Semantic Segmentation for Land Cover using Bhuvan 2D Satellite Images

Aryaman Gokarn
Electronics and Telecommunications
Sardar Patel Institute of Technology
Mumbai, India
gokarnaryaman09@gmail.com

Khushi Patni
Electronics and Telecommunications
Sardar Patel Institute of Technology
Mumbai, India
khushipatni18@gmail.com

Dr. Sukanya Kulkarni
Electronics and Telecommunications
Sardar Patel Institute of Technology
Mumbai, India
sukanya_kulkarni@spit.ac.in

***Abstract*—We present a novel approach for semantic segmentation of land cover using Bhuvan 2D satellite images. The proposed approach utilizes U-Net models for classification. The dataset consists of 2D satellite images covering a region in India. A pre-processing pipeline is applied to extract relevant features and generate ground-truth labels. The model is trained on the dataset, and the results demonstrate its effectiveness in accurately classifying different types of land cover, such as forest, water bodies, agriculture, roads and urban areas, using performance metrics such as Dice coefficient (F1-Score), and Intersection over Union (IoU). The model is made more accurate and robust through the implementation of data augmentation techniques. The proposed approach can have significant applications in environmental monitoring, disaster management, and urban planning, providing accurate and detailed information on land cover patterns and changes over time.**

***Index Terms*—Semantic Segmentation, Land cover Classification, Machine Learning, Deep Learning**

I. Introduction

Land cover classification involves the identification and categorization of land cover types in a specific geographic area using satellite view data. Satellite images provide valuable information about land cover characteristics, soil moisture content, and temporal changes. They offer a unique vantage point, allowing for the observation of large land areas simultaneously, which is not feasible with ground-based methods. Land cover classification is crucial for monitoring land use changes, enabling effective land resource management, and future planning. Additionally, it contributes to the study of environmental factors, such as vegetation distribution, water bodies, and their relationship to climate change and air pollution, facilitating environmental research and analysis.

Land cover classification involves the categorization of the land surface into various types, including forests, water bodies, agriculture, urban areas, and roads, as depicted in Fig. 1. The process aims to assign these distinct categories to different regions or pixels on a satellite view, providing a comprehensive understanding of the land cover composition in a specific area.

Land cover classification entails challenges related to the availability of accurate training data is crucial, as obtaining a comprehensive and well-labeled dataset can be time-consuming and challenging. To overcome this, techniques such as data augmentation and active learning can be employed to expand the training set and enhance its representativeness.

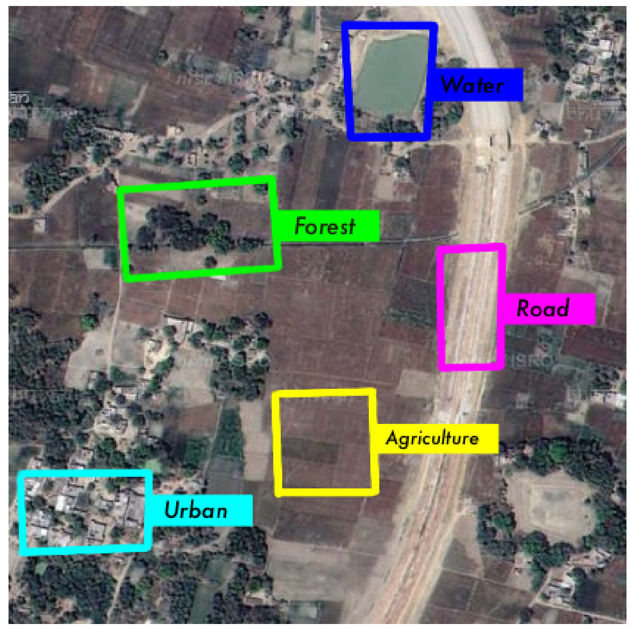

Fig. 1. Land cover classification based on Bhuvan satellite images. This visualization highlights agriculture, forest, road, water, and urban areas.

