Journal of Visual Language and Computing

journal homepage: www.ksiresearch.org/jv1c

A Comprehensive Analysis and Design of Land Cover Usage from Satellite Images using Machine Learning Algorithms

Rajeswara Rao Duvvada[1], Shaik Ayesha Fathima[2], Shaik Noorjahan[3]

1,2,3 Department of Computer Science & Engineering, VR Siddhartha Engineering College, India
1rajeshpitam@gmail.com, 2shaikfathima579@gmail.com, 3shaiknoorjahan2001@gmail.com

ARTICLE INFO

Article History.
Submitted: 2.9.2022
Revised: 5.9.2022
Accepted: 5.23.2022

Keyword,s:
Land Cover
Classification
DeepLabv3Plus
Deep Globe Land
Cover Classification
UNet
FPN

ABSTRACT

Real-time satellite images provide a true representation of Earth and its environment. Satellite images provide important remote sensing data that can be used in a variety of applications such as image fusion, change detection, land cover classification, agriculture, mining, disaster mitigation planning, and monitoring climate change. This project performs land cover classification, that is classification of satellite images into multiple predefined classes by means of several algorithms. In this study, we compared the accuracies of different segmentation models such as Unet, FPN, deeplabv3Plus with different encoders such as Resnet101, Resnet50, Resnet34 with different activation functions, such as softmax and sigmoid functions. Following the classification of satellite images, we will calculate the area of each land cover class. The dataset used here is the Deep Globe Land Cover Classification dataset. It contains seven pre-defined classes. The approach followed for each algorithm involves gathering and processing satellite imagery. The preliminary images are then preprocessed with data preprocessing techniques and the preprocessed images are fed into the appropriate algorithm. The obtained result will be then analyzed. After measuring the accuracy of each algorithm, the algorithm that gives the best results is identified based on the comparison of the accuracy obtained from each algorithm. Among the three segmentation models, DeepLabV3Plus segmentation model with resnet101 encoder gave more accuracy i.e, 90.17% when compared two remaining UNet, FPN segmentation models

1. Introduction

Many visual applications now use segmented images to get a deeper understanding of a scene. Segments or objects are created by segmenting video frames or images and, consequently, take on an important role in real-life applications. Examples include optical remote sensing, facial segmentation, autonomous driving, and computational photography. Literature has reviewed various techniques for image segmentation, including region-growing, thresholding, watersheds, Otsu, and K-means clustering. Also discussed are graph cuts and Markov random fields. Most of these older techniques utilize low-level features and cues to segment objects.

There are several neural network-based methods for detecting and classifying objects which have achieved notable success in recent years, showing improvement in terms of performance accuracy and performance time. A growing number of deep learning-based image segmentation models are being developed recently. These models can be used to segment a number of objects including humans, cars, horses, trees and the sky. In deep learning-based segmentation, most models recognize and characterize objects based on many images during the learning process. These models are developed primarily for identifying objects in frontal views.

In our previous study [1] we used Deeplabv3Plus with resnet50 encoder as our model to perform land cover classification in satellite imagery and now in this paper we are doing comparative analysis using different deep learning segmentation models, DeepLabV3, Unet, FCN for land cover classification. The proposed methodology will classify the images into predefined classes such as barren land, range land, forest land, agriculture land, water, urban land with different RGB values. Then we calculate the area of each predefined class based on the pixel count obtained from count_nonzero() method of numpy module. Deep Globe Land Cover Classification Dataset is the dataset used for this study as shown in figure 1. We calculate the accuracy of each segmentation model and identify the model that gives best accuracy.

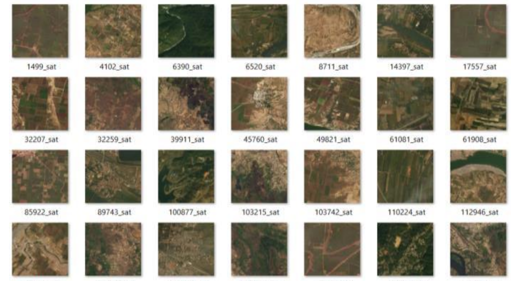

Figure 1: Sample image of the dataset [16]

DOI reference number:10-18293/JVLC2022-N1-014

1.1 DeepLabV3Plus Segmentation model

Assigning semantic labels to each pixel in a digital image is the goal of DeepLabv3+, a state-of-the-art deep learning model for semantic segmentation. It has been improved several times since it was first open sourced by Google in 2016; the latest version is DeepLabv3+ which has an encoding phase and decoding phase. Encoding an image employs convolutional neural networks in order to collect the essential information, whereas decoding reconstructs the output of desired dimensions using the acquired information. The DeepLab decoder modules support better segmentation along object boundaries using the Xception network backbone for training our DeepLab model. As an example, Google's Pixel smartphone is supposed to perform various image segmentation tasks using the trained model from DeepLab. DeepLab supports the following network backbones: MobileNetv2, Xception, ResNet, PNASNet, Auto-DeepLab.

1.2 Unet Segmentation Model

U-Nets are architectures that allow for semantic segmentation. They are constructed from a contracting path (sometimes called an encoder) and an expanding path (sometimes called a decoder). Typical convolutional network architecture is followed by the contracting path. A 2x2 max pooling step with stride 2 for downsampling is applied as well as two 3x3 convolutions (unpadded convolutions) repeatedly. By downsampling each feature channel twice, the number of channels is doubled. In the expanding path, upsampling is performed at every step, followed by a 2x2 convolution to halve the number of features channels, concatenation with a cropped map of all the features from the contracting path, and two 3x3 convolutions, each followed by a ReLU. In order to prevent the loss of border pixels, you must crop all of the convolutions. The final layer of the algorithm uses a 1x1 convolution to map each 64-component feature vector to the desired number of classes. 23 convolution layers make up the network.

