Comparative Analysis of U-Net Models Using ResNet34, InceptionV3, and VGG16 for Satellite Image Segmentation

Anubhav Yadav
Computer Science
KIET Group of Institutions
Delhi-NCR, Ghaziabad, India
anubhavy600@gmail.com

Sushil Kumar
Data Science and Deep Learning (DSDL) Lab
KIET Group of Institutions
Delhi-NCR, Ghaziabad, India
sushil.kumar@kiet.edu

Ajay Varshney
Computer Science
KIET Group of Institutions
Delhi-NCR, Ghaziabad, India
ajay2014.av1@gmail.com

Aayush Sharma
Computer Science
KIET Group of Institutions
Delhi-NCR, Ghaziabad, India
aayushsharma84477@gmail.com

Shivani
Computer Science
KIET Group of Institutions
Delhi-NCR, Ghaziabad, India
shivani.cs@kiet.edu

***Abstract*—Satellite image segmentation is a crucial component in remote sensing applications, facilitating the automated analysis of land cover and land use patterns. This research explores the effectiveness of the U-Net architecture and state-of-the-art deep learning techniques for satellite image segmentation. Utilizing diverse and well-annotated satellite image datasets, the capability of U-Net is explored in capturing intricate spatial features within images, making it a powerful tool for discriminating land cover classes. The experiments were conducted to analyze the results of the U-Net model with the ResNet34, InceptionV3, and VGG16 architectures. The U-Net-based ResNet34 model achieved the highest accuracy of 84.84% and a validation F1-score of 0.6455 after 50 epochs. The findings underscore the potential of U-Net and deep learning techniques to advance the field of remote sensing, providing solutions for real-world challenges such as environmental monitoring, urban planning, disaster management, and precision agriculture.**

Keywords: Satellite Image Segmentation, U-Net, Deep Learning, Transfer Learning, Land Cover Analysis.

I. Introduction

Deep learning (DL) is enabling new advances in satellite image segmentation, with the potential to revolutionize remote sensing in the 21st century [1]. DL models have been shown to achieve state-of-the-art results on a variety of satellite image segmentation tasks, including land cover segmentation, object detection, and scene classification. In this field, research delves into the realm of satellite image segmentation, where pixels represent land cover, unveiling Earth's transformations. The U-Net architecture, initially developed for medical imaging [2] is utilized as a powerful tool for satellite imagery analysis in this paper. This work builds upon previous research and tackles recent challenges in the field. Some challenges are data volume complexity, class imbalance, robustness to noise, and occlusions [3]. We integrate deep learning techniques, augmenting data and enhancing precision. Our approach holds implications for environmental monitoring, urban planning, disaster management, and agriculture. By automating land cover analysis [4], we propel remote sensing into the 21st century.

In our literature review, we explore key trends and advancements in this domain. Techniques such as image enhancement, employing histogram equalization, and adaptive filtering, have been pivotal [5]. Advanced segmentation algorithms [6] and machine learning-based approaches [7] are employed for accurate object recognition [8]. Deep learning, particularly convolutional neural networks (CNNs), plays a crucial role in feature extraction, achieving state-of-the-art results in tasks such as land-cover classification [9]. Satellite image segmentation is enhanced through machine learning algorithms like Support Vector Machines and Random Forests [10]. Temporal analysis, leveraging techniques like Principal Component Analysis, aids in effective change detection [11]. Addressing challenges posed by clouds and noise, various methods including machine learning-based cloud detection algorithms have been proposed [12]. Integration of satellite data with ground-based information [13] and initiatives for open data platforms [14] contribute to a holistic approach. Overall, the evolution of satellite image processing is characterized by advancements in segmentation, deep learning, classification, change detection, and increased data accessibility.

II. Methodology

A. Data Collection and Preprocessing

1) Data Collection: In satellite image segmentation, a dataset of satellite images was collected which was organized into two main categories: training and validation sets. Each category contained satellite images and their corresponding masks, representing the ground truth labels for the images.

2) Image Preprocessing: Several preprocessing steps were applied to the satellite images:

- **Resizing:** The images were resized to a consistent size of 256x256 pixels to ensure uniformity in the dataset.
- **Normalization:** Pixel values in the images were normalized to the range [0, 1] to facilitate training.
- **Augmentation:** Data augmentation techniques, including horizontal and vertical flipping, were applied to the training images to increase dataset diversity and improve model generalization.