The spatial and spectral resolution of satellite imagery can affect the classification accuracy, as certain land cover types may have similar spectral characteristics or appear differently at varying scales. Advanced classification algorithms and feature extraction methods, such as object-based classification and spatial-spectral analysis, can address these challenges by incorporating both spatial and spectral information. Additionally, class imbalance, where certain land cover classes are underrepresented, can hinder classification accuracy. Mitigation strategies such as class weighting, ensemble learning, or resampling techniques can alleviate this issue and improve classification performance.

Spectral variability due to atmospheric conditions, sensor characteristics, or seasonal changes can introduce noise and affect classification accuracy. Pre-processing techniques like radiometric calibration and atmospheric correction can help normalize and standardize satellite imagery, reducing spectral variations.

TABLE I
COMPARING AND ANALYSING

Author Name	Location/Region	Size of Dataset	Method	Accuracy
Gharbia	Nile and Delta	3500	AlexNet	90%
	Regions of Egypt	3500	VGG-16	94.6%
A. Chatterjee	West Bengal, India	-	VGG-16	89.65%
Alexander Rakhlin	Salt Lake City, UT	1000	U-Net	-
K. Sawant	Dehradun, India	105	Maximum Likelihood	85%

Land cover classification has many practical applications, including monitoring land use changes, assessing environmental impacts, and planning land management activities. Accurate land cover maps are essential for identifying areas that are vulnerable to environmental degradation, predicting ecological and environmental impacts, and developing effective land management strategies.

The paper is structured as follows: Following the introduction in Section I, a comprehensive review of the relevant literature is presented in Section II. Section III introduces the algorithm employed for land cover classification. The datasets utilized in this study are outlined in Section IV. The experimental procedure and performance evaluation are detailed in Section V. Finally, the paper concludes with Section VI.

II. Literature Review

To investigate the spatial characteristics of remotely sensed images for land cover classification, A. Chatterjee et. al. proposed an unsupervised approach in their paper [1], [2]. Their study focused on distinguishing agricultural, water bodies, urban settlements, and forest areas from the underlying scene using High-resolution satellite imagery of dual-polarized and hybrid polarimetric SAR images from ISRO RISAT-1 for West Bengal, India regions. To train the data, they utilized the VGG-16 [3] and transfer learning model and employed an entropy-based loss function to evaluate the performance of classification models [4]. Their approach achieved an accuracy of 86.08% and 89.65% for dual-polarized and hybrid SAR images, respectively, demonstrating the effectiveness of their proposed method.

In this paper, Alexandar Rakhlin proposed a segmentation model for land cover classification is proposed, specifically designed for the DeepGlobe Challenge [5], [6]. The model utilizes the U-Net architecture [7], known for its effectiveness in segmentation tasks with limited labeled data. Additionally, it incorporates the Lovász-Softmax loss function [8], which optimizes the Jaccard index for improved performance. The model was evaluated on the DeepGlobe Challenge dataset, consisting of 1000 satellite images containing pixel-level annotations. Comparative analysis with other contemporary segmentation models demonstrated that their model achieved the highest performance on test set. The authors highlight the ability of the model to handle challenges such as limited labeled data and imbalanced classes while achieving high accuracy. Overall, the findings suggest that the model holds promise as an effective approach for land cover classification using satellite imagery.

In recent years, Convolutional Neural Networks has become a popular tool for image classification due to their ability to extract complex features from images. In a paper by Gharbia et. al, two popular techniques for image classification, VGG-16 [3], [9], and AlexNet [10], [11] are proposed [12]. The models were trained using 3500 Landsat 5 images of the Egypt's North-eastern region along the Delta and Nile Valley regions, and categorized the image data into five classes: urban, roads, water body, deserts and vegetation. According to the authors, the accuracy achieved was 74.8% using AlexNet and 90.2% using VGG-16. To further enhance the accuracy, augmentation techniques were employed resulting in an accuracy of 90.0% using AlexNet and 94.6% using VGG-16.

