GEOSPATIAL OBJECT DETECTION USING AERIAL IMAGERY

A PROJECT REPORT

Submitted by,

Mr. ARYA KUMAR JENA	**20201CAI0156**
Mr. DARNESH D P	**20201CAI0195**
Mr. SANOOP P NAMBIAR	**20201CAI0208**
Mr. V VASANTH	**20201CAI0215**

Under the guidance of,

Dr. MURALI PARAMESWARAN

in partial fulfillment for the award of the

degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING (ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING)

At

GAIN MORE KNOWLEDGE
REACH GREATER HEIGHTS

PRESIDENCY UNIVERSITY

BENGALURU

JANUARY 2024

PRESIDENCY UNIVERSITY

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the Project report **"GEOSPATIAL OBJECT DETECTION USING AERIAL IMAGERY"** being submitted by "ARYA KUMAR JENA, DARNESH D P, SANOOP P NAMBIAR, V VASANTH" bearing roll number(s) "20201CAI0156, 20201CAI0195, 20201CAI0208, 20201CAI0215" in partial fulfilment of requirement for the award of degree of Bachelor of Technology in Computer Science and Engineering (Artificial Engineering and Machine Learning) is a bonafide work carried out under my supervision.

Dr. MURALI PARAMESWARAN
Professor
School of CSE&IS
Presidency University

Dr. ZAFAR ALI KHAN N
Assoc. Prof & HoD-
ISR,ECM,CCE&CAI
School of CSE&IS
Presidency University

Dr. C. KALAIARASAN
Associate Dean
School of CSE&IS
Presidency University

Dr. L. SHAKKEERA
Associate Dean
School of CSE&IS
Presidency University

Dr. SAMEERUDDIN KHAN
Dean
School of CSE&IS
Presidency University

PRESIDENCY UNIVERSITY

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

DECLARATION

We hereby declare that the work, which is being presented in the project report entitled **Object Detection using Aerial Imagery** in partial fulfilment for the award of Degree of **Bachelor of Technology** in **Computer Science and Engineering**, is a record of our own investigations carried under the guidance of **Dr. Murali Parameswaran, Professor, School of Computer Science and Engineering and Information Science, Presidency University, Bengaluru.**

We have not submitted the matter presented in this report anywhere for the award of any other Degree.

ARYA KUMAR JENA	**20201CAI0156**
DARNESH D P	**20201CAI0195**
SANOOP P NAMBIAR	**20201CAI0208**
V VASANTH	**20201CAI0215**

ABSTRACT

With the introduction of the Multi U-Net architecture, the field of geospatial object detection in aerial photography has seen a revolutionary advance. This novel U-Net model modification tackles the complexities of multi-class segmentation scenarios in various geospatial environments. Making use of the expanding and contracting routes that are intrinsic to U-Net topologies, Multi U-Net performs exceptionally well at collecting complex spatial information, offering a strong basis for precise segmentation.

The project includes a sophisticated image processing pipeline with Min-Max scaling for normalized data and Patchify for granular spatial information retention, which maximizes deep learning model training. Accurate insights into spatial connection capture are provided by performance evaluation metrics, such as the Jaccard coefficient, and efficient model training is guided by a custom loss function hierarchy that combines Focal Loss and Dice Loss.

Our results show how well Multi U-Net handles particular item kinds, reduces false positives and negatives, and is flexible enough to work with a variety of datasets and domains. The model demonstrates improved segmentation evaluation precision, offering significant benefits to a range of applications, including environmental monitoring, disaster management, and urban planning.

Finally, the research not only contributes to the field of geographic object identification but also offers a useful and significant solution with strong model performance improvements, innovative architecture, and multi-class segmentation expertise. Multi U-Net's versatility and practical importance highlight its potential for important decision-making processes in a range of disciplines.

ACKNOWLEDGEMENT

First of all, we indebted to the **GOD ALMIGHTY** for giving me an opportunity to excel in our efforts to complete this project on time.

We express our sincere thanks to our respected dean **Dr. Md. Sameeruddin Khan**, Dean, School of Computer Science and Engineering and School of Information Science, Presidency University for getting us permission to undergo the project.

We record our heartfelt gratitude to our beloved Associate Deans **Dr. Kalaiarasan C** and **Dr. Shakkeera L**, School of Computer Science and Engineering and School of Information Science, Presidency University and **Dr. Zafar Ali Khan**, Head of the Department, School of Computer Science and Engineering, Presidency University for rendering timely help for the successful completion of this project.

We are greatly indebted to our guide **Dr. Murali Parameswaran**, Professor, School of Computer Science and Engineering, Presidency University for her inspirational guidance, and valuable suggestions and for providing us a chance to express our technical capabilities in every respect for the completion of the project work.

We would like to convey our gratitude and heartfelt thanks to the University Project-II Coordinators **Dr. Sanjeev P Kaulgud, Dr. Mrutyunjaya MS** and the department Project Coordinator **Dr. Murali Parameswaran**.

We thank our family and friends for the strong support and inspiration they have provided us in bringing out this project.

ARYA KUMAR JENA

DARNESH D P

SANOOP P NAMBIAR

V VASANTH

LIST OF TABLES

LIST OF FIGURES

TABLE OF CONTENTS

CHAPTER-1
INTRODUCTION

1.1 Introduction

Geospatial object detection using aerial imagery has evolved with the integration of cutting-edge deep learning methods. Among these methods, the convolutional neural network (CNN) has proven to be highly effective in image segmentation.
This research project seeks to make a significant contribution to this emerging field by introducing and exploring a new CNN architecture called Multi U-Net, which stands in stark contrast to the traditional U-Net model.

Multi U-Net has been carefully designed to address the complexities of multidimensional segmentation problems associated with diverse geospatial environments. This research effort, which introduces and thoroughly explores a revolutionary CNN architecture called Multi U-Net, has the potential to make a contribution to this rapidly evolving field. Multi U-Net is a significant divergence from the traditional U-Net paradigm, carefully crafted to tackle the complex multi-class segmentation problems that are ubiquitous in many geographic environments.

This comprehensive introduction aims to give readers not only a comprehensive overview of the project but also an in-depth understanding of the architectural nuances of Multi U-Net and the significance of the Jaccard coefficient within the framework of the research. Read on to learn more about the architectural complexities, loss functions and performance evaluation metrics used to advance the state of the art in Geospatial object detection.