3) Mask Preprocessing: Similar preprocessing steps were applied to the masks:

- **Resizing:** The masks were resized to match the dimensions of the corresponding satellite images.
- **One-Hot Encoding:** The masks were converted from RGB format to one-hot encoded format. Each class in the masks was represented as a separate channel in the one-hot encoded mask, with each pixel assigned a value of 0 or 1 based on its class.

4) Data Patching: In satellite image segmentation, the implementation of data patching is crucial for augmenting the dataset and optimizing the training process. This technique involves subdividing large images and corresponding masks into smaller, overlapping patches of 256x256 pixels with a step size of 256 pixels. The primary purpose is twofold: firstly, to expand the training dataset significantly, addressing challenges associated with limited labeled data; and secondly, to preserve intricate spatial information by capturing local details within each patch. This approach not only facilitates the model in learning a more diverse set of features but also enhances its understanding of contextual relationships between adjacent regions. The overall result is improved generalization across diverse landscapes, contributing to more accurate and robust satellite image segmentation outcomes.

5) Data Generators: Data generators were used to efficiently load and preprocess the training and validation data in batches. These generators applied data augmentation and preprocessing steps on the fly, allowing us to work with large datasets without loading them entirely into memory.

B. Model Selection

For our satellite image segmentation task, we selected the U-Net architecture [15] due to its effectiveness in image segmentation tasks. U-Net is a convolutional neural network (CNN) architecture known for its ability to capture fine-grained details in images and perform precise segmentation.

C. Model Training

The U-Net model, designed for satellite image segmentation, underwent training using the provided dataset. The model was configured with the Adam optimizer and a custom loss function that seamlessly integrated categorical focal loss and Jaccard loss, strategically tailored for optimizing accurate segmentation. Throughout the training process, various evaluation metrics, including the intersection over union (IoU) score, F1 score, and accuracy, were diligently monitored. To enhance training efficiency and meticulously track the model's performance, a set of key callbacks was implemented. These included the ModelCheckpoint callback, which saved the best model weights based on the validation IoU score; the ReduceLROnPlateau callback, dynamically adjusting the learning rate when the validation IoU score plateaued; the EarlyStopping callback, which terminated training early if no improvement in the validation IoU score was observed; and the TensorBoard callback, enabling the logging of training metrics and visualization of the model's performance. This comprehensive approach to training, optimization, and performance monitoring collectively ensured the U-Net model's effectiveness in satellite image segmentation.

D. Model Evaluation

The trained U-Net model was evaluated on the validation dataset using the same evaluation metrics as during training. These metrics provided insights into the model's segmentation accuracy and generalization.

E. Inference and Prediction

Finally, we demonstrated how to use the trained model for inference and prediction on new satellite images. We provided code to load a new image, segment it using the model, and visualize the predicted masks alongside the original image. This methodology outlines the key steps involved in our satellite image segmentation project, from data collection and preprocessing to model selection, training, and inference. The U-Net architecture, coupled with data augmentation and preprocessing, enables accurate and efficient classification of objects within satellite imagery.

F. U-Net Architecture

The U-Net architecture is a convolutional neural network (CNN) model that has gained prominence in the field of image segmentation due to its effectiveness in capturing intricate details within images. It was originally introduced for biomedical image segmentation but has found applications in various domains, including satellite image segmentation.

The U-Net architecture is recognized for its unique U-shaped structure, comprising an encoder, bottleneck, and decoder components as shown in Figure 1. The encoder utilizes convolutional and max-pooling layers to extract contextual information, gradually decreasing spatial dimensions and simultaneously expanding feature channels. Concurrently, the bottleneck effectively preserves essential feature information to facilitate accurate object localization. In contrast, the decoder utilizes transpose convolutions to do up-sampling, resulting in the production of a segmentation mask. The efficiency of U-Net is attributed to the incorporation of skip connections, which establish connections between corresponding levels in both the encoder and decoder components. The preservation of high-resolution characteristics in these linkages contributes to the improvement of segmentation precision by effectively integrating comprehensive information from both components.