1.3 FPN Segmentation Model

With the help of fully-convolutional algorithms, the Feature Pyramid Network, or FPN, extracts features from any image at any scale. This approach, which creates feature pyramids without using the backbone convolutional architectures, could be used for object detection despite not being convolutional. Using this general solution, feature pyramids can be built in deep convolutional networks. Pyramids are constructed in two ways: top-down pathways and bottom-up pathways.

In bottom-up approaches, a feature hierarchy is derived by combining features at two scale levels using backbone ConvNets. For each stage of the feature hierarchy, feature pyramid levels are defined. The last layer of each stage's feature pyramid constitutes a reference set of feature maps. Feature activation maps generated by the final residual block of each stage are used for ResNets.

By upsampling more coarsely detailed, yet more semantically rich, feature maps at higher pyramid levels, and then combining them with features from the bottom-up pathway along lateral axes, the top-down pathway enhances lower resolution features. In contrast to the top-down pathway, which affects more regions, the bottom-up pathway was sampled less frequently, so its activations are more accurately located but the top-down pathway affects more regions. From the bottom-up pathway and from the top-down pathway, each lateral connection merges feature maps with the same spatial size.

1.4 Resnet

Residual neural networks (ResNet) consist of artificial neural networks that use shortcuts or skip connections, allowing them to jump over certain layers. In a typical ResNet model, there are two or three layers of skips with nonlinearities (ReLU) and batch normalization between them. A model known as HighwayNet uses a weight matrix in addition to skip weights in order to learn the skip weights. DenseNets are network models that have multiple parallel skips. New networks are non-residual networks that do not have parallel skips. In a suitably deep model, adding more layers can lead to a higher training error if it is subject to Degradation (accuracy saturation). The latter can be mitigated by adding skip connections. With training, the weights augment the previously-skipped layer by muting the upstream layer. Resnet 34, Resnet 50, Resnet 101 are the encoders used in this work.

1.5 Sigmoid Function

In mathematics, the sigmoid function, which is similar to a logistic function, a hyperbolic tangent, and an arctangent, is a function with a typical S-shaped curve. Logistic functions, also known as logistic sigmoid functions, are used in machine learning. Logistic functions are defined as follows:

$$S(x) = \frac{1}{1+e^{-x}}$$

The logistic function takes value between zero and one for both input and output.

1.6 Softmax Function

In mathematics, a Softmax is a function that converts numbers/logits into probabilities. Its output is a vector (we'll call it v) with probabilities of everything that can possibly happen. These probabilities are added for all possible outcomes or classes. Mathematically, Softmax is defined as,

$$S(y)_i = \frac{\exp(y_i)}{\sum_{j=1}^{n} \exp(y_i)}$$

2. Literature Survey

Rahul Neware [2] used three categories of algorithms: unsupervised, supervised and object-based classification. Supervised classification is a machine learning technique where inputs are given, and then classification is made according to the inputs. Unsupervised classification largely relies on cluster analysis method. To perform cluster analysis, pixels are segmented into numerous classes and grouped according to their common characteristics. Object-oriented classification is specifically applied to high resolution imagery. Objects are mainly segmented from satellite images in object-oriented classification. In supervised classification of high resolution multispectral LISS-IV images, two methods are employed: Maximum Likelihood (ML) and Support Vector Machine (SVM). The obtained optical image of the study area will be subjected to supervised classification methods of various algorithms. Obtain the optical image of the study area based on the LISS-IV measurement. Obtain the optical image of the study area based on the LISS-IV measurement. Subsequently, rectify the obtained optical image of the study area. Create a graphical representation of each band based on the classified image as output. This paper concludes by providing a step-by-step procedure for optical data classification, concluding that SVM classification was superior for optical data classification compared to ML classification. It proposes a step-by-step procedure for optical image classification. Efficient comparative analysis is made between SVM classification and ML classification.SVM classification gives a better result than ML classification technique with 77.5% accuracy.

Keerti Kulkarni, P. A. Vijaya [3] performed a comparative analysis of land classification is performed utilizing several land classification algorithms which are available in remote sensing methods, such as Maximum Likelihood, Multi-Label Classification (MLC) and Minimum Distance. Minimum distance classification is done by defining two classes and then calculating the mean vector for each class. As new objects (pixels) are classified, they are compared to the nearest mean vector calculated from those of known classes. The closest mean vector is then used to classify the new objects. Euclidian distance,Normalized Euclidian distance,Mahalanobis distance this distances are often used in this procedure Having the minimum distance equal to the maximum similarity as the index of similarity. Maximal likelihood classification's uses a probability based decision rule. The maximum likelihood procedure assumes that training classes in each band have a normal distribution (Gaussian distribution). An assignment of pixels to the class in which they have the highest probability is made in this paper. Calculating the probability assigned to each pixel assigns the pixels to the class with the highest probability. In traditional image classification, mixed pixels can cause problems because they belong to multiple classes but can only be assigned to one class at a time. Therefore, Multilabel Classification makes use of Spectral Mixture Analysis in order to estimate the likely composition of each pixel of the image. The methods used are Problem Transformations and Algorithm Adaptation Methods. It is concluded that multi-label method classifier would produce better results for hyperspectral remote sensing images and would be more appropriate for them. It is highly effective in overcoming the underlying dependency between labels. Multilabel classification gives a more natural representation, reduces complexity, and speeds up processing time.

Vineetha P., Smitha Asok Vijayamma ., Sandra Yesudasan Miranda [4] assessed Both the unsupervised image classification accuracy and the supervised classification-based predicted land use accuracy are evaluated with Kappa coefficients. Classification of images are undertaken in order to determine land use analysis for the selected area. LISS-IV 2015 image was categorized into six types of unsupervised classification classes: built up, paddy, water body, vegetation, plantation and scrub. Both the supervised and unsupervised classification methods utilized Maximum Likelihood Classification.The classes were resolved using the training sets from Google Earth that are selections of pixels which display a similar pattern. The supervised classification-based images were then recoded to produce the desired classes.For the training set identification stage, LISS-IV 2015 imagery was classified unsupervised using ArcGIS.

