iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 Issue: VI Month of publication: June 2021

DOI: https://doi.org/10.22214/ijraset.2021.35706

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

Damaged Building Detection from Satellite Multispectral Imagery

A Sangeetha[1], P Keerthi Reddy[2], O Deekshitha[3]
[1, 2, 3]Department of Computer Science and Engineering, Chaitanya Bharathi Institute of Technology(A),Gandipet,Hyderabad-75

Abstract: Damaged Building footprint detection in satellite and aerial imagery is crucial in city management. Building detection is a fundamental but a challenging problem mainly because it requires correct recovery of building footprints from high-resolution images. Buildings are one of the key pieces of cadastral information related to population and cities, and are fundamental to urban planning & policymaking. Critical infrastructures, such as public transport, electricity, water distribution networks, or postal and delivery services, rely heavily on accurate population and building maps. On top of that, it is essential to get real-life, up-to-date information about buildings each time there is a need for disaster risk management, risk assessment, or emergency relief. Accurate and fine-grained information about the extent of damage to buildings is essential for directing Humanitarian Aid and Disaster Response operations in the immediate aftermath of any natural calamity. Satellite and UAV (drone) imagery has been used for this purpose in recent years, sometimes aided by computer vision algorithms. Existing Computer Vision approaches for building damage assessment typically rely on a two stage approach, consisting of building detection using an object detection model, followed by damage assessment through classification of the detected building tiles. These multi-stage methods are not end-to-end trainable, and as well as suffer from poor overall results. We proposed the UNet segmentation model, a model that can simultaneously segment buildings and assess the damage levels to individual buildings and can be trainedend-to-end. We trained the model using X View 2 challenge dataset.
Keywords: building damage assessment, machine learning, convolutional neural networks, earthquake, multispectral Imagery

I. INTRODUCTION

Damaged buildings are often used as a proxy for affected population localization. Remote sensing is a powerful tool for identifying damaged buildings due to its wide coverage area and availability of data. However, humanitarian actors mostly rely on manual digitization of damaged structures, which remains the most reliable method. Manual digitization is labor-intensive, requiring trained image analysts, is unsuitable for large areas, and is prone to inconsistencies related to human errors due to fatigue or quality control.
Automating this process would greatly reduce the time required to produce damage assessment reports. It is essential to get real-life, up-to-date information about buildings each time there is a need for disaster risk management, risk assessment, or emergency relief. Damage estimation is done by identifying damaged and undamaged buildings.
Assessing damage levels for buildings can be focused, which is relevant to 3 types of natural disasters, and can have a significant impact on the efficacy of search and rescue operations in their aftermath. The ability to accurately detect and locate building footprints is a powerful tool at the service of many different applications such as illegal building detection , population mapping. It is essential to get real-life, up-to-date information about damaged buildings each time there is a need for disaster risk management, risk assessment, or emergency relief. At the start of a humanitarian crisis, it is critical for humanitarian agencies to know the locations of affected populations within the first few hours after a disaster in order to facilitate deployment of response activities. Damaged buildings are often used as a proxy for affected population localization. Remote sensing is a powerful tool for identifying damaged buildings due to its wide coverage area and availability of data. However, humanitarian actors mostly rely on manual digitization of damaged structures, which remains the most reliable method. Manual digitization is labor intensive, requiring trained image analysts, is unsuitable for large areas, and is prone to inconsistencies related to human errors due to fatigue or quality control.

II. LITERATURE REVIEW

In [1], it proposed a technique for extracting building footprints from high resolution satellite images containing only RGB bands using Convolutional Neural Network architecture based on U-shaped architecture like U-Net. The proposed method works irrespective of the geometric characteristics of buildings like shape, size and color. Moreover, the post-processing algorithms are useful in regularizing the building footprints. The method largely depends upon the training data provided and also the quality of the images in the training data. The proposed approach when provided with the training images from different countries can be used in the future to create a robust system that can detect and extract buildings from any part of the world.

