Deep learning models for natural disaster assessment using satellite imagery

DEEP LEARNING MODELS FOR NATURAL DISASTER ASSESSMENT USING SATELLITE IMAGERY

A Thesis by

Madhurika Sunil Patil

Bachelor of Engineering, SRM Institute of Science and Technology, 2018

Submitted to the School of Computing
and the faculty of the Graduate School of
Wichita State University
in partial fulfillment of
the requirements for the degree of
Master of Science

December 2022

DISASTER ASSESSMENT IN SATELLITE IMAGERY USING DEEP LEARNING MODELS

The following faculty members have examined the final copy of this thesis for form and content, and recommend that it be accepted in partial fulfillment of the requirement for the degree of Master of Science with a major in Computer Science.

Ajita Rattani, Committee Chair

Atri Dutta, Committee Member

Hongsheng He, Committee Member

Rajiv Bagai, Committee Member

ACKNOWLEDGEMENTS

I would like to express my gratitude and deepest appreciation to my advisor, Dr Ajita Rattani, for her guidance throughout my program. I extend my gratitude to members of my committee, Dr Atri Dutta, Dr Rajiv Bagai and Dr Hongsheng He for their suggestions and comments on this research work.

ABSTRACT

The current methods to assess damage which has occurred during a disaster are usually manual based assessments. This work talks about developing an automated damage assessment model using computer vision and deep learning techniques. This helps to provide rapid relief during disasters to the affected areas. Satellite images provide real time photos of the affected area and help in visualizing large areas. Integrating satellite images with computer vision and deep learning techniques for damage assessments reduces time and manual labor for damage assessment. The thesis focuses on creating a novel deep learning model which can classify the damage using pre and post satellite images of the disaster. To localize the damage we have used GRAD CAM visualization method. For the model to be deployed on satellites for real time damage detection and classification, pruning and quantization is proposed to compress the size of the model. The models are tested on XBD dataset and compared with existing state of the art models.

The thesis proposes basic model, patch based model, vision transformer, cross fusion and cross stitch models to test the damage classification. The XBD dataset is pre processed using image registration techniques. Multi scale inputs are used to train models over various scale of images. Further GRAD CAM is used to detect the damaged areas on post disaster images. Pruning and quantization methods are used to reduce size of models.

TABLE OF CONTENTS

Chapter Page

1. INTRODUCTION 1
   1.1 Problem Statement 1
   1.2 Background And Motivation 2
   1.3 Research Objective 2
   1.4 Damage Classification Using Deep Learning Methods 3
      1.4.1 Image Registration 3
      1.4.2 Multi Scale Inputs 4
      1.4.3 XBD Dataset 4
      1.4.4 Cross Fusion Model 6
      1.4.5 Cross Stitch Model 6
      1.4.6 Patch Based Models 7
      1.4.7 Vision Transformers 8
   1.5 Damage Visualization Of The Disaster 8
      1.5.1 Grad CAM Visualization 8
   1.6 Pruning And Quantization 9
   1.7 Contributions Of Thesis 11

2. LITERATURE REVIEW 12
   2.1 Damage Classification 12
   2.2 Patch based models and Vision Transformer 13

TABLE OF CONTENTS (continued)

Chapter Page

2.3 Multi scale Cross fusion and cross stitch models 14

2.4 GRAD CAM visualization 15

2.5 Pruning and quantization 16

3. METHODOLOGY 18

3.1 Proposed Method And Implementation 18

3.2 Pre Processing Images Using Image Registration 18

3.3 XBD Dataset 19

3.4 Implementation Details 21

3.5 Damage Classification 23

3.5.1 Basic Model With Registered Images 23

3.5.2 Patch Based Model 24

3.5.3 Vision Transformers Based Model 25

3.5.4 Cross Stitch Multi Scale Model 27

3.5.5 Cross Fusion Multi Scale Model 28

3.6 GRAD CAM Visualization 30

3.7 Pruning And Quantization 31

4. EVALUATION AND RESULT 34

4.1 Parameters For Evaluating The Disaster 34

4.1.1 Evaluating Disaster For Classification 34

4.1.2 Evaluating Damaged Regions After Disaster 35

TABLE OF CONTENTS (continued)

Chapter Page

4.1.3 Evaluating Pruning And Quantization ... 35

4.2 Results ... 35

4.2.1 Results For Basic Model With Registered Images ... 36

4.2.2 Result For Patch Model ... 38

4.2.3 Results For Vision Transformer ... 39

4.2.4 Results Of Cross Fusion Multi Scale Model ... 40

4.2.5 Result For Cross Stitch Multi Scale Model ... 41

4.2.6 Comparison With State Of The Art Models ... 44

4.2.7 Results For GRAD CAM Visualization ... 46

4.2.8 Results For Pruning And Quantization ... 46

5. CONCLUSION AND FURTHER WORK ... 50

REFERENCES ... 51

LIST OF FIGURES

Figure		Page
1	Image registration on hurricane harvey image of damage scale 3 - destroyed. The leftmost picture depicts the pre disaster image. The center image depicts the post disaster image and the rightmost image depicts the image after registration.	4
2	Disasters covered by the XBD dataset [1]. The column on left shows the pre disaster images and the right column shows the post disaster images.	5
3	Depicts the cross fusion model. Layer 1 and layer 2 depict the previous branch which is concatenated together	6
4	Depicts the cross stitch model. Layer 1 and layer 2 are concatenated individually and then passed further through other layers	7
5	Simple patch based model is depicted in the figure. The leftmost image shows us the patch extracted from image. The patches are then passed through DenseNet individually to extract the features.	8
6	Basic model of vision transformer [2] showcasing patches being created and processing it through vision transformer.	9
7	The image depicts Guatemala volcano disaster. Most of the regions are undamaged. (a) Pre disaster image is depicted for the disaster (b) Post disaster image has been depicted (c) Depicts the post registered image for disaster (d) Grad CAM visualization of the disaster	10

Figure Page

8 The model before and after pruning. (a) The model without pruning applied. (b) The model after pruning we can see some branches have been removed 10

9 Damage localization and classification [3] 13

10 Architecture of vision transformer[2] 14

11 GRAD CAM visualization on hurricane Florence 16

12 Before and after Pruning 16

13 Examples of post disaster images after image registration. (a) Post disaster image of hurricane Florence (b) Pre disaster image of hurricane Florence (c) Post image of hurricane Florence after image registration (d) Post disaster image of socal fire (e) Pre disaster image of socal fire (f) Registered image of post disaster image of socal fire 20

14 Class wise distribution of images for XBD dataset........ 22

15 Sample images from XBD dataset. The image pair from all categories 0 - undamaged, 1 - minor damage, 2 - major damage 3 - destroyed and 4 - unlabelled for various disasters 22

16 The workflow of basic model. The features of both pre and post images are extracted through DenseNet and then concatenated and passed through dense layers to get classified into one of 5 damage categories 24

Figure | Page

17 Describes the patch based models. The pre and post registered images are divided into 2 equal patches and paired together. One pair uses DenseNet to extract the features and other uses ResNet. Maximum of both the features is found. These are then passed through dense layers. 26

18 The pre and post registered images are passed through individual vision transformer modules to extract features. These features are the concatenated and passed through multiple reducing dense layers to generate output 27

19 Cross stitch model is in cooperated with combined features of pre and post disaster images. Dense layer added to each branch. Reduced dense layers prevents memorization of results. 29

20 In cooperation of cross fusion model to multi scale model. Each image is passed through DenseNet model to extract features and then through cross fusion model. The output from cross fusion model is then passed through dense layers. 30

21 Generation of GRAD CAM visualization for hurricane Florence. The damage level is 2 - major damage for the image. The pre and post disaster image are processed through cross fusion model. the weighted CNN are used to generate heatmap to visualize damage. 32

22 Visualizing class wise F-1 score, Precision and Recall for Basic model 37

23 Visualizing class wise F-1 score, Precision and Recall for Patch based model 39

LIST OF FIGURES (continued)

Figure Page

24 Visualizing class wise F-1 score, Precision and Recall for Vision Transformer model 40

25 Shows us the comparison of F-1, precision and recall of all classes for cross fusion model 42

26 Shows us the comparison of F-1, precision and recall of all classes for cross stitch model 43

27 F-1 score comparison between state of the art models and our models 45

28 Shows GRAD CAM visualization of images across all damage scale (a) GRAD CAM visualization of no damage Guatemala volcano (b) Minor damage GRAD CAM of Guatemala volcano (c) Major damage GRAD CAM visualization for hurricane Florence (d) GRAD CAM visualization for destroyed image of hurricane harvey (e) Unlabelled GRAD CAM visualization of hurricane Florence 47

29 Comparing model size for cross fusion before and after pruning 48

30 Comparing model size for cross stitch before and after pruning 48

LIST OF TABLES

Table	Page

1 Data split of XBD dataset [1] into train, test and validation 19

2 Accuracy and F-1 score of Basic model, patch based model and visual transformer 36

3 F-1 score, Precision and Recall for Basic model 37

4 F-1 score, Precision and Recall for Patch based model 38

5 F-1 score, Precision and Recall for Vision transformer 40

6 F-1 score , precision and recall for cross fusion model 41

7 Accuracy and F-1 score of best models 41

8 F-1 score , precision and recall for cross stitch model 43

9 Weighted F-1 score comparison 45

10 Size of models before and after pruning and combination of pruning and quantization 49

LIST OF ABBREVIATIONS

ADFF : Attention based deep feature fusion

CNN : Convolution Neural Network

COVID : Corona Virus diseases of 2019

GRAD CAM : Gradient-weighted Class Activation Mapping

MLP : Multi attention head

TFMOT : TensorFlow model optimization toolkit

TFLite : TensorFlow Lite

CHAPTER 1

1 INTRODUCTION

1.1 Problem Statement

Disasters have affected mankind for a very long time. They cause a loss of life and property. Humans have devised various methods to mitigate disasters. The methods used usually include manual ground based assessments which takes a lot of time and money leading to necessary relief not being provided on time. This causes delay in rescue of the ones affected by the disaster. Some disaster are difficult to be accessed by humans. Having an aerial view of the disasters makes it easy to detect damaged regions. Using high resolution satellite images, it is easy to capture the damaged area [1][3].

With the development of new technology, automated techniques are devised which can help in developing faster disaster detection techniques. With help of high quality satellites and drones, post disaster images can be captured. Methods which can detect changes in pre and post disaster are required. Classification and localization of the disaster is required so that proper mitigation can be done. If the technology were to be deployed on satellites it would be easier to detect the damaged regions in real time. It should be faster and efficient as well[1].