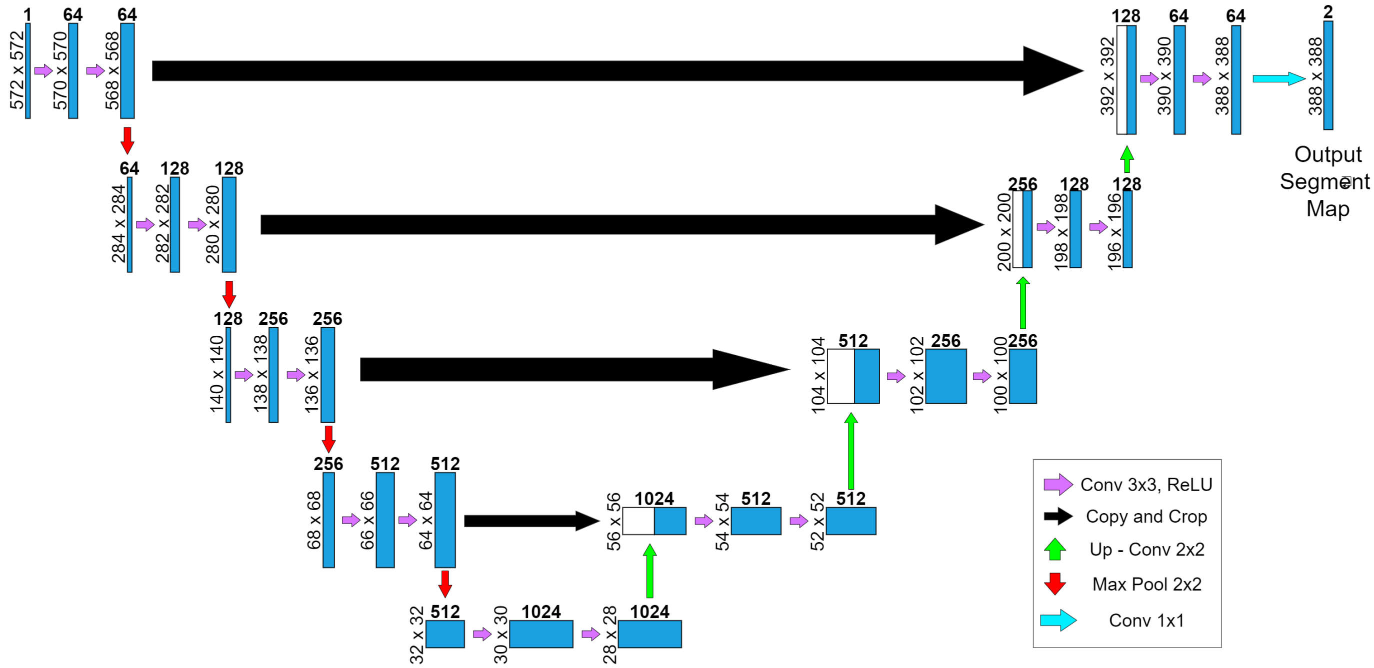

Fig. 1. The Basic Architecture of U-Net Model

G. Suitability for Image Segmentation

The U-Net architecture is particularly well-suited for image segmentation tasks for the following reasons:

- **Capturing Details:** The U-Net's symmetric design with skip connections enables it to capture both global context and fine details within an image, making it ideal for segmentation tasks where precise boundaries are crucial.
- **Hierarchical Features:** The encoder-decoder structure allows the network to learn hierarchical features, from low-level edges to high-level object representations, providing a holistic view of the image.
- **Adaptability:** U-Net can be adapted to handle multi-class segmentation tasks, making it suitable for scenarios where objects of different classes need to be identified within an image.

The U-Net architecture is well-suited for image segmentation due to its unique features. Its symmetric design, featuring skip connections, allows it to effectively capture both global context and intricate details, making it especially adept at tasks requiring precise boundary delineation. The encoder-decoder structure enables the learning of hierarchical features, ranging from low-level edges to high-level object representations, providing a comprehensive understanding of the image. Additionally, U-Net's adaptability makes it suitable for multi-class segmentation tasks, making it effective in scenarios where the identification of objects from different classes within an image is essential.

H. Modifications for Satellite Image Segmentation

To enhance U-Net's efficacy in satellite image segmentation [16], strategic modifications were introduced. Recognizing the multi-class nature of satellite imagery, U-Net was configured to produce multi-channel masks, aligning its architecture with the nuanced complexities present in satellite data.