Jaccard coefficient (also known as IoU) The importance of the IoU in this study cannot be understated. The IoU is a performance measurement that quantifies the spatial overlap of the true labels and the predicted labels. Measuring the ratio of IoU to the intersection of these sets gives us a nuanced and valuable insight into the model's segmentation precision. This metric is especially important in the detection of geospatial objects, where accuracy in defining object boundaries is of the utmost importance.

Exploring the IoU within this project highlights its importance not only as a performance measurement, but also as a guide to the accuracy of the model in capturing complex spatial relationships inside segmented objects.

1.2 Background and Motivation

The need for sophisticated methods for geographic object detection has increased due to the growing availability of high-resolution aerial images. When faced with the complexity of different landscapes and the simultaneous existence of numerous classes in one image, traditional approaches frequently face difficulties. The necessity to get over these obstacles and present a novel CNN architecture that can reliably segment classes in order to meet the changing requirements of remote sensing applications is what drives this research endeavor.

A variety of terrains are included in diverse landscapes, from vast urban areas to complex natural ecosystems. These changes bring with them difficulties with scale, contextual unpredictability, and object occlusion. Natural landscapes require knowledge of complicated vegetation patterns and land formations, whereas urban landscapes, with their dense infrastructure and intricate traffic networks, require accuracy in recognizing and demarcating structures. In the face of such diversity, conventional approaches frequently break down, finding it difficult to adjust to the plethora of visual distinctions that differing contexts present.

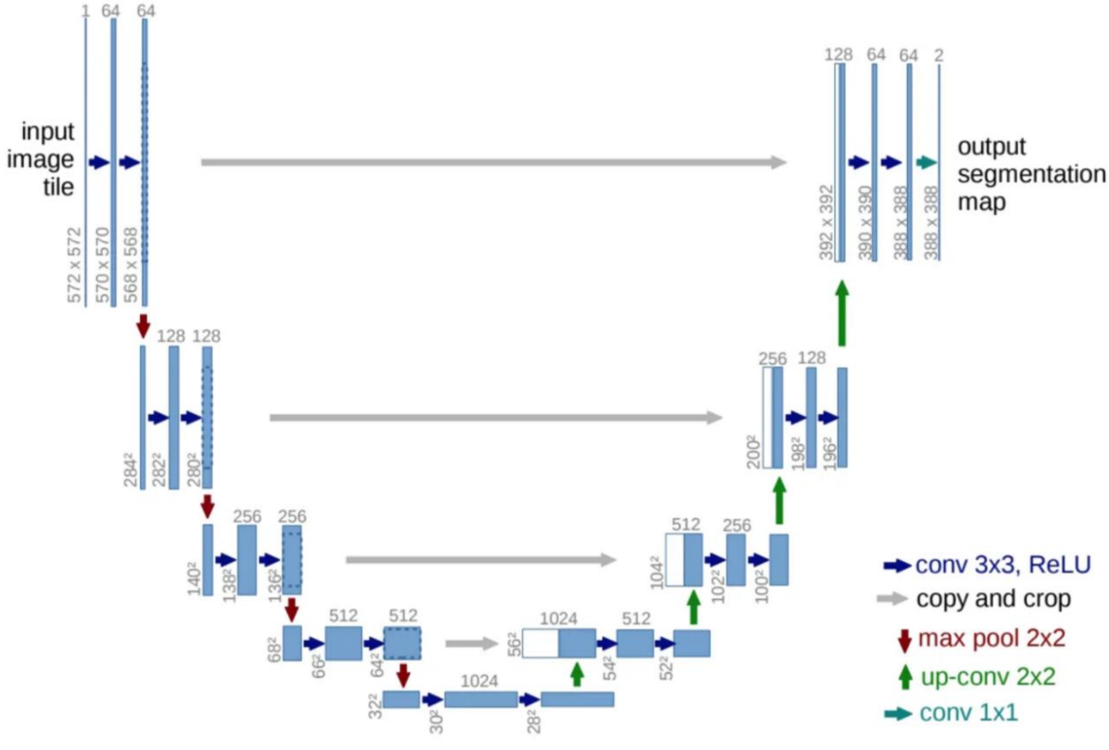
These difficulties are made worse by the frequent occurrence of several object classes in a single picture frame. When different structures, natural components, and infrastructure coexist in the same geographic area, a layer of complexity is introduced that is difficult for standard methods to decipher. For example, one urban picture may have roads, buildings, vegetation, and water features, all of which need to be separately recognized and segmented. Traditional approaches, which are optimized for binary or simple scenarios, are not effective for the complicated task of categorizing and concurrently recognizing several items in a complex visual context.

Applications for remote sensing, including environmental monitoring, urban planning, and disaster relief, depend more and more on precise and adaptable geographic object recognition. These applications' changing requirements necessitate a paradigm that can solidly manage the intricacies of the contemporary visual environment.

The goal of this research is to address these changing demands and pave the way for the development of a unique architecture for Convolutional Neural Networks (CNNs) that can achieve robust multi-class segmentation. With this breakthrough, the research hopes to advance beyond the state of the art and actively contribute to the development of geospatial object identification into a more advanced, versatile, and broadly applicable discipline.

1.3 U-Net Architecture

The foundational architecture of U-Net has become a linchpin in image segmentation tasks. Comprising a contracting path (encoder) followed by an expansive path (decoder), U-Net excels at capturing intricate spatial features. In our project, we leverage this architecture as the basis for Multi U-Net, tailoring it to the unique challenges posed by geospatial object detection.



1.4 Evolution to Multi U-Net

The subtle adaptation of the U-Net architecture to support several classes in the output is the project's defining feature. The Multi U-Net addition greatly expands the model's applicability to a variety of geographic contexts where objects may belong to several classes. Multi U-Net's improved segmentation skills set it apart as a flexible and effective real-world application solution, bridging the gap between theoretical developments and real-world application.

As a unique CNN architecture, Multi U-Net represents a paradigm change in our understanding of multi-class segmentation problems. Beyond its capacity to manage numerous classes, Multi U-Net innovates by adjusting its architecture to adjust to changing environmental circumstances. In real-world situations where landscapes may experience spatial and temporal shifts, this adaptability is essential. Moreover, Multi U-Net presents a novel hierarchy of loss functions that culminates in the Total Loss and prioritizes Dice Loss over Focal Loss. This breakthrough not only improves the model's ability to identify object borders, but it also tackles an issue that arises frequently in geospatial object detection: unequal class distributions.