K. Sawant et al. presented a study on the detection of changes in urban expansion and its impact on the land cover classes using Remote Sensing and GIS [13]. They utilized Landsat 8 images of Dehradun for the years 2013 and 2018, and trained a supervised machine learning algorithm to classify the images into four categories: Vegetation, Agriculture, Urban Forests, and Seasonal river beds. The classification accuracy exceeded 85%. The study analyzed the percentage change in these categories, but its applicability is limited due to a small dataset size.

Overall, the reviewed papers have made significant contributions to the field of land cover classification using various techniques such as machine learning and deep learning as shown in Table I. The papers have proposed models that can classify land cover into different categories such as vegetation, water body, urban area, and bare soil. However, there are some limitations to their proposed models. The models show classification errors due to the limited spectral reflectance properties inherent in satellite imagery. Some of the papers also have a limited dataset size, which can limit the generalizability of the models. Despite these limitations, the proposed models have the potential to aid in urban planning, disaster management, and environmental monitoring.

III. Algorithms

Satellite images often contain diverse and complex landscapes with various land cover types, such as agriculture, roads, forest, water bodies, and more. The classification of land cover involves the use of various algorithms and techniques to accurately identify and classify different land cover types.

Semantic segmentation is a popular algorithm used for land cover classification. Unlike traditional image classification techniques that assume a single object in an image, satellite image segmentation requires identifying and delineating

Fig. 2. Bhuvan Satellite 2D images of the Varanasi region in North India.

multiple objects or regions of interest within the image. This level of granularity allows for a detailed understanding of different objects and features present in the satellite image.

Semantic segmentation aims to categorize individual pixels within an image into predefined classes by assigning a semantic label to each pixel based on its contextual meaning. This paper focuses on two key algorithms for land cover classification: Resnet50 and VGG U-Net. Resnet50 U-Net incorporates the Resnet50 backbone for enhanced feature extraction, whereas VGG U-Net integrates low-level and high-level features efficiently via skip connections [14].

A. Base Model

The base model serves as the foundation for land cover classification algorithms. In this context, two popular base models are VGG16 and Resnet50.

1) VGG16: It is a convolutional neural network architecture widely used in computer vision tasks. It consists of 16 layers, including convolutional layers with narrow receptive fields (3x3), max pooling layers, and fully connected layers. VGG16 is known for its simplicity and uniform architecture, making it easy to understand and implement. As a result, VGG16 has been widely successful in diverse image classification tasks, including land cover classification.

2) Resnet50: Resnet50, also known as Residual Network 50, is an advanced architecture of convolutional neural network specifically designed to tackle the problem of vanishing gradients encountered during training. It brings a novel approach by incorporating residual connections, enabling the network to learn residual features instead of directly learning the desired underlying representation. Resnet50 is deeper and more complex than VGG16, enabling it to capture more intricate features and patterns in the input images. This makes Resnet50 well-suited for tasks that require modeling complex relationships, including land cover classification.

B. Segmentation Model - U-Net

U-Net is a popular architecture for semantic segmentation tasks. It consists of an encoder and a decoder, with skip connections between them. The encoder portion captures the spatial features of the input image through convolutional and pooling layers, gradually reducing the spatial dimensions. The decoder part of the model performs upsampling to restore the spatial resolution of the features and merges them with skip connections from the encoder. This integration enables the model to retain and leverage both fine-grained details and broader contextual information, enabling U-Net to achieve accurate and precise segmentation results.

The U-Net architecture, with its skip connections, is particularly well-suited for land cover classification as it helps preserve fine details and contextual information. It can effectively distinguish between different land cover types and produce high-quality segmentation maps.

By leveraging the base models (such as VGG16 or Resnet50) as the backbone of the segmentation model (U-Net), we can harness the strengths of both the architectures, resulting in robust and accurate land cover classification algorithms.

