Reconnaissance using image segmentation

Samaikhya Choppadandi (2145011), Hemanth Reddy Musipatla (2149799), Yashwanth Gurram (2145549), Sai Krishna Reddy Koppula (2151567)

Abstract

Space technologies can give operational advantage over potential adversaries. In predominant areas like defense and science, advance space engineering requires an automatic feature labelling system to accurately classify features in overhead imagery which helps in effective satellite intelligence and surveillance from space. Using computer vision, we can also orchestrate responses to future threats and opportunities with enhanced and affordable defense and security capability. The motive of our paper is to understand multiclass segmentation of satellite images for the surveillance of the area using convolutional networks like U-Net and Signet. These architectures are well known for their simple structures and fast learning. The dataset is prepared for training by generating masks and co-ordinates. The training dataset is also augmented for increasing the samples for training. These models are trained on the predefined classes of objects and were successfully able to segment the objects on the images.

I. INTRODUCTION

Satellite imagery has drastically improved our understanding of our planet. The interest in the object detection of satellite imagery is rapidly increasing today because of the recent improvements in computer vision and deep learning. In many applications, such as creating and keeping up-to-date maps, improving urban planning, environmental monitoring, and disaster relief, the ability to distinguish different types of objects in aerial images would be exceptionally helpful. In addition to the practical need for accurate aerial image interpretation systems, this domain presents a number of scientific challenges to computer vision. We examine the challenges due to the small dataset, the specific characteristics of the data, and supervised and unsupervised machine learning algorithms that are appropriate for this application.

Military personnel investigating a location in order to gather information about enemy forces, topography, and other actions using satellite view is called Reconnaissance. Military leaders receive images of enemy units and other intelligence from reconnaissance satellites. There are several major types of reconnaissance satellite. Missile early warning, nuclear explosion detection, Optical imaging surveillance and Radar imaging surveillance.

Detecting target patches in satellite pictures, such as airplanes, tankers, artillery, and other vehicles, is crucial in military applications that require accurate vehicle identification and tracking, such as surveillance and security. The presence of objects in various sizes, orientations, and extremely close positions are among the key obstacles in detecting targets in satellite images. Recognition systems for satellite imagery that can swiftly detect and categorize objects or targets in satellite photos. Automatic recognition of military objects in high-resolution satellite images, such as oil tankers, aircraft, and artillery, is critical in military applications.

II. LITERATURE REVIEW

The related work on satellite images and their classification is as below.

[1] The paper focuses on adaptation of fully convolutional neural network for multispectral data processing. The motive of this paper is to train the images with different augmentation techniques like boundary effect estimation. It doesn't rely on complex ensembling techniques and thus can be easily scaled for deployment.

[2] Images obtained from the Landsat -8 and Planet Scope satellites are used for estimation of automatic object detection quality. Three modification of convolutional neural network architecture for implementing the recognition algorithm was used. To analyze the accuracy of the object detection algorithm, the selected regions were compared with the areas by previously marked by experts. An important result of the study was the improvement of the detector for the class “Forest”.

[3] The results of two convolutional neural networks for building detection on satellite images of Planet database. To analyze the quality of developed algorithms, there was used Sorensen-Dice coefficient of similarity which compares results of algorithms with tagged masks. The masks were generated from json files and sliced on smaller parts together with respective images before the training of algorithms.

III. METHODOLOGY

The workflow of our project involves steps from Data preparation to modelling of the algorithm.

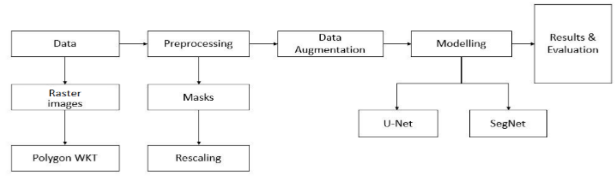


Fig. 1 Project Workflow

A. DATASET DETAILS

The dataset consists of 25 SAR images, with two versions of each image. i.e., 3-band, 16-band images covering 1km X 1km area. The 3-band images are the traditional RGB natural color images. The 16-band images contain spectral information by capturing wider wavelength channels. This multi-band imagery is taken from the multispectral (400 – 1040nm) and short-wave infrared (SWIR) (1195-2365nm) range. For reference, 3-channel – RGB band (3348 x 3403), 1-channel P-Band (3348 x 3403), 8 – channel M- Band (837 x 851) and 8 – channel A-Band (134 x 137). The dataset also provides file of Xmin, Ymax coordinates of the image and file containing image classes along with their respective multi polygon.

Fig. 2. Image samples in dataset

B. DATA AUGMENTATION

Data augmentation is a technique used to increase the diversity of training set by applying random at the same time realistic transformations, such as image rotation. Because of the small dataset, the data samples were augmented by randomly cropping and flipping the images before sending it to the model. Data augmentation may greatly improve data quality by reducing the amount of manual intervention which is necessary to produce useful information. The number of samples seen by the neural network model depends on the batch size and number of epochs during train, validation and test.

