5-2023

Ensemble Unsupervised Semantic Segmentation For Foreground-Background Separation On Satellite Image

Jaelen Tarry

ENSEMBLE UNSUPERVISED SEMANTIC SEGMENTATION FOR FOREGROUND-BACKGROUND SEPARATION ON SATELLITE IMAGE

A Thesis

by

JAELEN TARRY

Submitted to the Office of Graduate Studies of
Prairie View A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

May 2023

Major Subject: Electrical Engineering

ENSEMBLE UNSUPERVISED SEMANTIC SEGMENTATION FOR FOREGROUND-BACKGROUND SEPARATION ON SATELLITE IMAGE

A Thesis

by

JAELEN TARRY

Submitted to the Office of Graduate Studies of
Prairie View A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved as to style and content by:

Dr. Xiangfang Li Chair of Committee	
Dr. Lijun Qian Committee Member	Dr. Xishuang Dong Committee Member
Dr. Pamela Obiomon Committee Member	Dr. Annamalai Annamalai Head of Department
Dr. Pamela Obiomon Dean, Roy G. Perry College of Engineering	Dr. Tyrone Tanner Dean, Graduate Studies

May 2023

Major Subject: Electrical Engineering

ABSTRACT

Ensemble Unsupervised Semantic Segmentation for Foreground-background Separation on Satellite Image

(May 2023)

Jaelen Tarry, B.S., Prairie View A&M University, Prairie View, Texas

Chair of Advisory Committee: Dr. Xiangfang Li

Recently, computer vision has been promoted by deep learning techniques significantly, where supervised deep learning outperformed other methods such as in image segmentation. However, a large amount of annotated/labeled data is needed for training supervised deep learning models, while such big annotated data is typically unavailable in practice such as in satellite imagery analytics. In order to address this challenge, a novel ensemble unsupervised semantic segmentation method was proposed for image segmentation on satellite images. Specifically, an unsupervised semantic segmentation model was employed to implement foreground- background separation and then be placed within an ensemble model to increase the prediction accuracy further. Experimental results demonstrated that the proposed method outperformed baseline models such as k-means on a satellite image benchmark, the XView2 dataset. The proposed approach provides a promising solution to semantic segmentation in images that will benefit many mission critical applications such as disaster relief using satellite imagery analytics.

Index Terms - Convolution neural network (CNNs); deep learning; ensemble model; image segmentation; overhead imagery; unsupervised learning

ACKNOWLEDGMENTS

To start with, I want to thank the Lord for all he has given me and the opportunities he has placed in my hands, allowing me to achieve what I never thought would be possible. I am grateful for the grace and privilege I have received during this research and throughout my graduate program. My sincere gratitude also goes to my amiable supervisor and advisor, Dr. Xiangfang Li. Without your direction and steadfast support, this would not have been possible. Your guidance and support helped me to stay on track and helped me to figure out places that I needed to improve upon, allowing me the ability to grow.

Special thanks to Dr. Xishuang Dong and other members of the Center of Excellence in Research and Education for Big Military Data Intelligence (CREDIT Center) for their support and numerous direct and indirect contributions towards completing this research. I also immensely appreciate the members of my thesis committee, Dr. Lijun Qian and Dr. Pamela Obiomon, for their unrelenting support toward the timely completion of my thesis. Finally, I want to especially appreciate my family, who have always shown their support whenever I need it and have showered me with encouragement that I can achieve anything I set my mind to accomplish.

This research work is supported in part by IBM Master's Fellowship and by the U.S. Air Force Research Lab (AFRL). This material is based on research sponsored by the Air Force Research Laboratory under agreement number FA8650-20-2-5853. The U.S. Government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be

interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Research Laboratory of the U.S. Government.

TABLE OF CONTENTS

Page

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

CHAPTER 1

INTRODUCTION

1.1 Damage Assessment in Satellite Images using Machine Learning

The world is plagued by the onslaught of hurricanes, tornadoes, floods, tsunamis and many other natural disasters that can leave behind devastating destruction. Many workers and volunteers offer up their support to provide disaster relief for those affected, whether it is search and rescue or providing food and shelter. Disaster relief can come in many different shapes and forms; however, the most critical part of disaster relief is the determination of building damages. Currently, the process of inspecting the damages of a building after a natural disaster is slow and dangerous. In order to address this issue, damage assessment of buildings using satellite images is a promising approach.

Fig. 1.1. Example results for the [1] joint damage scale on a satellite image

This thesis follows the style of the *Institute for Electrical and Electronics Engineers (IEEE)*

With the recent rapid advancement of artificial intelligence (AI) and machine learning (ML), the Xview 2 challenge is designed to encourage researchers to develop effective machine learning solutions for damage assessment of buildings using satellite images. The Xview 2 challenge illustrated how to assess damages to affected areas by using computer vision algorithms to analyze satellite imagery. With the United States Department of Defense's Defense Innovation Unit (DIU) being the sponsor, many competitors used different ML techniques to assess the before and after damages from natural disasters [1]. The goal is to create an algorithm that could detect the buildings within the satellite imagery and assess the damages of those buildings according to the Joint Damage Scale.

An example of the damage scales can be seen in Fig. 1.1, where the colors represent a different level of damage. Here the green color represents buildings with little to no damage done, the red color represents severe damage has been done to the building, while the orange color is for buildings that have been damaged but not severe or urgent. The goal of the DIU is to help provide a faster and more reliable building damage estimation to produce a quicker response time to help the areas with the most severe damages from these disasters.

One of the most important steps in the process of damage assessment in satellite images is to detect the buildings within the satellite imagery, in other words, separate the buildings (foreground) from the background in the image.

1.2 Foreground-Background Separation

Xview 2 satellite imagery creates a significant data image that can provide competitors with vast information about the damages caused by disasters. However, not all the information within the image is usable, nor is its data essential.

The separation of image foreground-background was introduced to help remove

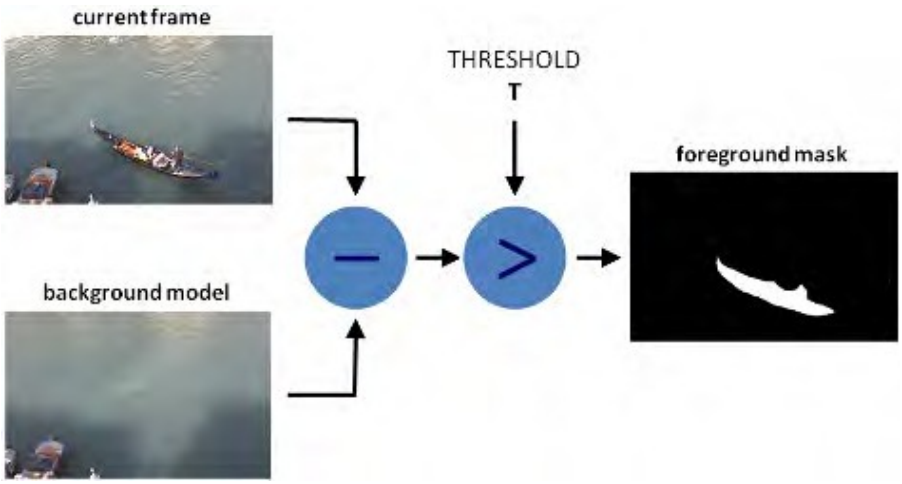


Fig. 1.2. Example of implementing [9] foreground-background separation

some excess data within the satellite imagery [9]. Foreground includes the parts of the images in which the data was needed for the Xview 2 challenge to assess the damage caused by natural disasters. This included types of large structures that the algorithm could detect, such as houses, factories, facilities, stores, malls, schools, and recreational centers. Background includes the data from the satellite imagery that is not used or needed to assess the damages for the Joint Damage Scale. An example of foreground separation is represented by Fig. 1.2, which takes the background from the image and only focuses on the foreground; in this case, it is represented by the canoe.

The foreground-background separation problem can be formulated as an image segmentation problem where the number of segments is two.

1.3 Image Segmentation

Image segmentation is the process of dividing an image into different regions based on the characteristics of pixels to identify objects or boundaries to simplify an image and more efficiently analyze it.

With the growth of Artificial Intelligence algorithms and their ecosystem, Digital Image Processing using Neural Networks has recently become popular. It has various application areas like security, banks, military, agriculture, law enforcement, manufacturing, and medical technologies [10]. For instance, the software behind green screens implements image segmentation to crop out the foreground and place it on a background for scenes.

In this thesis research, this technology has been applied to the overhead imagery from satellites to determine objects such as buildings, people, and vehicles. Currently, some datasets provided images that show the results from their use of image segmentation, like the PASCAL VOC dataset in Fig. 1.3. Other datasets have been used to test segmentation and object detection, such as PASCAL VOC, Cityscapes, and ADE20K.

There are many ways to implement image segmentation, such as Threshold Based Segmentation, Edge Based Segmentation, Region-Based Segmentation, Clustering Based Segmentation, and Artificial Neural Network Segmentation.

Fig. 1.3. [2] Pascal VOC dataset segmentation results

1.4 Challenges and Motivation

Recent years have witnessed that computer vision was significantly promoted by deep learning techniques, where supervised deep learning outperformed other methods but needed a mass of big, annotated data for training models. However, such big data is unavailable for emerging tasks such as satellite imagery analytics. Supervised learning-based semantic segmentation has been successful in background detection but not appropriate to the emerging tasks regarding limited annotated data available. Hence, unsupervised method is proposed in this thesis to address this challenge.

There exist additional challenges because of the availability and quality of satellite images. The available Xview 2 dataset allowed for the implementation of image segmentation and object detection onto satellite imagery. Before this, the use of segmentation or detection for satellite imagery was very little. This is partly due to the lack of data annotation for the satellite imagery, which hindered the use of segmentation evaluation metrics to determine the algorithms' efficiency. Additionally, the previous uses of segmentation and detection involved images consisting of clear and concise objects. These images had a different orientation from satellite imagery, similar to the difference between first person and third person viewpoints, which proved to be a learning hurdle for previously created segmentation techniques. The sizes of the images are also much larger compared to the images used in the PASCAL or Cityscapes datasets. This can prove to be harmful to the detection due to the large areas that the model needs to cover. In addition, due to the images being taken from a satellite, the contents are smaller and represented by fewer pixels. In order to address these additional challenges, an ensemble method is proposed in this thesis to improve the performance of the unsupervised image segmentation model.

1.5 Contributions

In this thesis research, a type of Artificial Neural Network Segmentation for the image segmentation of the satellite imagery from Xview 2 is proposed. Specifically, to address the challenge of limited labeled data, Unsupervised Semantic Segmentation (USS) is applied to use Deep Learning and create convolutional layers to segment the images into classes. This thesis provides more in-depth information about image segmentation and why USS was chosen to be implemented on the Xview 2 dataset.

Furthermore, ensemble unsupervised image segmentation and method validation are performed to improve the performance of the USS.

1.5.1 Ensemble Unsupervised Image Segmentation. Ensemble unsupervised semantic segmentation (USS) consists of methods used to detect background of images and perform foreground-background separation. Unsupervised semantic segmentation does not require huge amounts of annotated data to train models for label pixels of various objects in images, instead it accomplishes semantic segmentation through learning high-quality data representations and optimizing pseudo supervision to enhance performance. To further enhance performance of USS, the use of an ensemble model is proposed that consisting of multiple USS with different hyper-parameters and majority voting to combine the outputs from multiple USS into a final improved output.

1.5.2 Validation. The proposed performance metric of validation of the images is Intersection over Union (IOU) which is a popular metric of validation for segmented images [11]. By dividing the results of the intersecting pixels with the union of pixels of both images to produce a metric that could compare the produced results with those of other experiments.

1.6 Outline of the Thesis

The remaining portion of this thesis is structured as follows: Chapter 2 contains the literature review of different approaches and ideas of the competitors for the Xview2 challenge. The proposed method is given in Chapter 3. This chapter also included the different changes made and evaluation standard used. Chapter 4 provides the experimental results and analysis. Chapter 5 concludes this thesis and suggests some future work.