The resulting vectors were then transformed into maps. ArcGIS was used to convert this vector form into KML format and Google Earth was used to view the training sets for different classes, then the supervised classification was performed. The results of the study showed a moderate degree of accuracy in the classification of different land use categories based on area-based details. This study analyzed both supervised and unsupervised methods of remote sensing data classification to determine land use in sacred groves in Thiruvananthapuram district. The overall accuracy obtained was 69.09 %. A database of the types of land uses in these areas was also created as part of the study.

K. S. Ravichandran ; R. Sivagami; R. Krishankumar; [5] aimed at analysing the performance of different supervised learning algorithms to label pixels in images using the semantically labelled datasets obtained from (ISPRS). The workflow for every algorithm involves taking input data and obtaining the respective groundtruth image, generating the feature vector, and then obtaining a labeled thematic output using a classifier and finally assessing accuracy. For this comparative study, SVM based classification was used with Gaussian kernels with different kernel scale values which has been used in linearly non-separable data classification. With this approach, training data with associated class labels is used along with the attribute selection measure, in order to construct a top-down recursive divide and conquer approach to decision tree classification. This study uses the Gini index as an attribute selection measure among Information Gain, Gain ratio, and Gini index. Among the three attributes selection measures, the Gini index is chosen. KNN classification involves selecting the training and testing data and then selecting the distance metric. The best k value is detected and the knn model accuracy is evaluated after it has been built. Comparison of the results from the three state of the art techniques led to a conclusion that fine Gaussian support vector machines outperformed them all, with a classification accuracy of approximately 75.1448%. Workflow of every algorithm is same and easy to perform. Comparitive analysis is made between SVM Classification, KNN Classification and Decision Tree Classification.

3. Proposed System

3.1 Dataset

Deep globe land cover classification dataset:

- The dataset we used for this work is Deep Globe Land cover classification dataset which consists of nearly 1700 satellite images including masks shown in Figure 2.
- Resolution of each satellite image is 2448 x 2448 and training data consists 803 satellite images with pixel resolution of 50cm by DigitalGlobe satellite.
- The dataset consists of validation images(171) and test images(172) in total excluding mask images.

Figure 2*: Deep* globe land cover classification dataset [16]

3.2 Segmentation Models

In this paper we are using three segmentation models. They are Unet, FPN, DeepLabV3Plus.

Unet Segmentation Model:

U-Net is based on the idea of FCN segmentation. The U-Net architecture could be described as an encoder-decoder unit, divided into three parts. As shown in Figure 3, the down-sampling path is composed mainly of seven stages. The first stage consists of eight stages primarily applying three 3 x 3 convolutions with batch norms, followed by two 2 x 2 max-poolings. This horizontal bottleneck consists of two layers 3 x 3 convolution, followed by two layers 2 x 2 Up-Convolution, as illustrated in the Figure 3. Similarly, the upsampling path consists of four stages, each involving halving the features maps. There are two convolutional layers 3 x 3 representing a decoder, followed by two layers 2 x 2.

Figure.3 shows how the model skips paths between upsampling and downsampling paths in order to provide local and global information during upsampling. Last but not least, the convolutional layer provides segments at output 1 x 1 where the number of feature maps corresponds to the number of segments to be created. In the creation of the Unet segmentation model, we used Resnet34 as an encoder, softmax as an activation function, and sigmoid as an activation function.

FPN Segmentation Model:

Figure 4 shows a diagram of FPN. The system is composed of a top-down and bottom-up approach. We use Pretrained ResNet50 as a feature encoder for the bottom-up pathway of convolutional network extraction. Each module contains many convolution layers (i=1 to 5) and is composed of many convolution modules. During the top-down pathway, each convolutional module's output is labeled as Ci and then used in the top-down pathway. The spatial dimension reduces by 1/2 with each step up (in other words, it expands by double).

With each layer removed, the semantic value increases. With each step of the bottom-up pathway, the spatial resolution will decrease. In the top-down pathway, we apply a 1x1 convolution filter to reduce C5 channel depth to 256 so that we can create M5. Using two successive layers of 3x3 convolution, we create P5; this is the first feature map layer that is needed to segment objects. P5 has the same spatial resolution as conv5 and is a 128-channel layer. Then, we apply successively three times 3x3 convolution to output the next layer of features. This reduces the aliasing of the upsampling process. To reduce aliasing of the upsampling process we apply successively three 3x3 convolutions. This reduces the effect of the upsampling process.

By the end of the process, all Pi modules with 128 channels and 1/4 of the input image resolution have been concatenated. The next step is to apply 512 3x3 convolutional filters, batch normalization, and ReLU activation. Our output feature map consists of seven classes, so our output channel map consists of seven classes. Using convolutions again, we reduced the number of channels to acquire seven output channels. A spatial dropout operation is then performed and then the output image is upscaled by bilinear interpolation. to the original image size.

DeepDeepLabV3Plus Segmentation Model:

As shown in Figure.5, this model primarily relies on an encoder-decoder architecture. When encoding, features are encoded by the CNN model, then decoded by the decoder. The encoder extracts all the essential information from each channel separately and combines the information with point-wise convolutions to reconstruct the output, as shown in Figure 5. By using the spatial pyramid pooling (SPP) method, DeepLab divides feature maps separated by convolutional layers into spatial bins of a fixed number of bins. DeepLabV3 uses an array of convolutions with SPP to process images with multiple scales and to add context without requiring as many parameters as possible. This increases the field of view as opposed to the previous models that used max-pooling and batch normalization. The new model presented by uses more layers of Xception backbone with depth-wise dilated separable convolutions instead of using the previous models. FIGURE 5 illustrates the encoder module, which encodes multi-scale contextual information with the aid of arous convolution, and FIGURE 6 shows the decoder module, which refines the segmentation results by finely tuning the boundary conditions around each object. With our atrial convolution model, we can acquire multi-scale information by adjusting filters. With our atrial convolution model, the features are resolved by using atrial convolution.