C. DATA PRE-PROCESSING

To tackle the problem of changing image resolutions and image misalignment, a series of image pre-processing procedures are used. The RGB images were normalized in the range [0, 1] and scale percentile processing was used to improve visual perception by filtering out the lower 1 and higher 99 percentile pixel values. The low-resolution RGB images are up-scaled to match the high-resolution RGB photos in resolution. By data processing, the upscaling process for the low-resolution A-band photos produced visually good results. The high-resolution RGB image is utilized as a reference for image alignment and registration.

D. ARCHITECTURES

U-Net Architecture:

U-Net is one of the standard CNN architectures for image segmentation tasks. It is considered the best network for fast and precise segmentation of images. The network is based on fully convolutional layers whose architecture was modified and extended to work with fewer training images and yield more precise segmentation.

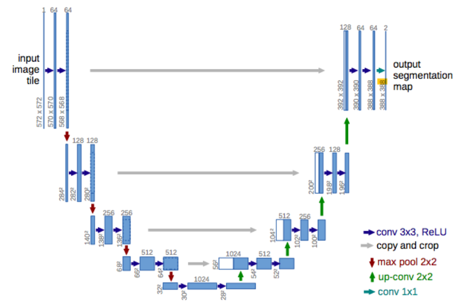


Fig. 3. U-Net Architecture

In general, U-Net architecture consists of encoder and decoder paths. The encoder path follows the typical convolution neural network architecture. We use batch normalization for convergence acceleration during training. In addition, we use rectified linear unit(ReLU) as the primary activation function, which helps overcome the vanishing gradient problem and allows models to learn faster and perform better. The number of feature channels is doubled at each down sampling step. The decoder path consists of the up sampling operation of the feature map followed by convolution with half number of feature channels, concatenation with the corresponding feature map from the encoder path, also followed by batch normalization and Sigmoid.

SegNet Architecture:

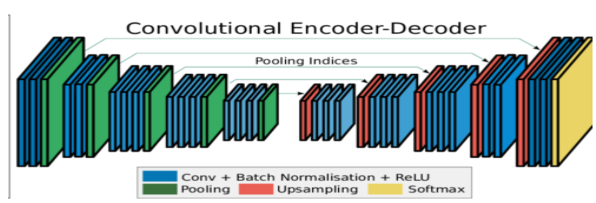


Fig. 4. SegNet Architecture

SegNet has an encoder, a decoder, and a final pixel classification layer. The architecture of this model is shown above. The encoder consists of 13 convolutional layers, 13 batch normalization, 13 ReLU activation functions, five max-pooling layers for down sampling, and five up sampling functions. The weights of this network initialize SegNet. Each layer of the encoder has a corresponding layer in the decoder. So, the decoder consists of the same layers as an encoder, except for max-pooling functions, which were exchanged for the same number of upsampling operations. The final output layer is a multiclass softmax classifier, which helps to predict each pixel's class probabilities independently.

Adam Optimizer:

Adaptive Moment Estimation is a technique for optimizing gradient descent algorithms. When working with problems involving a lot of data or parameters, the method is quite efficient and takes minimal memory. It's a hybrid of gradient descent with momentum and the RMSP algorithms.

Gradient descent with momentum:

This approach uses the 'exponentially weighted average' of the gradients to speed up the gradient descent algorithm. The technique converges faster to the minima when averages are used.

$$w_{t+1} = w_t - \alpha m_t$$

$$m_t = \beta m_{t-1} + (1 - \beta) \left[\frac{\delta L}{\delta w_t} \right]$$

mt = aggregate of gradients at time t [current] (initially, mt = 0)

mt-1 = aggregate of gradients at time t-1 [previous]

Wt = weights at time t

Wt+1 = weights at time t+1

αt = learning rate at time t

∂L = derivative of Loss Function

∂Wt = derivative of weights at time t

β = Moving average parameter (const, 0.9)

Binary Cross-Entropy:

In binary cross entropy, each of the predicted probabilities is compared to the actual output, which can either be 0 or 1. This score penalizes the probabilities according to their distance from the expected value. It describes the distance between the actual value and the expected value.

Binary Cross-Entropy is represented as:

$$L_{BCE} = -\frac{1}{n}\sum_{i=1}^{n}(Y_i \cdot \log \hat{Y}_i + (1 - Y_i) \cdot \log(1 - \hat{Y}_i))$$

E. EXPERIMENTAL RESULTS

The proposed algorithms are trained with the prepared datasets. The learning rate of 0.0001 was applied for both the algorithms, with 15 epochs each. The outputs generated were considerable as the maximum classes were forest and manmade structures.