CHAPTER 2

LITERATURE REVIEW

This thesis focused on foreground and background separation on satellite images. It can be viewed as a standard computer vision task *image segmentation* that is to classify pixels into binary classes including foreground and background. This literature review starts with introduction on image segmentation. Then, it discusses foreground-background separation. Finally, it presents a short summary of current work and highlight our motivation.

2.1 Image Segmentation

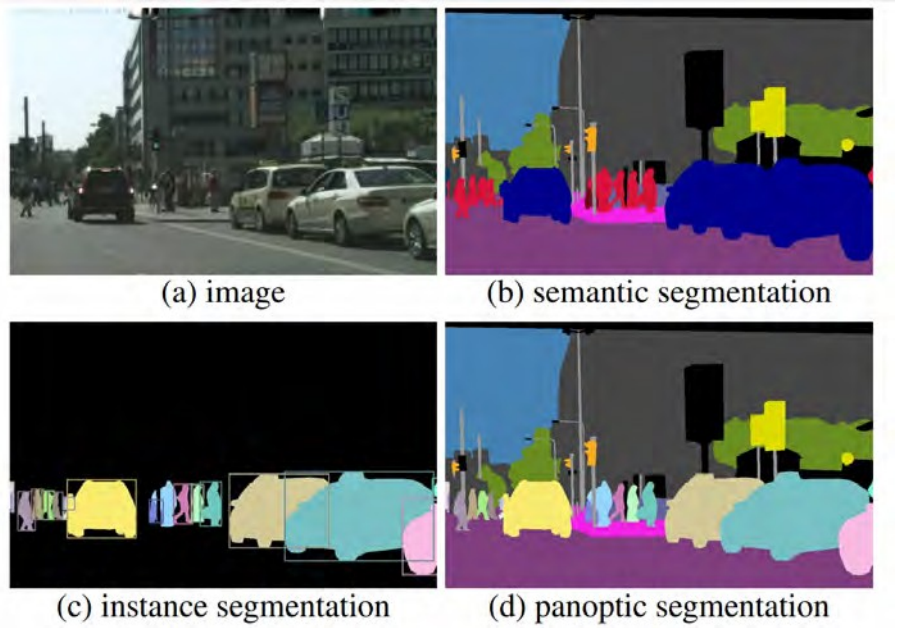


Fig. 2.1. [3] Different types of segmentation from the Cityscapes dataset

Image segmentation can also be broken down into three different types, such as instance segmentation, semantic segmentation, and panoptic segmentation. Through instance segmentation, each individual instance of each object is recognized [12]. The distinction between instance segmentation and semantic segmentation is that the latter does not categorize each individual pixel. When three identical objects of the same type (like bicycles) are present in an image, instance segmentation identifies each individual bicycle while semantic segmentation groups all three bicycles together as a single instance. Semantic segmentation divides picture pixels into one or more classes that can be used to represent actual objects in the real world rather than abstract concepts [13]. The conceptualization of tiny pixel clusters that are most likely to be associated with the same object, are regarded as a technique called region proposal. The process of region proposal and annotation involves classifying the pixel values into discrete groups using CNN. Panoptic Segmentation is the unification of both semantic and instance segmentation where is incorporates an efficient majority voting algorithm to com- bine the two types of segmentation [14]. This is produced by creating unique labels consisting of semantic labels and instance ids, to be placed on each pixel of an image. Once each pixel has been labeled it needs to have a corresponding class to identify each object in an image as shown in Fig. 2.1.

Fig. 2.2. Results from a [4] road detection software

There has been an abundance of applications for the use of image segmentation and different papers introduced with their findings. The majority of information discovered consisted of supervised semantic segmentation as opposed to unsupervised semantic segmentation that is used in this topic. Research involves expanding the field of view of filters to incorporate larger context without the need for increasing the number of parameters [15]. Or bring in a new approach for semantic segmentation, making use of an encoder and decoder framework followed by a pixel-wise classification layer [16]. Currently there have even been examples of using Deep Lab semantic segmentation hyper-spectral image classification to extract features pixel by pixel at multiple scales [17]. This can show the versatility and reliability of Deep Lab since there are so many different applications and statistics promoting it. Additionally, new research has also been introduced with the goal of producing a simpler, stronger, and faster system for image segmentation called panoptic segmentation.

When compared to using panoptic segmentation on the Cityscapes benchmarks, it had performed better than semantic segmentation and instance segmentation [18]. Deep Lab 2 introduces a few new models which make use of this panoptic segmentation to provide a faster and diverse applications for a wide range of uses [19] [20].

Applying image segmentation on satellite imagery consists of Modified U-Net Convolution networks, three stage segmentation, and two-stage algorithms. Modified U-Net Convolution networks have been introduced to show improvements on the detection of roads, buildings, and vehicles within satellite imagery [21]. Even from Fig. 2.2, what is showcased is the current results from the detection of roads within satellite imagery [22]. Unfortunately, this aspect could not be incorporated into the experiment due to the applications used but it still introduced a 10% increase when it comes to the detection of roads. Another technique used consist of three phases, where the purpose of the three-stage segmentation is to "divide image into set of non-overlapping regions based on special features" [23]. First phase includes the use of a filter method that can divide the remote sensing image into special blocks that are easier to segment. After this process the next phase entails using the watershed method of segmentation and the removal of noisy segmentation with the use of statistical threshold method. The last phase is the reduction of segmented areas using region-based segmentation, some of the results produced showcase the reduction of time, noise, and overall segmentation.

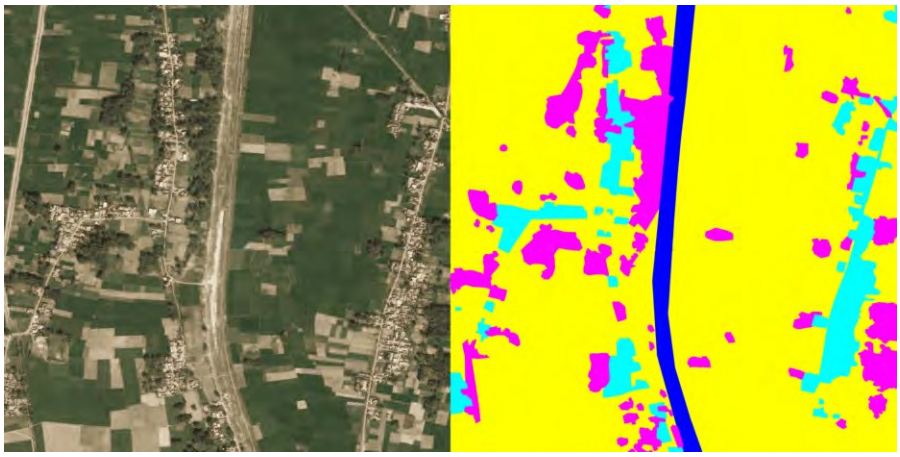


Fig. 2.3. Results from [5] satellite image segmentation

Two-stage algorithm can help with the segmentation of satellite images by detecting textural areas and small-sized objects within the images [24]. The study showed an impressive six percent increase in precision as compared to the k-means clustering [25]. An important component that help to increase the precision was due to the flexibility provided by convolutional neural networks. As compared to the current experiment, there is no distinction between the use of supervised or unsupervised elements so this technique cannot be used to further improve the evaluation results of this segmentation. Fig. 2.3, represents some of the previous results of unsupervised satellite image segmentation and shows the need for improvements. There are many different skills and techniques that could be used to further develop unsupervised image segmentation in computer vision [26].

2.2 Foreground-Background Separation

By removing the background from images it can help to further improve object detection or to help improve the accuracy of image segmentation and gets rid of unneeded information from images.

To further explain an example of an application that can incorporate background separation includes the Deep Lab software from TensorFlow, as it can differentiate the background of images and remove them [27]. Once the background is removed, the training can be done by placing the images with the removed backgrounds into new crowded backgrounds. The goal was to use the new crowed background images to improve object detection even in hard to tell images. The difference between this work and the current experimentation includes what detections are being made with through the experiments, with the use satellite imagery, compared to the article's use of natural images with only one object being detected such as an eagle or a plane [28]. The article's produced results have been proven to have a positive effect on the object detection of natural images with an 100% average increase in accuracy with the use of cluttered backgrounds [28]. Although, the application is very different compared to individual object detection of natural images, the use of Deep Lab software and its foreground-background image segmentation can help to discover new data about satellite imagery and their different applications.

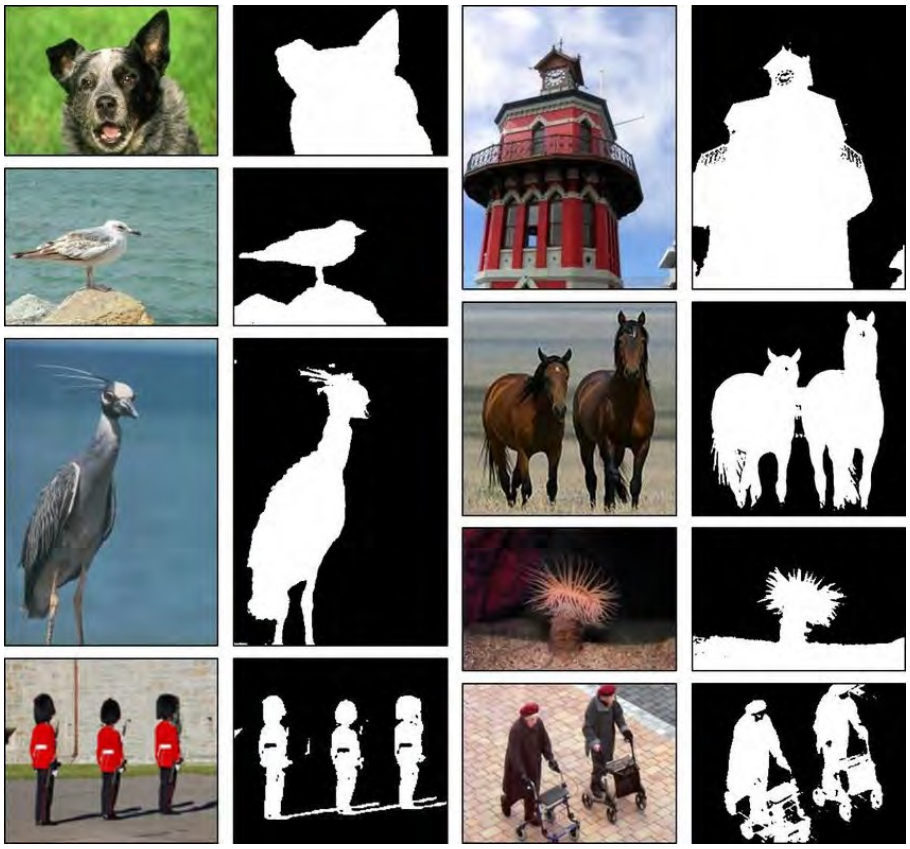


Fig. 2.4. Results from [6] Threshold Foreground Background Separation

Another experimentation based on the effectiveness of foreground-background separation, produced positive results to further support the effect that separation can have in segmentation and detection of computer vision [29]. There are even instances where experiments have applied foreground-background separation on videos as well by using a similar technique to superpixel detection [30]. The algorithms used for separation are also divers. This includes an algorithm based on active contour detection and an algorithm based on thresholding to implement the separation [31] [32]. Even from the example shown in Fig. 2.4 there are many different ways to incorporate and specify the importance of foreground-background separation within computer vision.

2.3 Summary

Deep learning techniques have promoted image segmentation significantly. The current research involves seems to focus on supervised image segmentation and few works have explored effective image segmentation on satellite imagery. However, supervised methods required detailed annotations on images to build large amounts of training data, which will be high-cost for satellite imagery applications. This research aimed to reduce efforts of data annotation for satellite imagery segmentation. It employed state-of-the-art unsupervised deep learning-based image segmentation methods to build baselines for fore-ground-back-ground separation. In addition, this thesis proposed a novel ensemble unsupervised segmentation method to enhance unsupervised segmentation performances. The proposed method was validated on a large satellite image dataset *xDB* and the performance evaluated via MIOU confirmed that the proposed method was able to separate the foreground and background with promising performance.

