TerraGrid: Harnessing Deep Learning Models for Satellite Image Segmentation

Nitin Pal
Vidyalankar Institute of Technology
Mumbai, Maharashtra, India

Soumya Ramkrishna
Vidyalankar Institute of Technology
Mumbai, Maharashtra, India

Harsh Patil
Vidyalankar Institute of Technology
Mumbai, Maharashtra, India

Nishant Choudhary
Vidyalankar Institute of Technology
Mumbai, Maharashtra, India

Rajashree Soman
Vidyalankar Institute of Technology
Mumbai, Maharashtra, India

ABSTRACT
Satellite imaging is a backbone of environmental monitoring and urban planning. State-of-the-art methodologies for image segmentation are extremely needed to process huge datasets in order to obtain substantial information out of it. This paper presents an extensive analysis on the current state of satellite image segmentation using deep learning with a special emphasis on some CNN architectures, namely InceptionResNetV2, InceptionResNetV2-UNet, Multi-UNet, VGG19, and VGG19-UNet. In this work, a number of pre-processing techniques were used namely advanced data augmentation and normalization. The set of experiments is done in the motivation to evaluate models in the best way based on a few performance metrics, such as accuracy, Dice coefficient, and validation loss among others, compared with the other models. The conducted experiments have shown that the InceptionResNetV2-UNet and VGG19-UNet provide higher segmentation accuracy compared to other models by approving a test accuracy of 89.90% and 89.41% with the dice coefficient of 85.5% and 84.9%, respectively. Further from this, proposed work introduce a Gradio-based web application for the end users to predict segmented images interactively, demonstrating real-world use cases for employed models. This bridges the gap between advanced machine learning research and satellite image evaluation, providing valuable insights for future work in this field.

Keywords
cnn, inceptionresnetv2-unet, image segmentation, satellite imagery, semantic segmentation, vgg19-unet

1. INTRODUCTION
Satellite imagery has become an indispensable tool in a variety of domains, including environmental monitoring and agriculture to disaster management and urban planning. The massive extent of data generated by Earth-observing satellites present both opportunities and challenges. To harness the potential of this data effectively, advanced processing techniques are essential [1]. Among these techniques is image segmentation. It is the method of partitioning an image into meaningful regions, is a critical step in extracting valuable information from satellite imagery [2]. Accurate segmentation enables the recognition of objects, land cover types, and changes over time, facilitating decision-making processes in numerous fields [3]. Image classification is a process in computer vision which classifies images based on their visual content based on predefined categories [4]. Image classification is a big part of machine learning and is already a well-established and highly applicable concept. Visual search APIs, image and facial recognition on social media platforms, object recognition models, remote sensing, medical diagnosis, and teaching are popular applications of this technology [5].

One can utilize a variety of machine learning algorithms to classify images. Data science provides a plethora of various algorithms that can be used for image classification [6]. One problem is that even though different classifier techniques can classify the images, determining which of these methods will produce more accurate findings is difficult.

The goal of this paper is to present a thorough examination of the most recent methods for segmenting satellite images using neural networks. It aims to explore the different neural network architectures, pre-processing techniques, loss functions, and evaluation metrics utilized in this field. Additionally, a comparative analysis is done of machine learning models to determine the accuracy of which model is best suited for the dataset.

The rest of the paper is structured as follows. Section 2 reviews the extant literature. Section 3 explains the research methodology. Section 4 presents the experimental results along with a discussion of the findings. Section 5 summarises the paper.

2. LITERATURE REVIEW
Deep learning is enabling phenomenal changes in the semantic segmentation of satellite data, improving information extraction from extensive satellite data sets. The ability to learn very fast and adapt to complex image patterns refines decision-making in many fields. In [7] two deep network architectures, UNet and Inception ResNet UNet are implemented and evaluated in automatic building detection from aerial imagery, they achieved highest accuracy of 97.95% and 0.96 in the dice metric. The InceptionResnet-UNet could detect buildings of different shapes, structures, textures, and colors in images in almost all regions, though UNet couldn't detect very large buildings. That is because the Inception ResNet UNet model is wide and deep with few variations in the number of parameters in comparison to UNet [8].

To boost the UNet performance, employing transfer learning from two different encoders, namely the VGG and ResNet. Moreover, using transfer learning, they developed a robust model unaffected by the similarity between different types of discontinuities in noisy seismic data which is improved by utilizing the skip connections strengths of ResNet model [9]. This paper [10] analyzes semantic segmentation using deep learning for autonomous driving, comparing 12 dataset along with 14 frameworks, and various data augmentation techniques. Study work in [11] categorizes classic segmentation algorithms like Edge Detection, Clustering Method, Random Walks, Co-Segmentation Methods like MRF-based and Object-Based, and the popular deep learning algorithms, highlighting their benefits and chronological progression in image segmentation technology.