U-NET OUTPUTS:

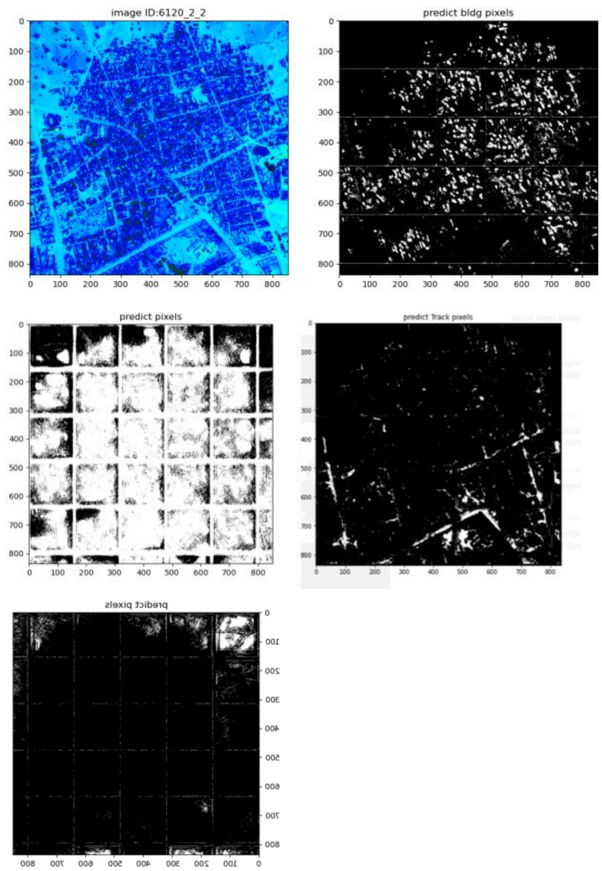


Fig. 5. Image segmentation outputs using U-Net

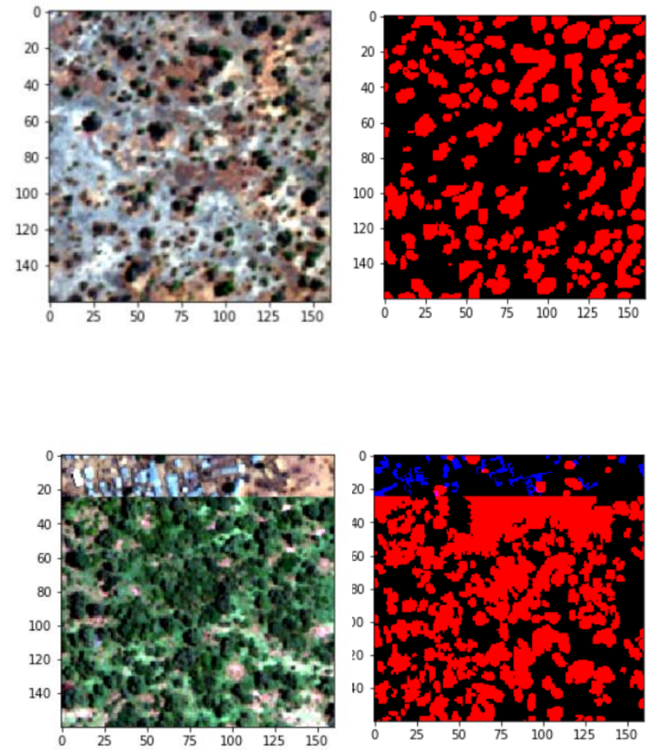
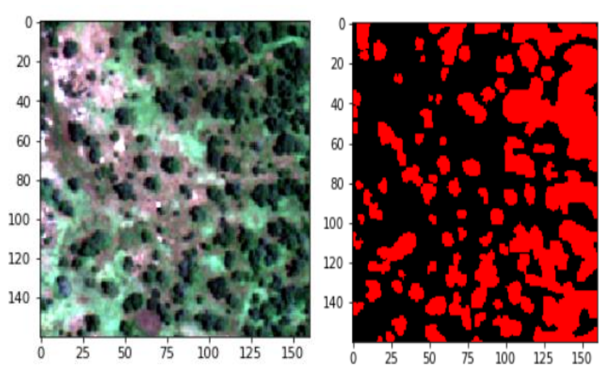


Fig. 6. Image segmentation outputs using segNet

SEGNET OUTPUTS:



IV. CONCLUSION

Given the small size of the dataset and the quantity of preprocessing required for the input photographs, this was a challenging project to work on. This necessitated data enhancement of the high-resolution photos in the form of random cropping and flipping. The process has been brute force for channeling the data and training models. The architectures worked well for huge objects but we need train on more models to check for smaller object segmentation.

REFERENCES

1. Vladimir Iglovikov, Sergey Mushinskiy, Vladimir Osin, Satellite Imagery Feature Detection using Deep Convolutional Neural Network, Kaggle Competition, 2018.

2. Comparison of Different Convolutional Neural Network Architectures for Satellite Image Segmentation, 2018 23rd Conference of Open Innovations Association (FRUCT)

3. Building Detection on Aerial Images Using U-NET Neural Networks, 2019 24th Conference of Open Innovations Association (FRUCT)

4. Dstl Satellite Imagery Feature Detection.https://www.kaggle.com/c/dstl-satellite-imagery-feature-detection/data

5. Dstl Satellite Imagery Competition, 1st Place Winner's Interview: Kyle Lee.http://blog.kaggle.com/2017/04/26/dstl-satellite-imagery-competition-1 st-place-winners-interview-kyle-lee