1.2 Background And Motivation

Manual damage assessment is very time consuming and can have human error in it. Disaster are not easy to be assessed manually. To make it more convenient for the first responders, aerial based technique is required to assess the large areas of disaster. Change detection techniques make it easier to classify the damaged regions. Classification of damages into categories helps in better understanding the degree of damage. If the models are deployed on satellites the damaged regions can be detected in real time.

Challenges faced during disaster classification and localization is developing an accurate algorithm which can understand the characteristics of all disasters. The regions which are damaged should be highlighted. The models deployed on the satellites should take very little memory and work efficiently at the same time.

1.3 Research Objective

The objective of this research is to detect the damaged regions of the disaster, classify them using deep learning models and reduce the size of these models. The work tries to build an improved model which can classify and detect damage as well. The pre and post disaster images are aligned on same coordinate system so that the change in features is studied accurately. Using these factors we try to build an efficient method which can help us assess damage. Various deep learning models are developed like patch based models, vision transformers, cross stitch models, etc. For features to be studied over various scales, multi scale models are created. Image Registration technique is used to align the pre and post disaster images on same coordinate system. This pre processing improves classification and localization as features on both images are one same coordinate system. To visualize damaged and undamaged regions, heat maps generated by GRAD CAM visualization method. Reduction in the size of models is done using unstructured magnitude based weighted pruning and post quantization methods.

1.4 Damage Classification Using Deep Learning Methods

The classification of damage is done using joint use of pre and post disaster images. Change between the pre and post disaster image is studied by extracting the features of the images and noting the change between them [4]. Most of the disasters are taken into consideration by using XBD dataset [1]. Proper alignment of pre and post disaster images is done using image registration techniques.

1.4.1 Image Registration

Transforming images from different sets onto one coordinate system is called image registration. There can be difference in the surroundings of pre and post disaster images when they were captured so alignment of the coordinates is done in order to detect exact change between them. Gradient descent is used for image registration. In Fig 1 we can see the post disaster image has been registered according to the coordinates of pre disaster image [5].

Figure 1: Image registration on hurricane harvey image of damage scale 3 - destroyed. The leftmost picture depicts the pre disaster image. The center image depicts the post disaster image and the rightmost image depicts the image after registration.

1.4.2 Multi Scale Inputs

Some information about the image can be lost if the scale of image is very small or very large. To understand the features of the image properly, extraction of complimentary information from the same image is done so that no information gets lost [6]. The pre and post images of the model are reduced to half the size as inputs to the models.

1.4.3 XBD Dataset

To understand the features of all kinds of disaster a dataset is required which covers all kinds of disasters. One such dataset is the XBD dataset [1] which contains pre and post disaster images of various disasters like floods, hurricanes, wildfires, earthquake and volcanoes. The damages caused during the disaster are classified on the joint damage scale of 0 - undamaged, 1 - minor damage, 2 - major damage and 3 - destroyed. There are some buildings

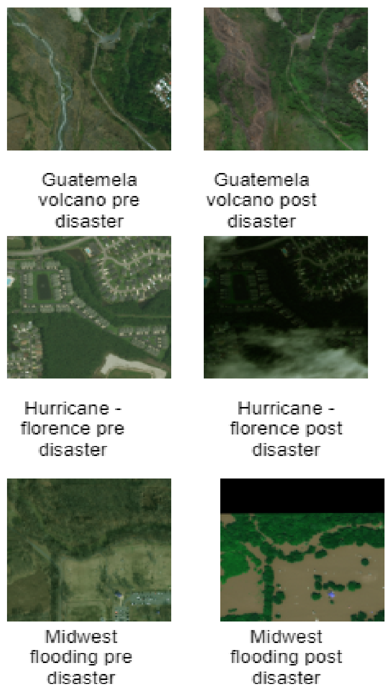

Figure 2: Disasters covered by the XBD dataset [1]. The column on left shows the pre disaster images and the right column shows the post disaster images.

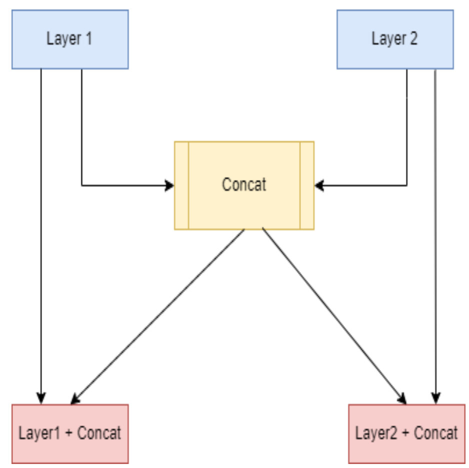

Figure 3: Depicts the cross fusion model. Layer 1 and layer 2 depict the previous branch which is concatenated together

which could not be classified as there were unlabelled. The Figure 2 shows an example of XBD dataset [1].

1.4.4 Cross Fusion Model

In this study a model is developed to understand the change in the features of pre and post disaster images effectively. Most models usually combine various features together. In cross fusion models, the features are combined from different branches within the model [7]. As shown in the Figure 3, the features from layer 1 and layer 2 are concatenated again with the concatenate layer of the model respectively. DenseNet is used to extract the features from both pre and post disaster images. These are the fed into the cross fusion model.

1.4.5 Cross Stitch Model

Another model which is developed to study the change between pre and post

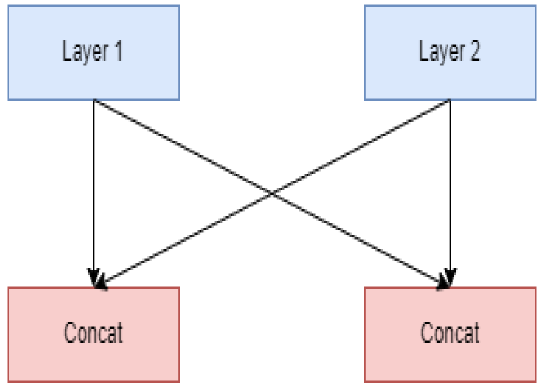

Figure 4: Depicts the cross stitch model. Layer 1 and layer 2 are concatenated individually and then passed further through other layers

disaster images is the cross stitch model [8]. In the field of image classification, the cross stitch model performs well. The features from various models are concatenated end to end. DenseNet is used to extract the features of pre and post disaster images and these are then concatenated together and then passed through the cross stitch model. The Figure 4 shows the cross stitch model. The layer 1 and layer 2 are concatenated together and dense layers are added to each concatenate layer at the end to generate better features.

1.4.6 Patch Based Models

Other model which is developed to classify disasters and understand change between pre and post disaster images is patch based model [9]. Patches of each pre and post disaster images are extracted and passed through DenseNet to extract the features individually. These are then concatenated together to classify images. Patch based models help to study image pixel by pixel which is then combined to help in understanding which category the image falls into.

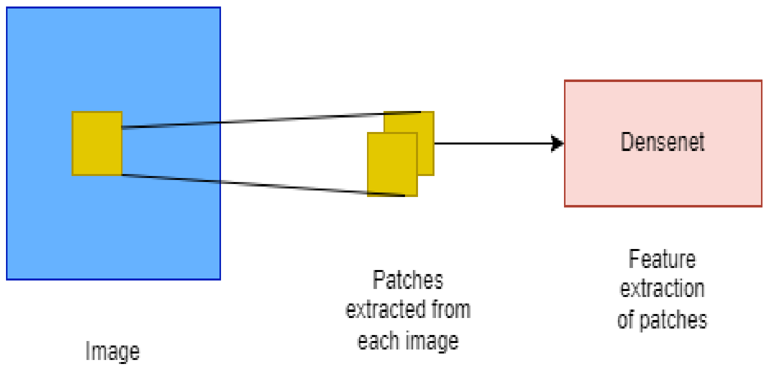

Figure 5: Simple patch based model is depicted in the figure. The leftmost image shows us the patch extracted from image. The patches are then passed through DenseNet individually to extract the features.

The Figure 5 shows how the model works. Image is divided into patches and each patch is then passed through DenseNet to extract the feature.

1.4.7 Vision Transformers

Vision transformer [2], a very popular model, is used to classify pre and post disaster images. Vision transformer [2] processes patches of images into transformer like model which converts each patch into a vector. Using the post disaster images and passing them through vision transformer helps in classifying the images. The Figure 6 shows the basic structure of vision transformer.

1.5 Damage Visualization Of The Disaster

1.5.1 Grad CAM Visualization

To visualize the damaged regions on post disaster image, GRAD CAM visualization [10] is used. Using the gradients of a target concept a gradient

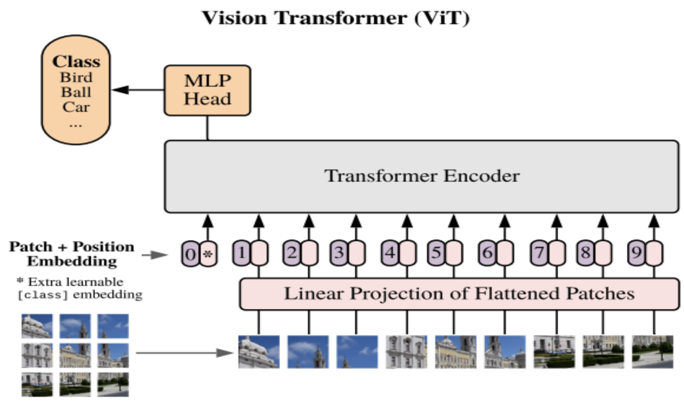

Figure 6: Basic model of vision transformer [2] showcasing patches being created and processing it through vision transformer.

weighted class activation map is developed. This map localizes the damage and undamaged area which makes it easy to visualize damage. The gradients with red hue are more damaged than the regions in blue hue. In the Figure 7, the damaged regions for the disaster are highlighted.