The VGG network has a good and simple structure with deficient number of layers and is used in many computer vision fields. The VGG19 network is an innovative object recognition model and is actually a deep CNN that performs very well in many tasks and datasets. The VGG19 algorithm was used to extract the roads [12], which is implemented in two steps. The first stage is segmentation, and the next stage is edge detection so that roads can be extracted. The authors in [13] also took CNN models of ResNet and InceptionNet along with the model AlexNet for the transfer learning in the classification of COVID-19 and found that they had the most prominent outcome of 93% accuracy. This research [14] proposed a method for detection of COVID-19 to make better diagnostic decisions using ResNet50v2 model with Bayesian CNN. This study [15] employed Auto-tuned inception-v4 (HPTI-v4) model for segmentation tasks, unleashing the power of inception model by achieving extraordinary results of up to 99% test accuracy.

The reported results shown in the studies by [16] were IoU with 55% and F1-Score 71%, measured on the Massachusetts building dataset. A Study [17] achieved 89% of F1 with the use of UNet architecture on the SpaceNet V2 dataset. Another case-study [18] accomplished land segmentation in the INRIA building dataset, achieving a value of 60%. The research [19] obtained 78% F1-Score by utilizing the DeepLabv3+ model for image segmentation. Another research [20] achieved 95% F1-Score using the UNet architecture. In [21], the authors state their result of 93.8% IoU metric for the Unet++ architecture with the SE-ResNeXt101 encoder, pre-trained on the ImageNet. In [22], the authors state their result of 93.8% IoU metric for the Unet++ architecture with the SE-ResNeXt101 encoder, pre-trained on the ImageNet. The research carried by Burcu et al. [23] on segmentation of building images employing UNet and UNet++ architectures achieved highest and lowest score of 0.9167 and 0.6140, respectively.

3. PROPOSED METHODOLOGY

The figure 1 depicts the proposed methodology showcasing a systematic approach to achieve defined research objectives. In this approach five distinct models were implemented and assessed to determine which model can thought to be most appropriate for the dataset. The dataset was acquired from a publicly available source namely Semantic Segmentation of Aerial Imagery Dataset on Kaggle community. The original dataset contained 72 images consists of 6 classes, which was further augmented using an image generator to generate a dataset containing 648 images. These images were trained in numerous models, with 70 images used for validation and 578 images for training.

3.1 Data pre-processing

The pre-processing of data is crucial for making the data more diverse and hence, robust for the model. With this aim in mind, various augmentations were done using the Albumentations Python Library: random cropping, flipping, rotation, alteration in brightness, or contrast. Albumentations efficiently performs multiple image transformation operations, all implemented for fast performance and having a clean but powerful interface for image augmentation, which is applied in all areas of computer vision, such as object classification, segmentation, and detection [24]. Such augmentations alleviate the problem of overfitting by enriching the number of image variations input into the models. Addition to this, other more complex ones were utilized besides the basic augmentations including Contrast-Limited Adaptive Histogram Equalization (CLAHE), grid distortions, and optical distortions. These further capture a wider range of complications within the dataset and hence improve the generalization capability of the model on unseen data.

Later, the normalization of pixel values within the range of (0, 1) was scaled by performing a division by 255 so that for whatever neural network it was fed to, the input was consistent. The data is further normalized by subtracting the dataset mean and then dividing by the standard deviation to bring more stability during the process of learning. RGB mask images are encoded into one-hot arrays so that categorical pixel labels are converted into a readable format for the model. This is also inclusive of histogram equalization if required to enhance the contrast in images. The images are resized to a standard dimension of (512, 512), which is the general size [25], therefore ensuring that the training process becomes computationally efficient.

3.2 Training and Evaluation

The training process utilized deep learning models namely InceptionResNetV2, InceptionResNetV2-UNet, Multi-UNet, VGG19, VGG19-UNet. These models were selected due to the strength of their architecture and proven success in relation to complex data distributions. More specifically, from the feature extracting capabilities of models like InceptionResNetV2 and VGG19, combined with UNet segmentation capabilities, very high accuracy was expected with semantic segmentation tasks as described in studies [26], [27].

The training was done with pre-trained weights most of the time, directly from ImageNet. Transfer learning techniques naturally freeze several of the layers, which are gradually unfrozen while fine-tuning the models for the segmentation task. Optimization was carried out using either stochastic gradient descent or the Adam optimizer, both with an initial learning rate of 0.0001. In practical training, learning rate schedulers were used for better model convergence and allowed techniques such as learning rate warm-up and decay. The regularization techniques used included dropout and batch normalization to avoid overfitting. Early stopping also had the added function of automatically stopping in case the performance on the validation set did not improve anymore, hence avoiding overfitting.

The performance of the trained models is evaluated using several metrics. Standard metrics for segmentation are the Intersection over Union (IoU), and pixel accuracy. Also, cross-validation is conducted in order to make sure the model generalizes well on different splits of data. Aside from these metrics, higher-order metrics such as the Dice coefficient and

accuracy metrics, and validation loss with respect to assessing the experimental results for further model performance validation, were also computed. The Dice coefficient is basically an overlap computation that is carried between the model-predicted outcomes against the ground truth masks [28] in a range (0, 1).