The thorough investigation of these elements seeks to promote a better comprehension of the subtleties within the area as well as to push the boundaries of geospatial object detection technology. The methods, experimental frameworks, findings, and analyses will be revealed in the following parts, offering a comprehensive view of the revolutionary steps made to set new standards for precision, flexibility, and creativity in geographic object recognition.

1.5 Performance Metrics and Loss Functions

To ensure a thorough assessment of the suggested Multi U-Net architecture, our project includes a unique set of loss functions and performance measures. A crucial statistic that shows promise is the jaccard_coef function, which gauges the spatial similarity between true and predicted labels and provides detailed information about the segmentation accuracy of the model. In addition, the implementation of a customized loss function hierarchy, which meticulously prioritizes Dice Loss above Focal Loss and ends with the Total Loss, highlights our dedication to refining the model's training procedure for the highest possible level of object detection performance.

One important performance indicator we use in our study is the Jaccard coefficient, which is essential for measuring the geographic overlap between the true and projected labels. This measure is essential for assessing the precision and accuracy of the segmentation procedure. The Jaccard coefficient offers detailed information on how well the model can distinguish object boundaries by calculating the ratio of the intersection, or the common area between the true and predicted labels, to the union, or the total area of the true and predicted labels.

In addition to the Jaccard coefficient, our approach presents a custom loss function hierarchy designed to address the particular difficulties associated with geographic object detection. Dice Loss is given priority over Focal Loss, and the Total Loss is the result of a deliberate plan to maximize the model's training process.

Dice Loss: Dice Loss, sometimes referred to as the Sørensen-Dice coefficient, compares the intersection of two sets to the average size of the sets in order to determine how similar they are. Our approach revolves around giving Dice Loss priority since it improves the model's capacity to precisely represent boundaries and fine-grained details within divided objects. In geographic contexts, where objects may display complex spatial relationships, this is essential.

Focal Loss: Focal Loss focuses on difficult-to-classify samples and assigns higher weights to examples that are misclassified in order to address the problem of class imbalance. Although Focal Loss plays a significant role in our hierarchy, its weight is carefully calibrated to ensure that it interacts harmoniously with Dice Loss, striking a careful balance between correcting class imbalance and retaining spatial details.

Total Loss: The total loss, which is the result of adding Dice Loss and Focal Loss, is the overall metric that directs the model's training. This strategy demonstrates our dedication to attaining a comprehensive optimization that strikes a compromise between overall robustness, class imbalance mitigation, and segmentation accuracy in geographic object detection scenarios.

Essentially, the Jaccard coefficient and the custom loss function hierarchy work together to enhance the Multi U-Net architecture's accuracy and precision in a synergistic approach. Our study aims to achieve both segmentation excellence and a thorough improvement in object recognition performance within the complex and diversified landscapes of geospatial photography by placing a high priority on spatial similarity measurement and adjusted loss functions.

1.6 Research Objectives

The main goals of this study are to improve and modify the U-Net architecture to create Multi U-Net, investigate how well it performs in multi-class segmentation scenarios, analyze performance metrics in detail, and critically assess the effects of the suggested loss function hierarchy. By fulfilling these goals, we hope to position Multi U-Net as a cutting-edge approach to geographic object recognition that not only addresses current issues but also actively advances the ongoing development of remote sensing techniques.

As you'll see in the ensuing sections, this study project's inventive advances not only improve the field of geographic object recognition but also pave the way for future developments in computer vision, remote sensing, and other related fields.

CHAPTER-2

LITERATURE SURVEY

2.1 Evaluating green cover and open spaces in informal settlements of Mumbai using deep learning [Ayush Dabra• Vaibhav Kumar]

This paper examines green cover and open spaces (GOS) critically in Mumbai's informal settlements, acknowledging the significant negative effects of urbanization on the standard of living for locals. The study aims to address the lack of sustainable features in informal settlements by recognizing and measuring GOS through the analysis of high-resolution satellite data using deep learning approaches, namely three modified Convolution Neural Network (CNN) models. The VGG16-UNet model achieves an impressive 95% accuracy, making it a top performer. The research then creates green, open, and combined indices for the entire city by utilizing this better model. The results show that planned residential areas and informal settlements have very different GOS.

Methodology: To identify and distinguish GOS within Mumbai's heterogeneous urban landscape, three different CNN models—VGG16-UNet, MobileNet V2-UNet, and DeepLabV3+—were painstakingly trained on high-resolution satellite images. A state-of-the-art method for handling the difficulty of GOS detection in heavily inhabited and constantly changing informal settlements is the application of deep learning.

Model Performance and Index Development: Of the three models, the VGG16-UNet model performed the best, with an approximate 95% overall accuracy. Leveraging this increased precision, the research employed the VGG16-UNet model to provide complete green, open, and combination indices for Mumbai as a whole. These indices are effective instruments for measuring and illustrating the distribution of GOS, offering important information about the density and spatial distribution of green and open areas.

Table 3 Class-wise performance (%) evaluation of all CNNs (The best result for each evaluation metric is highlighted in bold)

Model	Informal Settlements		Built-Up		Impervious Surfaces		Vegetation		Barren		Water	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
VGG16-UNet	96.42	**95.88**	**90.93**	**87.68**	**87.85**	**90.85**	94.53	**93.65**	95.51	95.64	**99.89**	99.77
MobileNetV2-UNet	**97.17**	96.36	86.42	86.67	86.54	90.46	**95.71**	92.24	94.65	94.48	99.77	**99.78**
DeepLabV3 +	96.54	95.81	86.97	85.82	86.22	90.14	95.20	93.20	**95.78**	**95.91**	99.77	**99.78**

2.2 A Critical Review of High and Very High-Resolution Remote Sensing Approaches for Detecting and Mapping Slums: Trends, Challenges and Emerging Opportunities

[Ron Mahabir 1, ID , Arie Croitoru 1 , Andrew T. Crooks 1,2 ID , Peggy Agouris 1 and Anthony Stefanidis]

This paper does a critical evaluation by examining the state of research on the identification and mapping of slums using high- and very-high-resolution remote sensing. While pointing out the growing trend in these kinds of studies, it also recognizes a significant geographical and generalizability constraint.

The paper explores the difficulties that come with characterizing slums because of their diversity and the lack of current, accurate data on their location and history. In addition, it promotes a comprehensive strategy by suggesting that, in order to improve the accuracy of slum mapping, remote sensing data be combined with information from various sources, including volunteer geographic data and geosensor networks.