Experimental results show that the proposed system can correctly extract the building footprints from satellite images with good accuracy. The proposed method can be further used for extracting other features like roads, water and other features from satellite images. In [2], it described a method to build convolutional neural networks that automatically detect damaged buildings in satellite images. We introduced a novel way to generate large numbers of negative training examples automatically in our data generation pipeline. It experimented with multiple model architectures and found the "two tower subtract" variant to perform the best at this task. Finally, it empirically showed that the model can generalize well to new regions and disasters if it is fine- tuned on a small set of examples from that region. As future work, they planned to investigate additional disaster types, especially hurricanes and armed conflicts. also planned to investigate techniques to make the model more robust to data flaws. In [3] presented a comprehensive analysis for the problem of large-scale damage detection using satellite imagery. It presented a novel use of hierarchical shape features in bags-of-visual words setting, and demonstrated its accuracy and efficiency advantages over multiple alternatives. Going forward, it improved the encoding scheme used in the current framework from hard quantization to one involving multiple soft-assignments. Furthermore, they planned on incorporating approximate sub-space learning mechanisms to further improve the efficiency of the unsupervised part of our framework. Finally, they applied damage-detection framework to a larger class of changes, such as detecting urbanization patterns. In [4] presented a semi supervised learning(SLL) method, because to give a live response for the disasters humanitarian requires accurate and timely data, which is collected from the satellites with the help of sensors. But most of the data in real disaster response scenarios its difficult to get a labeled data , which can be used to build a supervised machine learning algorithm for prediction. So [4] proposed a SSL method to train model for damage assessment with a minimal amount of labeled data and large amount of unlabeled data. The performance of SSL methods has been compare with so many supervised and semi supervised algorithms with a great accuracy. As future work, they planned to investigate how to effectively incorporate data from past disasters; there may be region-independent transformations caused by a disaster that the models do not sufficiently capture or different types of augmentations and losses that are more robust to the noise inherent to satellite imagery.

III.CONVOLUTIONAL NEURAL NETWORK

The technology used to solve this problem is Unet segmentation model. U- Net is a convolutional neural network that was developed for biomedical image segmentation. The network is based on the fully convolutional network and its architecture was modified and extended to work with fewer training images and to yield more precise segmentations. Segmentation of a 512 × 512 image takes less than a second on a modern GPU. U-Net consists of Convolution Operation, Max Pooling, ReLU Activation, Concatenation and Up Sampling Layers. The main idea is to supplement a usual contracting network by successive layers, where pooling operations are replaced by upsampling operators. Hence these layers increase the resolution of the output. One important modification in U-Net is that there are a large number of feature channels in the upsampling part, which allow the network to propagate context information to higher resolution layers. As a consequence, the expansive path is more or less symmetric to the contracting part, and yields a u- shaped architecture. The network only uses the valid part of each convolution without any fully connected layers. U-net uses a loss function for each pixel of the image. This helps in easy identification of individual cells within the segmentation map.

A. Image Mask Dataset

Spatial and colour augmentations to images in datasets is done. Augmentation is performed so as to expand the training data artificially. In order to maintain a reasonable amount of images, and above all to avoid overfitting by ensuring a sufficient invariance and robustness of the network, Data augmentation techniques is used for training set. The satellite images were shifted, flipped, rotated and changes done in the brightness of images, which allowed to train the model on a considerably larger set of images. This task was done using the “Albumentations” library which allowed to augment the data in real time when feeding the network with batches.