3.3 Model Architecture

3.3.1 InceptionResNetV2

The InceptionResNet architecture is a deep neural network design that blends aspects from both Inception and ResNet architectures. It aims to achieve superior performance in image classification tasks while maintaining computational efficiency and ease of training. InceptionResNet utilizes Inception modules [29], which include of many parallel convolutional branches with varying filter sizes. This allows the network to effectively collect data at different scales. These modules utilize 1x1 convolutions for reducing dimensionality and increasing non-linearity without significantly increasing computational cost, alongside 3x3 and 5x5 convolutions to capture spatial hierarchies. Additionally, pooling operations such as max pooling and average pooling are used for down-sampling feature maps and capturing global information.

A distinctive feature of InceptionResNet is the inclusion of residual connections, which address the vanishing gradient complication and expedite training of complex networks. These connections enable shortcuts bypassing network layers, with each block containing convolutional layers and element-wise addition with input. The architecture typically begins with a stem block for initial feature extraction and dimensionality reduction, followed by multiple Inception-ResNet blocks and reduction blocks that further process features and reduce spatial dimensions.

3.3.2 InceptionResNetV2-UNet

The InceptionResNet-UNet architecture merges features from InceptionResNet and UNet to improve semantic segmentation accuracy efficiently. It combines InceptionResNet's feature extraction with UNet's precise localization abilities [30]. The architecture consists of an encoder, based on InceptionResNet, which extracts high-level features using convolutional and pooling layers. Inception modules within the encoder capture features at different scales effectively. The decoder, inspired by UNet [31], performs up-sampling and concatenation operations to refine segmentation masks.

3.3.3 Multi-UNet

The Multi-UNet architecture is designed to improve semantic segmentation accuracy by integrating multiple UNet models into a single framework. It leverages the ensemble effect of multiple models to enhance segmentation performance. The architecture consists of multiple UNet-like networks, each serving as a branch within the model. These branches process input images independently and generate segmentation masks. Each UNet branch incorporated with encoder-decoder structure [31]. The encoder uses convolutional and pooling layers to extract features from input images, capturing both low-level and high-level aspects.

3.3.4 VGG19

VGG19 is a 19-layer deep convolutional neural network (CNN) designed for image classification tasks. It begins with simple 3x3 convolutional layers, followed by max-pooling layers to reduce spatial dimensions [9]. The network repeats this pattern multiple times, learning complex features from images, and eventually having fully connected layers for classification.

3.3.5 VGG19-UNet

The VGG19-UNet architecture merges the VGG19 CNN architecture with the UNet architecture, aiming to improve semantic segmentation accuracy. This fusion combines VGG19's strong feature extraction capabilities with UNet's precise localization abilities. At its core, VGG19-UNet consists of an encoder-decoder structure. The encoder part utilizes VGG19, featuring convolutional and max-pooling layers that progressively down-sample the image to capture hierarchical features. These layers utilize small 3x3 convolutional filters for max-pooling and feature extraction layers for spatial down sampling [9]. The decoder part, inspired by UNet [7], performs up-sampling and concatenation operations to recover spatial information and refine segmentation masks. Each decoder block includes up-sampling layers followed by concatenation with corresponding features from the encoder, preserving detailed information.

4. RESULTS AND DISCUSSION

Figure 2 details the model performance of InceptionResNetV2-UNet on an aerial image. The figure consists of three panels, where the first panel contains the input image, the aerial view of a residential and commercial area, very informative for Geographic Information Systems (GIS) tasks such as urban planning, infrastructure management, and environmental evaluation. It is the center panel, displaying the ground truth mask-a human-annotated representation of part of the image that only outlines separate regions or classes-a gold standard against which to evaluate the segmentation accuracy. The third is the predicted mask, produced by the model, essentially an approximation of ground-truth segmentation of an input image. A comparison between the predicted mask and the ground truth mask presents the efficacy of the model in covering a particular region and structure. Regions that the predicted mask closely fits with the ground truth demonstrate the respective strengths of the model in identifying such classes with precision. The perceived differences, typically manifesting as color or boundary differences between the ground truth and predicted masks, speak to the model's weakness in covering other classes with the same details. This really makes it necessary to verify the precision, recall, and accuracy of the model as against each class since it tends to vary wildly with the complexity and distinctiveness of features in each region. The figure also underscores that, while the model InceptionResNetV2-UNet does well at providing a segmentation for some classes, such as roads or water bodies, it will fail in others, possibly because of overlapping features, textures similar to each other, or even ambiguous boundaries.

The remaining figures (Figures 3, 4, 5, and 6), by which one may visually compare the Vgg19-UNet, InceptionResNetV2, Vgg19, and Multi-UNet models, respectively, together give an implication of the kind of weaknesses and the strengths of each architecture, and hence the importance of evaluation could be necessary before picking a model for applications requiring accurate class-specific segmentation.

Table 1 concludes the performance of several architectures with attention to accuracy, loss, and Dice Coefficient at both the training and validation phases on unseen datasets. Data for validation accuracy inform about predictive correctness that range from between 78.49% and 89.90% that was most notably recorded by InceptionResNetV2-UNet, which is almost a best fit for highly predictive performance. Loss values, that represents error rate, range from 0.290 to 0.600. For Vgg19-UNet the lowest value was reached, so the model with UNet in its structure can be less erroneous. The Dice Coefficient of how good the predicted samples look like true samples ranges

between 0.650 and 0.857. InceptionResNetV2-UNet and Vgg19-UNet demonstrated the highest values for the given coefficient, which means their goodness in detecting complex points and relations in the data. The comparison of detailed architectures presents some striking advantages UNet-augmented architectures constantly exhibit higher performances compared to their basic models in the sense of precision, error reduction, and similarity measures. Such results become radically important for guiding model selection for applications requiring high precision, reliability, and robust generalization capabilities.