CHAPTER 3

METHODOLOGIES

3.1 Unsupervised Semantic Segmentation

Unsupervised Semantic Segmentation (USS) is comprised of many building blocks that make up the core concepts within this experiment. The different components are further divided into: convolutional neural networks, learning modules, segmentation, and ensemble model. These topics are further introduced in the sections below.

3.1.1 Convolution Neural Networks. In the last decade, machine learning models have proven to be successful in many application areas, such as image recognition, times series forecasting, and sentiment analysis. Machine learning includes Deep Learning as a subset. The purpose of deep learning, which was developed in the year 2005 - 2006, was to solve the issues that Machine Learning currently faces. Deep Learning is a group of statistical machine learning methods used to build feature hierarchies that are frequently built using artificial neural networks. Deep Learning models can focus on the appropriate features independently, but they still need some programming assistance [33]. Additionally, these models address the dimensionality issue. The primary goal of deep learning is to create learning algorithms that closely resemble the human brain. It is put into practice with the aid of neural networks. Neural Networks (NN) are a vital piece of some of the most successful machine learning algorithms. The development of neural networks has been key to teaching computers to think and understand the world in the way that humans do [34].

Essentially, a neural network emulates the human brain cells, or neurons, are connected via synapses. This is abstracted as a graph of nodes (neurons) connected by weighted edges (synapses). Convolutional neural network (CNN), a class of artificial neural networks that have become dominant in various computer vision tasks, is attracting interest across various domains, including radiology. CNN is designed to automatically learn spatial hierarchies of features and provide flexibility through backpropagation using multiple building blocks, such as convolution layers, pooling layers, and fully connected layers [35].

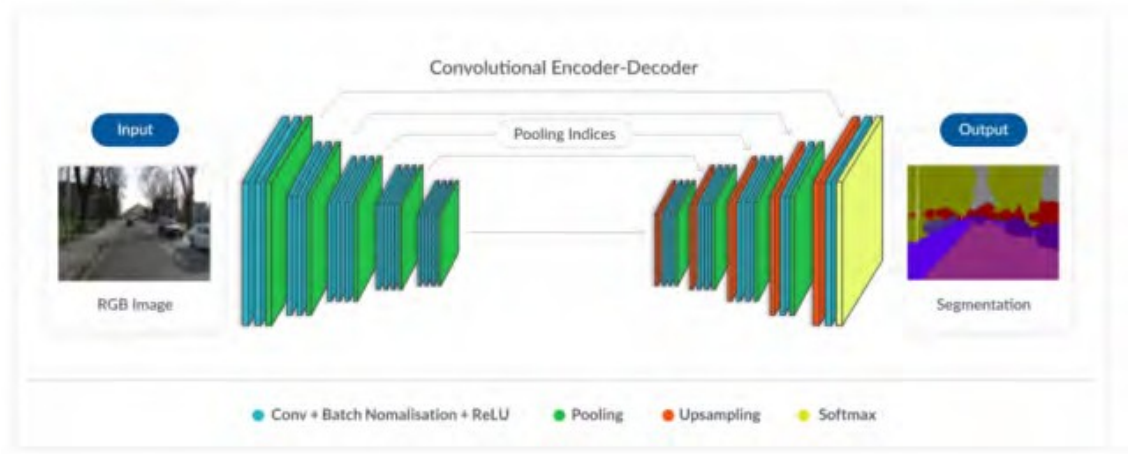


Fig. 3.1. The [7] SegNet model Architecture for Image Segmentation

Convolutional Neural Networks are deep learning models that are commonly used for vision applications and language processing tasks. A specialty of CNNs is that they can efficiently recognize patterns that occur in the input image, including lines, shapes, gradients, eyes, or even faces [36]. As a result of CNN's ability to work so well at recognizing details, the model is primarily used in computer vision applications. Unlike many other previous computer vision models, CNN's can work with a raw image, not needing any pre-processing.

What makes a CNN different from other neural networks is its special layer called the convolutional layer, and the pooling layer. In fact, the convolutional layers are the major building blocks used within CNNs and are the most important components of this network. Another aspect of the CNNs are the encoder and decoder layers, which are used in CNN architecture for segmentation as shown in Fig. 3.1. The input is encoded using encoders into a representation that can be transmitted across the network, and the representation is decoded using decoders [37]. For the goal of generating a segmentation map, the encoders might be convolution neural networks and the decoders could be based on deconvolutional or transposed neural networks.

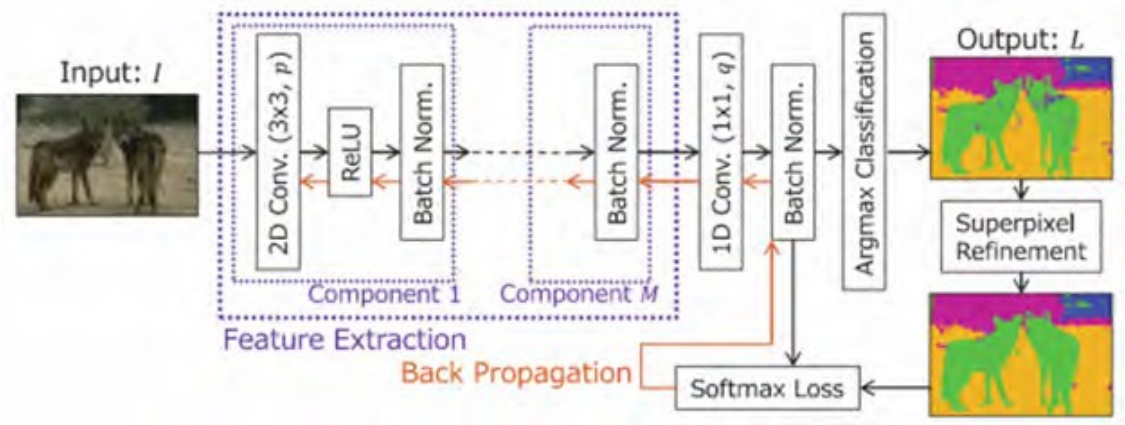


Fig. 3.2. [8] Unsupervised CNN with Backpropagation for Image Segmentation

Lastly, the implementation of the loss equation, can help in detecting the current progress of the segmentation even before the metric evaluation to allow a shallow understanding of how the segmentation is working. The Loss function used in the Cross Entropy Loss function is shown in Fig. 3.1. This function is provided by Pytorch and helps to give a basic understand of where the segmentation process is leading towards. When used, the lower the loss value the more accurate the

segmentation is, as a basic comparison when compared to other evaluation metrics. The loss can be produced by the equation below:

$$l(x, y) = L = \{l_1, \ldots, l_n\}^T, l_n = -\sum_{c=1}^{C} w_c \log \frac{\exp(x_{n,c})}{\sum_{c=1}^{c} \exp(x_{n,i})} y_{n,c}, \quad (3.1)$$

where x is the input, y is the target, w is the weight, C is the number of classes, and N is the number of pixels within the input images.

3.1.2 Learning Models. Segmentation can also infer more details from an image depending on the learning model used for the segmentation. There are three primary learning models: supervised learning, semi-supervised learning, and unsupervised learning. Supervised learning involves using annotated images of pictures to learn different patterns and similarities between the objects proposed. This learning method is the most commonly used module and produces a wide range of results from medical experiments to object detection and is known to provide the best results. However, with supervised learning, the amount of storage and the creation of models is the most powerful and time-consuming module of the three.

Unsupervised learning is where the model inspects an image without prior annotations used to learn and understand an image through the grouping of pixels within the image, called clustering. This learning has the least amount of power consumption and can produce results faster than the other modules. This is achieved through the grouping of pixels that closely resemble one another and those that are dissimilar to pixels within other clusters. However, it achieves worse results when compared to other learning.

Lastly, semi-supervised learning involves a combination of supervised and unsupervised learning to provide a more flexible and broader variety of applications. Semi-supervised consists of training on detecting objects within an image like the supervised model but includes additional unlabeled data, suggestions, and exampled labels. The model used for the segmentation is the Unsupervised Learning Model. Unsupervised machine learning can uncover patterns within data that have been previously hidden. Unsupervised learning has disadvantages such as interpretation and labeling results, lack of insight into the clustering, and the accuracy of outputs. However, unsupervised learning is often more convenient, quicker, and less expensive to implement than supervised learning since it does not require the manual classification of data that supervised learning does. Table 3.1 presents comparison of these three learning models.

TABLE I
COMPARISON OF THREE LEARNING MODELS

Models	Overview	Process	Examples
Supervised Learning	A majority of algorithms where the machine is trained using well-labeled data.	Mapping functions take the inputs to match the output to create a target function.	Linear Regression, Random Forest, SVM
Unsupervised Learning	Inputs of unlabeled data is analyzed and the machine begins to learn without supervision.	Input is used to create a model of the data.	K-Means, Hierarchical Clustering
Semi-Supervised Learning	Some data is labeled and some not. Best used with real world data. Goal: Have better results than labeled data alone.	Combination of above processes.	Self Training, Mixture Models, Semi-Supervised SVM

3.1.3 Image Segmentation. The image segmentation process consists of three steps that are described as follows: image pre-processing, Superpixilation, and Foreground-Background Separation. For the image pre-processing we assume that the image is an Red-Green-Blue (RGB) image that consist of a 3D array to represent the values for the RGB image. The representation of the image can be represented from the equation below:

$$\mathrm{I} = \{\mathrm{v}_n \in \mathrm{R}^3\}_{n=1}^{N}, \tag{3.2}$$

where each pixel value is normalized to [0, 1]. Then we send the image through the convolution components which consists of a 2D convolution, ReLU activation function, and a batch normalization function, where a batch corresponds to N, pixels of a single input image. The following step is the superpixilation which consists of creating a superpixel within the input image and using that as the center of the cluster to spread around and detect patterns to then classify the sections into different classes [38]. This can be represented by the following psudocode:

$$\{\mathrm{S}_k\}_{k=1}^{K} \leftarrow \mathrm{GetSuperPixel}(\{\mathrm{v}_n\}_{n=1}^{N}), \tag{3.3}$$

where S_k denotes a set of the indices of pixels that belong to the k_{th} superpixel. To allow for the creation of the clusters with the superpixels at the center starts with forcing all of the superpixels into having the same cluster label [39]. More specifically, by letting $|\mathrm{c}_n|_{n \in S_k}$ be the number of pixels in S_k that belong to the C_{nth} cluster helps in selecting the most frequent cluster label c_{max} [40]. Next, it is to obtain the response map using the equation: $\{\mathrm{y}_n = \mathrm{W}_c\mathrm{x}_n + \mathrm{b}_c\}_{n=1}^{N}$ by applying a linear classifier, where $\mathrm{W}_c \in \mathrm{R}^{q \times p}$ and $\mathrm{b}_c \in \mathrm{R}^q$. Once the response map has been obtained the next step is to apply the normalization process for the response map

y_n, before assigning cluster labels via argmax classification. Here, we use batch normalization which is described as follows:

$$y'_{n,©} = \frac{y_{n,©} - \mu_i}{\sqrt{\sigma_i^2 + \epsilon}}, \tag{3.4}$$

where μ_i and σ_i denote the mean and standard deviation of $y_{n,©}$, respectively. Note that ϵ is a constant that is added to the variance for numerical stability. This operation helps to convert the original responses y_n to y_n', where each axis has zero mean and unit variance. The information introduced into this section involves the steps taken to implement the segmentation for USS model.

Specifically, Algorithm 1 represents basic ideas of Unsupervised Image Segmentation model [8], while Algorithm 2 is the revised version of Algorithm 1 that implement foreground-background separation and the use of the evaluation model. However, it does not showcase the ensemble model or majority voting algorithm which is the major emphasis of this project and is introduced in Section 3.1.4.