Addressing the challenge of limited labeled data, an extensive data augmentation strategy was implemented, including rotation and scaling. This proved indispensable in mitigating the scarcity of labeled samples, empowering the model to generalize effectively. A customized loss function, combining focal and Jaccard loss, was employed to address accurate segmentation and class imbalance. The focal loss prioritizes underrepresented classes, while the Jaccard loss accentuates spatial accuracy, substantially elevating the model's performance.

In another study [17], U-Net's accuracy and robustness were enhanced through data augmentation, transfer learning with pre-trained CNNs, and regularization techniques like dropout and weight decay. These collectively fortify U-Net for accurate and robust classification in diverse satellite imagery scenarios. Data augmentation systematically expands the training dataset through diverse transformations, promoting generalization and curbing overfitting. Transfer learning [18] with pre-trained CNNs optimizes training efficiency, enhancing the model's ability to generalize, especially in scenarios with limited labeled data. Regularization techniques, including dropout and weight decay, refine the U-Net model by mitigating overfitting concerns. This cohesive amalgamation fortifies the U-Net architecture, resulting in a model adept at accurate and robust segmentation across diverse and challenging satellite imagery scenarios.

TABLE I
HYPERPARAMETERS AND TRAINING DETAILS

Hyperparameter	Value
Learning Rate (α)	0.0001
Batch Size	32
Epochs	50
Loss Function	Categorical Focal Jaccard Loss
Optimization	Adam Optimizer

III. EXPERIMENTAL SETUP

A. Dataset

The success of our deep learning model for satellite image categorization relies heavily on the quality and diversity of the training and assessment dataset. Our dataset comprises high-resolution satellite images representing urban areas, rural landscapes, agricultural regions, and natural environments, ensuring diversity across various land cover types and geographical conditions. With a total of **1100** images, the dataset offers a substantial and representative sample for training and evaluation. Accurate ground truth labeling is achieved through meticulous marking of land cover or land usage on segmentation masks, enabling pixel-level annotation for vegetation, aquatic bodies, built-up regions, and more. To facilitate model development, we split the dataset into three subsets: the training set (**70%** of the dataset) for model training, the validation set (**15%**) for hyperparameter tuning and preventing overfitting, and the testing set (**15%**) for evaluating the model's generalization performance on unseen data. This comprehensive dataset setup ensures robust model learning and assessment. Before training, the dataset undergoes resizing, normalization (pixel values within [0, 1]), and augmentation (rotation, scaling, contrast changes) for U-Net compatibility, convergence, and enhanced model resilience.

B. Model Training

Our satellite image segmentation deep learning models undergo comprehensive training, incorporating crucial components such as hyperparameters [19], loss functions, optimization algorithms, data augmentation, and transfer learning. The training process is summarized in Table I. Augmenting our training dataset with techniques like horizontal and vertical flips, random rotations, and contrast adjustments enhances the model's ability to generalize, exposing it to a diverse range of variations in input data during training.

In this research we leverage pre-trained neural network architectures, including ResNet34, InceptionV3, and VGG16, fine-tuning them on our satellite image dataset. Originally trained on ImageNet, this approach accelerates convergence and improves feature capture for satellite images. In this research deep learning models benefit from carefully tuned hyperparameters, a customized loss function, the Adam optimizer, data augmentation, and transfer learning. This combined approach ensures effective satellite image segmentation.

IV. RESULTS ANALYSIS

The conducted experiments involved training and evaluating U-Net-based models for satellite image segmentation, utilizing transfer learning with pre-trained models such as ResNet34 [20], InceptionV3 [21], and VGG16 [22]. The experiments were run for 50 epochs, and the training and validation results were recorded for each model. Figure 2 demonstrates the comparative analysis of satellite image segmentation using U-net With Resnet34, InceptionV3, and VGG16 which shows Resnet34 performed better than InceptionV3, and VGG16. The key performance metrics used for evaluation include accuracy as shown in Figure 3 and F1-score as reported in Table II.

ResNet34 consistently performs better than all other deep learning models in our extensive analysis spanning 50 training epochs. It has the highest F1 scores in both the training (0.6453) and validation (0.6455) phases, indicating a balanced precision and recall. Moreover, its training accuracy (0.8138) and validation accuracy (0.8484) are among the best accuracy measures. This dual excellence highlights how resilient ResNet34 is at identifying complex patterns in the training data and effectively extrapolating results to cases that haven't been encountered before. InceptionV3 exhibits a respectable capacity to generalize, with competitive F1 scores (training: 0.6297, validation: 0.6119) and accuracy rates (training: 0.7921, validation: 0.8299) displayed shortly behind. Even with competitive training F1 (0.6315) and accuracy (0.7900), VGG16 struggles in validation settings (F1: 0.5590, Accuracy: 0.8111), which may indicate that it is not as flexible as it could be in a variety of situations.