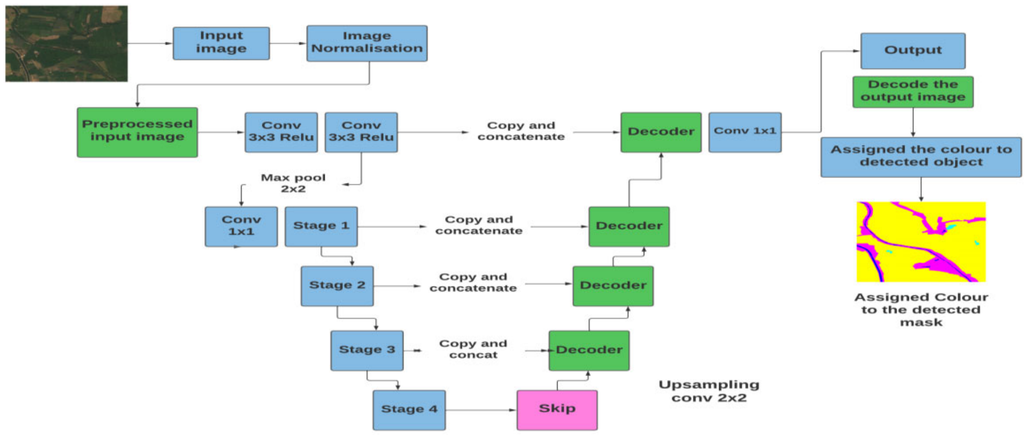

Figure 3: Unet Segmentation Model Architecture

Figure 4: FPN Segmentation Model Architecture

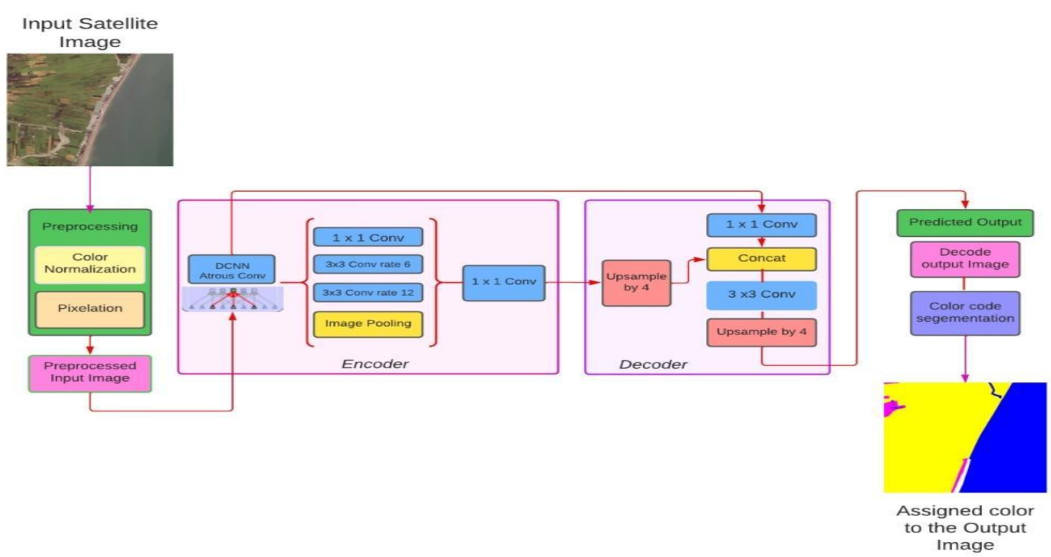

Figure 5: DeeplabV3Plus Segmentation Model Architecture

3.3 Proposed Methodology:

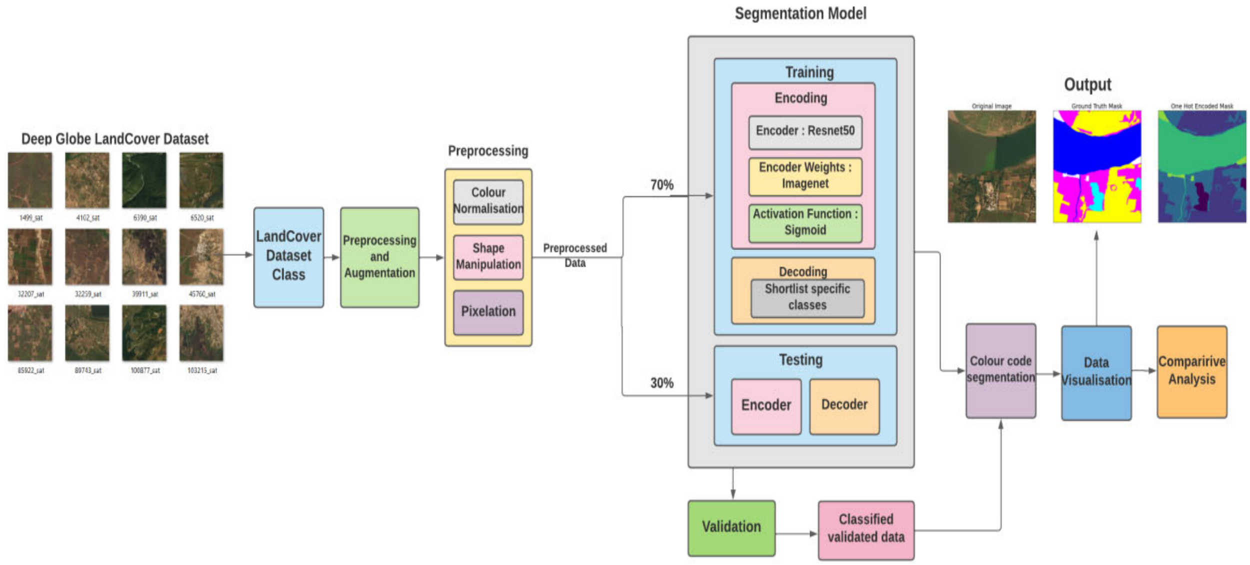

Figure 6: Flow Diagram for our approach

- Firstly we define LandCoverDataset class in which we read satellite images and then we apply augmentation on this images to create new copies of existing data using center_crop() and Random_crop() methods to increase the size of the dataset.

