**Extraction of Water Network from Satellite Image**

**Abstract**

The extraction of water bodies from satellite imagery has been widely explored in the recent past. Several approaches have been developed to delineate water bodies from different satellite imagery varying in spatial, spectral, and temporal characteristics. The approach we have taken for extraction is with Unet and Tensorflow i.e deep learning model. Visual interpretation of satellite data provides the best delineation of water bodies of varied sizes but is time-consuming, especially when working with high-resolution data.

Mapping of natural resources like forests and water bodies using satellite imagery has gained much importance in the recent past. Both forest and water resources are subject to intense exploitation and monitoring them at regular intervals is imperative for their sustainable management. Water bodies, which play a key role in the global carbon cycle and climate variations, are mapped in the spatiotemporal domain to analyze and assess the extent and rate of their degradation and disappearance. Geospatial tools are proving to be advantageous for such impact assessment for the implementation of conservation measures.

**The project is divided into the following steps:**

**i.Data Collection**

A collection of water bodies images captured by the Sentinel-2 Satellite. Each image comes with a black and white mask where white represents water and black represents something else but water. The masks were generated by calculating the NWDI (Normalized Water Difference Index) which is frequently used to detect and measure vegetation in satellite images, but a greater threshold was used to detect water bodies.

Two different directories contain 2269 images each where one represents original satellite images and the other masks generated images.

**ii. Data Exploration**

Before diving into the training model the most important step is to know about the data. So that we can know what things need to be done to get accurate results. But the number of data is a lot so to avoid going through each data we will take some images randomly and see how their original image and masked images are.

Below you can see the result of the images.