1.6 Pruning And Quantization

To make sure that the model can be deployed on satellite, the size of the model needs to be reduced. Pruning is a method to remove unnecessary branches of the model to increase the speed and efficiency of the model. Quantization is the method to approximate the floating point number of the tensors to create compact high performance models [11]. Pruning and quantization is implemented on damage classification models. TensorFlow model optimization toolkit [12] and TensorFlow Lite [12] is used for pruning the models. In the Figure 8 it shows how pruning works. The branches which are

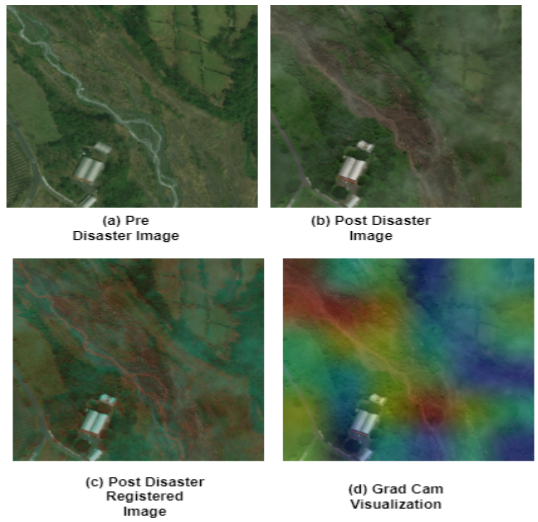

Figure 7: The image depicts Guatemala volcano disaster. Most of the regions are undamaged. (a) Pre disaster image is depicted for the disaster (b) Post disaster image has been depicted (c) Depicts the post registered image for disaster (d) Grad CAM visualization of the disaster

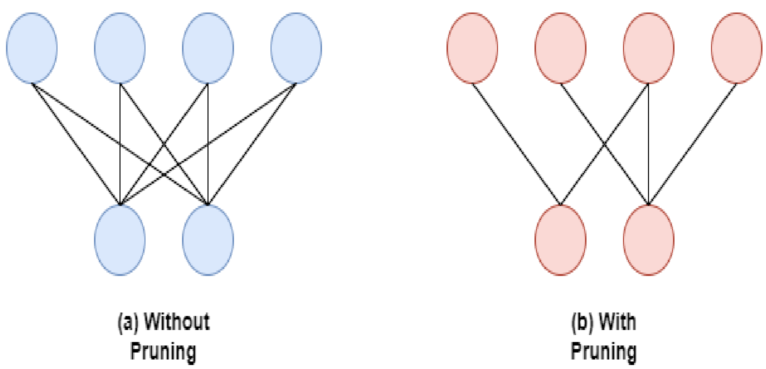

Figure 8: The model before and after pruning. (a) The model without pruning applied. (b) The model after pruning we can see some branches have been removed

not required are removed and the efficiency of the model is maintained. The pruning method used here is unstructured magnitude based weight pruning. The quantization method used is post quantization.

1.7 Contributions Of Thesis

The main contributions of this study is developing deep learning models based on joint use of pre and post disaster images for damage classification. Saliency maps are generated using GRAD CAM for damage localization. The model is compressed using pruning and quantization methods for deployment on edge devices.

CHAPTER 2

2 LITERATURE REVIEW

2.1 Damage Classification

Many techniques have been used to classify damage on XBD dataset. Commonly used methods for damage classification used Siamese networks, convolution networks, pyramid models. Hao et al [13] uses Siamese networks with attention mechanism to assess the damage. UNet model was in combination for damage segmentation and classification[13]. Gupta and Shah propose RescueNet [3], a unified model which can be used for segmentation and damage classification. DeepLabv3+ is repurposed with Dilated ResNet and Atrous Spatial Pyramid Pooling as backbone. This generates features from pre and post images and compares temporal change[3]. Navjot et al proposes DAHiTrA [14] which is a novel deep learning architecture with hierarchical transformers with UNet like architecture. The multi scale hierarchical spatial features is used find temporal difference between pre and post disaster images [14]. The XBD baseline model proposed by Gupta [1] processes the inputs through ResNet50 pre trained on ImageNet and shallow CNN. Combination of features from both models are used for segmentation and classification [1]. C. Wu et al [15] proposes a UNet model based on attention mechanism. On the layers from encoder attention is applied and fused with up sampled decoder layer. The features from pre and post disaster images are merged at

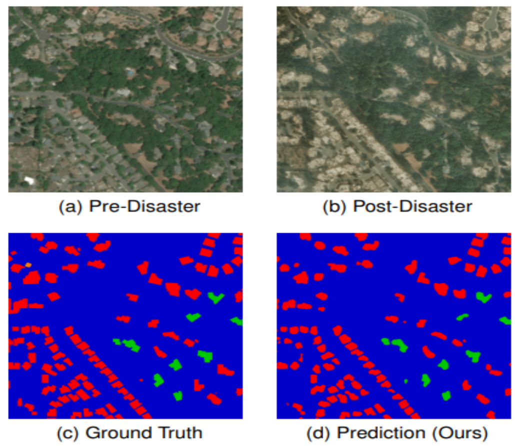

Figure 9: Damage localization and classification [3]

final classification layer [15].

Image registration is applied to the models which is inspired from the library Kornia in PyTorch. Edgar et al proposes Kornia [16],library similiar to OpenCv for image registration which uses CNN based modules to perform operations. Gradient descent and homography warp is used to generate coordinate based image registrations [16].

2.2 Patch based models and Vision Transformer

Dosovitskiy et al [2] proposes vision transformer seen in Figure 10 , a transformer based model for image classification. Patches of images are positionally and linearly embedded to pass through a transformer encoder to classify images [2]. Ethan and Weber et al [17] use the multi temporal information from pre and post disaster images to perform building localization and classification. The model uses ResNet50 with shared weights. The features

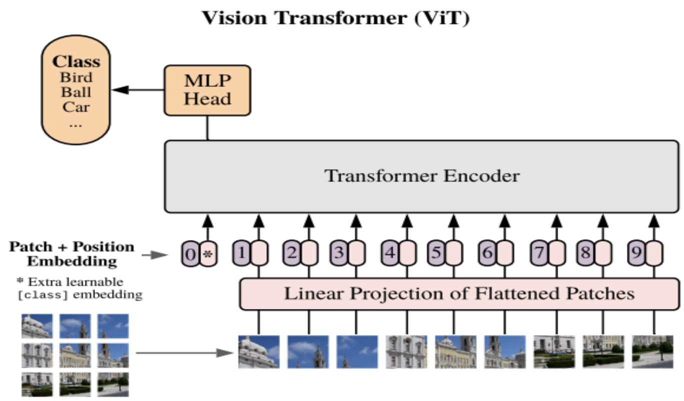

Figure 10: Architecture of vision transformer[2]

are concatenated together for localization and classification of damage [17]. Trockman et al [18] proposes a patch based model where patches are used as inputs and processed through spatial contrast[18].

2.3 Multi scale Cross fusion and cross stitch models

Multi scale models help in understanding images over various scales. Chun et al devised CrossVit [19], a multiscale dual branch vision transformer approach which combines the features from multi scale images. This model is used for image classification [19]. Chun fu et al [20] developed a novel method Big-Little Net which is CNN based network which has two branches where the big branch has same baseline structure and little branch with reduced convolution layers [20].

Multi task modules are used to train the model for two tasks. Commonly used task multi task modules are cross fusion and cross stitch which we have tried to integrate for change detection. Yu Shen et al [21] proposes BDANet ,a two stage UNet based framework for damage classification and localization on XBD dataset [1].The model integrates multi scale features to reduce scale influence with CDA module which aggregates features from pre and post disaster images using attention module to enhance representation of features [21]. Misra et al [8] developed cross stitch module to learn shared representations in convolution models. Features from various modules are trained end-to-end. Xio et al [22] proposes a dynamic cross fusion model DCFNet to access multi level features of images from same backbone network. This is used for damage localization and classification of damage on XBD dataset [1][22]. Alemida et al [23] proposes L-CNN, which combines multiple streams of CNN modules to perform classification [23].

2.4 GRAD CAM visualization

When proper annotations are not provided for a method is needed to localize images based on its features. Selvaraju et al [10] proposed GRAD-CAM (Gradient weighted class activation mapping) which uses gradient information of classes to localize regions. GRAD CAM can be used over any CNN module [10]. Panwar et al [24] proposes deep transfer learning algorithm with GRAD CAM visualization method to detect traces on COVID 19 in lungs based on images. Zhu et al [25] developed ADFF, attention-based deep feature fusion made up of three parts, for deep features multiplicative fusion ,attention maps used to generate GRAD CAM maps and cross entropy loss functions.

Figure 11: GRAD CAM visualization on hurricane Florence

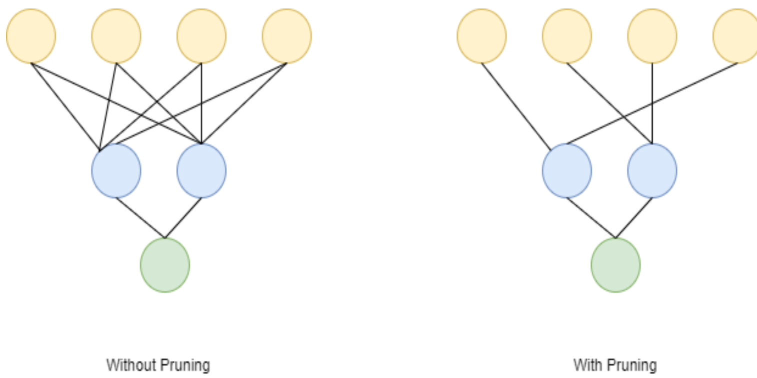

Figure 12: Before and after Pruning

The model is used to classify land cover for high resolution images[25]. Chen et al [26] proposes GRAD-CAM visualization technique on embedded networks. The model aggregates the weights for gradients from training samples and develop weight transfer methods which are efficient[26].

2.5 Pruning and quantization

To reduce the size of the models to be deployed in satellite images pruning and quantization methods are used. Zhu et al [27] proposes that baseline models used for experiments have too many parameters in them which are not required. Using gradual pruning methods, which require minimal tuning and easily used in training process [27]. Rao et al [28] proposes thermal control system using neural network and their comparison before and after pruning. After pruning the efficiency of model is maintained. Verma et al [29] proposes the comparison between two optimization techniques, TensorFlow Lite (TFlite) [12] and TensorFlow Tensor RT and the performance is compared. TFlite tends to perform better.

CHAPTER 3

3 METHODOLOGY

3.1 Proposed Method And Implementation

Proposed method talks about classification and visualization of damages happening during disaster. The methods to reduce the size of the models so that they can be deployed on edge devices is discussed. The models proposed for classification are cross stitch and cross fusion models. Other models included are patch based models and vision transformers for damage classification. To visualize the damaged regions GRAD CAM visualization technique [10] is used. XBD dataset [1] which contains of pre disaster and post disaster satellite images covering six of disasters like volcanoes, floods, wildfires, earthquakes, etc. The dataset contains 850,736 building polygons and covering over $4500km^2$ of area. To reduce the size of our models, pruning and quantization methods are used. TensorFlow model optimization toolkit and TensorFlow lite [12] are used for pruning the models and quantization as well. The models are compared with existing state of the art models on basis of their F-1 score.