Algorithm 1 Image Segmentation Algorithm

Input: $\mathrm{I} = \{\mathrm{v}_n \in \mathrm{R}^3\}_{n=1}^{N}$ ▷ RGB Image

Output: $\mathrm{L} = \{\mathrm{c}_n \in \mathrm{Z}\}_{n=1}^{N}$ ▷ Labeled Image

$\{\mathrm{W}_m, \mathrm{b}_m\}_{m=1}^{M} \leftarrow$ Init() ▷ Initialize

$\{\mathrm{W}_c, \mathrm{b}_c\} \leftarrow$ Init() ▷ Initialize

$\{\mathrm{S}_k\}_{k=1}^{K} \leftarrow$ GetSuperPixel ($\{\mathrm{v}_n\}_{n=1}^{N}$)

for $\mathrm{t} = 1$ **to** T **do**

$\{\mathrm{x}_n\}_{n=1}^{N} \leftarrow$ GetFeatures ($\{\mathrm{v}_n\}_{n=1}^{N}, \{\mathrm{W}_m, \mathrm{b}_m\}_{m=1}^{M}$)

$\{\mathrm{y}_n\}_{n=1}^{N} \leftarrow \{\mathrm{W}_c\mathrm{x}_n + \mathrm{b}_c\}_{n=1}^{N}$

$\{\mathrm{y}'_n\}_{n=1}^{N} \leftarrow$ Norm ($\{\mathrm{y}_n\}_{n=1}^{N}$) ▷ Batch Normalization

$\{\mathrm{c}_n\}_{n=1}^{N} \leftarrow \{$ argmax $\mathrm{y}'_n \}_{n=1}^{N}$ ▷ Class Assignment

for $\mathrm{k} = 1$ **to** K **do**

$\mathrm{c}_{max} \leftarrow$ argmax $||\mathrm{c}_n||_{n \in S_k}$

$\mathrm{c}'_n \leftarrow \mathrm{c}_n$ **for** $\mathrm{n} \in \mathrm{S}_k$

end for

$\mathrm{V} \leftarrow$ Cross-EntropyLoss $\{\mathrm{y}'_n, \mathrm{c}'_n\}_{n=1}^{N}$ ▷ Loss Function

$\{\mathrm{W}_m, \mathrm{b}_m\}_{m=1}^{M}, \{\mathrm{W}_c, \mathrm{b}_c\} \leftarrow \mathrm{Update}(\mathrm{V})$

end for

Algorithm 2 Revised Image Segmentation Algorithm

Input: $\mathrm{I} = \{\mathrm{v}_n \in \mathrm{R}^3\}_{n=1}^{N}$ ▷ RGB Image

Output: $\mathrm{L} = \{\mathrm{c}_n \in Z\}_{n=1}^{N}$ ▷ Labeled Image

$\{W_m, \mathbb{b}_m\}_{m=1}^{M} \leftarrow$ Init() ▷ Initialize

$\{W_c, \mathbb{b}_c\} \leftarrow$ Init() ▷ Initialize

$\{S_k\}_{k=1}^{K} \leftarrow$ GetSuperPixel ($\{v_n\}_{n=1}^{N}$)

for $\mathrm{t} = 1$ **to** T **do**

$\{x_n\}_{n=1}^{N} \leftarrow$ GetFeatures ($\{v_n\}_{n=1}^{N}, \{W_m, b_m\}_{m=1}^{M}$)

$\{y_n\}_{n=1}^{N} \leftarrow \{W_c x_n + b_c\}_{n=1}^{N}$

$\{y'_n\}_{n=1}^{N} \leftarrow$ Norm ($\{y_n\}_{n=1}^{N}$) ▷ Batch Normalization

$\{c_n\}_{n=1}^{N} \leftarrow \{$ argmax $y'_n \}_{n=1}^{N}$ ▷ Class Assignment

for $\mathrm{k} = 1$ **to** K **do**

$c_{max} \leftarrow$ argmax $||c_n||_{n \in S_k}$

$c'_n \leftarrow c_{max}$ **for** $n \in S_k$

end for

$\mathrm{Loss} \leftarrow$ SoftmaxLoss $\{y'_n, c'_n\}_{n=1}^{N}$ ▷ Loss Function

$f_c \leftarrow$ maxval (c'_n) ▷ Foreground Classification

$g_c \leftarrow \sum_{c=1}^{c_{max}-1} c'_n$ ▷ Background Classification

end for

MioU $\leftarrow$ GroundTruth + L ▷ Evaluation

3.1.4 Ensemble Model. After the creation of the Unsupervised Semantic Segmentation model from algorithm 2, next step is the implementation of an ensemble model. The basis of an ensemble model is to create, k, number of USS models consisting of different hyper-parameters [41]. Next, using the results produced from the k models, is passing them through the majority voting evaluation to produce the final output. Majority voting consist of comparing different components from the different models such as if $\mathrm{USS}_1 = 1$, $\mathrm{USS}_2 = 0$, and $\mathrm{USS}_3 = 1$ then the final output produced will be 1 due to the majority. Fig. 3.3 demonstrated majority voting ensemble model to our task on foreground-background separation.

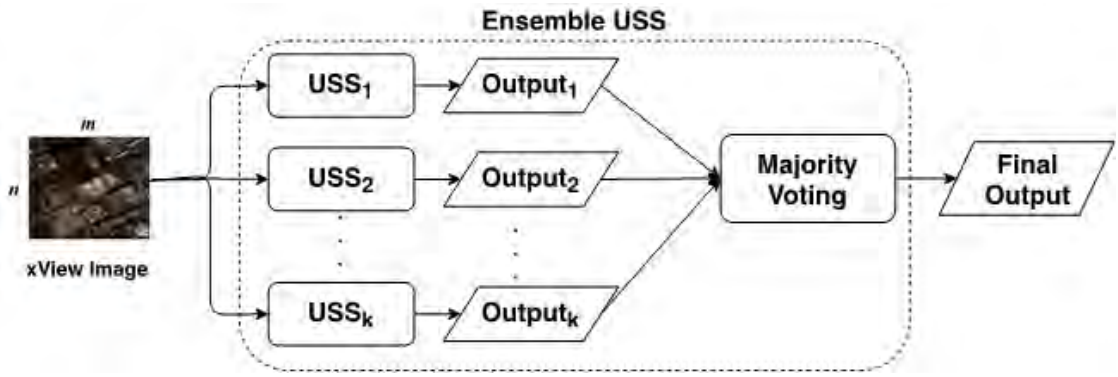


Fig. 3.3. Majority Voting Ensemble Model Block Diagram

To begin creating the ensemble model involves creating the needed models that have produced slightly different results from one another. This is done by changing some of the parameters from the model or by using different techniques to produce a different result from the USS model. This experiment creates the different models used by changing some of the parameter used in the USS model. The hyper- parameters that can be potentially changed are listed below:

- **nConv:** This parameter represents the number of convolutional layers used in the CNN, the base value is set to 2. This is the parameter that is changed throughout all of the different models.

- **nChannel:** This parameter provides a pivotal role in the convolutional layers and is what helps to label the clusters and to apply the represented color to the represented classes.

- **maxIter:** This parameter belongs to the maximum number of iterations that can be performed on an image, with each iteration the model becomes more refined to reduce the number of classes produced in the model.

- **minLables:** This parameter instructs the code of the minimum number of labels or classes that can be produced within an image. Once this number has been reached the code will stop running the iterations and output the current results.

The parameters that are consistent throughout the different models consist of nChannel, maxIter, and minLables. The value of nChannel is 8, this is due to each iteration that the image passes through the number is eventually reduced, so the value needs to be large enough to have a decent amount that can still be reduced through the iterations. Next, maxIter is set to a value of 100 because during testing the number of iterations did not surpass 70 except for a few outliers that used around 90 iterations, which makes 100 a great value for maxIter. With minLables the value is set to 4, This is due to the foreground-background separation and how there needs to be at least three classes to accurately produce the segmentation with the separation incorporated without errors. Lastly, nConv is set to 2 as a base value for the USS model however, for this experiment each models have a different number of convolutional layers to produce different results.

The values of 2, 4, and 5 are given to the USS_1, USS_2, and USS_3 models nConv parameter, respectively.

Once the model has been created, the next step involves the creation of the majority voting section for the ensemble model. The technique used for the majority vote compares the images pixel by pixel. Each step takes the position of a pixel by the x and y coordinates, and compares it between the two images and as stated earlier takes the majority representation of the three models. Each pixel is represented with either a 1 or a 0 to represent the foreground and background, respectively. Once the (1024x1024) 1, 048, 576 pixels have gone through the majority voting comparison it will create a(n) image based on the results. This will produce the output for the ensemble model as shown in figure 3.3, based on the three USS_k models.

3.2 Foreground-Background Segmentation

A challenging component of the implementation of the proposed method lies in the execution of the foreground-background separation to detect parts of an image and implement separation for critical items within the image. The challenges to implement are due to the unsupervised learning imposed on the project causing the detection to become more complex and less accurate [28]. To combat most of those challenges, the building blocks introduced in Section 3.1, by making use of the versatility of CNN-based models to further improve detection and accuracy. Currently, the USS model creates a segmentation with multiple classes, leading to multiple colors used to represent the corresponding classes. The goal of the foreground-background separation is to bring the number of classes produced to only two and to have those classes defined as black for the background and white for the foreground. This allows for using the ground truth provided by the xBD dataset

to produce a metric evaluation based solely on the foreground and background segmentation. To be implemented, a foreground-background detection statement that changes the assigned labels need to be created. Due to taking the number of classes proposed, we determined which class contains the most significant number of represented pixels, such as $f_c = \mathrm{maxvalue}(c'_n)$. Incorporate these ideas into the segmentation can be inferred from the pseudocode below:

$$g_c = \left(\sum_{c=1}^{C_{max}-1} c'_n\right) - f_c, \tag{3.5}$$

where f_c represents foreground class and g_c represents background class. This groups up the classes where the class that provided the most pixels will represent the foreground class, and the remaining classes will be grouped up to represent the background class. Once the classes have been determined the foreground and background will be represented with the colors white and black, respectively.

To summarize, the implementation of the proposed method revolves around the use of Convolution Neural Networks (CNN) to create an Unsupervised Semantic Segmentation (USS) model to perform images segmentation on satellite images. This includes the use of backpropagation incorporated into the CNN to help increase the detection of features for the USS model [8]. This experiment consists of using an unsupervised learning model applied to the CNN in response to the small amounts of annotated satellite image data which limits the amount of training that can be completed. Following is the creation of the model which is described by the algorithm 2, that tells the different steps taken to complete the segmentation. This includes the creation of superpixels to center the clustering and allows for the better determination of classes and the use of foreground-background separation for additional support.

The last step involves creating the majority voting ensemble model which is used as a way to further the improvement of the model, by collecting data from multiple results and fusing those results together to form a better output.