One of the key advantages of deep learning-based semantic segmentation for land cover classification is its ability to handle complex and heterogeneous landscapes. For example,

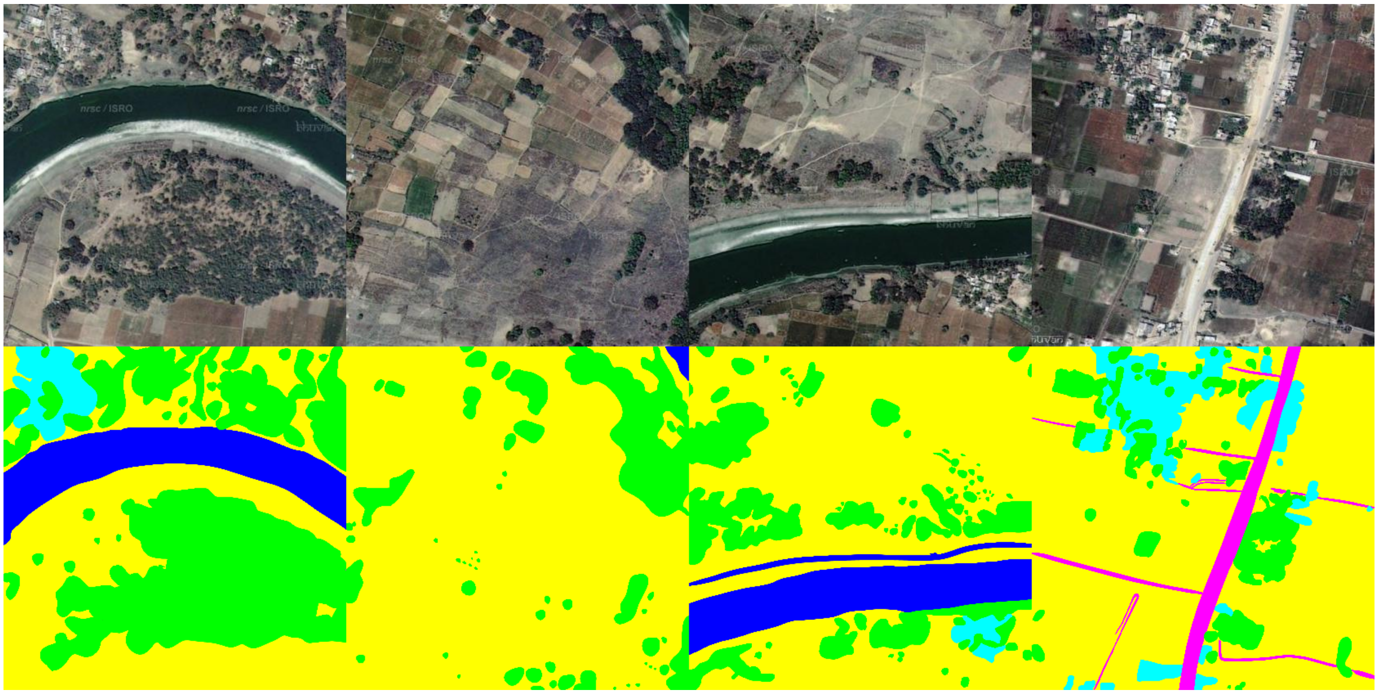

Fig. 3. Satellite Images and their corresponding segmentation masks.

it can differentiate between different types of vegetation or identify water bodies with different shapes and sizes. Another benefit is the capacity to efficiently handle substantial volumes of data within a limited time frame, enabling the generation of high-resolution land cover maps at regional or global levels within reasonable time constraints.

IV. Dataset

A. Study Area

The dataset, depicted in Fig. 2, consists of satellite 2D images of Varanasi, a city located in the northern part of India, in the state of Uttar Pradesh, with coordinates ranging from 25.3° to 25.5° N latitude and 83° to 83.2° E longitude.