3.2 Pre Processing Images Using Image Registration

As the images were captured from a satellite, its not possible to capture the post disaster image at exact coordinates as pre disaster images due to various

Table 1: Data split of XBD dataset [1] into train, test and validation

Data Split	Image Pairs
Train	11322
Validation (Hold)	1526
Test	1504

technical and weather conditions. For better feature extraction and change detection, the pre and post disaster images should be on the same coordinate system. To achieve this image registration method is used. Kornia [16], a PyTorch library which uses differential operators to perform transformations on the images so that the coordinates match. Transformations like rotations, scalings and shearings are used. These are then warped by a homo graphic warper which warps it to reference pre disaster image and clears distortions. Gradient descent is used to perform image registration. To reduce the error the pixels are optimized at each level of multi level pyramid. Loss function is also optimized at each level of multi level pyramid. The Figure 13 shows the registered images for post disaster image. After image registration there is an alignment between pre and post disaster images [16].

3.3 XBD Dataset

For the models to understand and detect changes effectively a dataset is required which covers all kinds of disaster. XBD dataset [1] contains pairs of pre and post disaster images of six kinds of disasters like floods, earthquakes, volcanoes, wildfires, hurricanes and tsunami. The dataset contains 850,736 buildings and covers $4500km^2$ of area. The disaster is classified on the basis

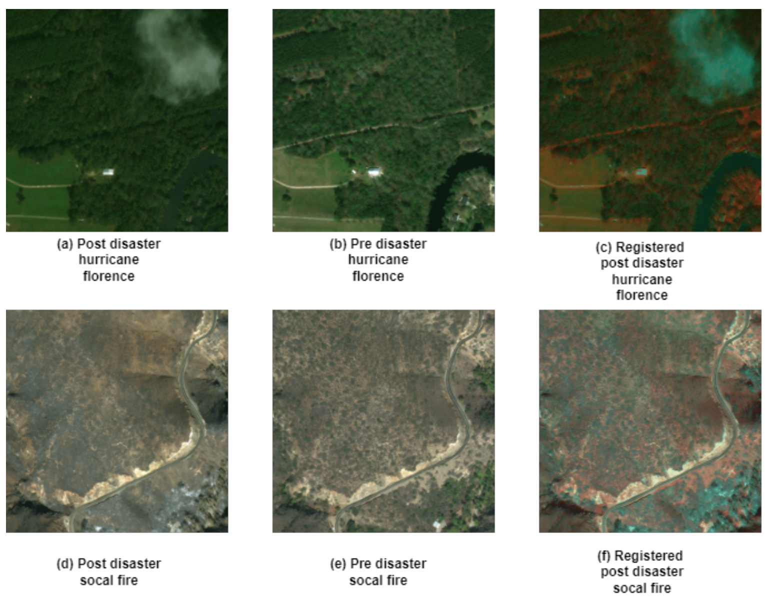

Figure 13: Examples of post disaster images after image registration. (a) Post disaster image of hurricane Florence (b) Pre disaster image of hurricane Florence (c) Post image of hurricane Florence after image registration (d) Post disaster image of socal fire (e) Pre disaster image of socal fire (f) Registered image of post disaster image of socal fire

of joint damage classification scale which is created to evaluate damage across all kinds of disasters. 0 stands for no damage, 1 stands for minor damage, 2 stands for major damage 3 stands for destroyed. Some buildings in the dataset have not be labelled so they are classified category 4. The dataset tries to cover disasters from all over the world. Data of 15 disasters have been collected from countries like Nepal, Guatemala, Portugal, etc. The size of the images are 1024 x 1024. The Figure 15 represents image pairs from XBD dataset [1]. The class wise distribution of data is shown in Figure 14. Class 0 has most images and class 4 has the least. There is uneven distribution of images in all classes.

For training the models, the images which have no labeling given at all, are removed. The average of all labels of the disaster of the buildings is taken and used as a unified label for the images. To create the dataset, tier 1 and tier 3 are combined together to create the train dataset. From the table 1 the dataset after cleaning has 11322 pairs of train images, 1526 pairs of validation images and 1504 pairs of test images.

3.4 Implementation Details

All the experiments were performed on NVIDIA Quadro RTX 4000 GPU. Each model was trained for 100 epochs. The learning rate is 0.00004 and the models were trained on batch size of 32, 64 and 128. The loss function used for all models was categorical cross entropy. Adam is the optimizer that we used for all models. For multi scale models images of size 400 x 400 and 200

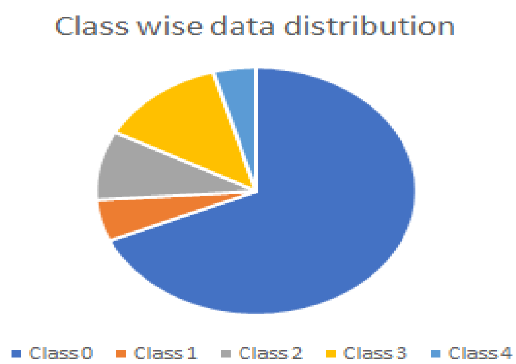

Figure 14: Class wise distribution of images for XBD dataset

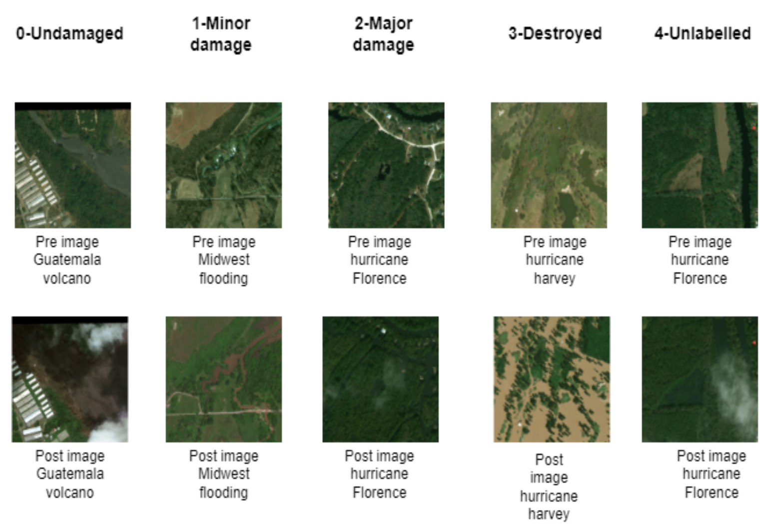

Figure 15: Sample images from XBD dataset. The image pair from all categories 0 - undamaged, 1 - minor damage, 2 - major damage 3 - destroyed and 4 - unlabelled for various disasters

x 200 are used.

3.5 Damage Classification

For damage classification, few models were tested. The models which are used - patch based model, cross fusion model, cross stitch model and vision transformer [2]. These models are classified into five categories of damage. All the post disaster images have registered. For training all the models, XBD dataset [1] has been used. To train the models 11322 pairs of pre and post images are used. For testing 1504 pairs of images are used. For validation 1526 pairs of images are used.

3.5.1 Basic Model With Registered Images

To classify images and calculate the change between pre and post disaster images a basic model is developed. The features of both pre disaster and post disaster images are extracted using DenseNet model. The DenseNet models are pre trained on Imagenet. The features are extracted from DenseNet model on the average pool layer (1024,). The features are then concatenated together and passed through number of dense layers. The dense layers added are (2048, 256, 32 and 5). Dropout (0.1) to prevent over fitting. The final dense layer of (5,) is added so that the model classifies the images into one of the five categories of damage. Decreasing dense layers make sure results are not memorized.

In the Figure 16 the workflow of the model is shown. Pre and post registered images are the inputs. DenseNet blocks are added to extract features

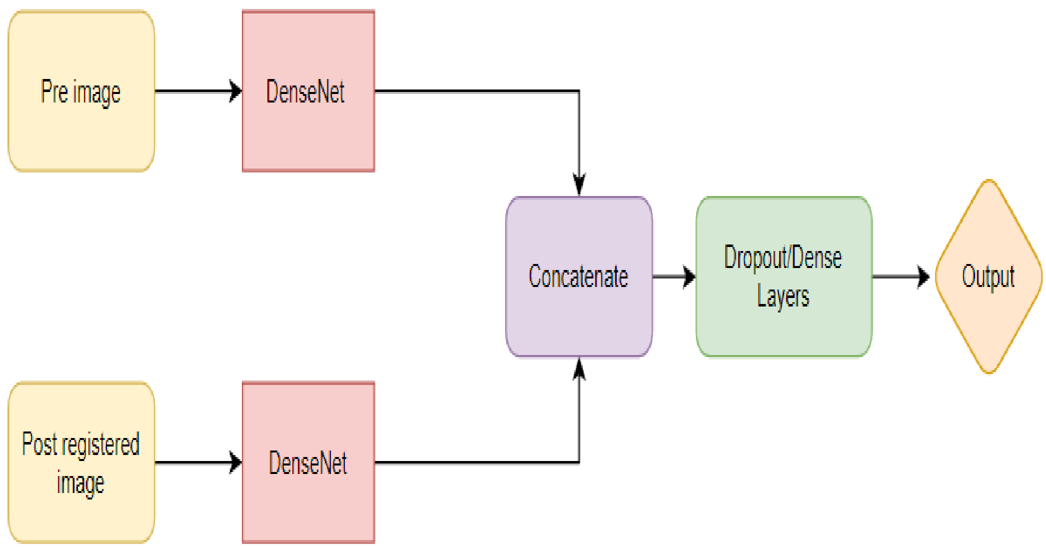

Figure 16: The workflow of basic model. The features of both pre and post images are extracted through DenseNet and then concatenated and passed through dense layers to get classified into one of 5 damage categories

which are then concatenated and passed through dense layers. This model creates a base to develop future models which provide better results.The size of images used as inputs is 224 x 224.

3.5.2 Patch Based Model

When processing individual images to classify them some details can be missed. To study each image pixel by pixel in detail the images are broken down into patches. For the model, the images are vertically divided into two equal patches. From the Figure 17, the patches of left side of both pre and post disaster images are paired together and same is done for the right side of images.

Once the patches are created one pair of patch is individually passed through DenseNet which are pre trained on Imagenet. The other pairs are trained individually on ResNet. The ResNet is also pre trained on Imagenet. On both sides the features are concatenated together. For ResNet the features after concatenation are (4096,) and for DenseNet the features after concatenation are (2048,). On both sides dense layers of shape (1024, 64) are added to reduce number of inputs. Maximum layer is used to calculate maximum value out of all inputs. At the end a dense layer (5,) is added in order to get output as one of the five categories. The maximum helps us choose the best value out of all for classification hence used to get better results. Using DenseNet and ResNet models helps us understand features of the efficiently.

3.5.3 Vision Transformers Based Model

The vision transformer [2] is the model used for image classification where images are converted into multiple patches of equal size and then processed through a transformer like architecture to convert them into vectors. Positional and linear embedding are added to each patch and then processed through a transformer encoder. To classify images an extra token is added at the end.