CHAPTER 4

EXPERIMENTAL RESULTS

4.1 xDB Dataset

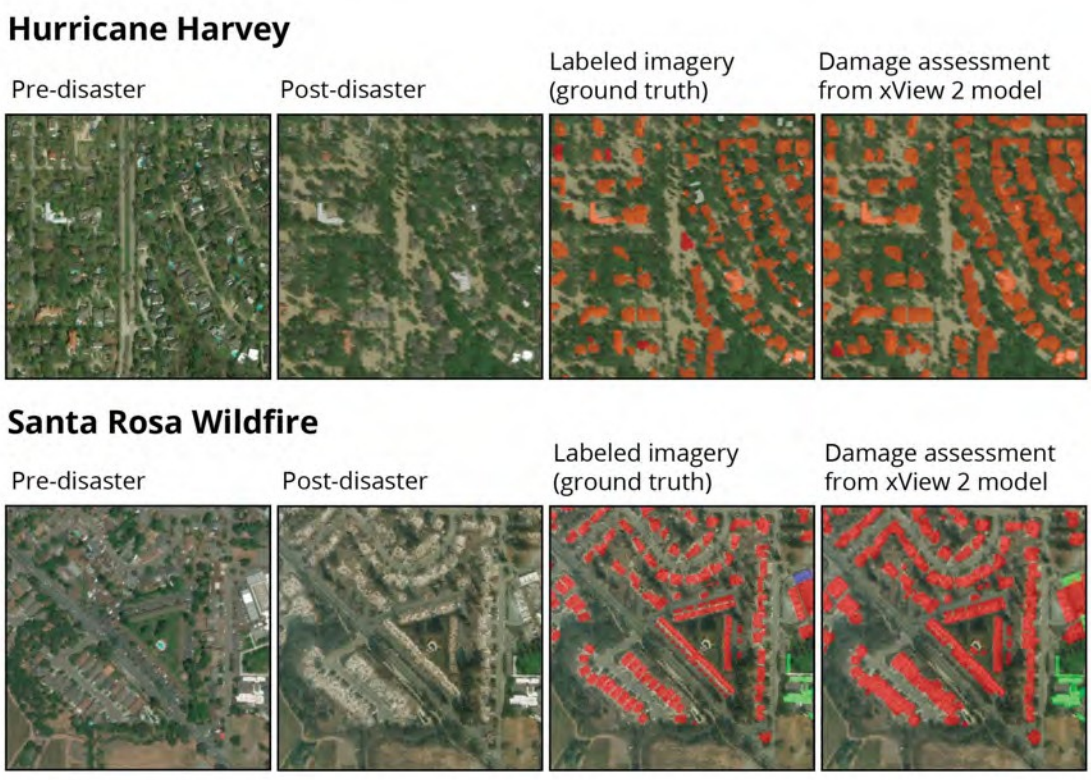


Fig. 4.1. Samples of [1] damage scale Image Segmentation for Xview 2

The Xview 2 Challenge, used a dataset created specifically for the competition by Software Engineering Institute (SEI) researchers of labeled satellite images. This is one of the largest, most comprehensive, and highest-quality public datasets of

annotated, high-resolution satellite imagery available online. It contains “before” and “after” satellite images from disasters around the world, such as wildfires, landslides, dam collapses, volcanic eruptions, earthquakes, tsunamis, storms, and floods [1]. By working with disaster response experts, SEI was able to create the joint damage scale that could accurately represent the real-world damage conditions. Almost every image included building outlines, bounding boxes, damage levels, and labels for environmental factors such as fire, water, and smoke.

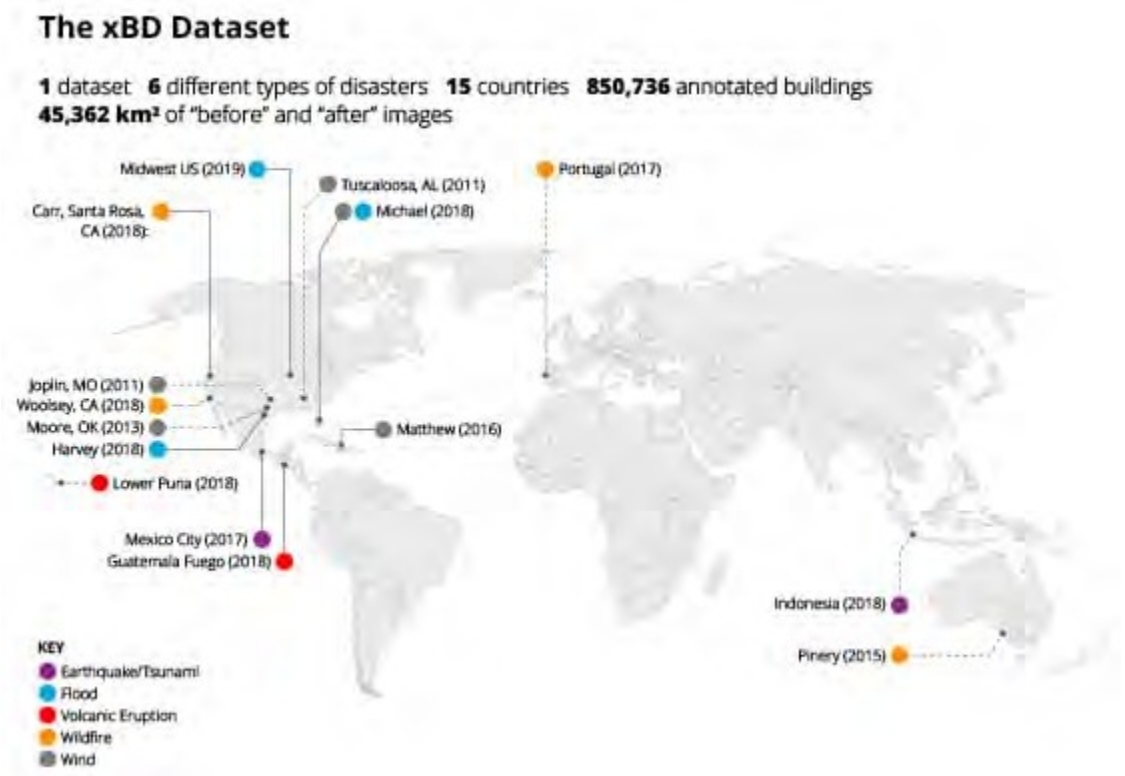


Fig. 4.2. World map disaster locations from [1] xBD Dataset

The xBD dataset contains nine folders that contain the “before” disaster satellite images and includes some of the images that include the building outlines on a

majority of images to help with evaluation of image segmentation [1]. These nine folders include images from Joplin, Portugal, Woosley, Tuscaloosa, Sunda, Puna, Pinery, Nepal, and Moore. After gaining an understanding of the different tools, software, and data used for the segmentation of satellite imagery form the xDB dataset we begin our evaluation of the different models used the experiments.

4.2 Evaluation Metrics

To produce results for the segmentation of the satellite images there are many different routes that could be used for this experiment. These different methods include Pixel Accuracy, Intersection-Over-Union (IOU), and Dice Coefficient. Pixel accuracy involves the use of comparing the segmented image to the ground truth however, there is a problem when it comes to using the pixel accuracy. Some disadvantages include the over-estimation that gives the results an inflated result as that does not provide and accurate depiction of the results. The goal of this experiment was to provide more accurate results to allow for better comparison with other segmentation techniques to see the improvements being implemented, which is why pixel accuracy will not be used. The Dice Coefficient is one of the most commonly used evaluation techniques used for semantic segmentation that provides straightforward metrics and is exceedingly accurate. Dice Coefficient involves the use of pixels, where the overlapping pixels are doubled and divided by the sum of the pixels from both images to produce its evaluation metrics.

Lastly, another popular choice for segmentation is Intersection-Over-Union (IOU) which includes determining the area of overlap between the segmentation and the ground truth divided by the area of union between the same images, as shown in equation 4.1 [11]. The current range used for IOU metric evaluation is 0 – 1 where zero represents no overlap and 1 represents the max overlap where there are no

discrepancies between the two images. Additionally, since this experiment is using a multi-class segmentation to produce more accurate results takes the Mean of IOU to produce the Mean Intersect over Union (MIOU) as seen in equation 4.2. When used in a controlled study the results from Dice Coefficient provided an almost identical metric score to Mean Intersect over Union (MIOU) to show that both methods can work for segmentation. The metric currently used for this experimentation is the MIOU evaluation. This is due to its ease of use and wide range of comparisons provided by other segmentation experiments and can be represented by the following equation.

$$\mathrm{IoU}(y_t, y_p) = \frac{\mathrm{I}}{y_t + y_p - \mathrm{I}}, \tag{4.1}$$

$$\mathrm{MIOU} = \frac{\mathrm{IoU}_1 + \ldots + \mathrm{IoU}_c}{\mathrm{c}}, \tag{4.2}$$

where I represents the intersecting elements between y_t, the truth values representing the ground truth and y_p, the prediction values produced by the results from the segmentation models. Then the values gained from the IoU calculations are placed into the MioU equation where c, represents the number of classes. For this experimentation the number of classes is set to 2 representing the foreground and background classes. This is also due to the way that the ground truth is an- notated and only shows the corresponding building while everything else is blacked out, this is similar to the representation produced after the foreground-background separation and an example is shown in Fig. 4.3.

(a) Original Image (b) Ground Truth

Fig. 4.3. An example of fore-ground back-ground separation from *Joplin 35* within the xDB dataset. In the ground truth image, the black color is for back-ground while the white color is for fore-ground.

(a) Original Image (b) Ground Truth

Fig. 4.4. An example of fore-ground back-ground separation from the folder *Joplin 40* within the xDB dataset. In the ground truth image, the black color is for back-ground while the white color is for fore-ground.

4.3 Experiment setup

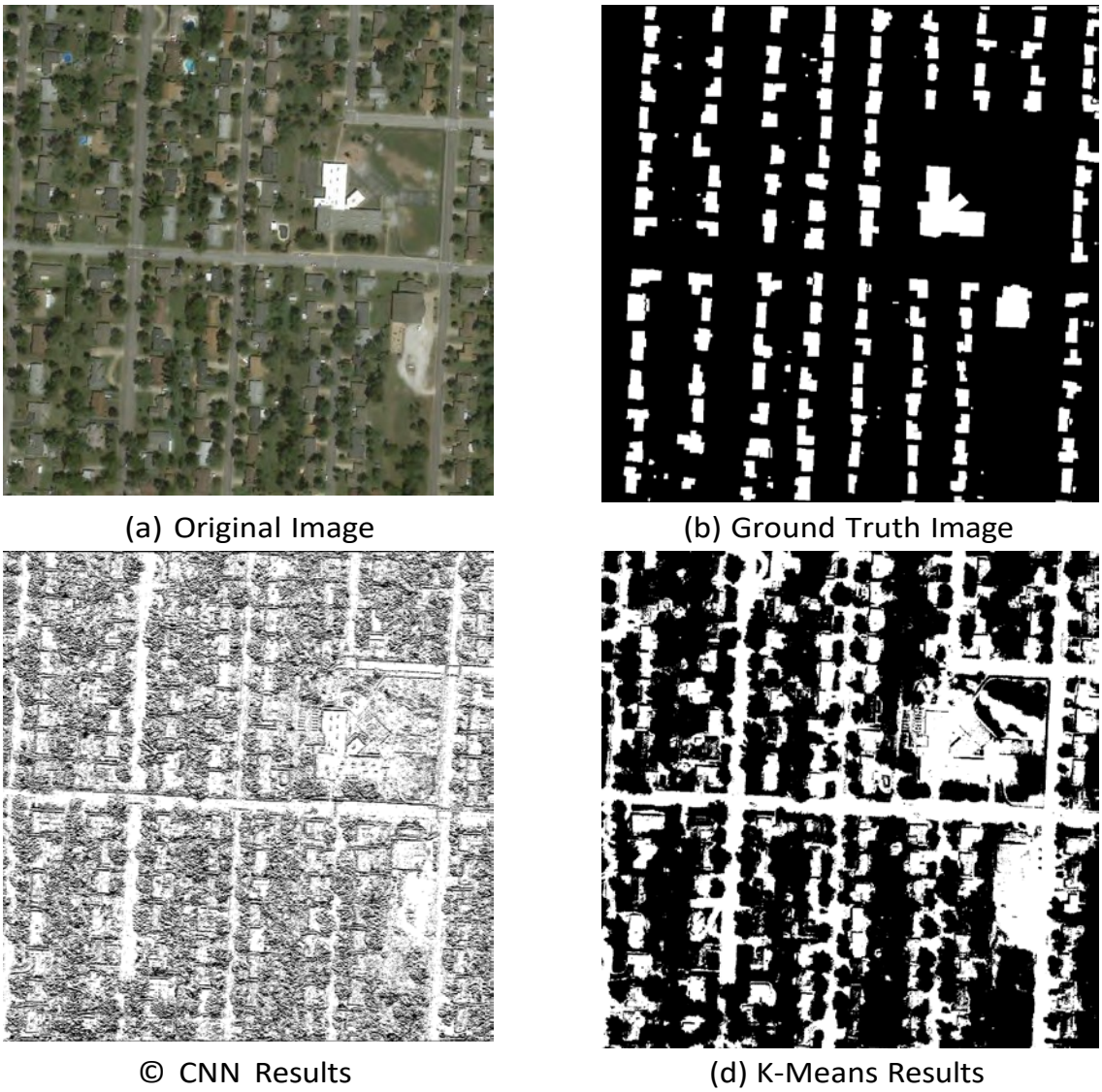
Throughout the process of this experiment multiple models were created to test and compare the results produced on some of the more popular unsupervised methods out there. The first model consisted of a base CNN model that was based off of a backpropagation type to help with the detection of elements within an image. This is one of the most common types of models used for image segmentation that has been used on Cityscapes and PASCAL datasets and has great flexibility. The next model is one of the most popular unsupervised image segmentation models and has been implemented in a plethora of studies, K-Means model. A majority of comparisons involving the use of an unsupervised model mostly make use of K-Means due to its accuracy and consistency. The third model is the USS model that incorporates the techniques created to improve the detection and segmentation of satellite images and to produce better results than some unsupervised models. Lastly, is the Ensemble model which creates multiple USS models with different parameters to combine based on majority voting 3.1.4. This technique can help to improve the results produced from the different USS models to output the best possible results possible.