B. U-Net Segmentation

Pre-trained B4Encoder and pre-trained UNetDecoder is used. Original U-net uses ReLU which is over ridden with Swish activation function as it is produces more accurate results. The dimensions of the input images has been changed , as the original U-Net was designed for images of size 572 × 572 , but this implementation has for 512 × 512. U-Net channels used are [32, 64, 128] for encoding and [128, 64, 32] for the decoding part respectively. To avoid overfitting of the model dropout value set to 0.1. F1 score used for metrics calculation of the model as the loss based on the F1 score (Dice coefficient) is commonly used for image segmentation as it allows coping with class imbalance. Batch normalization after each Swish activation to speed-up training. The size of the batch for training purposes is set as 5 and used Adam optimizer.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

C. Prediction

Predict function takes pre disaster , post disaster images and model as input and classifies damaged buildings . The images are passed through model and it returns the output image. Metrics- confusion matrix by passing ground truth and the output image as arguments. This is a multi-class model.so, each diagonal elements represent the true postive values of each class in confusion matrix .

D. Architecture of the Model

Instead of developing a model from scratch, an existing model of Convolutional Neural Network for image segmentation is used . Namely, it turned to the U-Net, originally developed for biomedical image segmentation. Once trained, the network was able to output an image segmentation with good accuracy. Basically, the U-net builds upon the Fully Convolutional Network . A contracting path extracts features of different levels through a sequence of convolutions, ReLU activations and max poolings, allowing to capture the context of each pixel. A symmetric expanding path then upsamples the result to increase the resolution of the detected features. In the U-Net architecture, skip-connections (concatenations) are added between the contracting path and the expanding path, allowing precise localization as well as context. The expanding path therefore consists of a sequence of up-convolutions and concatenations with the corresponding feature map from the contracting path, followed by ReLU activations. The number of features is doubled at each level of down sampling. A figure of the sample U-Net is presented below.

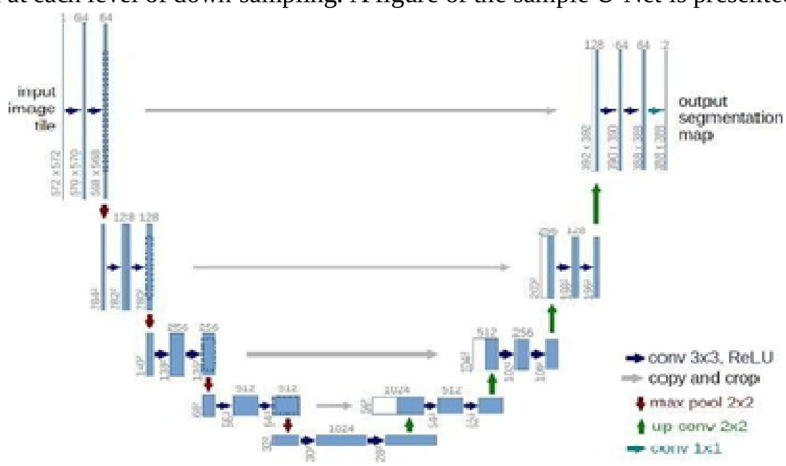

Fig. 1 Sample UNet segmentation model

The *UNET* was developed by Olaf Ronneberger et al. for Bio Medical Image Segmentation. The architecture contains two paths. First path is the contraction path (also called as the encoder) which is used to capture the context in the image. The encoder is just a traditional stack of convolutional and max pooling layers. The second path is the symmetric expanding path (also called as the decoder) which is used to enable precise localization using transposed convolutions. Thus it is an end-to-end fully convolutional network (FCN), i.e. it only contains convolutional layers and does not contain any Dense layer because of which it can accept image of any size. The main advantage of U-Net is it is able to localize and distinguish borders is by doing classification on every pixel, so the input and output share the same size.

1) *Encoder:* The encoder is the first half in the architecture diagram of UNet. It usually is a pre- trained classification network like VGG/ResNet where you apply convolution blocks followed by a max pool down sampling to encode the input image into feature representations at multiple different levels.
2) *Decoder:* The decoder is the second half of the architecture. The goal is to semantically project the discriminative features (lower resolution) learnt by the encoder onto the pixel space (higher resolution) to get a dense classification. The decoder consists of upsampling and concatenation followed by regular convolution operations.
3) *Upsampling:* Upsampling in CNN might be new to those of you who are used to classification and object detection architecture, but the idea is fairly simple. The intuition is that we would like to restore the condensed feature map to the original size of the input image, therefore we expand the feature dimensions. Upsampling is also referred to as transposed convolution, upconvolution, or deconvolution. There are a few ways of upsampling such as Nearest Neighbor, Bilinear Interpolation, and Transposed Convolution from simplest to more complex.