In cooperating the vision transformer [2] module Figure 6 into the model as shown in the Figure 18. The pre and post registered images are processed through individual blocks of vision transformer model [2]. The final layer of vision transformer model generates (1024) features for both pre and post

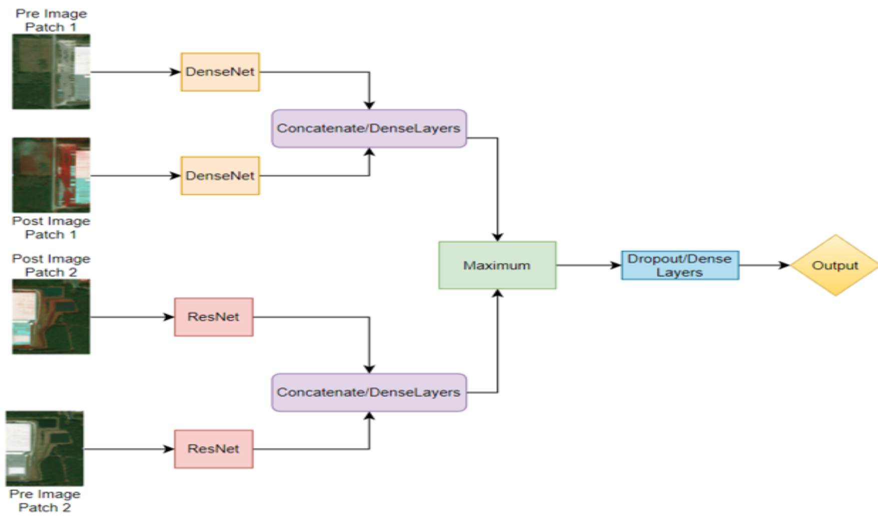

Figure 17: Describes the patch based models. The pre and post registered images are divided into 2 equal patches and paired together. One pair uses DenseNet to extract the features and other uses ResNet. Maximum of both the features is found. These are then passed through dense layers.

disaster images. The patches of size 15x15 are created for the images. MLP (Multi attention head) is added at the end for classification. These features are concatenated (2048) and passed through set of decreasing dense layers (1024, 128, 32). The final dense layer is (5,) to classify the images in one of the 5 categories of damage. Three dropout (0.1) layers are alternatively added between dense layers to reduce over fitting. A batch normalization layer is added at the end to to make training faster and stabilized by normalizing the layers.

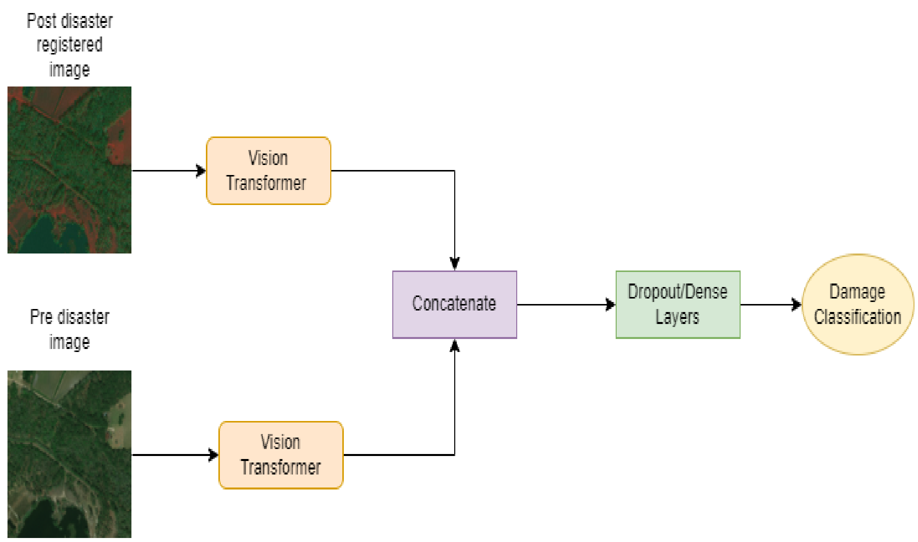

Figure 18: The pre and post registered images are passed through individual vision transformer modules to extract features. These features are the concatenated and passed through multiple reducing dense layers to generate output

3.5.4 Cross Stitch Multi Scale Model

Cross stitch models are usually used for multi task learning but we try to in cooperate them into the model to understand change detection. These models are good at understanding optimal linear combinations for the shared tasks. Cross stitch model [8] help in deciding how much information actually needs to be shared. Combining the features from pre and post disaster images. From the Figure 20, the multi scale features of both pre and post disaster images pairs are passed through cross stitch network and different sets of dense layers are added at each end of the cross stitch model. This helps in understanding the features better. Cross stitch models are trained end to end.

The images are scaled to the size of 400 X 400 and 200 X 200. All the images are passed through DenseNet to extract features. DenseNet is pre trained on Imagenet and the final layer avg pool (1024) is used to extract the features. The features of pre disaster images are combined together in pairs using concatenate. Same is done for post disaster images. For the cross stitch model, the combined features of pre disaster images are concatenated with combined features of post disaster images. Two branches are created. Dense layers (2048, 128) are added to one branch and dense layer (2048) is added to another branch. The features of these two branches are further combined and passed through decreasing set of dense layers (2176, 1024, 32). The final dense layer of (5,) is added to classify images according to damage scale. Dropout layer helps us reduce the overfitting.

3.5.5 Cross Fusion Multi Scale Model

Multi scale inputs have been used for this model in order to study features on various scales. The features from these multi scale images can be studied effectively if they are fused together. In cooperating one of the commonly used multi task model, cross fusion into the model can help us understand features of images on various levels [7]. Combining these results eventually can help in achieving better results. The cross fusion model combines features generated by various layers within the model.

The pre and post disaster images of size 400 X 400 and 200 X 200 are taken and paired together. Each image is passed through DenseNet which

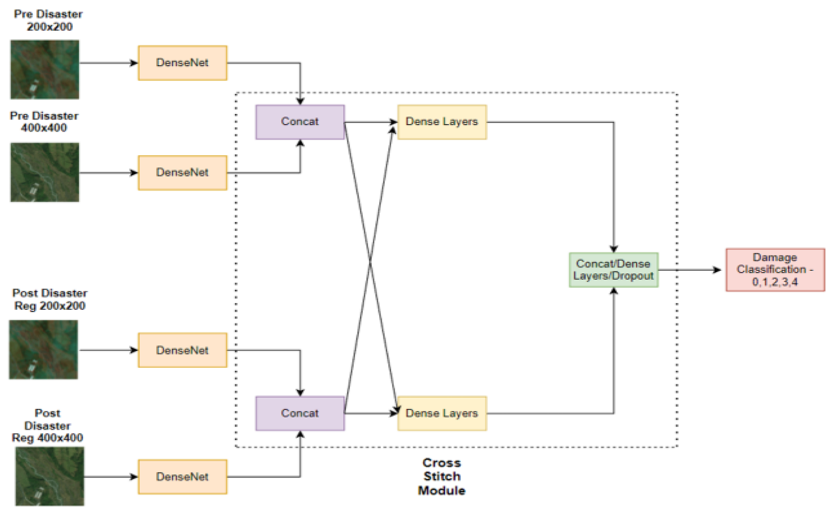

Figure 19: Cross stitch model is in cooperated with combined features of pre and post disaster images. Dense layer added to each branch. Reduced dense layers prevents memorization of results.

is pre trained on Imagenet. The features are extracted on average pool layer (1024). The features are concatenated (2048) together for pre and post disaster images respectively. These are further concatenated (4096) together and passed through a dense layer (2048). These are then combined with features from previous layers of combined pre and post disaster images. This block of cross fusion model helps us understand the change even better. Two separate branches which are further combined and passed through reducing dense layers and dropout layer to reduce over fitting. Dense layer (5,) is added in end to classify according to the damage scale.

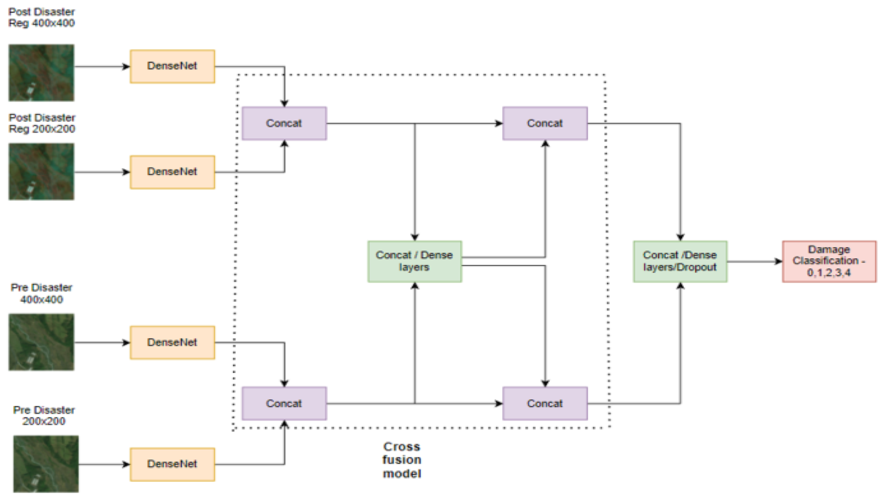

Figure 20: In cooperation of cross fusion model to multi scale model. Each image is passed through DenseNet model to extract features and then through cross fusion model. The output from cross fusion model is then passed through dense layers.

3.6 GRAD CAM Visualization

To be able to visualize damaged regions effectively GRAD CAM [10] is used. The gradients are calculated with respect to the feature maps of the convolution layers. The GRAD CAM [10] generates heat maps which are weighted combination of all convolution layers present in the model. GRAD CAM [10] can be used on any CNN model and can easily calculate weighted combination of layers. GRAD CAM stands useful when local annotations are not provided and images are classified as whole. They generate weakly localized heat maps. In case of disaster the damage across the whole region which includes the damage done to terrain, man made objects, roads, etc needs to be understood. The heat maps generated by GRAD CAM [10] help in visualizing them as well. The more warmer hue in the heat map, more damaged the region is.

One of the best models to classify damage is the cross fusion model. DenseNet modules extract the features of the image at conv5−block16−concat. The features of pre and post images are concatenated on the convolution layers. Cross fusion model is applied in the same way followed by reducing convolution layers. GRAD CAM visualization [10] generates the heat map from last weighted combination of last convolution layer (6,6,32). This is followed by global pooling layer and a dense layer (5,). The gradients are calculated with respect to the convolution layers and heat map is generated. The figure 21 shows example of GRAD CAM visualization. The damaged regions are highlighted in red.