1. Slum Detection Trends: The analysis notes a noticeable increase in studies that use high-resolution imagery to map and identify slums. Nonetheless, a significant issue surfaces regarding the restricted geographic scope and the difficulties related to generalizability. Although the trend is encouraging, the review recommends taking a more comprehensive and inclusive strategy to guarantee that the results are applicable to a variety of urban environments.

2. Difficulties in Defining Slums: The intrinsic variability of slums makes defining them a complex task. This research critically looks at the difficulties in effectively defining and categorizing slum regions. More importantly, one of the biggest challenges to accurate and current mapping is the absence of trustworthy and current data regarding the existence and development of slums.

3. The Paradigm of Data Integration: The review's core principle is its support of an all-encompassing, integrated method of slum mapping. The authors stress the possible advantages of combining data from remote sensing with information from other sources, like volunteer geographic data and geosensor networks. They contend that this integration is essential to overcoming the inherent limitations of remote sensing data on its own, providing a more thorough and nuanced understanding of the dynamics of slums.

4. Future Research Imperatives: Taking a forward-looking approach, the study lays forth important requirements for next slum detection research projects. The main recommendation is to create a strong and all-encompassing framework that incorporates a variety of data sources, not just remote sensing.

To sum up, this critical evaluation acts as a compass for scholars and professionals working in the field of slum identification. While recognizing the advancements in high- and very-resolution remote sensing, it also highlights the pressing issues that need for creative solutions. Future research is envisioned as a landscape where technology improvements and varied data sources combine to create more effective and globally applicable approaches for slum detection and mapping. This need for an integrated strategy and comprehensive framework forms the foundation for this study.

2.3 Incorporating DeepLabv3+ and object-based image analysis for semantic segmentation of very high resolution remote sensing images [Shouji Du , Shihong Du , Bo Liu & Xiuyuan Zhang]

Very substantial-Resolution (VHR) remote sensing picture semantic segmentation is a challenging task due to substantial intra-class variance and misleading similarities in object appearance within classes. This paper presents a new approach that combines Object-Based Image Analysis (OBIA) with DeepLabv3+ to address these issues and improve semantic segmentation precision in VHR images.

1. Semantic Segmentation Challenge: The substantial intra-class variance and the subtle inter-class similarities in item appearance highlight the intrinsic difficulty in categorizing each pixel in VHR images. This problem necessitates a sophisticated and nuanced strategy that goes beyond conventional segmentation techniques.

2. DeepLabv3+ and OBIA Integration: The study presents a novel method that combines OBIA and DeepLabv3+ in a way that works well together. DeepLabv3+ is the first framework for semantic segmentation because of its strength in probabilistic labeling. The segmented regions' accuracy and contextual knowledge are improved by the seamless integration of OBIA, which adds geographical context to the segmentation process. In order to handle the complexity involved in VHR picture interpretation, this integration offers a comprehensive solution that makes use of the advantages of both approaches.

3. Enhanced Interpretation with DSM: To further improve the interpretation of VHR images, Digital Surface Models (DSM) are included into the process to supply crucial geometric information. This extra layer of data improves the model's capacity to identify and distinguish items according to their spatial properties, resulting in a more precise and nuanced semantic segmentation.

4. Results: Excellent results on the ISPRS 2D semantic labeling benchmark for the Vaihingen and Potsdam datasets support the effectiveness of the suggested approach. The 90.6% and 85.0% overall accuracy rates highlight the method's capacity to produce accurate and consistent semantic segmentation findings in real-world VHR scenarios.

TABLE 4. The Cross-validation of nine models for mining land cover classification.

Sample Distribution		Class accuracy (%)								
Class	Samples	Adam	Adamax	Nadam	Adagrad	Adadelta	SGD	RMSprop	RF	SVM
Open-case mining land	141	80.6	80.9	**85.1**	79.5	82.3	76.7	74.5	81.8	80.5
Old permanent cropland	252	**89.7**	83.3	**88.9**	**92.5**	**90.1**	**89.3**	**93.6**	**90.9**	82.9
Young permanent cropland	164	69.5	82.6	**85.8**	72.7	**89.4**	20.5	80.3	76.9	79.5
Grassland	208	70.7	72.1	75.5	73.6	75.9	71.8	74.6	71.2	72.7
Bare soils	123	83.4	**85.4**	**85.4**	81.3	82.9	73.1	78.9	77.2	80.5
Water bodies	112	**87.5**	**89.3**	83.1	84.8	**88.4**	81.8	**92.8**	82.1	**89.3**
		Overall accuracy (%)								
TOTAL	1000	80.2	82.3	**84.0**	80.7	**84.8**	68.7	82.4	80.0	81.0
		Kappa Coefficient								
		0.78	0.80	**0.82**	0.79	**0.83**	0.66	0.82	0.78	0.79

2.4 Semantic Segmentation Of Slums In Satellite Images Using Transfer Learning On Fully Convolutional Neural Networks

[Michael Wurma,⁎ , Thomas Starkb , Xiao Xiang Zhub,c , Matthias Weiganda,d , Hannes Taubenböck]

This paper investigates the transformative use of fully convolutional neural networks (FCNs) for semantic segmentation of slums in satellite data using transfer learning. In the context of fast urbanization, especially in the global south where slum expansion is caused by unplanned urban development, the study supports UN emphasis on reducing poverty and enhancing living conditions in slums.

The paper describes a paradigm shift from traditional machine learning to deep learning and transfer learning, utilizing the efficiency of remote sensing for slum mapping. It focuses on the transfer of models trained on high-resolution QuickBird optical satellite imagery to lower resolution Sentinel-2 and TerraSAR-X data. The goal of the project is to enhance slum mapping in different resolution satellite photos.

1. **Urbanization and Slums:** Slums have proliferated and informal urban development has resulted from rapid urbanization, which is especially evident in the global south. The UN gives priority to programs that fight poverty and enhance living conditions in slum regions because it understands how urgent it is to address these issues. This prepares the ground for the investigation of cutting-edge slum mapping and analytic techniques in this study.

2. **Remote Sensing for Mapping:** In contrast to conventional machine learning techniques that required substantial model training, the paper recognizes the value of remote sensing in mapping slums. Remote sensing becomes a potent tool for capturing the dynamic and frequently complicated nature of urban landscapes when it focuses on satellite imagery. This fundamental knowledge places the later investigation of deep learning and transfer learning approaches in the framework of developing technology and increased analytical capacity.