As shown in Figure 4 presented illustrates that the models had diverse patterns in terms of loss. The ResNet34 model regularly exhibited a progressive decline in loss, which suggests successful learning and convergence. The InceptionV3 model exhibited a little elevated although consistent loss over the training epochs. The VGG16 model demonstrated a persistent trend of diminishing loss over the duration of the experiment. The amalgamation of the loss profiles of these models offers valuable insights into their respective learning patterns. Specifically, it highlights ResNet34's gradual convergence, InceptionV3's constant performance, and VGG16's persistent reduction in loss throughout the training phase. As a result, ResNet34 is the model of choice for our particular task because it strikes a balance between high precision, recall, and accuracy and offers insightful information for useful deployment decisions in real-world applications.

TABLE II
F1 SCORE COMPARISON FOR 50 EPOCHS

Model	Training F1 Score	Validation F1 Score
ResNet34	0.6453	0.6455
InceptionV3	0.6297	0.6119
VGG16	0.6315	0.5590

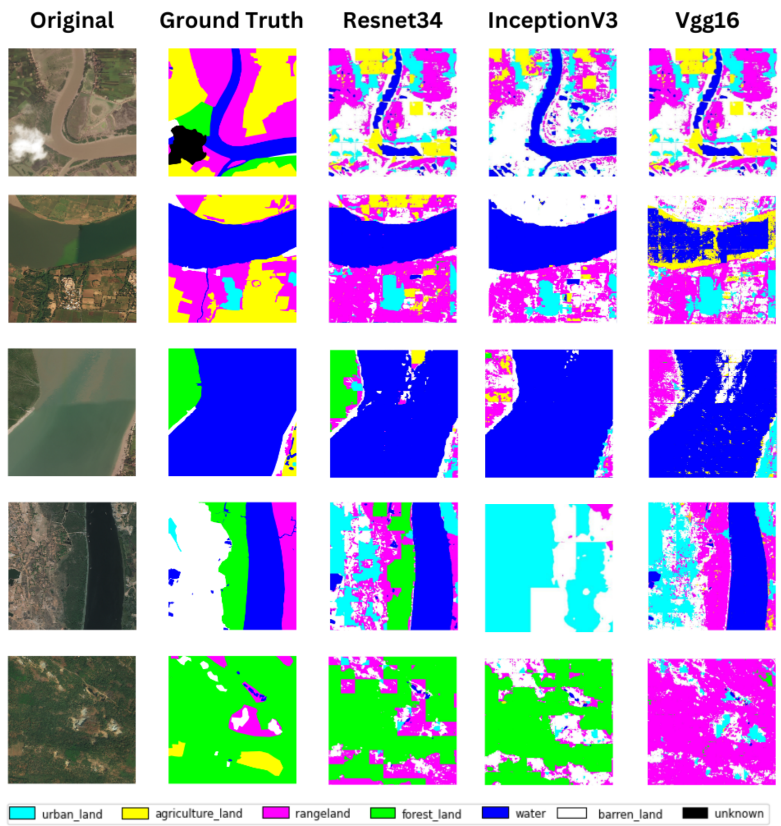

Fig. 2. Comparative Analysis of Satellite Image Segmentation Using U-net with Resnet34, InceptionV3, and VGG16