3.7 Pruning And Quantization

For the models to be deployed on satellite and edge devices the size of the model should be reduced and the accuracy should be maintained. These models take up a lot of space as they perform many computations. With more layers and branches more computations need to be performed and build up multiple add operations. This increased computational cost make it difficult the models to be deployed on edge devices. This can be achieved using pruning and quantization. Pruning is the process of eliminating weighted tensors and achieve computationally efficient model in order to achieve less

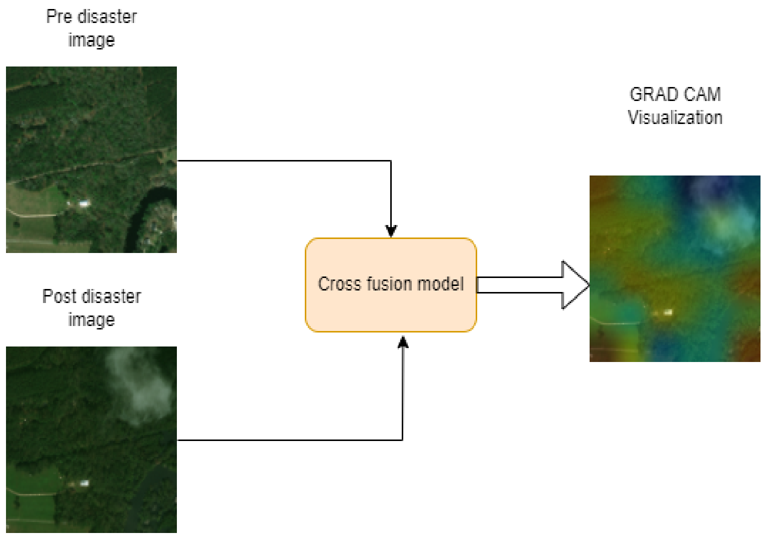

Figure 21: Generation of GRAD CAM visualization for hurricane Florence. The damage level is 2 - major damage for the image. The pre and post disaster image are processed through cross fusion model. the weighted CNN are used to generate heatmap to visualize damage.

training time [27]. Quantization is a process to convert weights into integer types so that they use less memory and make faster calculations. Using these methods we reduce the size of our models significantly to deploy them on satellites [27]. Unstructured pruning is used here where the magnitude of weights of tensors is reduced. For quantization, post quantization technique is used where quantization is applied after the model has been trained.

TensorFlow has in built libraries, TensorFlow optimization toolkit (TF-mot) [12] and TensorFlow lite (TFlite) [12] which help in pruning the cross fusion and cross stitch models. Pruning is performed using TFmot and TFlite respectively and pruning and quantization using TFlite. The initial sparsity of the model is set to 0.5 and final sparsity is set to 0.80. The efficiency of the model is checked on how much reduction is there in the accuracy. Post training quantization is applied to the models to achieve better results.

CHAPTER 4

4 EVALUATION AND RESULT

This thesis talks about classifying the damage which has occurred during disaster using cross fusion and cross stitch models. The dataset used is the XBD dataset [1]. The models were compared on the basis of weighted F-1 score with other existing models. Accuracy was also calculated. GRAD CAM visualization [10] method was used to detect the disaster affected regions. The size of the models was also reduced using pruning methods.

4.1 Parameters For Evaluating The Disaster

The damage is evaluated on the basis of category it falls in. The XBD dataset [1] has four categories: 0 - undamaged, 1 - minor damage, 2 - major damage, 3 - destroyed, 4 - unclassified. The detection of damaged regions is evaluated using GRAD CAM visualization. The size of model is reduced using pruning and quantization methods. TFlite and TFmot [12] is used to do so. The size of models and their accuracy's are compared.

4.1.1 Evaluating Disaster For Classification

The models developed are compared on the basis of F-1 score, precision and recall. These metrics are used instead of accuracy as the data distribution of the XBD dataset [1] are uneven. Weighted average is the metric set by the baseline paper to compare the efficiency of the models in other papers. The accuracy of the models is calculated as well. Other baseline models were also compared like XBD baseline model [1], Ethan weber [17], Dahitra [14] and Rescuenet [3] with both of our models on the basis of weighted average.

4.1.2 Evaluating Damaged Regions After Disaster

The damaged regions are highlighted on post disaster images using GRAD CAM visualization [10] methods. The gradients of final convolution layer are compared with gradients of classification score. This helps in identifying which region is affected the most. In the cross fusion model the gradients are extracted on $conv2d_3$ layer. The damaged regions are highlighted using warmer colours. The more lighter colours represent the less damaged regions. The images are then viewed using matplotlib library.

4.1.3 Evaluating Pruning And Quantization

Pruning and quantization methods are used for reducing the size of the models. The main aim is to reduce the size of the model without compromising the accuracy of the models. Reduced bytes size of the pruned and quantized models shows that the models have become smaller. When quantization is applied, the persistent accuracy is checked and whether there is drop in the accuracy of the pruned model as compared to the orignal model [27].

4.2 Results

In this section the results of individual models are discussed,comparing the results with other existing models, the results obtained on GRAD CAM visualization [10] and the results of the pruning and quantized models. All

Table 2: Accuracy and F-1 score of Basic model, patch based model and visual transformer

Model	Accuracy			F-1 Score		
	Train	Test	Validation	Train	Test	Validation
Basic Model	100	76.86	69.46	1.0	0.75	0.65
Patch based model	99.89	61.03	52.55	1.0	0.53	0.43
Visual Transformer	97.98	61.56	58.32	0.98	0.56	0.52

the model are trained on XBD dataset [1]. The dataset is already divide into test, validation , tier1 and tier3. Combination of tier1 and tier3 id done to make it into train dataset.

4.2.1 Results For Basic Model With Registered Images

F-1 score, precision, recall and accuracy are calculated for the model. From the table 2 the training accuracy achieved by the model is 100 percent, test accuracy is 76.86 percent and validation accuracy is 69.46 percent. The weighted F-1 score achieved by the model is 1.0 for training, 0.75 for test and 0.65 for validation. The table 3 shows us individual F-1 score for all classes.

From the table 3 Class 0 has the highest F-1 score considering the uneven dataset which has more undamaged data. Class 2 and class 3 have performed relatively well in terms of other classes followed by class 4. Class 1 performs the worst out of all classes. This may be due to uneven distribution of data in each classes. In terms of accuracy the model has performed well. The visualization of these classes is done in Figure 22

Table 3: F-1 score, Precision and Recall for Basic model

Classes	Precision	Recall	F-1 Score
Class 0	0.87	0.96	0.91
Class 1	0.29	0.04	0.08
Class 2	0.67	0.64	0.66
Class 3	0.54	0.39	0.46
Class 4	0.15	0.27	0.19

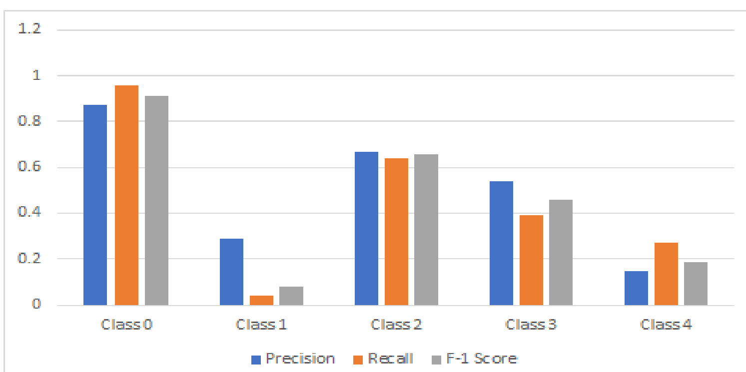

Figure 22: Visualizing class wise F-1 score, Precision and Recall for Basic model

Table 4: F-1 score, Precision and Recall for Patch based model

Classes	Precision	Recall	F-1 Score
Class 0	0.66	0.92	0.77
Class 1	0.08	0.02	0.03
Class 2	0.21	0.04	0.06
Class 3	0.32	0.1	0.15
Class 4	0.0	0.0	0.0

4.2.2 Result For Patch Model

From table 2 the train accuracy achieved by patch based models is 99.89 percent. The test accuracy is 61.03 percent and the validation accuracy is 52.55 percent. The weighted F-1 score for train is 1.0, the test is 0.53 and validation is 0.43. The model has performed fairly well on the test data but on comparing overall accuracies with other models seems to be a bit low. This may be because the dataset does not have precise label for each patch. Various regions in same image may have different levels of damage hence model could not understand it properly.

For class wise distribution from table 4 class 0 shows better performance as compared to other classes due to uneven data. Class 3 gave second best performance. Other classes seem to perform poorly.This may be due to unbalance in the dataset and model performing poorly on the data. Recall scores of the classes tell us most detected class is class 0 and least detected is class 4. Precision value of class 0 is highest so it was able to detect the class very well. The visualization is done in Figure 23.

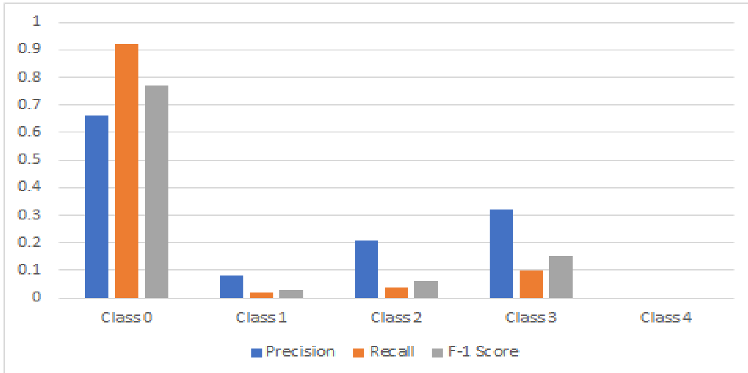

Figure 23: Visualizing class wise F-1 score, Precision and Recall for Patch based model

4.2.3 Results For Vision Transformer

As shown in the table 2, the model does not perform well as compared to other models developed. There is a significant drop in the train accuracy. The train accuracy is 97.98 percent. The test accuracy achieved is 61.56 percent and validation accuracy is 58.32 percent. The test and validation accuracy is low for a good result. The model scores low on weighted F-1 scores. There is a drop F-1 scores for train, test and validation. The train score is 0.98, the test score is 0.56 and the validation score is 0.52. These results tell us that the model doesn't perform well. It may be due to variety of features present in the image, proper embedding could not be made resulting in poor classification.

For class wise distribution, from the table 5 it is observed that the model performs best to classify class 0 images due unbalanced dataset having huge chunk of undamaged images. Class 1 seems to not be detected by the model.