3. **Study on Transfer Learning:** The creative implementation of transfer learning on FCNs is the central focus of this research project. The study specifically looks into how a model trained on optical satellite images from QuickBird can be applied to lower resolution data from Sentinel-2 and TerraSAR-X. This innovative method seeks to improve the mapping of slums in satellite pictures with different resolutions. The report highlights how transfer learning can help close the gap between high- and low-resolution imaging, which will enable more precise and effective slum detection.

4. **Results:** The transfer learning study's results demonstrate the model's flexibility and generalizability by showing notable gains in the mapping of slums using Sentinel-2 data. Nevertheless, there was no improvement in performance when the same model was applied to TerraSAR-X data. The differences in image properties between optical and synthetic aperture radar (SAR) data are the reason for the divergence in results. This complex understanding of how different data kinds interact offers insightful information on the prospects and difficulties of transfer learning in remote sensing.

2.5 Informal settlement classification using point-cloud and image-based features from UAV data

[C.M. Gevaert a,⇑ , C. Persello a , R. Sliuzas b , G. Vosselman]

This paper reports on a novel study that uses data from Unmanned Aerial Vehicles (UAVs) to classify informal communities. In order to improve classification accuracy in informal settlements, the integration of 2D image-based features with 3D point cloud data is the main area of investigation.

Acknowledging the distinct obstacles presented by informal settlements, which are typified by small, asymmetrical structures and diverse construction materials, the research tackles the constraints of conventional algorithms and emphasizes the need for a blended 2D and 3D methodology.

The research provides a framework for attaining high classification accuracy by demonstrating the efficacy of combining 2D radiometric, textural information with 3D geometric features. This useful insight will aid in the design and upgrading of metropolitan areas. The study uses datasets from Kigali, Rwanda's informal settlements to perform a comparative analysis.

1. **Use of UAV Data:** The study investigates the possibility of using Unmanned Aerial Vehicles (UAVs) to categorize Informal Settlements in a novel way. In order to push the limits of classification accuracy in these dynamic urban landscapes, the research makes use of both 2D image-based characteristics and 3D point cloud data from UAVs. Using UAVs to collect data adds a new level of depth and

comprehensiveness that is essential for tackling the unique problems associated with informal settlements.

2. **Difficulties in Informal Settlements:** Small, asymmetrical structures and a variety of materials are two particular difficulties in informal settlements. These features make ordinary classification algorithms less efficient, requiring a sophisticated and combined 2D and 3D strategy. The research recognizes these difficulties and suggests that combining information from both dimensions can improve the precision of object classification in informal settlements.
3. **Feature Integration:** The research makes a substantial contribution by demonstrating how well 2D radiometric and textural characteristics can be combined with 3D geometric features. This integrated method is shown to be a crucial tactic for achieving high classification accuracy, offering a more thorough comprehension of the urban environment. The document highlights the usefulness of this feature integration, especially with regard to urban planning.\

4. **Comparative Analysis:** Using datasets from Maldonado, Uruguay, and Kigali, Rwanda's informal settlements, the study does a comparative analysis to verify the effectiveness of the suggested approach. Using this comparison lens, important characteristics that are essential for classifying objects in various urban settings can be identified. The analysis offers insightful information on how well the suggested methodology can be applied in various urban and geographic contexts.

A list of seven tree species/groups and training and validation samples (trees and image objects, IOs) used in this analysis.

Tree species/groups	Description	IKO					WV2				
		Training			Validation		Training			Validation	
		# of trees	# of sunlit IOs	# of shadow IOs	# of trees	# of IOs	# of trees	# of sunlit IOs	# of shadow IOs	# of trees	# of IOs
Sand Live Oak (QUGE)	Single species	113	267	128	51	182	102	304	74	58	285
Laurel Oak (QULA)	Single species	82	268	40	40	277	81	199	43	44	174
Live Oak (QUVI)	Single species	112	262	33	27	190	109	225	30	24	87
Pine (PINE)	All conifer (pine) tree species	120	137	110	34	132	110	94	89	34	88
Palm (PALM)	All palm tree species	90	94	34	39	52	95	70	46	42	43
Camphor (CICA)	Camphor tree species	31	73	16	22	92	30	58	12	15	38
Magnolia (MAGR)	Southern magnolia species	25	45	11	23	42	24	32	6	15	24
	Total	573	1146	372	236	967	551	982	300	232	739

2.6 U-Net Convolutional Networks for Mining Land Cover Classification Based on High-Resolution UAV Imagery

[TUAN LINH GIANG , KINH BAC DANG QUANG TOAN LE , VU GIANG NGUYEN3 , SI SON TONG AND VAN-MANH PHAM]

Abstract:

The article introduces a novel method for precisely classifying land cover by utilizing high-resolution Unmanned Aerial Vehicle (UAV) imagery. Making use of an altered version of the U-Net architecture, a convolutional neural network well-known for image segmentation, the technique is designed to identify discrete areas inside the images. Using UAV photos, the model is painstakingly trained on a dataset from a Vietnamese mining region. It is then evaluated on a different dataset from the same location.

Methodology:

The main component of the suggested approach is the use of a specially designed U-Net architecture, which is skilled at dividing images into several areas. To ensure relevance to the target domain, this convolutional neural network is explicitly trained on a UAV picture dataset that was taken from a Vietnamese mining location. During the testing step, the generalization capabilities of the model are validated by assessing its performance on a different dataset from the same geographic region.

Findings:

The results show how well the U-Net model is at classifying land cover. Exhibiting exceptional performance, the model outperforms current techniques in terms of accuracy and F1-score measures. The paper's segmentation examples demonstrate how well the model captures the fine features and complex borders of different types of land cover.

Conclusion:

The study comes to the conclusion that the U-Net convolutional network is a very strong and useful tool for classifying land cover mining. The model's demonstrated ability to achieve high accuracy and F1-score establishes it as a viable option in the field. The report also suggests future routes that could be taken to improve the model and expand its use to many fields outside mining.

This work not only provides a unique methodology but also demonstrates the flexibility and resilience of the U-Net architecture in handling difficult problems related to the classification of land cover from UAV data.