B. Dataset Used

The images was obtained by Bhuvan - India Geo Platform of ISRO. Bhuvan 2D visualization interface is designed to provide map view, satellite image view, hybrid and terrain view [15]. Bhuvan 2D satellite images comprise a collection of high-resolution images capturing the Earth's surface. These images were obtained from the Indian Remote Sensing Satellite (IRS) and were processed and made available through the Bhuvan Geo Platform, which is managed by the Indian Space Research Organization (ISRO).

To construct the dataset, we resized the original images to a resolution of 420x420 pixels, which allowed to make the dataset more manageable and reduce the computational load during subsequent processing. Additionally, we employed data augmentation techniques, including rotation, flipping, and scaling, which will be elaborated upon in Section IV-C. These augmentations were applied to enhance the dataset's diversity and enhance the accuracy of the land cover classification model. The goal was to introduce variations in the dataset that would better represent real-world scenarios and improve the model's ability to generalize to different conditions and orientations.

In order to train the land cover classification model, we generated a set of masks manually for each image in the dataset, as depicted in Figure 3. These masks served the purpose of labeling each pixel in the image according to its respective land cover class, encompassing categories such as forest, water, urban areas, roads, and agricultural land. Although the process of manually creating these masks is laborious and time-consuming, it is a crucial step that allows us to attain higher accuracy and reliability in our results. By undertaking this meticulous approach, we contribute to the advancement of land cover mapping, enabling more precise analysis and understanding of land cover patterns and changes.

C. Augmentation

Data augmentation is a technique which is particularly useful when working with limited datasets. It increases the size of a dataset by producing new, altered versions of the existing data which helps to improve model performance by providing additional training examples that capture a wider range of variation in the data.

There are several ways to augment a dataset, including:

1) Rotation: Rotating an image by a certain angle can create new training examples with different orientations. This is particularly useful when working with objects or scenes that can appear in different orientations.

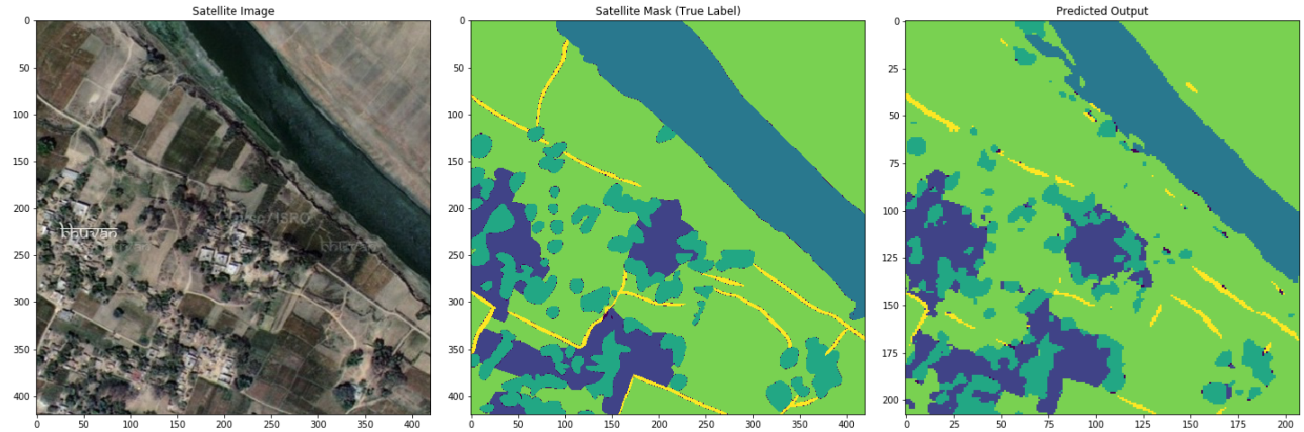

Fig. 4. Input Satellite Image along with its corresponding mask, and the predicted land cover output produced by the Resnet50 U-Net machine learning model

2) Flip: Flipping an image horizontally or vertically can create new examples with different orientations or perspectives. This is useful for object detection or scene understanding tasks.
3) Translation: Moving an image in the horizontal or vertical direction can create new examples with different positions in the frame. This is useful for object detection tasks.
4) Shearing: Applying a shear transformation to an image can create new examples of different shapes or orientations, which can help the model learn to recognize objects in different perspectives.
5) Color augmentation: Altering the brightness, contrast, or saturation of an image can create new examples with different lighting conditions, which can assist the model in learning to recognise objects in various lighting conditions.