Table 1. Validation and Training Result Comparison across Models

Model	Type	Evaluation Metrics		
		Accuracy (%)	Loss	Dice Coefficient
Inception ResNetV 2	Training	75.12	0.687	0.650
	Validation	78.49	0.600	0.697
Inception ResNetV 2-UNet	Training	91.61	0.269	0.846
	Validation	89.90	0.290	0.855
Multi-UNet	Training	83.50	0.491	0.731
	Validation	84.73	0.448	0.763
Vgg19	Training	76.38	0.648	0.670
	Validation	78.98	0.584	0.695
Vgg19-UNet	Training	91.60	0.251	0.857
	Validation	89.41	0.308	0.849

The Gradio web app depicted in Figure 7 provides an interface for users to predict segment

learning models. Based on the results depicted in Table 1, the best performing models are considered for user interface web app. The interface allows users to make a selection between the best model namely InceptionResNetV2-UNet or Vgg19-UNet and upload an image. Once an image and model are chosen, the application pre-processes the image, resizing it to fit the model's input requirements. The selected model then predicts the mask for the image consisting of 6 classes namely land, water, vegetation, building, road, and unlabeled and were used for training, the prediction will show predicted mask (road, building, unlabeled and land classes will be combined into land class) highlighting the 3 classes namely land, vegetation, and water. The predicted mask is converted into an RGB image, visually representing the segmentation results. Additionally, the application calculates the percentage area of each class and generates a pie chart for visualization. Users can interact with the interface to view the predicted masked image and explore the distribution of class areas in the satellite image.

Gradio is a Python library that simplifies the creation of customizable and interactive interfaces for deep learning models. The interface offers a user-friendly interface for interacting with machine learning models via web applications, eliminating the need for web development expertise. Its intuitive design enables users to input data, visualize model predictions, and interpret results in real-time, enhancing model understanding and usability. With Gradio, deploying machine learning models becomes effortless, empowering researchers and practitioners to share and demonstrate their models with a wider audience. The tool's seamless integration with existing Python frameworks significantly enhances machine learning access and promotes collaboration within the AI community. The interactive tool illustrated in figure 7 depicts a valuable insight into the performance of different segmentation models and facilitates informed decision making for environmental monitoring and urban planning.