4) *Convolution operation:* Convolution is a mathematical operation on two functions that produces a third function that expresses how the shape of one is modified by the other. The term convolution refers to both the result function and to the process of computing it. There are two inputs to a convolutional operation-
   a) A 3D volume (input image) of size (nin x nin xchannels)
   b) A set of „k" filters (also called as kernels or feature extractors) each one of size (f x f x channels), where f is typically 3 or 5.
5) *Max Polling:* In simple words, the function of pooling is to reduce the size of the feature map so that we have fewer parameters in the network. Basically from every 2x2 block of the input feature map, we select the maximum pixel value and thus obtain a pooled feature map.
6) *Modifications to the existing system:* Some of the UNet features are modified. Here pre-trained B4Encoder and pre- trained UNetDecoder are used. Original U-net uses ReLU. We've over ridden the ReLU activation function with Swish activation function as it is produces more accurate results. Then we changed the dimensions of the input images, as the original U-net was designed for images of size 572 × 572 , We've set it for 512 × 512. UNet channels we used are [32, 64, 128] for encoding and [128, 64, 32] for the decoding part respectively. To avoid overfitting of the model we've set dropout value to 0.1. We used F1 score for metrics calculation of the model as the loss based on the F1 score (Dice coefficient) is commonly used for image segmentation as it allows coping with class imbalance. We added batch normalization after each Swish activation to speed-up training. The size of the batch we've set for training purposes is 5. We've used Adam optimizer.

E. Training and Testing Results

Model has been trained using NVIDIA GPU. We've given a batch size of 5 and carried out training for 100 epochs and with a learning rate of 0.01. At the end of training the model ,.i.e., for the 100th epoch Dice score reached to 0.7454. Below are the samples of the predicted data.

Fig. 2 (a) (b) (c) (d)

In fig. 2 , the image(a) represents pre-disaster satellite image, the image(b) represents post-disaster satellite image , the image(c) represents ground truth and the final image(d) represents predicted output after passing it through the trained model. Red represents damaged buildings and yellow represents slightly damaged buildings.

Fig. 3 (a) (b) (c) (d)

The image(3.a) is the pre-disaster image and image (3.b) is the post-disaster image of Joplin tornado (disaster event). Red represents damaged buildings , yellow represents slightly damaged buildings and green represents undamaged buildings.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

Fig. 4 (a) (b) (c) (d)

The image(4.a) is the pre-disaster image and image (4.b) is the post-disaster image of hurricane florence (disaster event) . Red represents damaged buildings , yellow represents slightly damaged buildings and green represents undamaged buildings.

Pre-disaster Post-disaster Ground Truth Predicted output

Fig. 5 (a) (b) (c) (d)

The image(5.a) is the pre-disaster image and image (5.b) is the post-disaster image of hurricane Florence (disaster event yellow represents slightly damaged buildings and green represents undamaged buildings.

F. Observations

a) Confusion Matrix: A confusion matrix is a tabular summary of the number of correct and incorrect predictions made by a classifier. It is used to measure the performance of a classification model. It can be used to evaluate the performance of a classification model through the calculation of performance metrics like accuracy, precision, recall, and F1-score.

This basically has 4 values.

a) True Positive: The number of pixels in an image that model has predicted True, and the actual value is also true. In our model it means, building is present and is actually detected.
b) True Negative: The number of pixels that model has predicted False, and the actual value is also False. In our model it means, building is not present and is actually not detected.
c) False Positive: The number of pixels that model has predicted True, and the actual value is False. It is also called a Type-I error.
d) False Negative: The number of pixels in an image that model has predicted False, and the actual value is True. It is also called Type-II error.