- We perform data preprocessing on our satellite images using Color Normalization, Shape Manipulation, pixilation.

- Color normalization is used to normalize the rgb values in satellite images on the basis of pixel values, Shape Manipulation is to resize the satellite images in our dataset to unified dimensions and apply Pixelation to not exceed the original or actual dimensions of our satellite images .

Segmentation Model:

We divide our dataset in the ratio 7:3 for training and testing 70% of the data is used to train our segmentation models and 30% of the data is used for testing.

Create Segmentation model with pretrained encoder:

Create Unet Segmentation model

- ‘Resnet34’ is used as encoder, ‘imagenet’ as encoder- weights and activation functions as sigmoid and softmax.

- From ‘segmentation_models_pytorch’ module we used Unet() method.

- Perform training and testing of the model.

Create FPN Segmentation model

- ‘Resnet50’ is used as encoder, ‘imagenet’ as encoder- weights and activation functions as sigmoid and softmax.

- From ‘segmentation_models_pytorch’ module use FPN() method.

- Perform training and testing of the model.

Comparision of Unet, FPN and DeepLabV3Plus Segmentation Models:

- Compare the accuracies of segmentation models with different encoders and different activation functions.

- Determine the best segmentation model.

Decoding:

- In decoding we Shortlist specific classes to segments using ‘tolist()’ and ‘values.tolist()’ methods to get the classes and values.

Testing the model:

- In order to test the model test dataloader is created for each segmentation model.

- We display the results of testing using Data visualization method.

Color code Segmentation:

- We perform color code segmentation by coloring the classes with their corresponding rgb values which are mapped to their class keys.

Data Visualization:

- In data visualization we display our original satellite images, classified image, converted image.
- Classified image is obtained from color_code_segmentation() method and converted image is obtained from reverse_one_hot() method.
- In order to plot the results we use matplotlib module.

Area Calculation:

- We perform area calculation of each class by obtaining the pixel count of each predefined class in that satellite image using 'count_nonzero()' method of numpy module.

4. Visualization Results Of Segmentation Models For Satellite Images

For each segmentation model we give original image, classified image and converted image. We classify the satellite images into predefined multiple classes such as Urban Land, Agriculture Land, Range Land, Forest Land, Water, Barren Land, Unknown with rgb values 0,255,255; 255,255,0 ; 255,0,255 ; 0,255,0 ; 0,0,255 ; 255,255,255 ; 0,0,0 respectively and with colors cyan/aqua, yellow, pink, green, blue, white, black respectively.

Unet Model with Resnet 34 encoder and sigmoid activation function:

Figure 7 depicts the data visualization output of Unet Model. The Encoder used is Resnet34 and the activation function is sigmoid activation function.

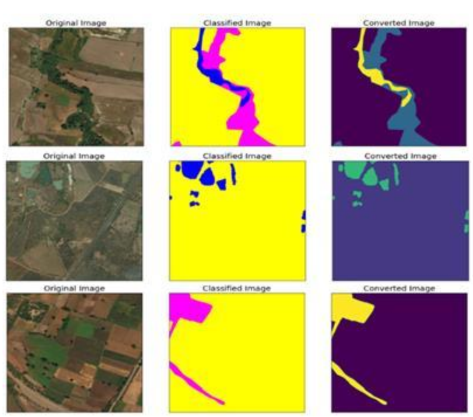

Figure 7: Unet Model with Resnet34 encoder and sigmoid activation function

Unet Model with Resnet 34 encoder and softmax activation function:

Figure 8 depicts the data visualization output of Unet Model. The Encoder used is Resnet34 and the activation function is softmax activation function.

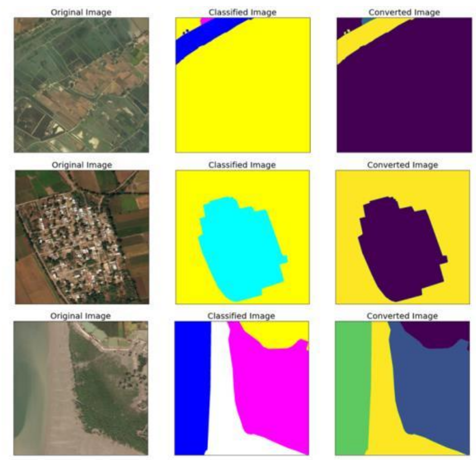

Figure 8: Unet Model with Resnet 34 encoder and softmax activation function

FPN Model with Resnet 50 encoder and softmax activation function:

Figure 9 depicts the data visualization output of FPN Model. The Encoder used is Resnet50 and the activation function is softmax activation function.

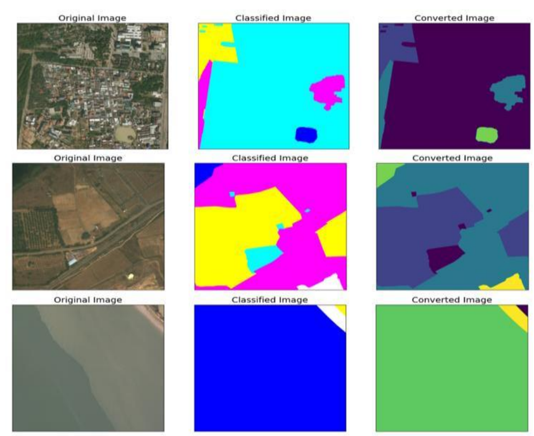

Figure 9: FPN Model with Resnet 50 encoder and softmax activation function

FPN Model with Resnet 50 encoder and sigmoid activation function:

Figure 10 depicts the data visualization output of FPN Model. The Encoder used is Resnet50 and the activation function is sigmoid activation function.