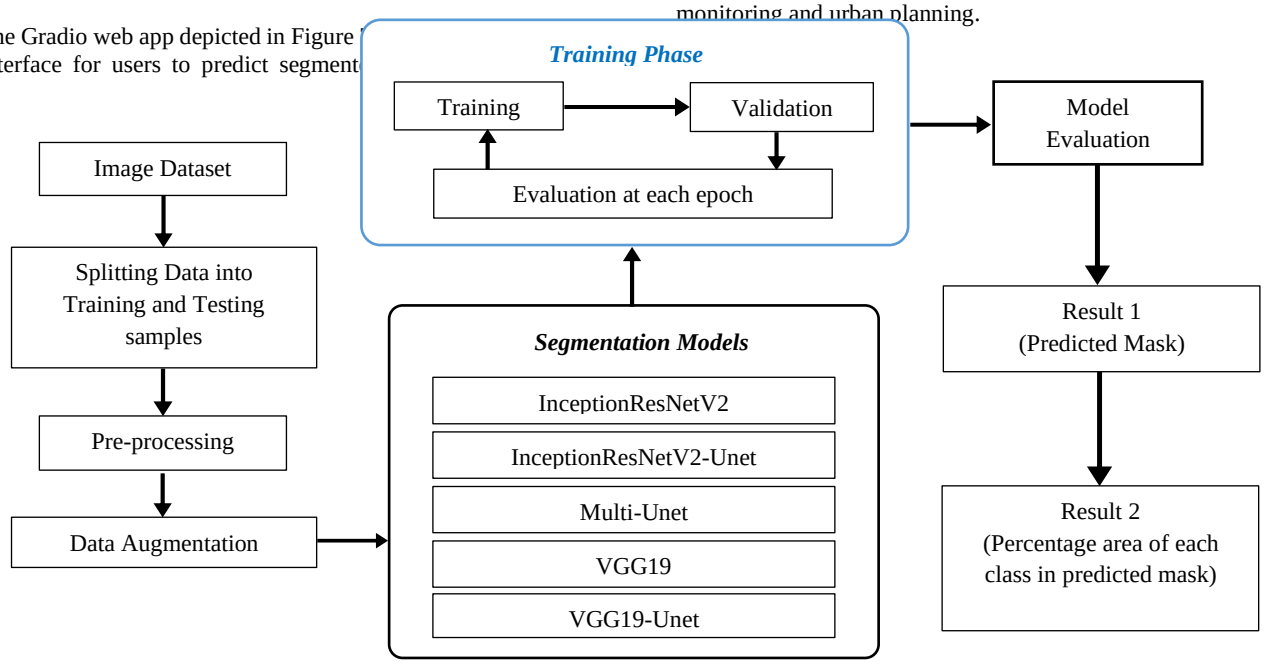

Figure 1: The Workflow of proposed methodology

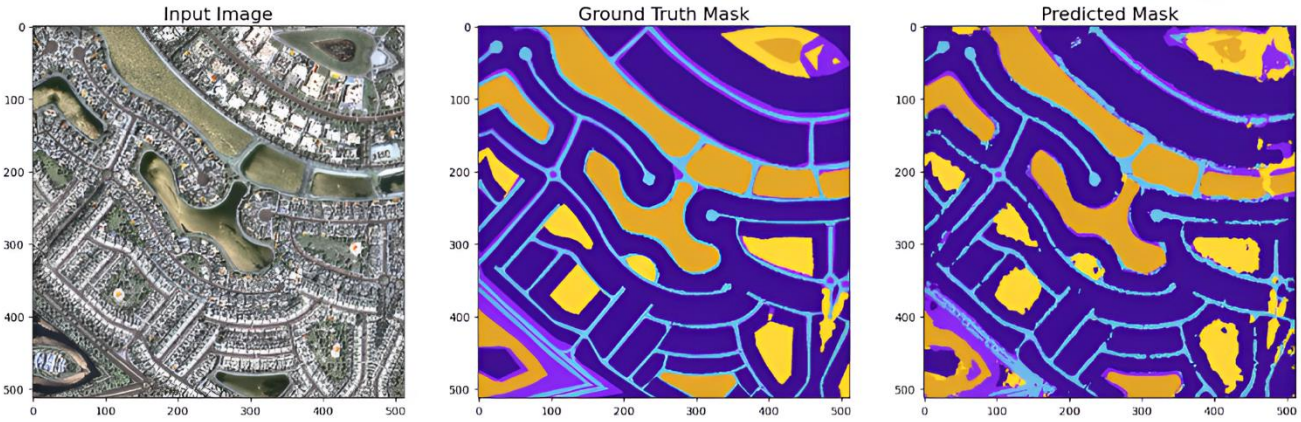

Figure 2: Difference in Ground Truth Mask and Predicted Mask for InceptionResNetV2-UNet Model

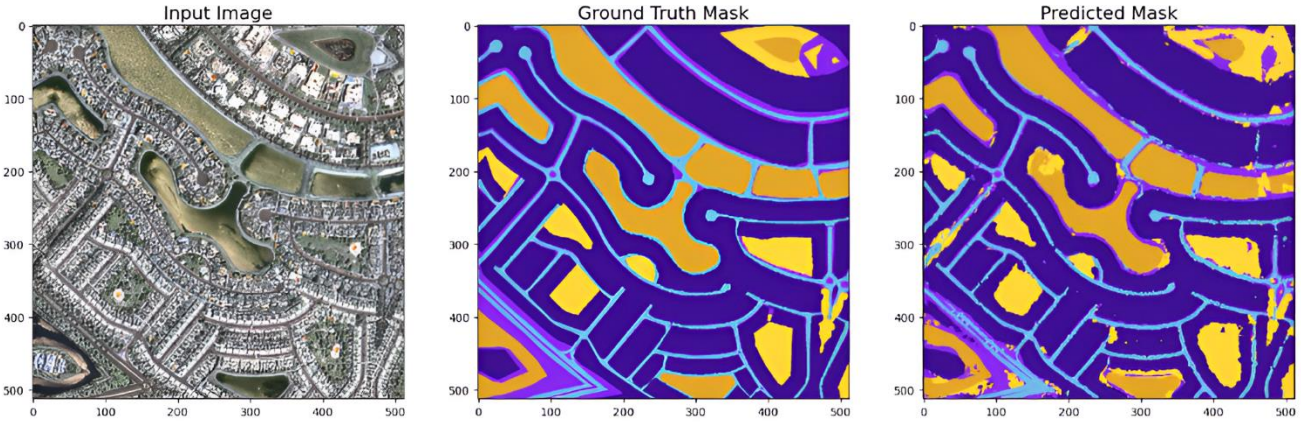

Figure 3: Difference in Ground Truth Mask and Predicted Mask for VGG19-UNet Model