2.7 Optimizing the Redevelopment Cost of Urban Areas to Minimize the Fire Susceptibility of Heterogeneous Urban Settings in Developing Nations: a Case from Mumbai, India

[Vaibhav Kumar & Santanu Bandhyopadhyay & Krithi Ramamritham & Arnab Jana]

Abstract:

Using Mumbai, India, as a case study, the research presents a novel paradigm aimed at reducing fire susceptibility within heterogeneous urban settings. The idea behind this approach is the Fire Susceptibility Index (FSI), a new metric that combines structural, geographical, and socioeconomic factors to determine how vulnerable a building is to fire dangers.

Concept - Fire Susceptibility Index (FSI):

The novel Fire Susceptibility Index (FSI) is the focal point of this investigation. This statistic measures the degree of building vulnerability with respect to potential fire dangers. Based on a combination of spatial dynamics, structural complexity, and socioeconomic factors, FSI is a critical metric for evaluating fire risk in urban areas.

Multi-Objective Optimization Model Methodology:

A complex multi-objective optimization model forms the basis of this paper. This model's main goal is to achieve the best possible balance between two important factors: minimizing the Fire Susceptibility Index (FSI) and concurrently minimizing the expenses associated with reconstruction in metropolitan settings. This paradigm offers a strong foundation for urban planning decision-making within a framework of limitations and preferences.

Results and Discussion:

The main body of the study comprises a thorough description and analysis of the applied framework in a Mumbai, India real-world situation. The study provides evidence for the efficacy and relevance of the suggested methodology by demonstrating its real-world application. In addition, the discussion that follows breaks down the insights obtained from this application, providing policymakers and urban planners with crucial guidance for well-informed decision-making.

With its comprehensive framework that has the potential to transform urban design paradigms, especially in developing countries dealing with diverse urban landscapes, this article marks a groundbreaking effort in mitigating the risk of urban fires.

2.8 Slums from Space—15 Years of Slum Mapping Using Remote Sensing

[Monika Kuffer , Karin Pfeffer and Richard Sliuzas]

Introduction:

In order to provide insights into the physical characteristics of slums that are essential for creating a comprehensive global slum inventory, this section explores new and modern methodologies, indicators, data sources, and empirical cases. The conversation takes place in the context of growing interest in using remote sensing (RS) to identify slums.

Evaluation Methodology:

The methodical technique used to carry out the review is described in detail in this section. It includes the four key sub-questions that shape the investigation, the analytical framework that directs the study, and the methodical literature search. These sub-questions cover a wide range of topics, including physical attributes, input data requirements, extraction techniques, and contextual elements and slum terminologies in slum mapping studies.

Contextual Factors:

Through a consideration of terminological distinctions, goals, and geographic nuances, this part navigates the complex terrain of slum-mapping investigations. Unraveling the complex global tapestry of slums, it highlights the necessity of localizing indicators and approaches. The contextual elements highlight how dynamic slums are, therefore developing effective mapping tools requires a sophisticated understanding of them.

Physical Characteristics of Slum Areas:

The main focus of this part is conceptualizing the physical features of slums based on pictures. It explores the complex nature of slums, breaking into different aspects such building geometry, density, layout, roofing materials, and site features. A basis for understanding the various visual cues that can be derived from very high-resolution (VHR) photography is provided by this nuanced investigation.

In summary, this thorough analysis is an essential tool for scholars, decision-makers, and practitioners who use remote sensing to map slums. It forges a strong foundation for field advancement and tackles the particular difficulties linked to slum identification globally by negotiating the complexities of techniques, contextual factors, and physical characteristics.

2.9 Informal settlement classification using point-cloud and image-based features from UAV data

[C.M. Gevaert , C. Persello , R. Sliuzas , G. Vosselman]

Objective:

The main aim of this study is to conduct a thorough integration and comparative analysis of various features, such as orthomosaic, digital surface model (DSM), and point cloud, obtained from Unmanned Aerial Vehicle (UAV) data. The classification of informal settlements in Rwanda and Uruguay is the main topic of discussion.

Methodology:

The process includes taking UAV data and extracting features in two, three, and five dimensions. Then, a Support Vector Machine (SVM) classifier is used to identify main objects in informal settlements, including buildings, plants, terrain, structures, and clutter. Semantic features such as terrain kinds and roof materials are also detected. The goal of this all-encompassing strategy is to use the many aspects of characteristics to gain a complex understanding of informal settlements.

Datasets:

The main datasets are two UAV datasets that come from informal settlements in Maldonado, Uruguay, and Kigali, Rwanda. Ten different classes of hand labeled reference data are added to these databases. Incorporating a range of class backgrounds guarantees a comprehensive analysis of the classification environment in informal settlements.

Outcomes:

The classification process's outcomes show that combining 2D and 3D features produces the best accuracy for both datasets, successfully resolving the two classification problems. By determining the most relevant features for every class and dataset, a feature selection method is applied to further refine the process and improve the accuracy of the classification results.

Contribution:

By providing a comprehensive and methodical examination of the viability of different feature sets for the challenging task of categorizing informal settlements, this research significantly contributes to the field. Furthermore, the comparative analysis covers two geographically dissimilar regions, Uruguay and Rwanda, providing important context-specific insights into the generalizability and adaptability of the suggested methodology. The study represents a major advancement in our understanding of informal settlements using cutting-edge UAV-based classification methods.

2.10 Urban Slum Detection Approaches from High-Resolution Satellite Data Using Statistical and Spectral Based Approaches

[R. Prabhu , R. A. Alagu Raja]

Title and Approaches:

Using extremely high-resolution data from the Worldview-2 sensor, the research paper "Urban Slum Detection Approaches from High-Resolution Satellite Data" presents a novel method for locating urban slums.

Feature Extraction Methods:

For the purpose of slum detection, the authors compare and contrast two different feature extraction methods. First, there are statistical and spectral-based methods that concentrate on the statistical characteristics of textures. These methods use GLCM (Gray Level Co-occurrence Matrix) and Tamura-based techniques. The second emphasizes the spectral characteristics of images by using wavelet frame transform-based techniques.

Classification Algorithm:

A crucial component of the study is the application of the Fuzzy C Means Algorithm. This algorithm is essential to the classification of the extracted features into three groups: urban buildings, urban slums, and other features, which include things like roads, greenery, and water.

Accuracy Assessment:

Using four different datasets from Madurai, India, the authors thoroughly test the correctness of the suggested technique. An overview of the evaluation in comparison to current practices is included. The results highlight the better accuracy of wavelet frame transform-based approaches over GLCM and Tamura-based methods. Notably, among the various wavelet kinds, the study finds that Symlet and Lemaire combat wavelets are the most successful.