Overall, image augmentation is a powerful technique, that helps to prevent the model from over-fitting and improves its accuracy on previously unseen data. It allows us to generate more data from existing images, making our model more robust and capable of handling a wider range of real-world scenarios.

V. Experiment and Performance Evaluation

We divided the dataset into a training set and a test set, using an 80:20 ratio. The training set comprised 80% of the dataset, while the remaining 20% constituted the test set. This division ensured that the models were trained on a sufficiently large and diverse set of images, while also allowing for an unbiased evaluation on unseen data. We trained the Resnet50 and VGG U-Net model using the augmented training set. The training process involved feeding the Bhuvan satellite images and their corresponding masks into the model, optimizing the model parameters, which in turn predicts the land cover map, i.e. predicted output, as shown in Fig. 4.

To comprehensively evaluate the performance of the trained models, we measured several metrics, including accuracy, Intersection over Union (IoU), Dice coefficient (F1-Score), and classification report.

1) IoU: IoU, also known as the Jaccard index, calculates the overlap between predicted and ground truth masks. It measures the similarity between the predicted and actual land cover regions.

$$\text{IoU} = \frac{TP}{TP + TN + FN} \tag{1}$$

2) Dice Coefficient: The Dice coefficient is another measure of similarity between the predicted and ground truth masks. It quantifies the agreement between the two masks, emphasizing the common regions while penalizing false positives and false negatives.

$$\text{Dice} = \frac{2TP}{2TP + TN + FN} \tag{2}$$

3) Classification Report: The classification report offers a comprehensive overview of performance metrics for individual land cover classes, providing details such as precision, recall, F1-score, and support (occurrence count).

These metrics offer valuable insights into the model's capability to accurately classify diverse land cover types and provide a comprehensive assessment of its overall performance.

TABLE II
U-Net Models: Performance Metrics for Different Epochs

Epochs	Epochs	Mean IoU	Mean Dice Coefficient	Pixel Accuracy
VGG16	5	0.56	0.68	0.88
	10	0.58	0.70	0.88
	15	0.56	0.68	0.89
Resnet50	5	0.58	0.70	0.90
	10	0.54	0.67	0.88
	15	0.56	0.68	0.88

Performance metrics of VGG16 and Resnet50 with U-Net Model for different epochs are listed in Table II. The highest performance in terms of Dice coefficient is achieved by the Resnet50 model at epoch 5, while the VGG16 model achieves its highest performance at epoch 10, both with a score of 0.70. It can be observed that the Resnet50 model outperforms the VGG16 model with higher accuracy achieved at the 5th epoch, indicating more efficient training time. It is important to highlight that even though epoch 15 has a high pixel accuracy, it does not necessarily imply superior segmentation ability.

TABLE III
U-NET MODEL: BEST PERFORMANCE METRICS FOR DIFFERENT LAND COVER CLASSES

	VGG16		Resnet50	
Classes	**IoU**	**Dice Coefficient**	**IoU**	**Dice Coefficient**
Urban	0.425	0.597	0.381	0.551
Water	0.785	0.879	0.806	0.893
Forest	0.601	0.751	0.633	0.775
Agriculture	0.864	0.927	0.894	0.944
Road	0.215	0.354	0.200	0.333

Table III presents the performance metrics of the VGG16 and Resnet50 with U-Net model for different land cover classes. Scores for the "Road" and "Urban" classes are comparatively lower than the scores for "Water," "Forest," and "Agriculture". This difference in scores can be attributed to the frequency of occurrence of each class in the dataset. The model may not have had sufficient examples or exposure to accurately classify the "Road" and "Urban" classes. To further improve the scores of IoU and Dice coefficient for these classes, we can train the model on a larger dataset that includes more instances of these specific classes. This will provide the model with more representative samples and enhance its ability to accurately segment and classify the "Road" and "Urban" classes.