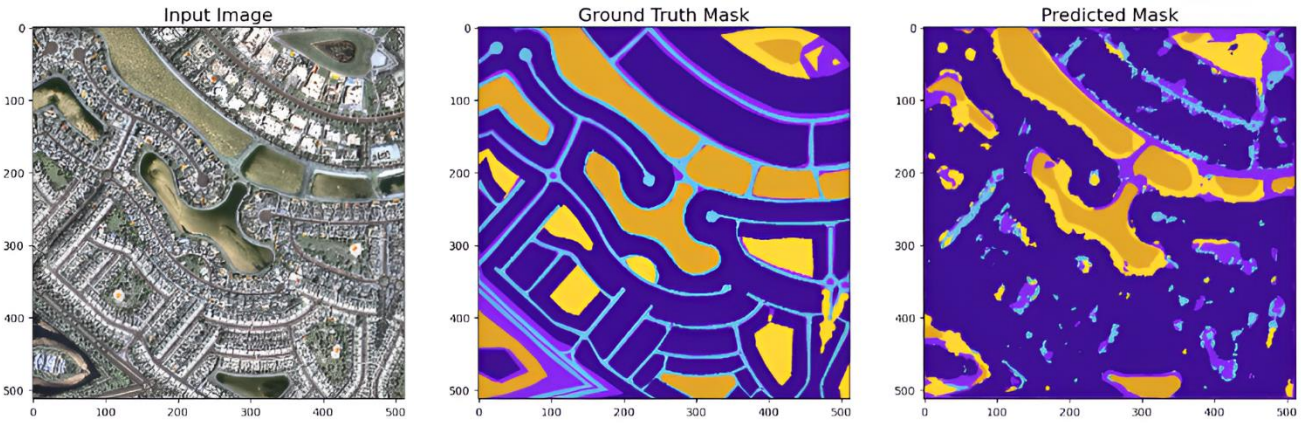

Figure 4: Difference in Ground Truth Mask and Predicted Mask for InceptionResNetV2 Model

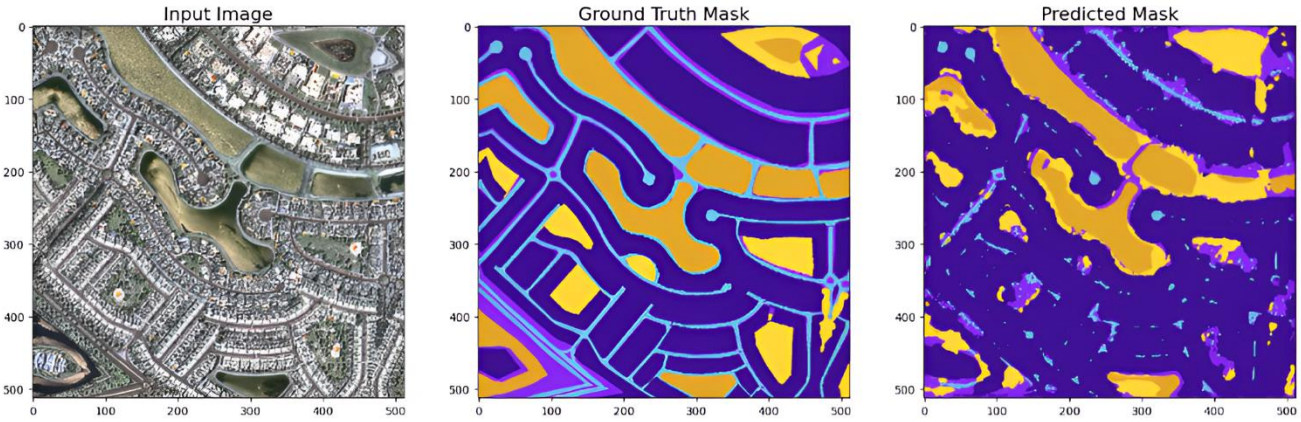

Figure 5: Difference in Ground Truth Mask and Predicted Mask for VGG19 Model

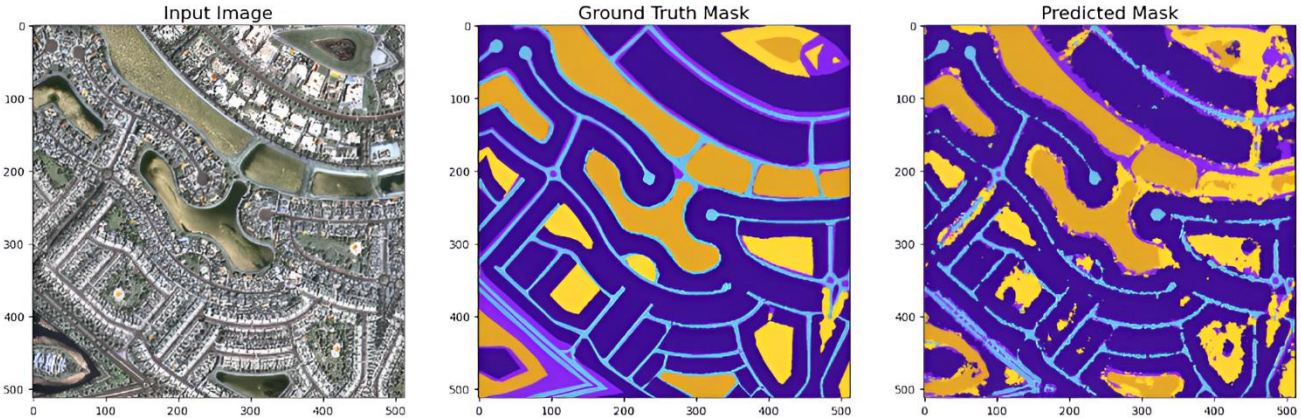

Figure 6: Difference in Ground Truth Mask and Predicted Mask for Multi-UNet Model