4.4 Results & Discussions

This section presents and discusses experimental results and corresponding observations.

4.4.1 Performance comparison between baselines. Firstly, is producing a baseline result that can be used as a reference point for other models to determine if the model has improved by comparing results. The baseline model consists of the unsupervised CNN shown in Fig. 3.2. This produced a result that had an average MIOU value of 0.28. However, adding one more model to provide a comparison can

also help to provide a more accurate description of the performance that the model can provide. The additional model used is a popular type of unsupervised image segmentation model called K-Means. After testing the K-Means model had produce a better result than the unsupervised CNN model at an average MIOU value of 0.31. Fig. 4.5 presented visualization comparison between K-Means and unsupervised baseline (unsupervised CNN), which indicated that the unsupervised baseline outperformed K-Means on this case.



(a) Original Image (b) Ground Truth Image

© CNN Results (d) K-Means Results

Fig. 4.5. Comparing unsupervised CNN and K-Means on fore-ground and back-ground separation.

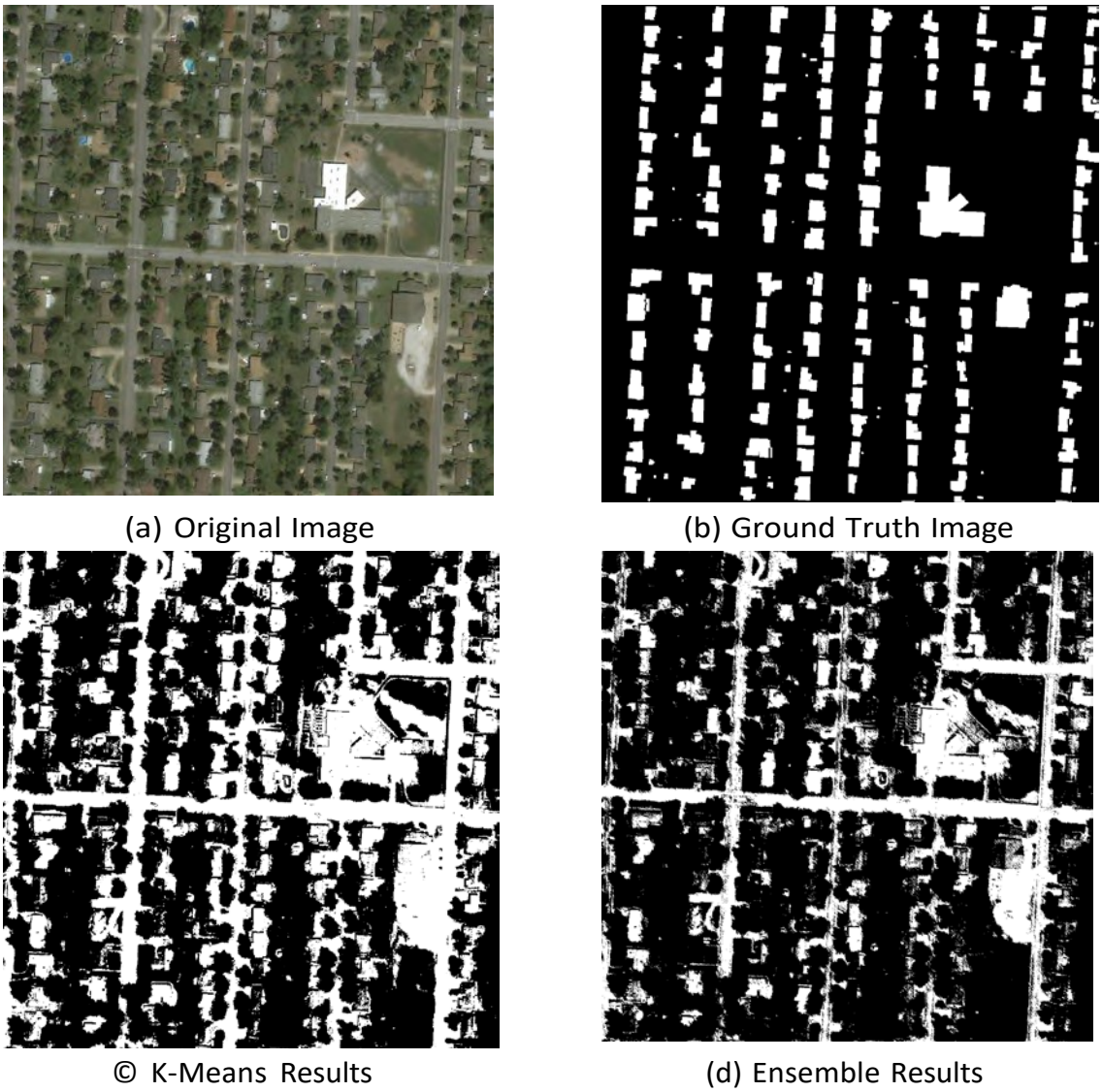
4.4.2 Performance comparison between baselines and proposed method.

After creating and testing of the baseline unsupervised CNN model and the k-means model, the next step is to begin testing on the USS model, produced by following the steps of the algorithm 2. Once the model was created, the test began to produce

an average MIOU value of 0.43. This shows that the USS model is an improved model that can produce a better MIOU result than that of the baseline CNN and K-Means models. Lastly, to further improve the results of the USS model the majority voting ensemble was employed. By creating the USS_k models that were introduced in Section 3.1.4, the results after testing produced an average MIOU increase of 0.20, which means that the proposed ensemble model can effectively improve performance for our task. The results from the ensemble can be seen by the Table below.

TABLE II
MIOU RESULTS FOR THE ENSEMBLE MODELS

Data	USS_1 Model	USS_2 Model	USS_3 Model	Ensemble(n=3)
Joplin	0.43	0.433	0.425	0.453
Moore	0.43	0.427	0.424	0.45
Nepal	0.437	0.437	0.438	0.474
Puna	0.438	0.438	0.438	0.478
Sunda	0.43	0.438	0.434	0.465
Tuscaloosa	0.434	0.435	0.434	0.467
Woolsey	0.44	0.436	0.44	0.477
Portugal	0.436	0.437	0.439	0.475
Pinery	0.438	0.438	0.437	0.477



(a) Original Image (b) Ground Truth Image

© K-Means Results (d) Ensemble Results

Fig. 4.6. K-Means & Ensemble visualization results for Joplin Tornado Image 35

Even from Fig. 4.6, there is a clear difference between the K-Means model and the ensemble model from their segmentation performance. However, there are also times when the ensemble can underperform due to it being heavily reliant on the different models, but this is only on a small percentage of the images. Now that all of the models have been tested all of the results can be seen in Table 4.1, that

Can show the results of the models when used on all nine folders from the xBD dataset.

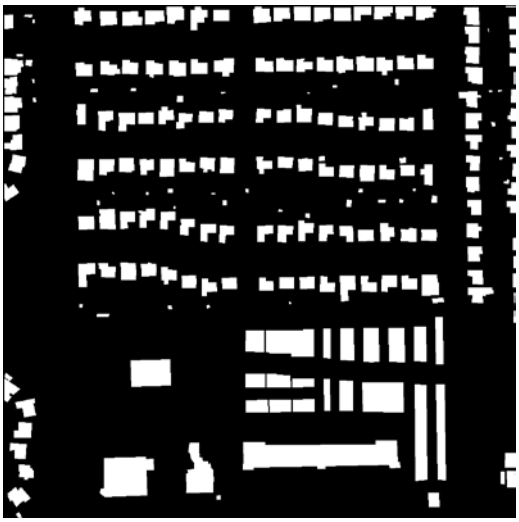
TABLE III
COMPARING MIOU RESULTS FOR THE 4 MODELS ON SATELLITE IMAGERY

Data	Base CNN	K-Means	USS	Ensemble
Joplin	0.286	0.314	0.43	0.453
Moore	0.284	0.316	0.43	0.45
Nepal	0.286	0.3	0.437	0.474
Puna	0.283	0.292	0.438	0.478
Sunda	0.269	0.366	0.43	0.465
Tuscaloosa	0.283	0.32	0.434	0.467
Woolsey	0.281	0.309	0.44	0.477
Portugal	0.277	0.312	0.436	0.475
Pinery	0.264	0.302	0.438	0.477

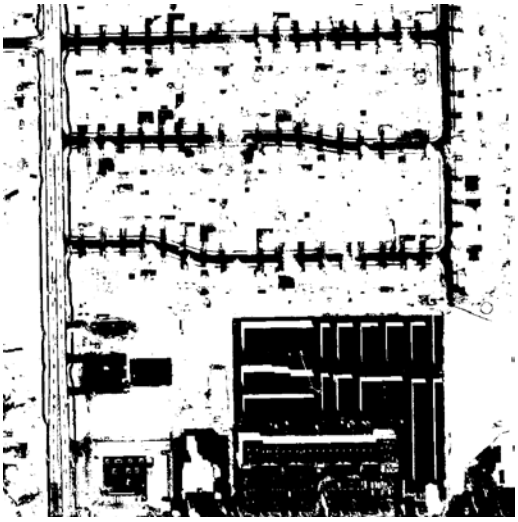
4.4.3 Visualization of fore-ground and back-ground separation across different folders.

