained for 216 epochs, the convolutional neural networks used the ReLU activation function with the Adam optimiser and the vision transformer used the GELU activation function with AdamW optimiser. Finally, the results were compared. The process flow block diagram is shown in Figure 10.

6 RESULTS AND ANALYSIS

As can be seen in Table 1, the vision transformer architecture successfully exploits large amounts of augmented image data, to provide an accuracy of 88.49%, for the task of classifying images from the dataset. The pretrained CNN models ResNet, InceptionNet, EfficientNet and VGG, all provide a lower accuracy of 71.02% each.

The vision transformer (ViT) architecture has several key differences compared to the convolutional neural network (CNN) architecture. In terms of computational efficiency, ViTs outperform CNNs by nearly four times. ViTs also have a weaker inductive bias, which means they rely more on data augmentation to produce accurate results. At lower levels, ViT incorporates more global data than CNN, resulting in quantitatively different features. Additionally, the skip connections in ViT have a stronger impact on performance and representation similarity than in CNN. CNNs require additional layers to compute equivalent representations to a smaller set of ViT lower layers. Furthermore, CNNs use a global average pooling step in their classification training, whereas ViT uses a separate classification token.

7 CONCLUSION

In this paper, we have performed a comparative study of the performance of convolutional neural networks and vision transformers. We have also used generative adversarial network in order to augment the dataset before they are input into the models. We can conclude that, on the augmented XBD dataset, the state-of-art vision transformers achieve a higher accuracy (88.49%) than the convolutional neural networks (71.02%).

REFERENCES

[1] Anju Asokan, J. Anitha, Monica Ciobanu, Andrei Gabor, Antoanela Naaji, and D. Jude Hemanth. 2020. Image Processing Techniques for Analysis of Satellite Images for Historical Maps Classification—An Overview. Applied Sciences 10, 12 (2020). https://doi.org/10.3390/app10124207

[2] Polina Berezina and Desheng Liu. 2022. Hurricane damage assessment using coupled convolutional neural networks: a case study of hurricane Michael. Geomatics, Natural Hazards and Risk 13 (12 2022), 414–431. https://doi.org/10.1080/19475705.2022.2030414

[3] Quoc Dung Cao and Youngjun Choe. 2020. Building damage annotation on post-hurricane satellite imagery based on convolutional neural networks. Natural Hazards 103 (09 2020). https://doi.org/10.1007/s11069-020-04133-2

[4] Yifan Da, Zhiyuan Ji, and Yongsheng Zhou. 2022. Building Damage Assessment Based on Siamese Hierarchical Transformer Framework. Mathematics 10 (06 2022), 1898. https://doi.org/10.3390/math10111898

[5] Saramsha Dotel, Avishekh Shrestha, Anish Bhusal, Ramesh Pathak, Aman Shakya, and Sanjeeb Panday. 2020. Disaster Assessment from Satellite Imagery by Analysing Topographical Features Using Deep Learning. 86–92. https://doi.org/10.1145/3388818.3389160

[6] Ishita Dutta, Sreeparna Banerjee, and Mallika De. 2013. An Algorithm for Pre-Processing of Satellite Images of Cyclone Clouds. International Journal of Computer Applications 78 (09 2013), 13–17. https://doi.org/10.5120/13598-1317

[7] Rafik Ghali, Moulay A. Akhloufi, and Wided Souidene Mseddi. 2022. Deep Learning and Transformer Approaches for UAV-Based Wildfire Detection and Segmentation. Sensors 22, 5 (2022). https://doi.org/10.3390/s22051977

[8] Cengis Hasan, Ross Horne, Sjouke Mauw, and Andrzej Mizera. 2022. Cloud removal from satellite imagery using multispectral edge-filtered conditional generative adversarial networks. International Journal of Remote Sensing 43, 5 (2022), 1881–1893. https://doi.org/10.1080/01431161.2022.2048915.

[9] Swapandeep Kaur, Sheifali Gupta, and Swati Singh. 2021. Hurricane Damage Detection using Machine Learning and Deep Learning Techniques: A Review. IOP Conference Series: Materials Science and Engineering 1022, 1 (jan 2021), 012035. https://doi.org/10.1088/1757-899X/1022/1/012035

[10] Swapandeep Kaur, Sheifali Gupta, Swati Singh, Deepika Koundal, and Atef Zaguia. 2021. Convolutional Neural Network based Hurricane Damage Detection using Satellite Images. Natural Hazards 103, 3 (09 2021), 3357–3376. https://doi.org/10.21203/rs.3.rs-934531/v1

[11] Furkan Luleci, F. Necati Catbas, and Onur Avci. 2022. Generative adversarial networks for labeled acceleration data augmentation for structural damage detection. Journal of Civil Structural Health Monitoring (sep 2022). https://doi.org/10.1007/s13349-022-00627-8

[12] Xue Rui, Yang Cao, Xin Yuan, Yu Kang, and Weiguo Song. 2021. DisasterGAN: Generative Adversarial Networks for Remote Sensing Disaster Image Generation. Remote Sensing 13, 21 (2021). https://doi.org/10.3390/rs13214284

[13] Ritwik Gupta, Richard Hosfelt, Sandra Sajeev, Nirav Patel, Bryce Goodman, Jigar Doshi, Eric Heim, Howie Choset, Matthew Gaston, 2019. xBD: {A} Dataset for Assessing Building Damage from Satellite Imagery, Computing Research Repository (2019). https://doi.org/10.48550/arXiv.1911.09296

[14] Christine Dewi, Rung-Ching Chen, Yan-Ting Liu, Hui Yu. 2021. Various Generative Adversarial Networks Model for Synthetic Prohibitory Sign Image Generation. Applied Sciences 2021, 11, 2913. https://doi.org/10.3390/app11072913