Figure 10: FPN Model with Resnet 50 encoder and sigmoid activation function

DeepLabV3Plus with resnet101 encoder:

Figure 11 depicts the data visualization output of DeepLabV3Plus Model. The Encoder used is Resnet101 and the activation function is sigmoid activation function.

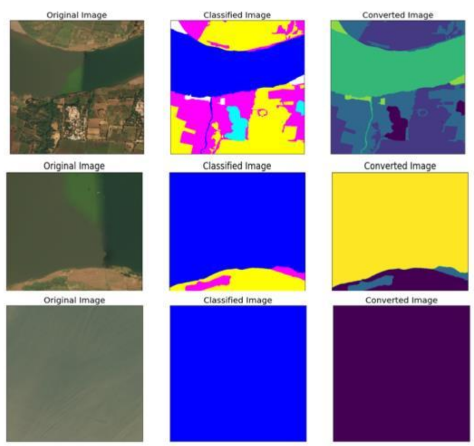

Figure 11: DeepLabV3Plus with resnet101 encoder

Performance Evaluation

IOU Score:

A method's ability to predict bounding boxes based on a set of input data is measured by this statistic. Methods for generating anticipated bounding boxes can be evaluated using this calculator. The difference between the expected and ground truth bounding boxes is represented by this percentage between 0 and 1.

$$IOU=\frac{Area\ of\ overlap}{Area\ of\ Union}$$

Unet Model:

IOU Scores of Unet segmentation model with sigmoid and softmax activation functions for different epochs are plotted in Figure 12.

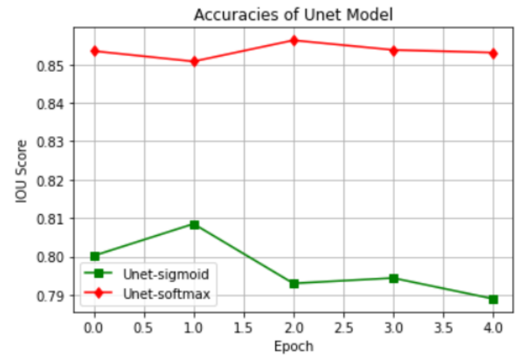

Figure 12: Accuracies of Unet Model

FPN Model:

IOU Scores of FPN segmentation model with sigmoid and softmax activation functions for different epochs are plotted in Figure 13.

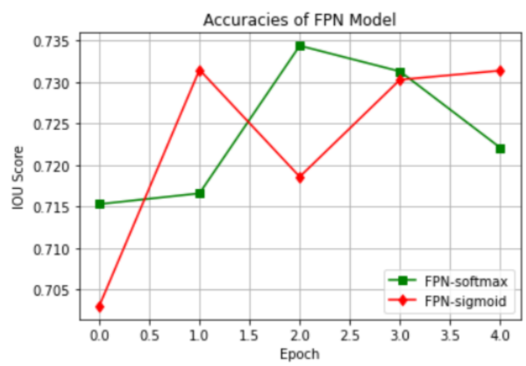

Figure 13: Accuracies of FPN Model

IOU Scores of DeepLabV3Plus segmentation model with resnet50 and resnet101 encoders for different epochs are plotted in Figure 14.

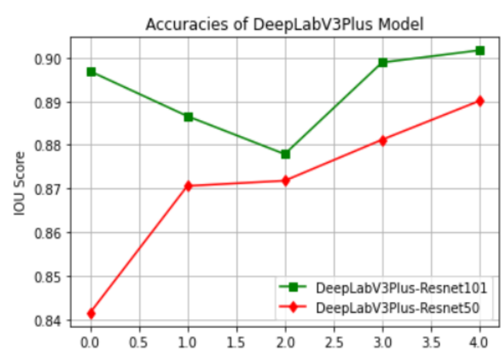

Figure 14 : Accuracies of DeepLabV3Plus Model

IOU scores of the 3 different segementation models are as follows,

S. No	Segmentation Model	Encoder	Activation Function	M-IOU Score
1	Unet	Resnet34	Sigmoid	0.7890
2	Unet	Resnet34	Softmax	0.8531
3	FPN	Resnet50	Sigmoid	0.7314
4	FPN	Resnet50	Softmax	0.7221
5	DeepLabV3Plus	Resnet50	Sigmoid	0.8713
6	DeepLabV3Plus	Resnet 101	Sigmoid	0.9017

Table 2 : Mean IOU Scores of Segmentation Models

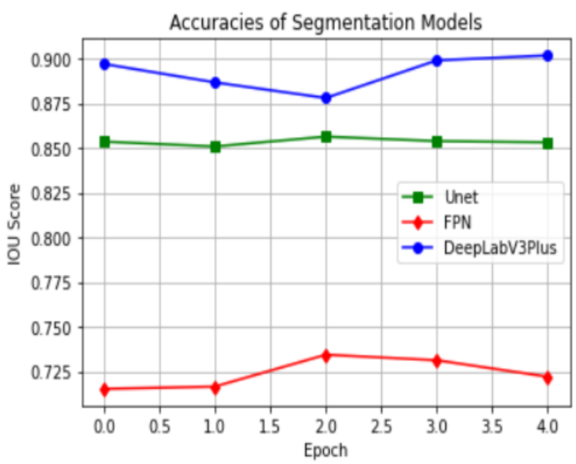

Figure 15: Accuracies of Segmentation Models

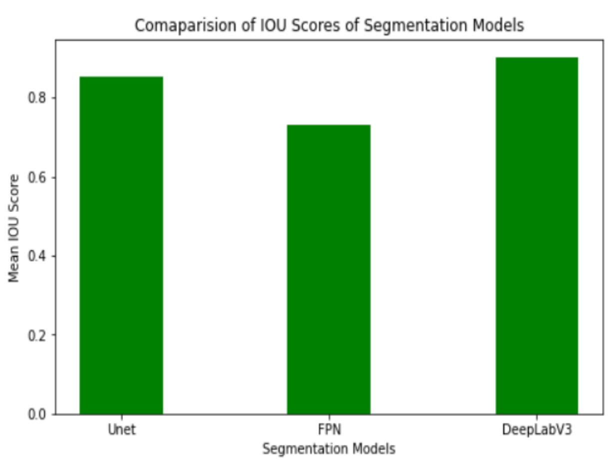

Figure 16: Comparison of IOU Scores of Segmentation Models

Among the three segmentation models, DeepLabV3Plus gives more iou score with resnet101 encoder. Unet gives more iou score with softmax function than with sigmoid function. Similarly, FPN also gives more iou score with softmax function than with sigmoid function. In case of Unet, resnet34 encoder is used whereas for FPN resnet50 encoder is used. DeepV3Plus gives more iou score with resnet101 encoder than with resnet50 encoder [1].