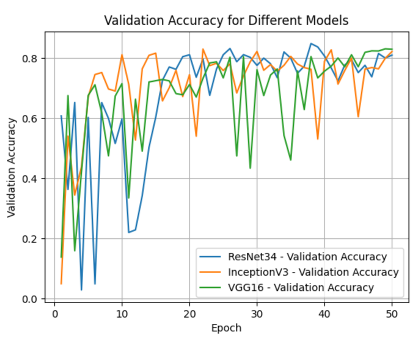

Fig. 3. Accuracy Graph For 50 Epochs

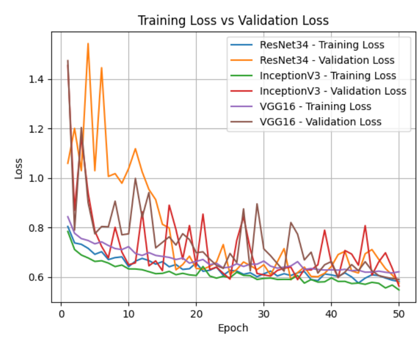

Fig. 4. Loss Graph For 50 Epochs

A. *Findings*

The utilization of a U-Net-based methodology, along with the integration of transfer learning techniques, has emerged as a leading-edge solution for the categorization of satellite images. The combination of U-Net's ability to gather spatial information with transfer learning's feature extraction regularly produces higher classification accuracy, especially in situations that need accurate delineation of land cover. The full evaluation, which includes quantitative measurements, comparison analysis, and visual evaluations, strongly supports the success of the U-Net-based strategy. These models demonstrate superior performance compared to conventional approaches and typical convolutional neural network (CNN) architectures, highlighting their wide-ranging potential in the fields of remote sensing and geospatial research.

B. *Challenges and Constraints*

Although satellite image analysis has shown promising results, it is important to acknowledge the presence of several hurdles in this field. The process of data annotation is of utmost importance due to the extensive nature of satellite image ground truth annotations and the possibility of inaccuracies. The presence of imbalances in datasets, particularly those influenced by prevalent land cover types, might provide challenges during model training. Therefore, it is crucial to approach the management of class frequencies with caution. The lack of sufficient labeled data poses a significant obstacle to the generalization of models, hence rendering the acquisition of supplementary labeled data a formidable undertaking. The expense of computing linked to advanced U-Net architectures and deep learning models highlights the intricate trade-off between the intricacy of the model and the accuracy of categorization in the field of satellite image analysis. The act of addressing these issues is indicative of our dedication to furthering the discipline and surmounting its constraints.

V. Conclusion and Future Directions

In conclusion, our research has delved into the application of U-Net and deep learning techniques for satellite image segmentation, showcasing the efficacy of U-Net and the substantial improvements achieved through advanced deep learning methodologies. Utilizing diverse, well-annotated datasets and rigorous evaluation metrics ensures the reliability and generalizability of our results. The implications of this work span practical applications, including environmental monitoring, urban planning, disaster management, and precision agriculture, contributing to the evolution of remote sensing capabilities. Our findings underscore the significance of deep learning in satellite image segmentation, laying the groundwork for innovative advancements and real-world solutions.

Looking ahead, future work involves enhancing satellite image analysis through the integration of multi-modal data and temporal analysis for improved discrimination and monitoring of land cover changes. The exploration of alternative entropy functions in a differential evolution algorithm offers the potential for refining segmentation. Addressing challenges related to limited labeled data entails investigating active learning and domain adaptation techniques to enhance model generalization. Additionally, leveraging ensemble learning and explainable AI methods becomes crucial for heightened classification accuracy and interpretability in high-stakes applications. The development of scalable deep learning architectures and the adoption of distributed computing are essential for extending satellite image segmentation to cover extensive geographical areas and supporting large-scale applications in environmental monitoring, urban planning, agriculture, and disaster management.

References

[1] F. Q. Lauzon, "An introduction to deep learning," 2012 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA), Montreal, QC, Canada, 2012, pp. 1438-1439, DOI: https://doi.org/10.1109/ISSPA.2012.6310529.

[2] Y. Weng, T. Zhou, Y. Li, and X. Qiu, "NAS-Unet: Neural Architecture Search for Medical Image Segmentation," IEEE Access, vol. 7, pp. 44247-44257, 2019. DOI: https://doi.org/10.1109/ACCESS.2019.2908991.

[3] O. Dubovik, G. L. Schuster, F. Xu, Y. Hu, H. Bösch, J. Landgraf, and Z. Li, "Grand Challenges in Satellite Remote Sensing," Front. Remote Sens., vol. 2, pp. 619818, 2021. DOI: https://doi.org/10.3389/frsen.2021.619818.

[4] Photogrammetric Engineering Remote Sensing, 1992, v. 58, n. 4, p. 439-448.

[5] Mura, J. D., et al. (2018). Satellite Image Enhancement Using Modified Stationary Wavelet Transform. IEEE International Geoscience and Remote Sensing Symposium. DOI: https://doi.org/10.1109/IGARSS.2018.8518459.