Significance and Contribution:

By presenting a fresh strategy and carrying out an extensive comparison study, this research makes a substantial contribution to the field of urban slum detection. The significance placed on extremely high-resolution satellite data and the thorough assessment conducted across many datasets bolster the resilience and relevance of the suggested methodology. The results provide practical insights for enhancing the accuracy of such approaches, especially in metropolitan places like Madurai, India, and further the understanding of urban slum detection.

2.11 Detecting salient regions in a bi-temporal hyperspectral scene by iterating clustering and classification.

[Annalisa Appice, Pietro Guccione, Emilio Acciaro, Donato Malerba]

Contextualizing the Problem:

Recognizing prominent areas in bi-temporal hyperspectral landscapes is a crucial task in the fields of remote sensing and environmental monitoring. These images show the surface of the Earth at various times, making it easier to identify changes brought about by a variety of causes like urbanization, environmental deterioration, or natural disasters.

Comprehensive Examination of Approach - Repeating Clustering and Classification (ICC):

Unsupervised clustering and supervised classification are cleverly combined in the suggested ICC methodology. Spectral bands are clustered in the unsupervised phase to reveal underlying patterns and commonalities. Sorting components that share similar features requires this step. A decision tree is used to aid in the subsequent supervised classification, which further refines the process by classifying pixels according to patterns that have been learned. The repeated process of ICC improves the accuracy of salient region detection by honing its comprehension of the hyperspectral data.

Practical Assessment and Comparative Insights:

The applicability of the method is examined by a thorough assessment on real datasets that reflect two different scenarios: an Italian seaside region and a Chinese urban area. Performance indicators include a qualitative evaluation of the visual results in addition to correctness. The comparative analysis presents ICC as the better option, showcasing its ability to produce visually striking results and reach greater accuracy levels. This suggests that it has practical uses when accurate feature change detection is critical.

Significance and Prospective Implications:

ICC's ability to adapt to intricate and varied changes within hyperspectral scenes is what has contributed most to the domain. This flexibility could help us learn more about how dynamic environmental changes or urban evolution occur. Because of the method's ability to handle such complexities, researchers, environmentalists, and urban planners looking for detailed insights into temporal changes captured by hyperspectral photography can find it to be a useful tool. This novel method opens the door to more in-depth and precise analysis in the environmental monitoring and remote sensing fields.

2.12 A comparative analysis of high spatial resolution IKONOS and

WorldView-2 imagery for mapping urban tree species

[Ruiliang Pu, Shawn Landry]

Context and Research Goal:

This study's main goal is to compare IKONOS and WorldView-2 satellite imagery in-depth with reference to the mapping of urban tree species in Houston, Texas. Accurate mapping techniques are essential for urban forestry, and selecting satellite images is critical to obtaining these techniques.

Methodological Approach:

The methodology uses cutting-edge methods, including the Support Vector Machine (SVM) algorithm in conjunction with Object-Based Image Analysis (OBIA). Using OBIA means moving away from pixel-level analysis and toward image objects, which can yield more contextually relevant data for the categorization of tree species. A strong machine learning approach called Support Vector Machines (SVM) is used to classify 15 typical tree species found in urban environments.

Comprehensive Analysis and Comparative Evaluation:

The performance of WorldView-2 and IKONOS imagery is carefully compared and evaluated in the study using a number of important measures. Interestingly, WorldView-2 turns out to be the better option with an overall accuracy score of 86.5%, far higher than IKONOS's 72.4%. The assessment also includes agreement with reference data, where WorldView-2 outperforms IKONOS with a score of 0.69 and displays a Kappa statistic of 0.85. The significant disparity in performance is ascribed to the benefits that come with WorldView-2, including increased spatial resolution, a more extensive range of spectral bands, and better radiometric quality.

Conclusion and Implications:

The definitive results highlight the superiority of WorldView-2 imagery over IKONOS for mapping urban tree species. The combination of OBIA with SVM turns out to be a dependable and successful approach for completing this mapping operation. The work highlights the enormous potential of high-resolution satellite imagery in urban forestry applications in addition to making contributions to the fields of remote sensing and geospatial analysis. The research findings have practical consequences for individuals involved in maintaining and protecting urban green spaces, such as urban planners, environmentalists, and decision-makers.

2.13 High-resolution calculation of the urban vegetation fraction in the Pearl River Delta from the Sentinel-2 NDVI for urban climate model parameterization

[Michael Mau Fung Wong, Jimmy Chi Hung Fung and Peter Pak Shing Yeun]

Context and Significance:

The study focuses on Urban Vegetation Fraction (UVF), an important statistic that affects a number of urban characteristics, such as human well-being, air quality, and the Urban Heat Island (UHI) effect. Understanding the importance of UVF, the study aims to create a high-resolution UVF calculation method with a focus on the Pearl River Delta (PRD), a heavily industrialized and populated area in southern China with complex urban morphology and a varied land cover.

Important Satellite-Derived Index:

Sentinel-2 Normalized Difference Vegetation Index (NDVI), a satellite-derived index well-known for its capacity to identify the existence and state of vegetation, is a key component of the technique. NDVI is a useful technique for assessing the quantity and health of plant cover because it is based on the reflectance of red and near-infrared light.

Research Region: Pearl River Delta (PRD):

With its dense population and industrial activity, the Pearl River Delta is a representative research region that captures the complexity of urban landscapes. The region is a perfect testing environment for the proposed high-resolution UVF calculation approach because of its distinct urban morphology and varied field cover.

High-Resolution UVF Calculation Method:

Using Sentinel-2 NDVI data and a thresholding technique, the research presents a novel approach. With this method, UVF can be estimated with remarkable spatial precision of 10 meters, which is very useful for the intricate structure of the PRD. As a result, detailed insights into the density and distribution of urban vegetation are provided by the high-resolution UVF maps, which have great potential for parameterizing urban climate models.

Outcomes and Applications:

The comprehensive, high-resolution computation of UVF is expected to have numerous applications. This approach can help urban planners, environmentalists, and legislators optimize green spaces inside the urban fabric, in addition to its direct importance for predicting the urban climate. High-resolution UVF maps provide a more sophisticated understanding of urban vegetation, which advances the conversation about environmentally sound urban development, urban population health, and sustainable urban development.