Dice Loss:

In the 1940s, Srensen-Dice created a statistic to measure how similar two samples are. It is named dice loss after the Srensen-Dice coefficient. A 3D segmentation technique for medical images was introduced by Milttari et al. in 2016 for computer vision research.

$$D = \frac{2\sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2}$$

Pi and gi, as seen above, are Dice coefficients which represent pixels that are both associated with prediction and ground truth, respectively. The boundaries are detected by either displaying pi and gi values of 0 or 1, which indicate whether the pixel is a boundary (value 0) or not. If pi and gi are the same, when the sum advances, the numerator is the

sum of successfully predicted boundaries, while the denominator is the sum of ground truth boundaries (both 1 in value).

Unet Model:

Dice Loss of Unet segmentation model with sigmoid and softmax activation functions for different epochs are plotted in Figure 17.

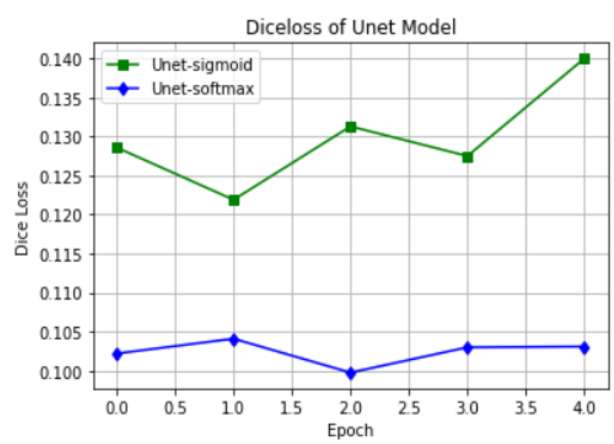

Figure 17: Dice Loss of Unet Model

FPN Model:

Dice Loss of FPN segmentation model with sigmoid and softmax activation functions for different epochs are plotted in Figure 18.

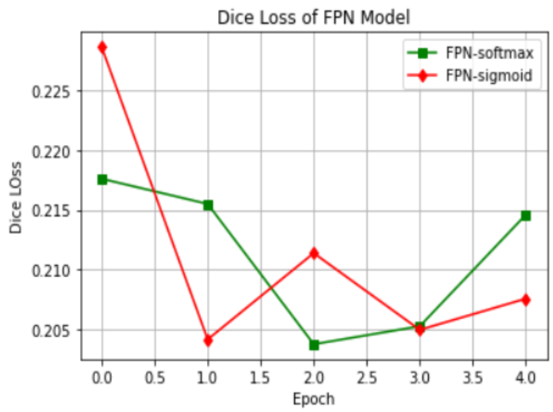

Figure 18 : Dice Loss of FPN Model

Dice Loss of DeepLabV3Plus segmentation model with resnet50 and resnet101 encoders for different epochs are plotted in Figure 19.

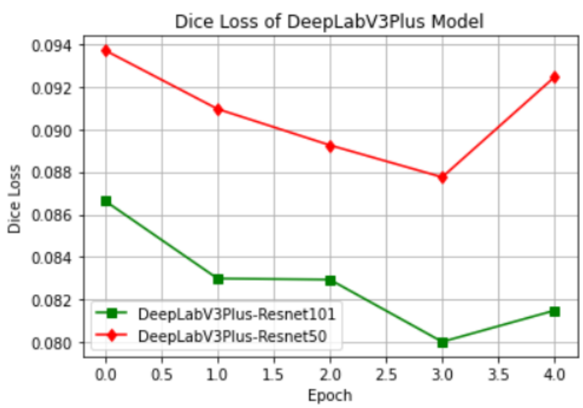

Figure 19 : DiceLoss of DeepLabV3Plus Model

Dice Loss of the 3 different segementation models are as follows,

S.NO	Segmentation Model	Encoder	Activation Function	Mean Dice Loss
1	Unet	Resnet34	Sigmoid	0.1400
2	Unet	Resnet34	Softmax	0.1031
3	FPN	Resnet50	Sigmoid	0.2076
4	FPN	Resnet50	Softmax	0.2146
5	DeepLabV3Plus	Resnet50	Sigmoid	0.0908
6	DeepLabV3Plus	Resnet101	Sigmoid	0.0815

Table 3 : Mean Dice Loss of Segmentation Models

Among the three segmentation models, DeepLabV3Plus gives more less dice loss with resnet101 encoder. Unet gives more dice loss with sigmoid function than with softmax function. FPN gives more dice loss with softmax function than with sigmoid function. In case of Unet, resnet34 encoder is used whereas for FPN resnet50 encoder is used. DeepV3Plus gives less dice loss with resnet101 encoder than with resnet50 encoder.