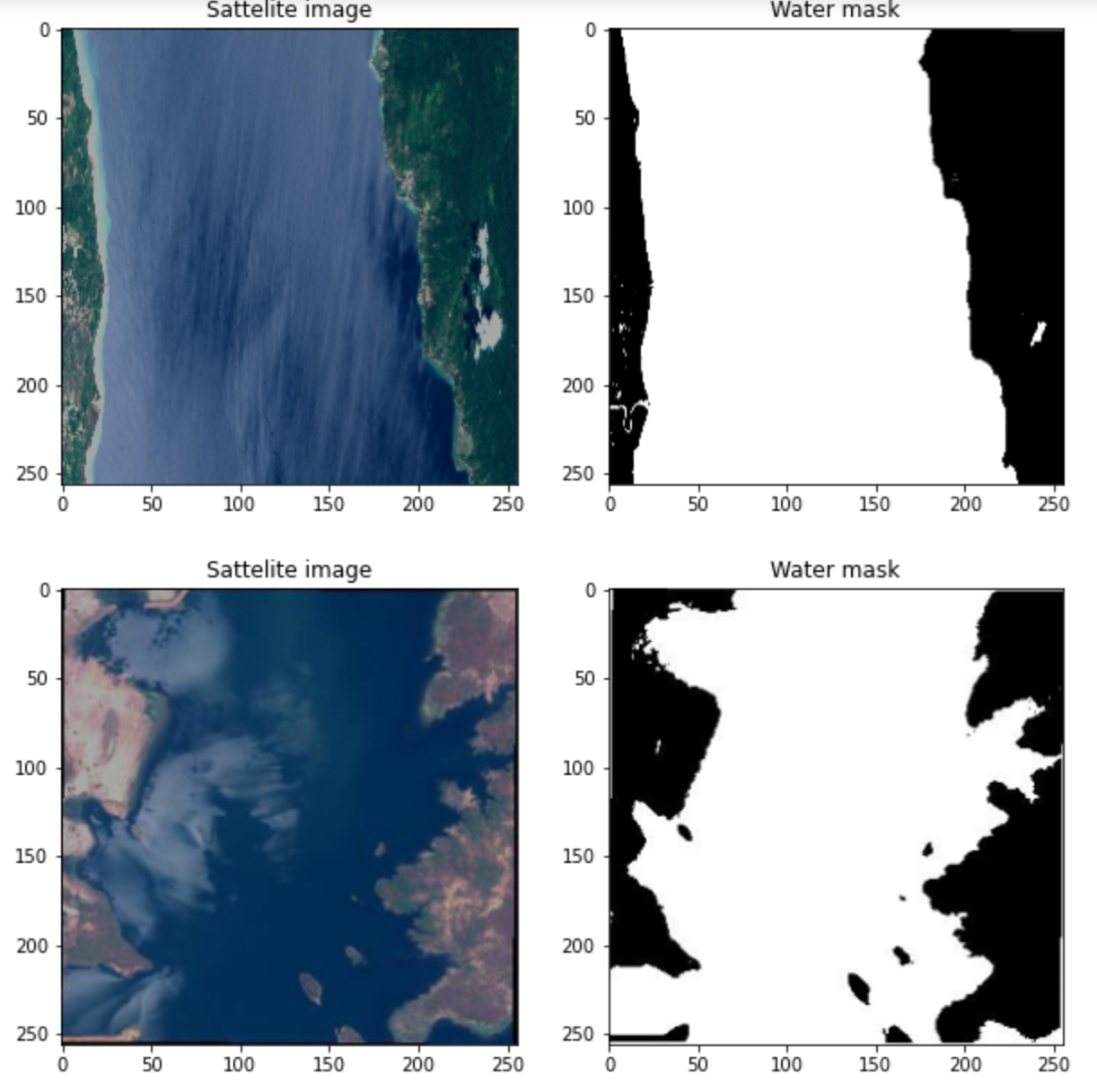
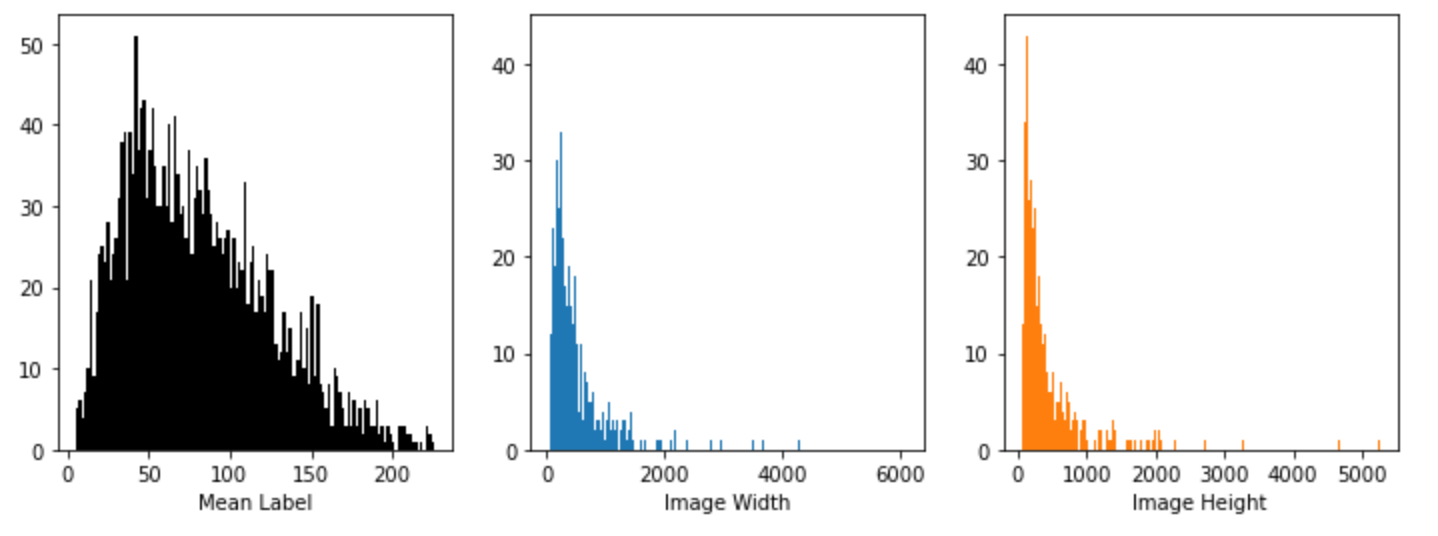


Fig: Visualization of satellite images and its masked region

The masked images are quite accurate but we cannot depend upon them to make sure whether we can rely on the data so we will measure the data quality.