If a satellite image is taken from other source and passed as an input by the user, we generate the ground truth by using the segmentation tool in matlab. and then pass the output image and ground truth image as parameters for finding confusion matrix.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

G. Confusion Matrix and Precisions

X axis represents number of pixels corresponding to predicted label. Y axis represents number of pixels corresponding to True label. Diagonal elements represents True Positive values.

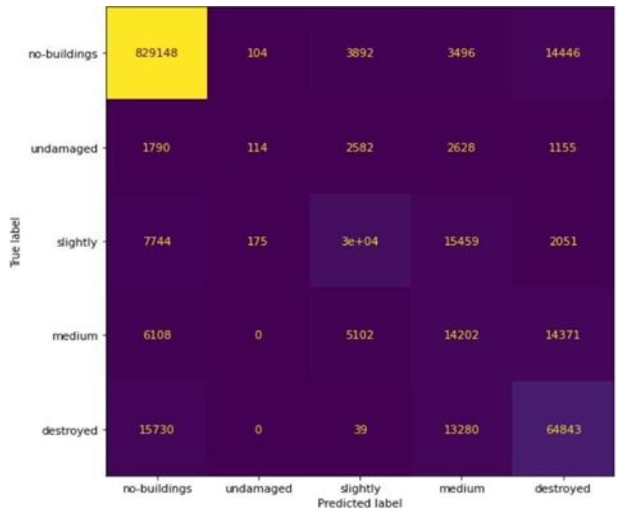

Fig. 6 (a)

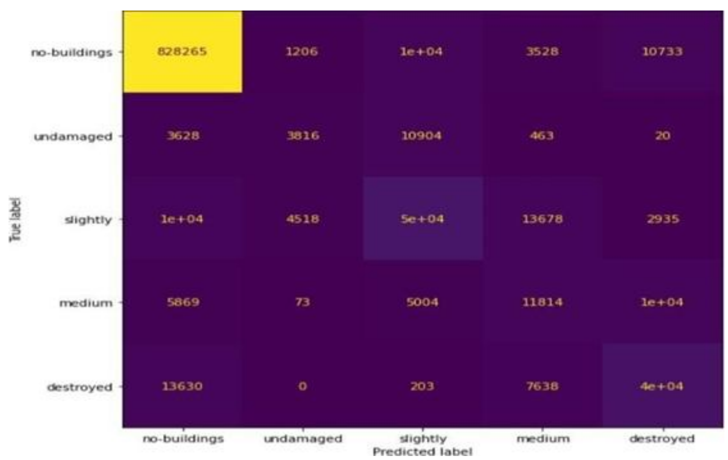

Fig. 6 (b)

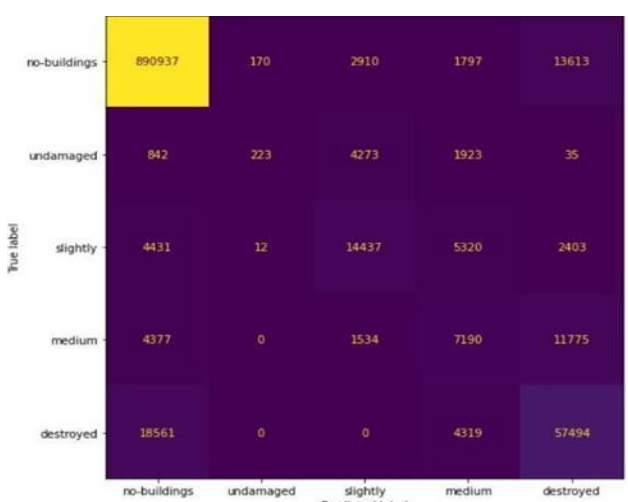

Fig. 6 (c)