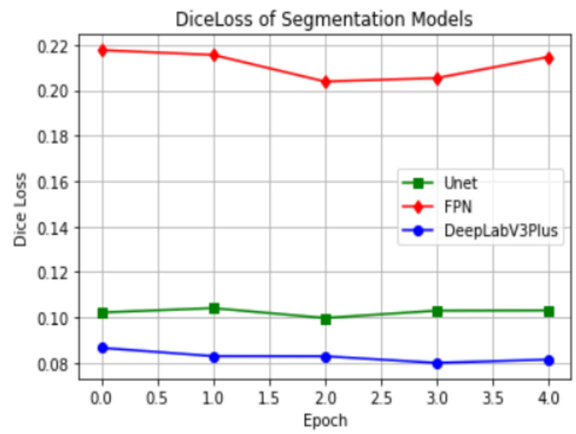

Figure 20 : Dice Loss of Segmentation Models

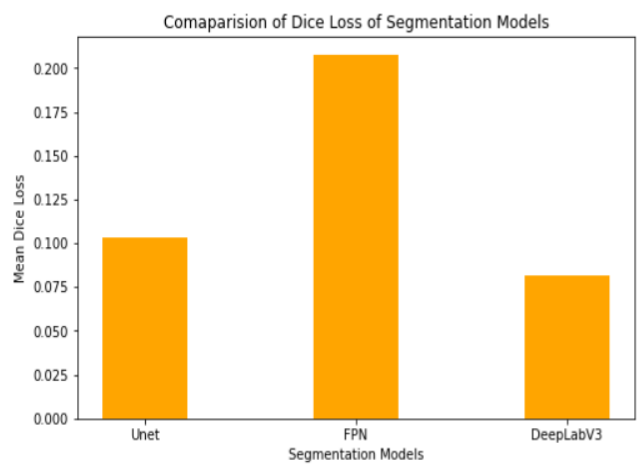

Figure 21: Comparison of Dice Loss of Segmentation Models

Accuracies:

Accuracies of the 3 different segementation models are as definded above.

Among the three segmentation models, DeepLabV3Plus gives more accuracy with resnet101 encoder. Unet gives more accuracy with softmax function than with sigmoid function. Similarly, FPN also gives more accuracy with softmax function than with sigmoid function. In case of Unet, resnet34 encoder is used whereas for FPN resnet50 encoder is used. DeepV3Plus gives more accuracy with resnet101 encoder than with resnet50 encoder.

S.NO	Segmentation Model	Encoder	Activation Function	Accuracies
1	Unet	Resnet34	Sigmoid	78.90%
2	Unet	Resnet34	Softmax	85.31%
3	FPN	Resnet50	Sigmoid	73.14%
4	FPN	Resnet50	Softmax	72.21%
5	DeepLabV3Plus	Resnet50	Sigmoid	87.13%
6	DeepLabV3Plus	Resnet101	Sigmoid	90.17%

Table 4 : Accuracies of Segmentation Models

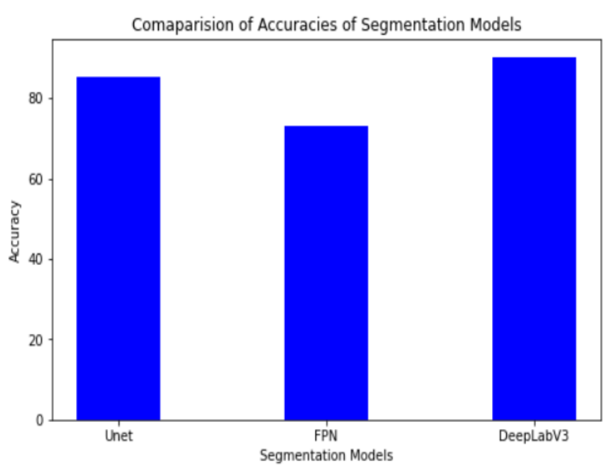

Figure 22: Comparison of Accuracies of Segmentation Models

5. Conclusion

Our work explores three semantic segmentation models, FPN, U-Net, and DeepLabV3Plus, for the classification of Land cover. Even though the satellite images show considerable variation, the pre-trained deep learning-based segmentation models still offer encouraging results. Further testing of all three models with a valid set of satellite images has further enhanced their accuracy compared to FPN. Overall, DeeLab3Plus and U-Net are more accurate than FPN. As a result, the deep learning-based segmentation models achieve Mean IoU of 90%, 85% and 73%, respectively. These models outperform conventional methods significantly by a large margin. We may extend the work to include other deep learning-based segmentation models using different datasets in the future.

6. References

[1] D. R. Rao, S. Noorjahan and S. A. Fathima, "Classification of Land Cover Usage from Satellite Images using Deep Learning Algorithms," in ICEARS, 2022.

[2] Rahul Neware ” Comparative Analysis of Land Cover Classification Using ML and SVM Classifier for LISS-iv Data”, 2019.

[3] Keerti Kulkarni; P. A. Vijaya ” A comparitive study of land classification using remotely sensed data” in IEEE, 2017.

[4] Sandra Yesudasan Miranda, Vineetha P., Smitha Asok Vijayamma, “A Comparative Study of Land use Classification using Remote Sensing Techniques, in and around Selected Sacred Groves of Thiruvananthapuram District “in 2016.

[5] R. Sivagami; R. Krishankumar; K. S. Ravichandran, ” A Comparative Analysis of Supervised Learning Techniques for Pixel Classification in Remote Sensing Images” in IEEE, 2018.

[6] Selim S. Seferbekov, Vladimir I. Iglovikov, Alexander V. Buslaev, Alexey A. Shvets, "Feature Pyramid Network for Multi-Class Land Segmentation" in 2018.

[7] Python, https://www.python.org/doc/essays/blurb/

[8] Pandas, https://pandas.pydata.org/docs/pandas.pdf/

[9] UML diagrams includes use case, activity diagrams, sequential diagrams, https://modeling-languages.com/

[10] RajibMall, Fundamentals of Software Engineering. 2ed, PHI.

[11] Use case, activity diagrams, sequential diagrams, https://modeling-languages.com/

[12] Understanding of the LSTM Networks, https://colah.github.io/posts/2015-08-Understanding-LSTMs/

[13] TensorFlow, https://www.tensorflow.org/about.

[14] Connor Shorten, Taghi M. Khoshgoftaar, “A survey on Image Data Augmentation for Deep Learning” in 2019.

[15] “Segmentation-models-pytorch 0.2.1” in 2021.

[16] Deep Globe Land Cover Classification Dataset in Kaggle, 2018.