It seems that there is no completely black or completely white images data in the masked dataset, hence we aren't required to perform filtering. These cases seem to be labeled right in many cases. Hence we are good with the dataset.

**iii. Data Augmentation**

The quantity and diversity of training data determine the success of most machine learning models and deep learning models in particular. However, one of the most prevalent problems in applying machine learning in the organization is a lack of data. This is due to the fact that gathering such information can be costly and time-consuming in many circumstances.

Here in our case, we have just 2269 masked and 2269 satellite image datasets. As we are using a Deep learning algorithm to train our model, it is insufficient. Hence we are using the data augmentation technique to increase the number of datasets for our project. Let's start with the simple introduction:

Data augmentation is a set of techniques for producing additional data points from current data in order to artificially increase the amount of data available. Making modest adjustments to data or utilizing deep learning models to produce additional data points are examples of this.

## **iii.a. Why is it important now?**

Machine learning applications, particularly in the deep learning domain, are rapidly diversifying and expanding. Techniques for data augmentation could be useful in combating the issues that the artificial intelligence sector faces.

By creating fresh and varied instances to train datasets, data augmentation can help improve the performance and results of machine learning models. A machine learning model performs better and is more accurate when the dataset is rich and sufficient.

## **iii.b. How is it done?**

For data augmentation, making simple alterations to visual data is popular. In addition, generative adversarial networks (GANs) are used to create new synthetic data. Classic image processing activities for data augmentation are:

1. padding
2. random rotating
3. re-scaling,
4. vertical and horizontal flipping
5. translation ( image is moved along X, Y direction)
6. cropping
7. zooming
8. darkening & brightening/color modification
9. grayscaling
10. changing contrast
11. adding noise
12. random erasing

## 

## **iii.c.How is it done in our Project?**

For our project, we are using rotation techniques to increase the size of the dataset. Using this method, the dataset can be increased from 2x to 4x. We are using rotation angles of 90, 180, 270. Hence the dataset will be increased by 4 times.