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

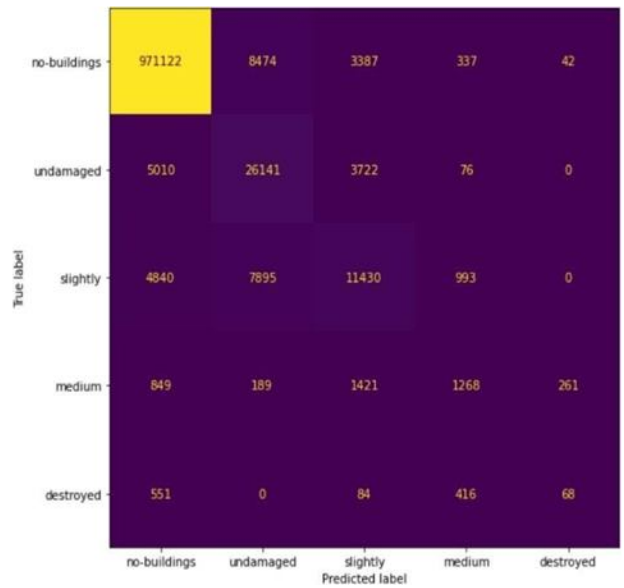

Fig. 6 (d)

Figure 6 (a),(b),(c) and (d) are based on the values of confusion matrix , Precision score is calculated. Precision score for the above observation is 0.96 . No. of True Positive pixels for no – buildings , undamaged buildings , slightly damaged buildings, medium damaged buildings and destroyed buildings.

IV. CONCLUSIONS

In this paper, we described a method to build convolutional neural networks that automatically detect damaged buildings in satellite images. Identifying buildings in satellite, aerial, and drone imagery can be done quickly and easily. By using deep learning approach we can detect damaged buildings that reduces cost, time, and other resources spent on analyzing geospatial data.

For future work, we plan to investigate additional disaster types, especially hurricanes and armed conflicts. We also plan to investigate techniques to make the model more robust to data flaws. For example, we can introduce random translations in the training images to make the model more robust to misalignment between pre- and post-disaster satellite images.

REFERENCES

[1] Chandan Chawda1,Jagannath Aghav and Swapnil Udar,"Extracting Building Footprints from Satellite Images using Convolutional Neural Networks", International Conference on Advances in Computing, Communications and Informatics, 2018
[2] Joseph Z. Xu, Zebo Li and Wenhan Lu," Building Damage Detection in Satellite Imagery Using Convolutional Neural Networks", 2019
[3] Lionel Gueguen and Raffay Hamid," Large-scale damage detection using satellite imagery", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015
[4] Gupta, Ritwik; Goodman, Bryce; Patel, Nirav; Hosfelt, Richard; Sajeev, Sandra; Heim, Eric; et al. (2019): "Creating xBD: A Dataset for Assessing Building Damage from Satellite Imagery". Carnegie Mellon University. Preprint. https://doi.org/10.1184/R1/8135576.v1
[5] Min Ji, Lanfa Liu, and Manfred Buchroithner. "Identifying collapsed buildings using postearthquake satellite imagery and convolutional neural networks: A case study of the 2010 haiti earthquake. Remote Sensing", 10:1689, 2018
[6] David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, and Colin Raffel." Mixmatch: A holistic approach to semi-supervised learning". In NeurIPS, 2019
[7] Kihyuk Sohn, David Berthelot, Chun-Liang Li, Zizhao Zhang, Nicholas Carlini, Ekin D Cubuk, Alex Kurakin, Han Zhang, and Colin Raffel. "Fixmatch: Simplifying semi-supervised learning with consistency and confidence". arXiv preprint arXiv:2001.07685, 2020.
[8] Ritwik Gupta, Richard Hosfelt, Sandra Sajeev, Nirav Patel, Bryce Goodman, Jigar Doshi, Eric Heim, Howie Choset, and Matthew Gaston. "xbd: A dataset for assessing building damage from satellite imagery", 2019.
[9] Joseph Z. Xu, Wenhan Lu, Zebo Li, Pranav Khaitan, and Valeriya Zaytseva. "Building damage detection in satellite imagery using convolutional neural networks". CoRR, abs/1910.06444, 2019.