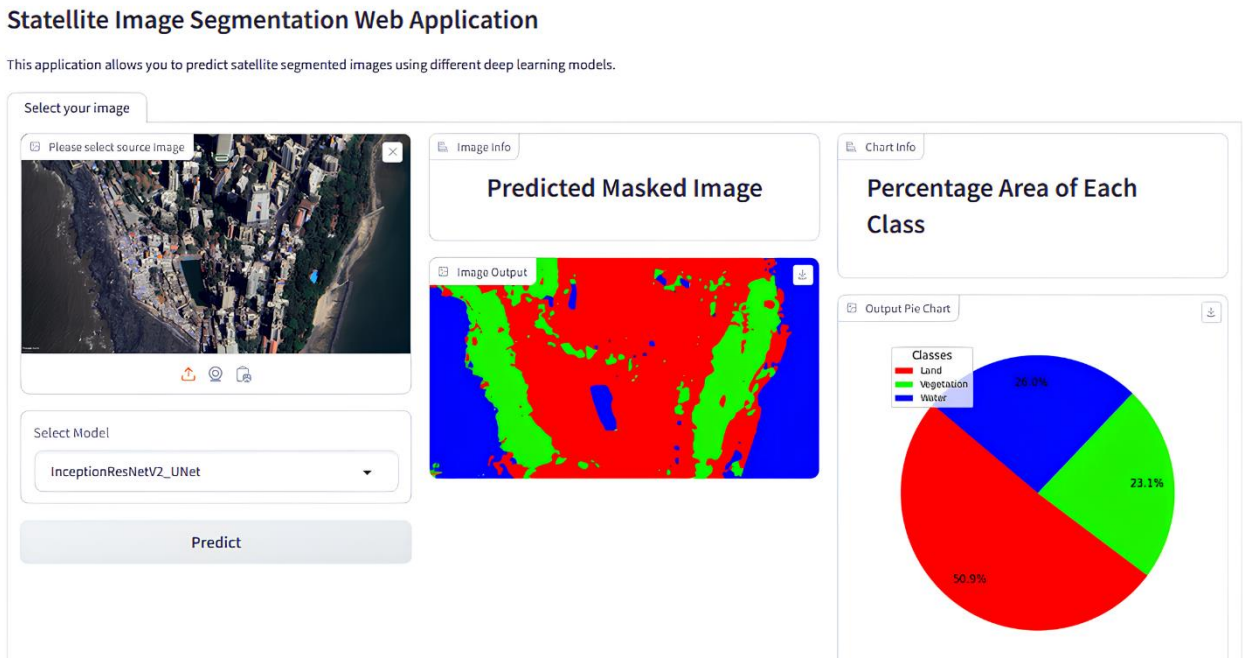

Figure 7: Gradio based User Interface for satellite image segmentation

5. CONCLUSION

The study evaluates five CNN models for satellite image segmentation, with InceptionResNetV2-UNet achieving the highest segmentation accuracy and Dice coefficient, attributed to the strong combination of InceptionResNet's feature extraction and UNet's localization capabilities. Pre-processing techniques like image augmentation and normalization improve model robustness. Key metrics like pixel accuracy, IoU, and Dice coefficient confirm the architecture's potential for precise segmentation tasks. These models are able to leverage prior information from huge datasets using pre-trained weights and transfer learning, but regularization methods avoid overfitting and make good generalizations for unseen data. The practical applicability of these models is demonstrated by the Gradio web application, which enables prediction by making the interactive, visualized segmented images. The gap between advanced machine learning research and real-world applications, especially those involving urban planning, disaster management, and environmental monitoring, will be narrowed by this work, highlighting the transformative potential of AI in satellite image analysis.

Future directions involve coherent advanced architectures like transformer-based models, along with attention mechanisms and self-supervised learning techniques. Other improvements also include the addition of Gradient-Weighted Class Activation Mapping (Grad-CAM) to further enhance the explain-ability of the CNN models. Challenges include dataset limitations, model complexity, and computational resources, which can be addressed by acquiring larger, diverse datasets, developing lightweight architectures, and implementing uncertainty estimation and interpretability techniques. Embracing these challenges and advancements in machine learning can expedite progress towards accurate, efficient, and interpretable satellite image segmentation solutions.

6. ACKNOWLEDGMENTS

Our thanks to the experts who have contributed towards development of this research project.

7. REFERENCES

[1] Vance, T. C., Huang, T., and Butler, K. A. 2024. Big data in Earth science: Emerging practice and promise. Science.

[2] Garea, S. A. and Das, S. 2024. Image Segmentation Methods: Overview, Challenges, and Future Directions. 2024 Seventh International Women in Data Science Conference at Prince Sultan University (WiDS PSU).

[3] Tong, X., Xia, G., Lu, Q., Shen, H., Li, S. You S., and Zhang, L. 2020. Land-cover classification with high-resolution remote sensing images using transferable deep models. Remote Sensing of Environment.

[4] Toennies, K. D. 2024. Image Classification: A Computer Vision Task. In: An Introduction to Image Classification. Springer.

[5] Pearce, W., Özkula, S. M., Greene, A. K., Teeling, L., Bansard, J. S., Omena J. J., and Rabello, E. T. 2018. Visual cross-platform analysis: digital methods to research social media images. Information. Communication & Society.

[6] Sheth, V., Tripathi, U., and Sharma, A. 2022. A Comparative Analysis of Machine Learning Algorithms for Classification Purpose. Procedia Computer Science.