Table 5: F-1 score, Precision and Recall for Vision transformer

Classes	Precision	Recall	F-1 Score
Class 0	0.71	0.89	0.79
Class 1	0.0	0.0	0.0
Class 2	0.37	0.16	0.22
Class 3	0.2	0.13	0.16
Class 4	0.08	0.05	0.06

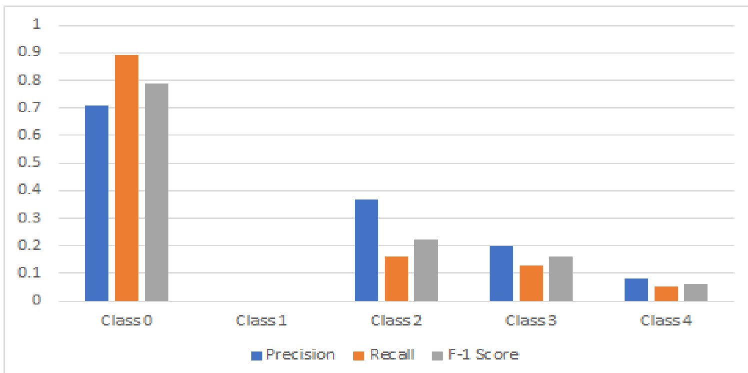

Figure 24: Visualizing class wise F-1 score, Precision and Recall for Vision Transformer model

The model has performed fairly well to detect images in class 2 and class 4. Class 4 is not well detected by the model. The precision of class 0 is highest making it most detected class for the model. Recall score also tells us that class 0 is the most detected class. The graph in Figure 24 shows us comparison of each class.

4.2.4 Results Of Cross Fusion Multi Scale Model

The F-1 score, precision,recall and accuracy for the model is as shown in table 7. The overall train accuracy of the model is 100 percent, the test accuracy of the model is 80.98 percent and the validation accuracy is 72.60

Table 6: F-1 score , precision and recall for cross fusion model

Classes	Precision	Recall	F-1 Score
Class 0	0.89	0.97	0.93
Class 1	0.20	0.02	0.04
Class 2	0.71	0.76	0.73
Class 3	0.65	0.57	0.61
Class 4	0.19	0.23	0.21

Table 7: Accuracy and F-1 score of best models

Model	Accuracy			F-1 Score		
	Train	Test	Validation	Train	Test	Validation
Our Model (cross stitch)	100	82.44	74.04	1.0	0.80	0.70
Our Model (cross fusion)	100	80.98	73.52	1.0	0.79	0.70

percent. The weighted average F-1 score of our model for test accuracy is 0.79. F-1 score of individual class is calculated. The individual F-1 score for each model and performance of each class is shown in table 6.

Class 0 had the highest F-1 score, recall and precision value because the model was able to detect it the best. The model was not able to classify class 1 that well as the precision and recall values are very less. Class 2 and class 3 damages are fairly well detected as well. As there are more images for undamaged class, the model was able to detect it better than other classes. The model has performed fairly well as we have achieved test accuracy of 80.98 percent. The graphical representation is shown in the Figure 25.

4.2.5 Result For Cross Stitch Multi Scale Model

The model was trained on XBD dataset as well with the same data split used for the previous model.

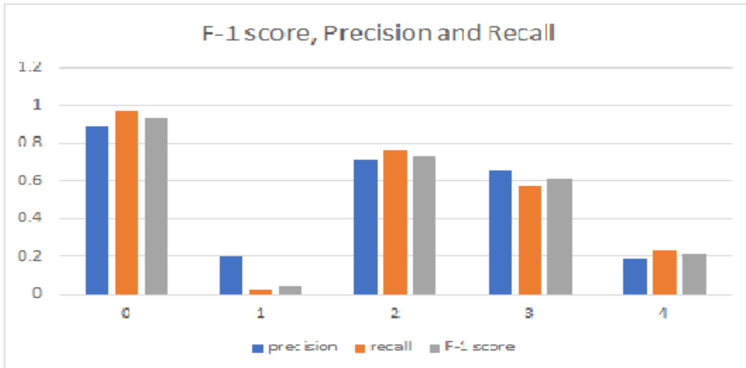

Figure 25: Shows us the comparison of F-1, precision and recall of all classes for cross fusion model

From table 7 the weighted average F-1 score for the model is 0.80 which is better than most of the existing models. The training accuracy of the model is 100 percent, test accuracy is 82.44 percent and validation accuracy is 74.04 percent. From the table 8 individual F-1 score, recall and precision for each class is calculated as well.

The highest F-1 score, precision and recall is achieved for class 0. The model is best able to detect class 0. Class 1 has the least F-1, precision and recall making the least detected class. Class 2 and class 3 have been detected well too scoring and F-1 score around 0.76 and 0.66 respectively. The model has scored good accuracy of 82.44 percent and has performed well. Due to dis-balance in the dataset, class 0 performs better on the model than other classes. The graph representation is seen in the Figure 26.

Table 8: F-1 score , precision and recall for cross stitch model

Classes	Precision	Recall	F-1 Score
Class 0	0.89	0.98	0.93
Class 1	0.00	0.00	0.00
Class 2	0.74	0.77	0.76
Class 3	0.72	0.61	0.66
Class 4	0.19	0.23	0.20

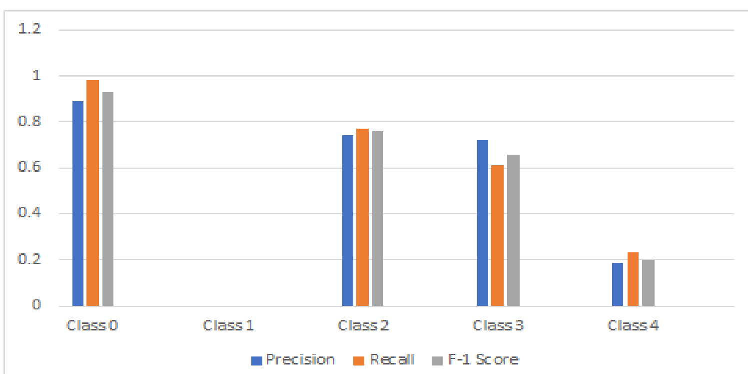

Figure 26: Shows us the comparison of F-1, precision and recall of all classes for cross stitch model

4.2.6 Comparison With State Of The Art Models

Few existing models which use XBD dataset for damage classification are compared on basis of their weighted average F-1 score.

From table 9 the XBD baseline model [1] is uses 2 streams of - Resnet50 and shallow CNN and concatenates their result to classify the damage. The weighted average score achieved was 0.26. The Rescuenet [3] model uses dilated ResNet and atrous spatial pyramid pooling to re purpose DeepLabv3+ for generating multi scale features. These are used for damage classification. The weighted F-1 score achieved is 0.77. Dahitra [14] model proposes using novel transformer based architecture to detect damage. The weighted f-1 score generated by the model is 0.79. The Ethan weber model[17] uses multi temporal model generates f-1 score of 0.74.

Comparing the models created for the study it is observed that the best performing model is the cross stitch model and least performing model is patch model. The patch model performs bad because features of each patch were not classified effectively. Cross stitch model was able to perform well because the multi scale values and shared features help us understand the model better hence giving a better result. The basic model performs fairly well on the data because the architecture being similar to our best models. Vision transformer also tends to perform not well on the dataset. On an average the cross stitch model performs better than state of the art models on basis of weighted F-1 score by 25 percent.

Comparing the models with existing state of the art models, the cross

Table 9: Weighted F-1 score comparison

Models	F-1 scores
Cross stitch (our model)	0.80
Cross fusion (our model)	0.79
Dahitra[14]	0.79
Rescuenet[3]	0.77
Basic model (our model)	0.75
Ethan Weber [17]	0.74
Vision transformer (our model)	0.56
Patch model (our model)	0.53
XBD baseline model[1]	0.26

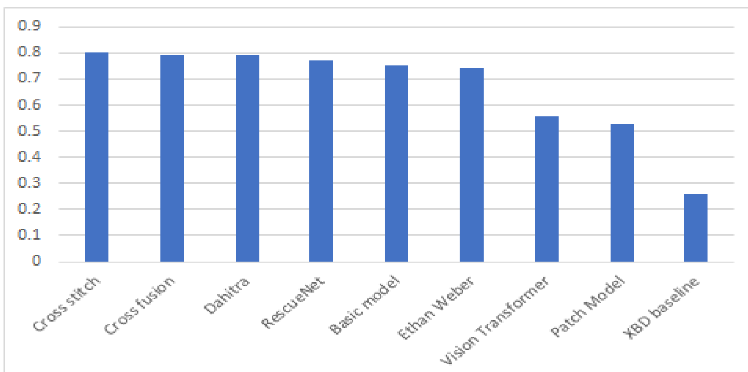

Figure 27: F-1 score comparison between state of the art models and our models

stitch and cross fusion model outperforms the state of the art models. Dahitra [14] has results close to our cross fusion model with a tie at F-1 score of 0.79. The basic model outperforms the Ethan Weber model [17]. The vision transformer and patch based model do not perform well as compared to all models. All of the models beat the weighted F-1 score of XBD baseline model [1].

4.2.7 Results For GRAD CAM Visualization

GRAD CAM visualization [10] can be performed on any CNN model to create heat map to visualize the desired region. Here cross fusion model is used to visualize damaged regions. The heat map generated depicts the damaged regions in warmer tones and as the intensity of damage decreases,the colours change to cooler tones. Using these method damage visualization is done on terrain in the background as well.

From the Figure 28, the damage visualization for images across various classes. The damage scale for each building was associated as average of all labels of building to get basic idea of the damage level. For the image with no damage most of the area around the building isn't damaged. The damage is observed in few spots of terrain. For minor damage some yellow hue around the buildings is observed. The damage to terrain is seen clearly. For major and destroyed a red hue is observed around the damaged building. The model is even able to detect damaged terrain as well. The unlabelled image damaged regions are clearly observed. On comparing pre and post images the difference between the images is highlighted properly.

4.2.8 Results For Pruning And Quantization

To make sure two of the best models, cross stitch and cross fusion need to be deployed on satellites and edge devices, pruning and quantization methods to reduce the size of the models. Pruning of both the models using TensorFlow optimization toolkit (TFmot) and TensorFlow Lite (TFLite) [12] is done.