VI. CONCLUSION

In this paper, we have proposed an approach for land cover classification utilizing VGG16 and Resnet50 with U-Net architectures, leveraging Bhuvan 2D satellite images and manually created masks for accurate classification. Our models achieves a training accuracy of 96.63%, test accuracy of 90% with dice coefficient of 0.70. The findings of this study showcase the efficacy of the VGG16 and Resnet50 U-Net models in semantically segmenting satellite images, emphasizing the significance of developing precise and efficient algorithms for land cover classification. The models can produce high-resolution land cover maps that can be used for a wide range of applications, including environmental monitoring, urban planning, and natural resource management.

REFERENCES

[1] A. Chatterjee, J. Saha, J. Mukherjee, S. Aikat, and A. Misra, "Unsupervised land cover classification of hybrid and dual-polarized images using deep convolutional neural network," *IEEE Geoscience and Remote Sensing Letters*, vol. 18, no. 6, pp. 969–973, 2020.

[2] Y. Zhou, H. Wang, F. Xu, and Y.-Q. Jin, "Polarimetric sar image classification using deep convolutional neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 12, pp. 1935–1939, 2016.

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

[4] J. Saha and J. Mukherjee, "Recal: Reuse of established cnn classifier apropos unsupervised learning paradigm," in *Computer Vision, Pattern Recognition, Image Processing, and Graphics: 7th National Conference, NCVPRIPG 2019, Hubballi, India, December 22–24, 2019, Revised Selected Papers 7*. Springer, 2020, pp. 174–184.

[5] A. Rakhlin, A. Davydow, and S. Nikolenko, "Land cover classification from satellite imagery with u-net and lovász-softmax loss," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 262–266.

[6] DeepGlobe, "Deepglobe: A large-scale dataset for semantic segmentation of urban areas," 2018. [Online]. Available: http://deepglobe.org/resources.html

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

[8] M. Berman, A. R. Triki, and M. B. Blaschko, "The lovász-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4413–4421.

[9] H. Qassim, A. Verma, and D. Feinzimer, "Compressed residual-vgg16 cnn model for big data places image recognition," in *2018 IEEE 8th annual computing and communication workshop and conference (CCWC)*. IEEE, 2018, pp. 169–175.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[11] M. Papadomanolaki, M. Vakalopoulou, S. Zagoruyko, and K. Karantzalos, "Benchmarking deep learning frameworks for the classification of very high resolution satellite multispectral data," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 3, p. 83, 2016.

[12] R. Gharbia, N. E. M. Khalifa, and A. E. Hassanien, "Land cover classification using deep convolutional neural networks," in *International Conference on Intelligent Systems Design and Applications*. Springer, 2020, pp. 911–920.

[13] K. Sawant, R. Prakash, A. Vidyarthi, A. Ramola, and A. K. Shakya, "Analysis of metropolitan land use classification for agriculture, forest, and river beds: A case study of dehradun, uttarakhand, india," in *2020 International Conference on Advances in Computing, Communication & Materials (ICACCM)*. IEEE, 2020, pp. 71–75.

[14] R. Gupta and M. Shah, "Rescuenet: Joint building segmentation and damage assessment from satellite imagery," in *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2021, pp. 4405–4411.

[15] "Bhuvan - Indian Geo-Platform of ISRO," https://bhuvan-app1.nrsc.gov.in/bhuvan2d/bhuvan/bhuvan2d.php, accessed: Sept 10, 2022.