[7] Aghayari, S., Hadavand, A., Mohamadnezhad Niazi, S., and Omidalizarandi, M. 2023. Building Detection from Aerial Imagery using Inception Resnet UNET and UNET Models. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences.

[8] Peng, C., Liu, Y., Yuan X., and Chen, Q., 2022. Research of image recognition method based on enhanced inception-ResNet-V2. Multimedia Tools Application.

[9] Alfarhan, M., Deriche, M., and Maalej, A. 2020. Robust Concurrent Detection of Salt Domes and Faults in Seismic Surveys Using an Improved UNet Architecture. IEEE Access.

[10] Haneen, A., and Ahmad. M. B. 2022. Deep Learning-Based Frameworks for Semantic Segmentation of Road Scenes. Electronics.

[11] Busra, E. S., Guzel, M. S., Bostanci, G. E., Ekinci, F., Asuroglu, T., and Acici, K. 2023. Deep-Learning-Based Approaches for Semantic Segmentation of Natural Scene Images: A Review. Electronics.

[12] Bouhsissin, S., Sael N., and Benabbou, F. 2021. Enhanced VGG19 Model for Accident Detection and Classification from Video. 2021 International Conference on Digital Age & Technological Advances for Sustainable Development (ICDATA).

[13] Maghdid H. S., Asaad A. T., Ghafoor K. Z., Sadiq A. S., and Khan M. K. 2020. Diagnosing COVID-19 pneumonia from X-ray and CT images using deep learning and transfer learning algorithms. CoRR ArXiv.

[14] Ghoshal, B., and Tucker, A. 2020. Estimating uncertainty and interpretability in deep learning for coronavirus (COVID-19) detection. ArXiv.

[15] Shankar, K., Zhang, Y., Liu, Y., Wu, L., Chen, C. H. 2020. Hyperparameter tuning deep learning for diabetic retinopathy fundus image classification. IEEE Access.

[16] Ji, S., Wei, S., and Lu, M. 2019. Fully Convolutional Networks for Multisource Building Extraction from an Open Aerial and Satellite Imagery Data Set. In IEEE Transactions on Geoscience and Remote Sensing.

[17] Jiwani, A., Ganguly, S., Ding, C., Zhou, N., and Chan, D. M. 2021. A Semantic Segmentation Network for Urban-Scale Building Footprint Extraction Using RGB Satellite Imagery. ArXiv.

[18] Maggiori, E., Tarabalka, Y., Charpiat, G., and Alliez, P. 2017. Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark. 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS).

[19] Chen, Q., Wang, L., Wu, Y., Wu, G., Guo, Z., and Waslander, S. L. 2019. Aerial imagery for roof segmentation: A large-scale dataset towards automatic mapping of buildings. ISPRS Journal of Photogrammetry and Remote Sensing.

[20] Cai, Y., He, H., Yang, K., Fatholahi, S. N., Ma, L., Xu, L., and Li, J. 2021. A Comparative Study of Deep Learning Approaches to Rooftop Detection in Aerial Images. Canadian Journal of Remote Sensing.

[21] Glinka, S., Owerko, T., and Tomaszkiewicz, K. 2022. Using Open Vector-Based Spatial Data to Create Semantic Datasets for Building Segmentation for Raster Data. Remote Sensing.

[22] Bakirman, T., Komurcu, I., and Sertel, E. 2022. Comparative analysis of deep learning based building extraction methods with the new VHR Istanbul dataset. Expert Systems with Applications.

[23] Amirgan, B., and Erener, A. 2024. Semantic segmentation of satellite images with different building types using deep learning methods. Remote Sensing Applications: Society and Environment.

[24] Alexander, B., Vladimir, I., Eugene, K., Alex, P., Mikhail, D., and Alexandr, K. 2020. Albumentations: Fast and Flexible Image Augmentations, Information.

[25] Saponara, S., and Elhanashi, A. 2021. Impact of Image Resizing on Deep Learning Detectors for Training Time and Model Performance. In: Saponara, S., De Gloria, A. (eds) Applications in Electronics Pervading Industry, Environment and Society. ApplePies 2021, Springer.

[26] Siciarz, P., and McCurdy, B. 2022. U-net architecture with embedded Inception-ResNet-v2 image encoding modules for automatic segmentation of organs-at-risk in head and neck cancer radiation therapy based on computed tomography scans. Physics in Medicine & Biology.

[27] Anaya-Isaza, A., Mera-Jiménez, L., Cabrera-Chavarro, J. M., Guachi-Guachi, L., Peluffo-Ordóñez, D., and Rios-Patiño, J. I. 2021. Comparison of Current Deep Convolutional Neural Networks for the Segmentation of Breast Masses in Mammograms. In IEEE Access.

[28] Zhang Y., and Guindon, B. 2016. Application of the Dice Coefficient to Accuracy Assessment of Object-Based Image Classification. Canadian Journal of Remote Sensing.

[29] Szegedy, C., Wei, L., Yangqing, J., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. 2015. Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[30] Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. A. 2017. Inception-v4, inception-ResNet and the impact of residual connections on learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI'17), AAAI Press.

[31] Ronneberger, O., Fischer, P., and Brox, T. 2015. UNet: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention. Springer.