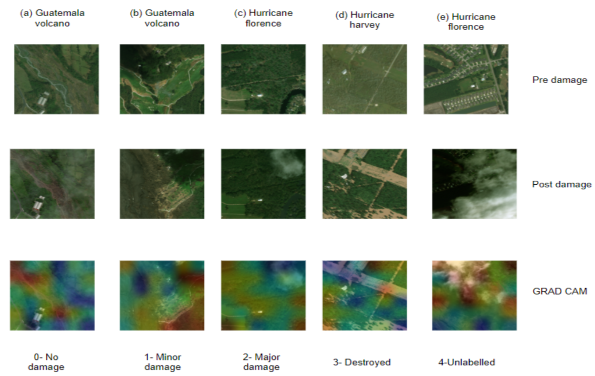

Figure 28: Shows GRAD CAM visualization of images across all damage scale (a) GRAD CAM visualization of no damage Guatemala volcano (b) Minor damage GRAD CAM of Guatemala volcano (c) Major damage GRAD CAM visualization for hurricane Florence (d) GRAD CAM visualization for destroyed image of hurricane harvey (e) Unlabelled GRAD CAM visualization of hurricane Florence

Pruning and quantization is performed using TFlite [12]. The results are seen the table 10.

From the table 10 it is observed that there is no decrease in the model sizes when the models is pruned using TFMot[12]. On pruning the model with TFlite [12] there is a slight drop in the model size by 0.02 percent for cross fusion and cross stitch. On using quantization with TFlite [12] much better results are observed. The size of the models drop down by 76 percent for cross fusion and 72 percent for cross stitch model. In case of efficiency of models after pruning, there is not much decrease in the accuracy. The accuracy of cross fusion model is decreased by 1 percent for TFmot [12] and TFlite [12]. On adding quantization in TFlite [12], 1 percent decrease is seen. For cross stitch model, the decrease in accuracy for TFmot [12] and TFlite

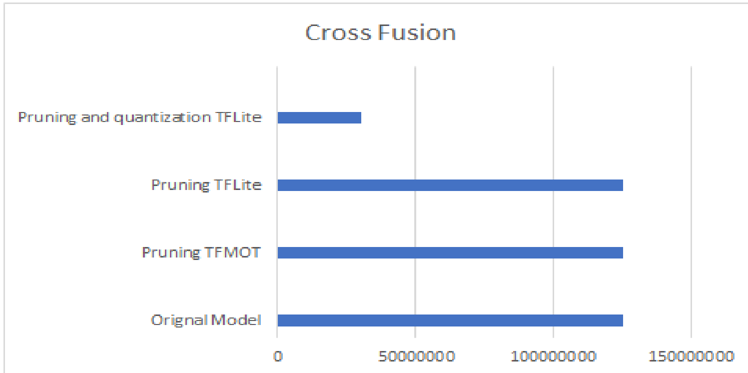

Figure 29: Comparing model size for cross fusion before and after pruning

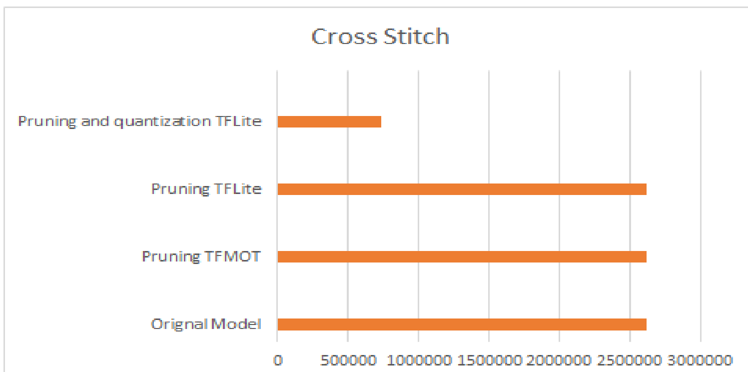

Figure 30: Comparing model size for cross stitch before and after pruning

[12] pruned models is 3 percent. On quantization the decrease is 3 percent. The graphical representation of size of the models can be observed in the Figure 29 and Figure 30.

Table 10: Size of models before and after pruning and combination of pruning and quantization

Model	Cross Fusion		Cross stitch	
Pruning and quantization methods	Size of model (bytes)	Accuracy	Size of model (bytes)	Accuracy
Orignal Model	125225663.00	81.11	2612307.00	82.44
Pruning TFMOT	125225662.00	80.71	2612306.00	79.52
Pruning TFLite	125223040.00	80.71	2611458.00	79.52
Pruning and quantization TFLite	30036667.00	80.45	729147.00	79.65

CHAPTER 5

5 CONCLUSION AND FURTHER WORK

The proposed models is successfully able to perform damage classification and damage localization for variety of disasters. The size of models was reduce significantly. The cross fusion and cross stitch models are able to out perform many state of the art models and provide a good result. The damaged areas were successfully visualized with gradient heat maps on cross fusion models. These models can be used in future on detecting damage from drone and UAV images as well. These models can also be used to detect damages made by other kinds of man made disasters as well. Applying Image registration techniques and using multi scale models has significantly improved our results. Dividing the image into patches does not necessarily provide great results for damage classification. Further improvement can be made to improve classification and localization of disaster by taking care of unbalanced dataset. Method to remove some obstructions like clouds, smoke, etc from images should be integrated to get better results and understand changes between pre and post disaster images. The models of smaller size can be deployed on satellites for real time detection.

REFERENCES

REFERENCES

[1] R. Gupta, R. Hosfelt, S. Sajeev, N. Patel, B. Goodman, J. Doshi, E. Heim, H. Choset, and M. Gaston, "xbd: A dataset for assessing building damage from satellite imagery," *arXiv preprint arXiv:1911.09296*, 2019.

[2] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[3] R. Gupta and M. Shah, "Rescuenet: Joint building segmentation and damage assessment from satellite imagery," in *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2021, pp. 4405–4411.

[4] K. R. Nia and G. Mori, "Building damage assessment using deep learning and ground-level image data," in *2017 14th conference on computer and robot vision (CRV)*. IEEE, 2017, pp. 95–102.

[5] L. G. Brown, "A survey of image registration techniques," *ACM computing surveys (CSUR)*, vol. 24, no. 4, pp. 325–376, 1992.

[6] Z. Gong, P. Zhong, Y. Yu, W. Hu, and S. Li, "A cnn with multiscale convolution and diversified metric for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 6, pp. 3599–3618, 2019.

[7] C. Peng, K. Zhang, Y. Ma, and J. Ma, "Cross fusion net: A fast semantic segmentation network for small-scale semantic information capturing in aerial scenes," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2021.

[8] I. Misra, A. Shrivastava, A. Gupta, and M. Hebert, "Cross-stitch networks for multi-task learning," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3994–4003.

[9] J. Lin, W.-M. Chen, H. Cai, C. Gan, and S. Han, "Memory-efficient patch-based inference for tiny deep learning," *Advances in Neural Information Processing Systems*, vol. 34, pp. 2346–2358, 2021.

[10] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, "Grad-cam: Why did you say that?" *arXiv preprint arXiv:1611.07450*, 2016.

[11] T. Liang, J. Glossner, L. Wang, S. Shi, and X. Zhang, "Pruning and quantization for deep neural network acceleration: A survey," *Neurocomputing*, vol. 461, pp. 370–403, 2021.

[12] T. TensorFlow, "Model optimization tensorflow lite," Oct 2021. [Online]. Available: https://www.tensorflow.org/lite/performance/model$_{o}ptimization$

[13] H. Hao, S. Baireddy, E. R. Bartusiak, L. Konz, K. LaTourette, M. Gribbons, M. Chan, E. J. Delp, and M. L. Comer, "An attention-based system for damage assessment using satellite imagery," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. IEEE, 2021, pp. 4396–4399.

[14] N. Kaur, C.-C. Lee, A. Mostafavi, and A. Mahdavi-Amiri, “Dahitra: Damage assessment using a novel hierarchical transformer architecture,” *arXiv preprint arXiv:2208.02205*, 2022.

[15] C. Wu, F. Zhang, J. Xia, Y. Xu, G. Li, J. Xie, Z. Du, and R. Liu, “Building damage detection using u-net with attention mechanism from pre-and post-disaster remote sensing datasets,” *Remote Sensing*, vol. 13, no. 5, p. 905, 2021.

[16] E. Riba, D. Mishkin, D. Ponsa, E. Rublee, and G. Bradski, “Kornia: an open source differentiable computer vision library for pytorch,” in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 3674–3683.

[17] E. Weber and H. Kané, “Building disaster damage assessment in satellite imagery with multi-temporal fusion,” *arXiv preprint arXiv:2004.05525*, 2020.

[18] A. Trockman and J. Z. Kolter, “Patches are all you need?” *arXiv preprint arXiv:2201.09792*, 2022.

[19] C.-F. R. Chen, Q. Fan, and R. Panda, “Crossvit: Cross-attention multi-scale vision transformer for image classification,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 357–366.

[20] C.-F. Chen, Q. Fan, N. Mallinar, T. Sercu, and R. Feris, “Big-little net: An efficient multi-scale feature representation for visual and speech recognition,” *arXiv preprint arXiv:1807.03848*, 2018.

[21] Y. Shen, S. Zhu, T. Yang, C. Chen, D. Pan, J. Chen, L. Xiao, and Q. Du, "Bdanet: Multiscale convolutional neural network with cross-directional attention for building damage assessment from satellite images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2021.

[22] H. Xiao, Y. Peng, H. Tan, and P. Li, "Dynamic cross fusion network for building-based damage assessment," in *2021 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2021, pp. 1–6.

[23] A. P. G. de Almeida and F. de Barros Vidal, "L-cnn: a lattice cross-fusion strategy for multistream convolutional neural networks," *Electronics Letters*, vol. 55, no. 22, pp. 1180–1182, 2019.

[24] H. Panwar, P. Gupta, M. K. Siddiqui, R. Morales-Menendez, P. Bhardwaj, and V. Singh, "A deep learning and grad-cam based color visualization approach for fast detection of covid-19 cases using chest x-ray and ct-scan images," *Chaos, Solitons & Fractals*, vol. 140, p. 110190, 2020.

[25] R. Zhu, L. Yan, N. Mo, and Y. Liu, "Attention-based deep feature fusion for the scene classification of high-resolution remote sensing images," *Remote Sensing*, vol. 11, no. 17, p. 1996, 2019.

[26] L. Chen, J. Chen, H. Hajimirsadeghi, and G. Mori, "Adapting grad-cam for embedding networks," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 2794–2803.

[27] M. Zhu and S. Gupta, "To prune, or not to prune: exploring the efficacy of pruning for model compression," *arXiv preprint arXiv:1710.01878*, 2017.

[28] V. V. R. M. K. Rao, M. Rapp, J. Henkel, H. Amrouch, and M. Wolf, "On the effectiveness of quantization and pruning on the performance of fpgas-based nn temperature estimation," in *2021 ACM/IEEE 3rd Workshop on Machine Learning for CAD (MLCAD)*. IEEE, 2021, pp. 1–7.

[29] G. Verma, Y. Gupta, A. M. Malik, and B. Chapman, "Performance evaluation of deep learning compilers for edge inference," in *2021 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*. IEEE, 2021, pp. 858–865.