The produced results from the remaining folders are as follows:

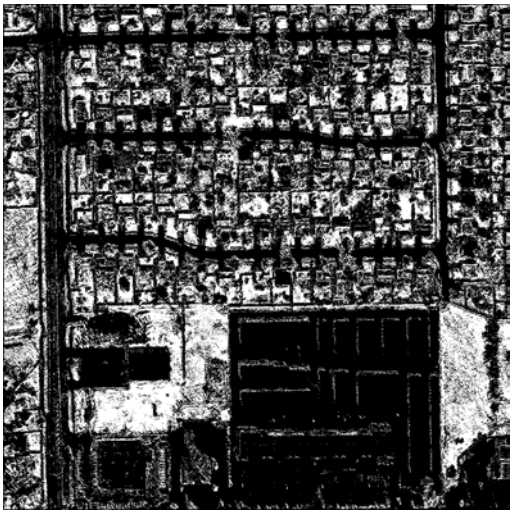
(a) Original Image



(b) Ground Truth Image



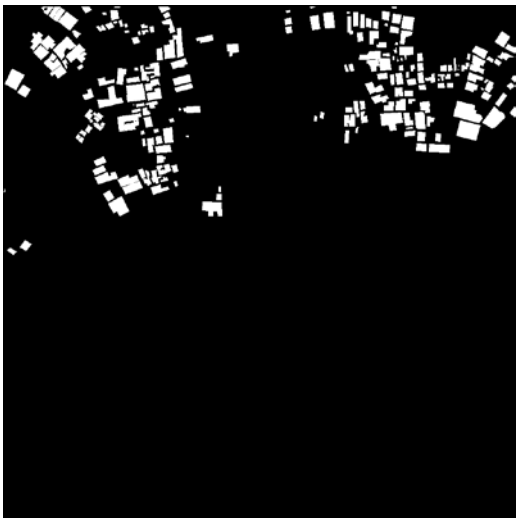
(c) K-Means Results



(d) Ensemble Results

Fig. 4.7. Moore Tornado Image 54 Results

(a) Original Image



(b) Ground Truth Image

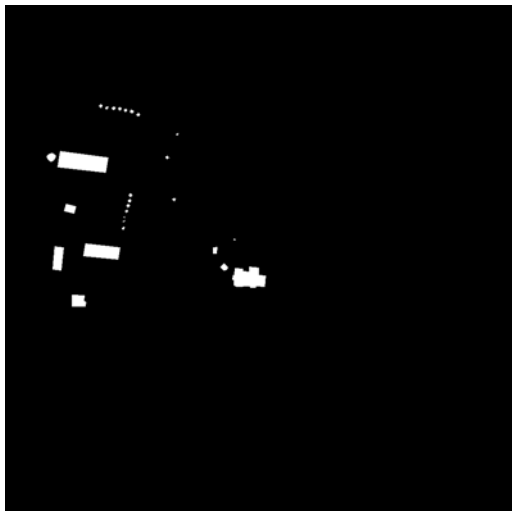
(c) K-Means Results

(d) Ensemble Results

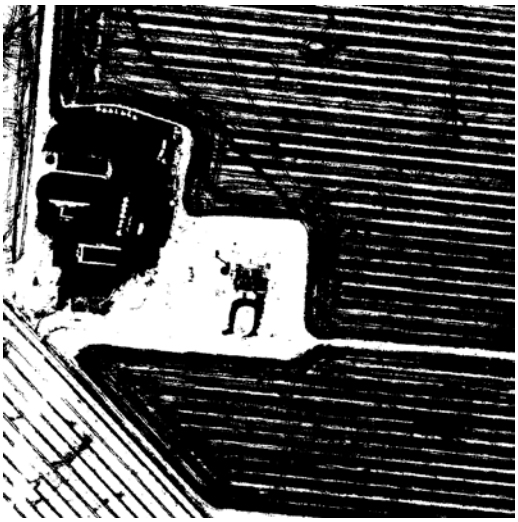
Fig. 4.8. Nepal Flooding Image 31 Results

Within the first two Figs. there are noticeable areas where the model struggled with its detection and did not produce the best results. Both examples in Figs. 4.7 and 4.8, there is a discontinuity with the elevation and greenery which had caused the model to struggle with the detection and did not produce the best results.

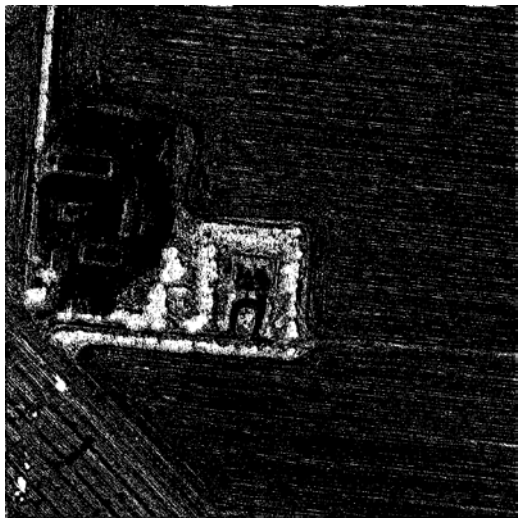
(a) Original Image



(b) Ground Truth Image



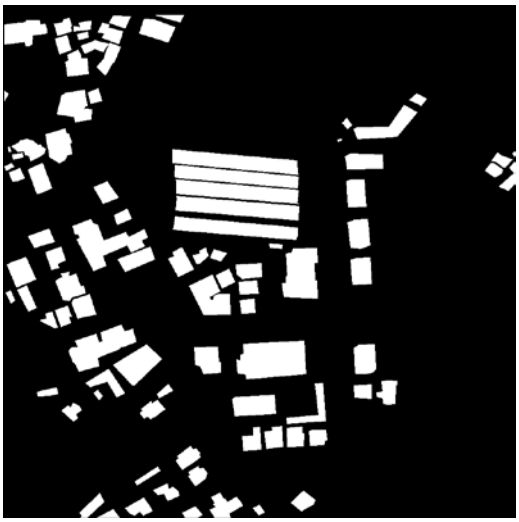
(c) K-Means Results



(d) Ensemble Results

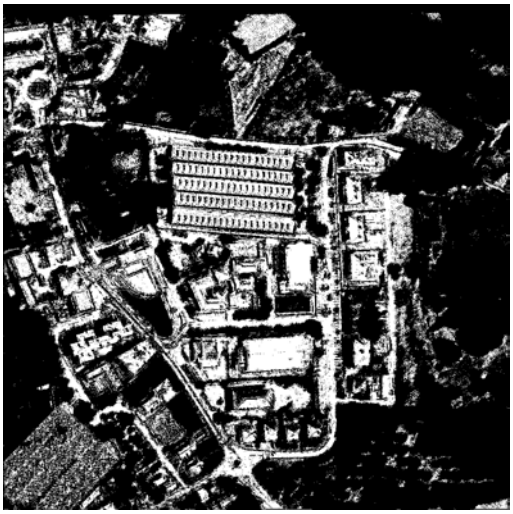
Fig. 4.9. Pinery Bushfire Image 782 Results

(a) Original Image



(b) Ground Truth Image

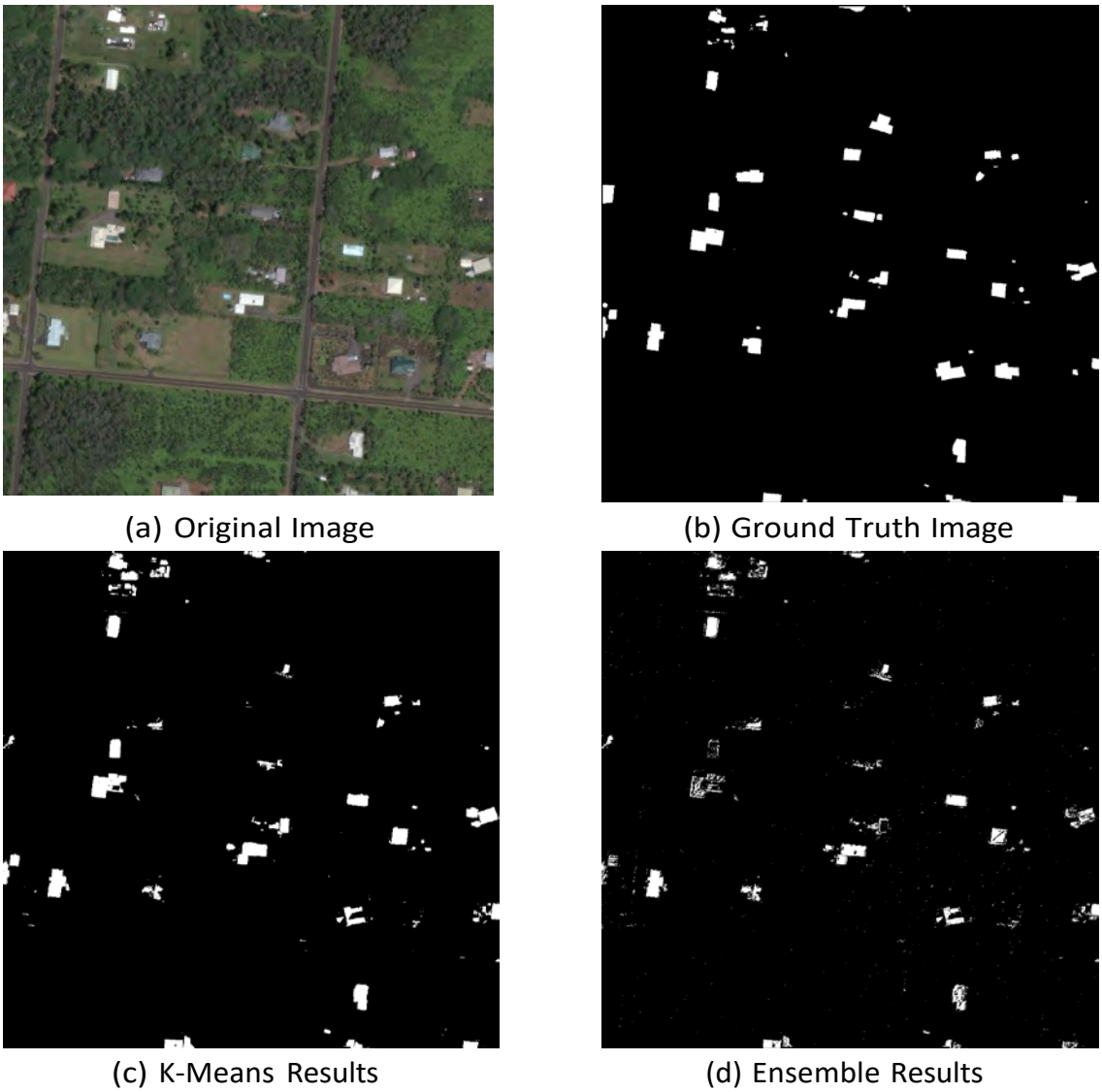
(c) K-Means Results



(d) Ensemble Results

Fig. 4.10. Portugal Wildfire Image 81 Results

Within Figs. 4.9 and 4.10, there are changes that occur to the images once implemented through the ensemble model. Because of the majority voting being implemented, additional buildings were either lost or added to the detection in the ensemble results. This shows that the ensemble segmentation can sometimes help to increase, but also sometimes can hurt the detection results.



(a) Original Image (b) Ground Truth Image

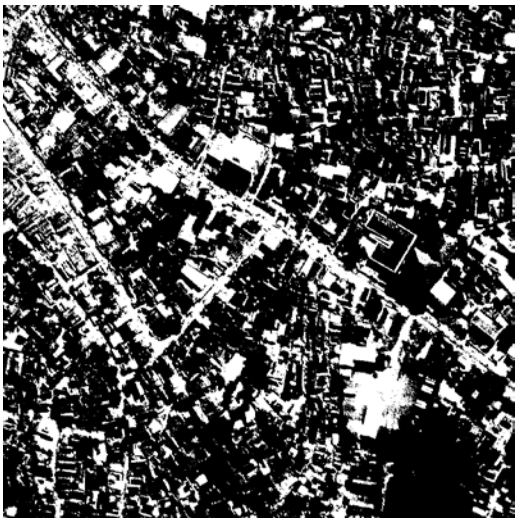
(c) K-Means Results (d) Ensemble Results

Fig. 4.11. Lower Puna Volcano Image 91 Results

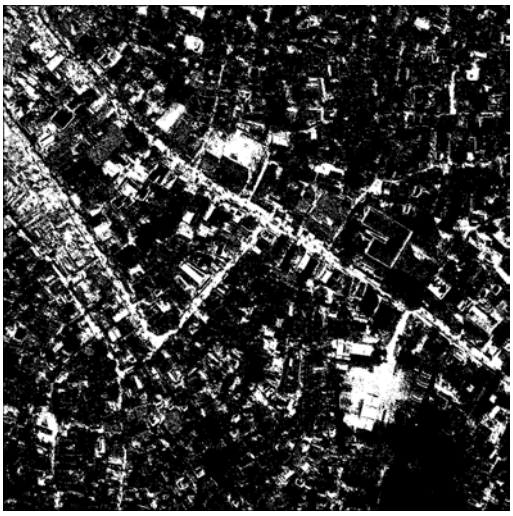
The segmentation produced in Fig. 4.11 is what shows more of the possibilities that the EUSS models can produce. Even when compared to the ground truth it is notable that some building have an almost perfect detection and this is due to the brighter coloring of the roofs which allow for an easier detection. This technique can even be incorporated in Fig. 4.12, as the model struggled with the cluster detection of the smaller buildings.