Fig: Augmentation Process

**Proposed Model**

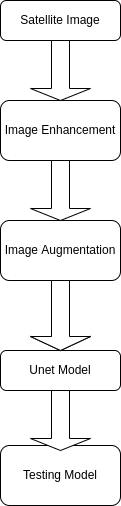


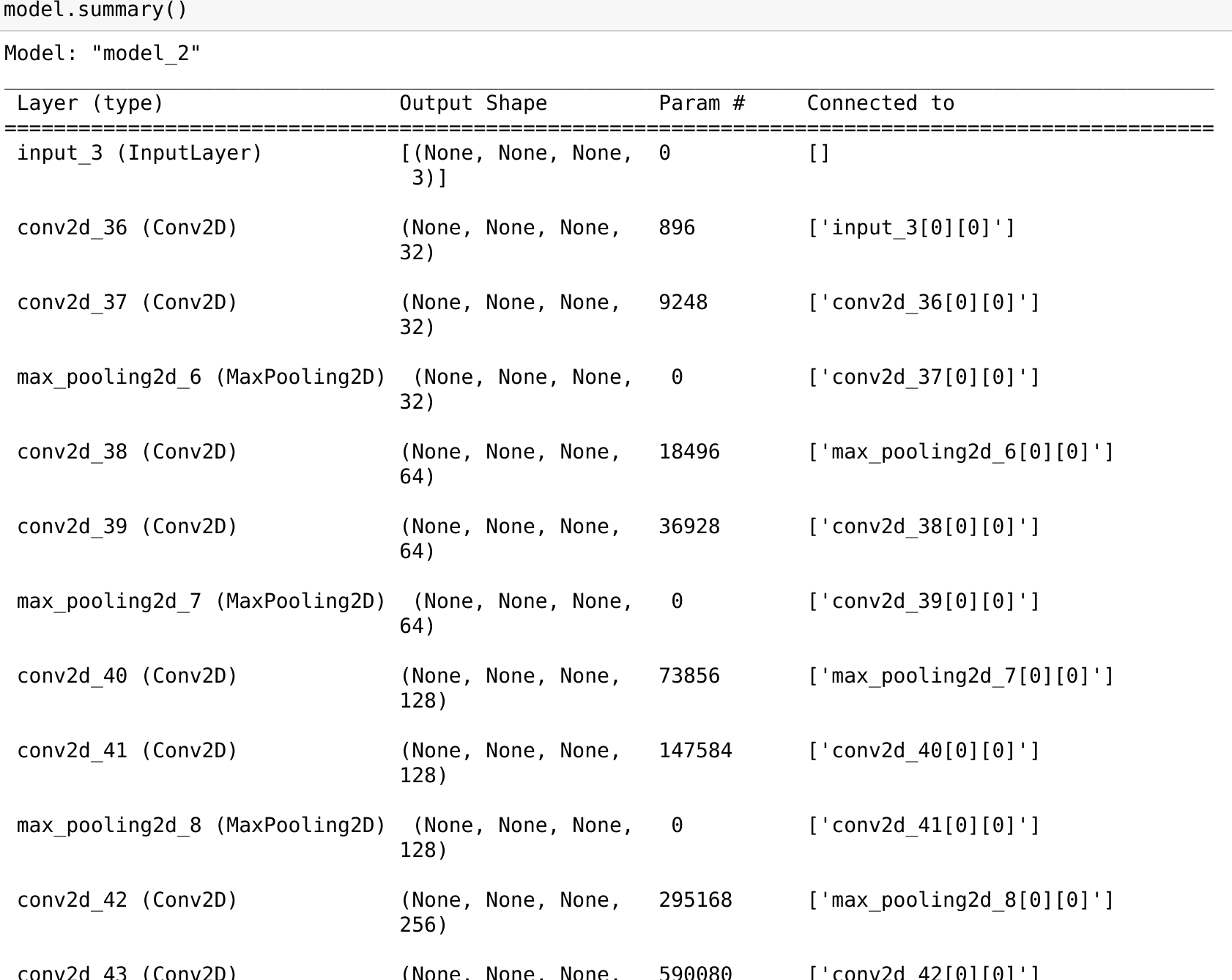
Fig: Flowchart of Proposed model

**iv. Model Creation**

As the number of datasets becomes 18.2k and the model is totally based on CNN. And for the better accuracy of the model, a complex model is needed to create.

In the training model, we have divided it into three different parts: Encoder, Center, Decoder. The U-net model is performing very well in image segmentation tasks. The implementation has a lower number of channels due to memory restrictions. It also may be simple enough for binary classification.

Below can be seen the layers and the parameter of the U-net model.



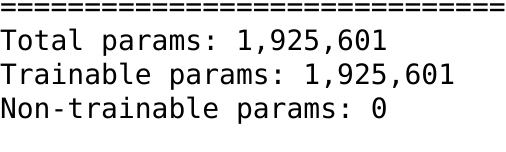
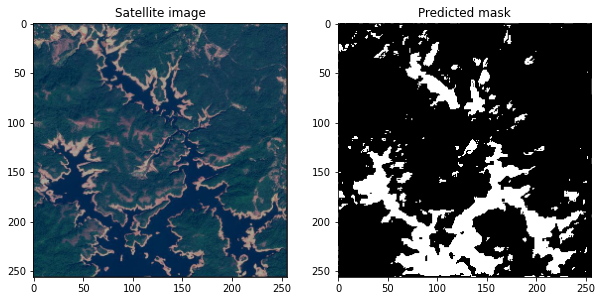
****

Fig: U-net layers and parameters

**Testing**

Here are a few random pictures of satellite images which have predicted the water network.





**Accuracy & Conclusion**

We presented our initial model for detecting rivers and watersheds from satellite pictures in this study, which was based on image processing methodology. The methodology was tested on a set of images obtained from the Sentinel-2 Satellite. For better recognition of rivers and watersheds, a new level of segmentation was used. We obtained a good accuracy using our proposed model, which is significantly higher than other U-net and Tensorflow implemented models available to date.

This study could pave the way for future studies on developing water resource management, which is critical for future generations. The suggested methodology is generic in the sense that it can be used to extract various arboreal networks in medical pictures, such as blood vessels.

**Specification Required:**

Ram: 64 GB

GPU: Geforce 8GB or higher (GTX1060ti or higher)

SSD: 2 TB

Operating system: Windows, Linux, and Mac