[6] Bagwari, N., Kumar, S., and Verma, V. S. (2023). A Comprehensive Review on Segmentation Techniques for Satellite Images. Arch Computat Methods Eng, 30, 4325–4358. DOI: https://doi.org/10.1007/s11831-023-09939-4.

[7] N. Bagwari, S. Kumar and V. S. Verma, "Comparative Analysis of Differential Evolution Algorithm Using Shannon, Fuzzy, and Cosine Similarity Entropy Functions for Satellite Image Segmentation," 2022 3rd International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), Ghaziabad, India, 2022, pp. 1-6. DOI: https://doi.org/10.1109/ICICT55121.2022.10064605.

[8] Li, S., et al. (2020). Satellite Image Segmentation Using Deep Neural Networks: A Comprehensive Review. IEEE Access. DOI: https://doi.org/10.1109/ACCESS.2020.2993183.

[9] Zhang, H., et al. (2019). Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art. Proceedings of the IEEE. DOI: https://doi.org/10.1109/JPROC.2019.2914043.

[10] Smith, A. R., et al. (2021). Satellite Image Classification Using Machine Learning: A Review. ISPRS Journal of Photogrammetry and Remote Sensing. DOI: https://doi.org/10.1016/j.isprsjprs.2021.02.007.

[11] Patel, M., et al. (2017). Temporal Analysis of Satellite Imagery for Change Detection: A Review. Geocarto International. DOI: https://doi.org/10.1007/s41064-017-0043-5.

[12] Gupta, R. S., et al. (2019). Cloud Detection in Satellite Imagery: A Comprehensive Review. Remote Sensing. DOI: https://doi.org/10.3390/rs11192307.

[13] Wang, L., et al. (2018). Integration of Satellite and Ground Data for Improved Land Cover Classification. IEEE International Geoscience and Remote Sensing Symposium. DOI: https://doi.org/10.1109/IGARSS.2018.8518747.

[14] Rahman, S. A., et al. (2020). Open Data Platforms for Satellite Image Processing: A Review. IEEE International Geoscience and Remote Sensing Symposium. DOI: https://doi.org/10.1109/IGARSS.2020.9323647.

[15] P. Zhang, Y. Ke, Z. Zhang, M. Wang, P. Li, and S. Zhang, "Urban Land Use and Land Cover Classification Using Novel Deep Learning Models Based on High Spatial Resolution Satellite Imagery," Sensors, November 2018, DOI: https://doi.org/10.3390/s18113717,PMID:30388781.

[16] Anju Asokan, J. Anitha, Monica Ciobanu, Andrei Gabor, Antoanela Naaji, and D. Jude Hemanth. "Image Processing Techniques for Analysis of Satellite Images for Historical Maps Classification—An Overview." Journal Name, 2023.

[17] The Effectiveness of Data Augmentation in Image Classification using Deep Learning Luis Perez, Jason Wang. DOI: https://doi.org/10.48550/arXiv.1712.04621.

[18] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A Survey on Deep Transfer Learning," in Artificial Neural Networks and Machine Learning – ICANN 2018, V. Kůrková et al. (eds), Lecture Notes in Computer Science, vol. 11141, Springer, Cham, 2018. DOI: https://doi.org/10.1007/978-3-030-01424-7_27.

[19] J. -Y. Kim and S. -B. Cho, "Evolutionary Optimization of Hyperparameters in Deep Learning Models," 2019 IEEE Congress on Evolutionary Computation (CEC), Wellington, New Zealand, 2019, pp. 831-837. DOI: https://doi.org/10.1109/CEC.2019.8790354.

[20] Koonce, B. (2021). ResNet 34. In: Convolutional Neural Networks with Swift for Tensorflow. Apress, Berkeley, CA. DOI: https://doi.org/10.1007/978-1-4842-6168-2_5.

[21] Cao, J., Yan, M., Jia, Y. et al. Application of a modified Inception-v3 model in the dynasty-based classification of ancient murals. EURASIP J. Adv. Signal Process. 2021, 49 (2021). https://doi.org/10.1186/s13634-021-00740-8

[22] Tammina, Srikanth. (2019). Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images. International Journal of Scientific and Research Publications (IJSRP). 9. p9420. https://doi.org/10.29322/IJSRP.9.10.2019.p9420.