(a) Original Image

(b) Ground Truth Image



(c) K-Means Results



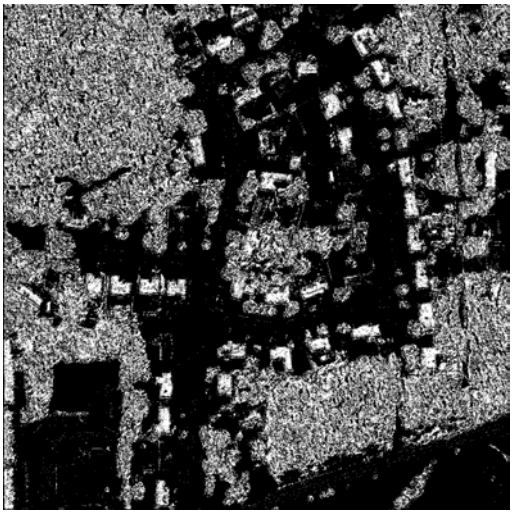
(d) Ensemble Results

Fig. 4.12. Sunda Tsunami Image 16 Results

(a) Original Image

(b) Ground Truth Image

(c) K-Means Results

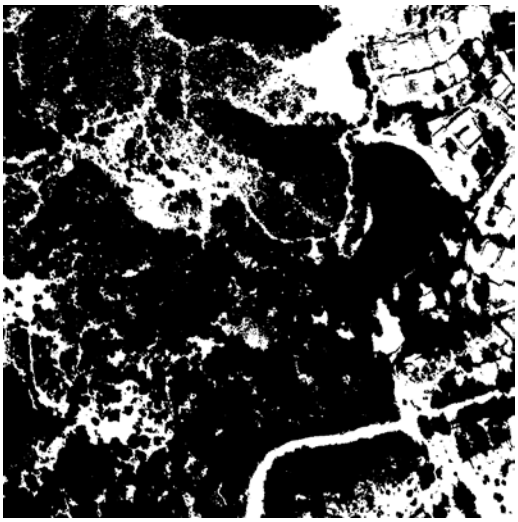


(d) Ensemble Results

Fig. 4.13. Tuscaloosa Tornado Image 80 Results

(a) Original Image

(b) Ground Truth Image



(c) K-Means Results

(d) Ensemble Results

Fig. 4.14. Woosley Fire Image 320 Results

The results shown from Fig. 4.13 and Fig. 4.14 produced some of the best visual results and begin to showcase what the goal of this research is striving to achieve. These images have a higher detection and produce a higher MIOU result of 0.57 and 0.53, respectively. This showcases that the path of this experiment is correct by producing results that more than double the accuracy of the base models results.

Although, the consistency could be further worked upon to be able to provide better results for satellite image segmentation compared to more popular unsupervised models, it is showcased that there is room for improvements for ensemble unsupervised segmentation's.

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Conclusion

Damage assessment using satellite images is a promising approach after natural disasters and it plays a critical role in disaster relief efforts. This requires performing image segmentation to identify objects of interests and separate the target objects (foreground) from background. However, there exists very limited annotated satellite images in practice. Thus, machine learning models using supervised learning could not be applied for image segmentation in satellite images. Furthermore, the size of the satellite image is very large while the contents are very small and represented by a few pixels such as in XView 2 dataset.

In order to address this challenge, a novel ensemble unsupervised semantic segmentation method was proposed for image segmentation on satellite images. Specifically, an unsupervised semantic segmentation model was proposed to implement foreground-background separation and then be placed within an ensemble model to increase the prediction accuracy further. Experimental results demonstrated that the proposed method outperformed baseline models such as k-means on a satellite image benchmark, the XView2 dataset. The experiments produced results that showcased the versatility and flexibility of the ensemble unsupervised semantic segmentation. The improvements from the base CNN model to the final MIOU results of the ensemble USS model showcased that there can still be improvements made to produce an even greater result that can possibly challenge supervised segmentation models. The experiments also demonstrated the future possibility for unsupervised

segmentation, by reducing the amount of time it takes to run seg- mentation and also reducing the amount of annotations that need to be made to images prior to running a segmentation model. This can allow for more testing to be accomplished in a shorter amount of time and even show cases a steady result throughout the different folders from the xBD dataset as shown in Section 4.4.3.

In sum, the proposed approach provides a promising solution to semantic segmentation in images that will benefit many mission critical applications such as disaster relief using satellite imagery analytics. This research could potentially save lives and benefits the entire society.

5.2 Future Work

There are some areas that could be further improved upon to produce an even greater result. The first is splitting of the images. This is because the large images provided by the satellite makes it hard to determine smaller building or vehicles. Moreover, it would be interesting to implement the opposite approach of that in article [21] and try to reduce the detection of roads to better focus on the detection of large buildings.

REFERENCES

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

[2] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International Journal of Computer Vision*, vol. 88, pp. 303–338, June 2010.

[3] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollar, "Panoptic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[4] I. Demir, K. Koperski, D. Lindenbaum, G. Pang, J. Huang, S. Basu, F. Hughes, D. Tuia, and R. Raskar, "Deepglobe 2018: A challenge to parse the earth through satellite images," 06 2018.

[5] S. S. Seferbekov, V. I. Iglovikov, A. V. Buslaev, and A. A. Shvets, "Feature pyramid network for multi-class land segmentation," *CoRR*, vol. abs/1806.03510, 2018.

[6] Y. Gandelsman, A. Shocher, and M. Irani, ""double-dip": Unsupervised image decomposition via coupled deep-image-priors," *CoRR*, vol. abs/1812.00467, 2018.

[7] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[8] A. Kanezaki, “Unsupervised image segmentation by backpropagation,” in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1543–1547, 2018.

[9] L. U. Ambata and E. P. Dadios, “Foreground background separation and tracking,” in *2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*, pp. 1–6, 2019.

[10] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopou- los, “Image segmentation using deep learning: A survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3523–3542, 2022.

[11] A. Ahmadzadeh, D. J. Kempton, Y. Chen, and R. A. Angryk, “Multiscale iou: A metric for evaluation of salient object detection with fine structures,” in *2021 IEEE International Conference on Image Processing (ICIP)*, pp. 684–688, 2021.

[12] A. M. Hafiz and G. M. Bhat, “A survey on instance segmentation: state of the art,” *International journal of multimedia information retrieval*, vol. 9, no. 3, pp. 171–189, 2020.

[13] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, “A review on deep learning techniques applied to semantic segmentation,” *arXiv preprint arXiv:1704.06857*, 2017.

[14] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollár, “Panoptic segmentation,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9404–9413, 2019.

[15] G. Papandreou, L.-C. Chen, K. P. Murphy, and A. L. Yuille, "Weakly- and semi-supervised learning of a deep convolutional network for semantic image segmentation," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 2015.

[16] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[17] Z. Niu, W. Liu, J. Zhao, and G. Jiang, "Deeplab-based spatial feature ex- traction for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 2, pp. 251–255, 2019.

[18] B. Cheng, M. D. Collins, Y. Zhu, T. Liu, T. S. Huang, H. Adam, and L. Chen, "Panoptic-deeplab: A simple, strong, and fast baseline for bottom-up panoptic segmentation," *CoRR*, vol. abs/1911.10194, 2019.

[19] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2018.

[20] M. Weber, H. Wang, S. Qiao, J. Xie, M. D. Collins, Y. Zhu, L. Yuan, D. Kim, Q. Yu, D. Cremers, *et al.*, "Deeplab2: A tensorflow library for deep labeling," *arXiv preprint arXiv:2106.09748*, 2021.

[21] N. Subraja and D. Venkatasekhar, "Satellite image segmentation using modified u-net convolutional networks," in *2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*, pp. 1706–1713, 2022.

[22] T. Li, M. Comer, and J. Zerubia, “Feature extraction and tracking of cnn segmentations for improved road detection from satellite imagery,” in *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 2641–2645, 2019.

[23] T. Sai Krishna and A. Y. Babu, “Three phase segmentation algorithm for high resolution satellite images,” in *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, pp. 2217–2223, 2016.

[24] M. Pogudin and E. Medvedeva, “Two-stage algorithm for segmentation of satellite images,” in *2022 24th International Conference on Digital Signal Processing and its Applications (DSPA)*, pp. 1–4, 2022.

[25] R. M. Esteves, T. Hacker, and C. Rong, “Competitive k-means, a new accurate and distributed k-means algorithm for large datasets,” in *2013 IEEE 5th International Conference on Cloud Computing Technology and Science*, vol. 1, pp. 17–24, 2013.

[26] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopou- los, “Image segmentation using deep learning: A survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3523–3542, 2022.

[27] Z. Yang, X. Peng, Z. Yin, and Z. Yang, “Deeplab v3 plus-net for image seman- tic segmentation with channel compression,” in *2020 IEEE 20th International Conference on Communication Technology (ICCT)*, pp. 1320–1324, 2020.

[28] V. Varatharasan, H. Shin, A. Tsourdos, and N. Colosimo, “Improving learning effectiveness for object detection and classification in cluttered backgrounds,” *CoRR*, vol. abs/2002.12467, 2020.

[29] B. E. Moore, C. Gao, and R. R. Nadakuditi, “Panoramic robust pca for foreground-background separation on noisy, free-motion camera video,” 2017.

[30] Y. Wang, H. Wei, X. Ding, and J. Tao, “Video background/foreground separation model based on non-convex rank approximation rpca and superpixel motion detection,” *IEEE Access*, vol. 8, pp. 157493–157503, 2020.

[31] S. Pei, L. Li, L. Ye, and Y. Dong, “A tensor foreground-background separation algorithm based on dynamic dictionary update and active contour detection,” *IEEE Access*, vol. 8, pp. 88259–88272, 2020.

[32] L. U. Ambata and E. P. Dadios, “Foreground background separation using genetic algorithm,” in *2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*, pp. 1–5, 2019.

[33] D. Aksu and M. Ali Aydin, “Detecting port scan attempts with comparative analysis of deep learning and support vector machine algorithms,” in *2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT)*, pp. 77–80, 2018.

[34] N. Jin and D. Liu, “Wavelet basis function neural networks for sequential learning,” *IEEE Transactions on Neural Networks*, vol. 19, no. 3, pp. 523–528, 2008.

[35] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

[36] B. Oh and J. Lee, “A case study on scene recognition using an ensemble convolution neural network,” in *2018 20th International Conference on Advanced Communication Technology (ICACT)*, pp. 351–353, 2018.

[37] G. Şahin and O. Susuz, “Encoder-decoder convolutional neural network based iris-sclera segmentation,” in *2019 27th Signal Processing and Communications Applications Conference (SIU)*, pp. 1–4, 2019.

[38] G. Chen, C. He, T. Wang, K. Zhu, P. Liao, and X. Zhang, “A superpixel-guided unsupervised fast semantic segmentation method of remote sensing images,” *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[39] H. Zhang, C. Wu, L. Zhang, and H. Zheng, “A novel centroid update approach for clustering-based superpixel methods and superpixel-based edge detection,” in *2020 IEEE International Conference on Image Processing (ICIP)*, pp. 693–697, 2020.

[40] W. Kim, A. Kanezaki, and M. Tanaka, “Unsupervised learning of image segmentation based on differentiable feature clustering,” *CoRR*, vol. abs/2007.09990, 2020.

[41] H. Temiz, “An experimental study on hyper parameters for training deep convolutional networks,” in *2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, pp. 1–8, 2020.

CURRICULUM VITA

Jaelen Tarry

Department of Electrical and Computer Engineering
Roy G. Perry College of Engineering
Prairie View A&M University, Texas
Email: jtarry22@gmail.com
LinkedIn: https://www.linkedin.com/in/jtarry/

Education

M.S. Electrical Engineering, Prairie View A&M University, 2023.

B.Sc. Computer Engineering, Prairie View A&M University, 2020.

Professional Experience

Prairie View A&M University, Graduate Research Assistant, 2021–2023.

Prairie View A&M University, Dean Assistant, Summer 2020–2021.

SAIC, Huntsvill, Alabama, Software Development Intern, Summer 2018–2019.

Award & Recognition

IBM Masters Fellowship Award, 2022