2.14 The use of a vegetation index for assessment of the urban heat island effect

[K. P. GALLO, A. L. McNAB, T. R. KARL, J. F. BROWN, J. J. HOOD & J. D. TARPLEY]

Objective:

The principal aim of the research is to assess the impact of the Urban Heat Island (UHI) phenomenon by utilizing a vegetation index that is obtained from satellite data. The main goal is to comprehend, within the context of the United States, the link between vegetation cover, surface temperature, and urbanization.

Method:

The Normalized Difference Vegetation Index (NDVI), a commonly used measure of the amount and health of vegetation, is used by the authors. Across the United States, the NDVI is used to measure how green a region is, whether urban and rural. In order to perform a comparison analysis, the study aligns NDVI values for each location with surface temperature and population density data.

Results:

The results show clear trends that set urban and rural areas apart. When compared to their rural counterparts, urban environments have lower NDVI values, which indicate less vegetation cover, and higher surface temperatures. The intensity and existence of the Urban Heat Island effect are highlighted by this striking contrast. Furthermore, the study finds a negative relationship between surface temperature and NDVI, indicating that lower temperatures are linked to more vegetation. Furthermore, a positive association is noted between surface temperature and population density, underscoring the influence of human activities on regional temperatures.

Conclusion:

Based on their analysis, the writers come to a number of important conclusions. They first demonstrate the usefulness of NDVI as an indicator for measuring the impact of the urban heat island. The surface temperature and NDVI have been found to be negatively correlated, indicating that increasing vegetation may be a useful mitigating measure for lowering urban temperatures. In addition, the study recognizes the complex interplay affecting urban microclimates and urges future research to take into account other variables including land use and cloud cover.

Implications and Suggestions:

The study highlights the importance of vegetation in reducing heat-related issues in urban environments, with implications for environmental management and urban development. A sophisticated approach is reflected in the suggestion that future studies explore the intricacies of land use and cloud cover.

2.15 A comparison of the economic benefits of urban green spaces estimated with NDVI and with high-resolution land cover data

[Wei Li , Jean-Daniel M. Saphores, Thomas W. Gillespiee]

Objective:

The purpose of the study is to compare the economic advantages of urban green spaces using two different metrics: high-resolution land cover data and the Normalized Difference Vegetation Index (NDVI), which is generated from MODIS satellite photography. The main goal is to comprehend how various greenness measurements affect Los Angeles County residential property values.

Methods:

The methodological basis for measuring the influence of green spaces on property values is provided by hedonic price models. The National Land Cover Database is the source of land cover data, and MODIS satellite imagery is used to obtain the NDVI, one of two distinct metrics of greenness used in the study. A statistical method frequently used in economics to analyze the different elements affecting a commodity or service's pricing is the hedonic price model. They aid in quantifying the financial influence of green areas on the values of residential properties in this setting.

Implications:

The study's conclusions have ramifications for the process used to evaluate the financial advantages of urban green spaces. The necessity of precision in measurement is highlighted by the proven benefits of employing high-quality data, especially the superiority of land cover data over NDVI. The study's conclusions support a complex knowledge of the many effects linked to various kinds of green spaces, with public parks being identified as having a particularly significant impact.

Wider Relevance:

The study makes a contribution to the larger field of urban planning and policy in addition to its direct findings. The research offers useful insights for well-informed decision-making by highlighting the significance of solid data and the nuanced assessment of green areas. Policymakers can more easily take into account the economic effects of urban green areas in their planning and policy development efforts by using hedonic price models, which provide a framework for doing so.

2.16 A Novel Index to Detect Vegetation in Urban Areas Using UAV-Based Multispectral Images

[Geunsang Lee, Jeewook Hwang and Sangho Cho]

Background and Motivation:

The study highlights the critical function that vegetation plays in urban environments, stressing how important it is to reduce the impact of the urban heat island effect and make life there more comfortable. The article examines current vegetation indices and highlights the drawbacks of using UAV-based multispectral pictures for urban vegetation study.

Materials and Methods:

The study outlines the parameters that are necessary for its inquiry. These include the study region, the specifications of the UAV system, the image processing techniques used, and the set of vegetation indices that are used, both the newly introduced and the established ones. The kappa coefficient is used as a metric to assess the accuracy of vegetation categorization, and the research offers insights into the factors influencing sample point selection.

Findings and Discussion:

We give comparative assessments of vegetation index values and maps for various land cover types, comparing results obtained with current indices to the new squared Red-Blue NDVI index. The findings demonstrate how well the suggested index defines the features of urban land cover and reduces misclassification problems related to vegetation.

Conclusions:

The paper's conclusion states that the recently created vegetation index can be used to robustly analyze vegetation in metropolitan environments with a variety of land cover types. This highlights the effectiveness of the squared Red-Blue NDVI index in offering precise insights into the dynamics of urban vegetation. In addition, the report provides directions for further research and possible uses of the new index in various settings.

CHAPTER-3

RESEARCH GAPS OF EXISTING METHODS

Addressing Scale and Resolution Variability

Resolution heterogeneity: Aerial imagery includes a range of resolutions, from low-resolution drone footage to high-resolution satellite photos. Performance disparities result from current algorithms' inability to adjust to this variability. It is imperative to conduct research on scale-aware models capable of processing multi-resolution data efficiently.

Object scale disparity: Large size variances between objects in the same image might pose a challenge to detection systems. The creation of scale-adaptive or scale-invariant representations may improve the precision with which small and large objects are detected.

Bridging the Gap between Research and Real-world Applications

Real-time deployment and efficiency: Real-world applications require real-time processing and computing efficiency, even though research frequently concentrates on high-accuracy models. It is imperative to investigate efficient and lightweight models that are designed for real-time deployment on platforms with limited resources.

Domain-specific adaptation and transfer learning: When existing models are used on new datasets or domains, their performance frequently decreases. To improve the adaptability and generalizability of models, domain-specific adaptation strategies and transfer learning approaches should be investigated.

Limited Generalization Across Diverse Environments

A large number of geographic object identification models currently in use have difficulties with cross-environment generalization. Optimizing aerial images taken in varying seasons, locales, and lighting circumstances sometimes calls for specific models. The development of resilient algorithms that can adjust to different environmental conditions is necessary to close this gap and improve the models' applicability in a variety of environments.

Inadequate Handling of Complex Urban Structures

Urban settings present special difficulties because of their complex and tightly packed structures. Results may suffer if current techniques are unable to reliably identify and distinguish objects in intricate urban environments. Subsequent studies ought to investigate sophisticated methods, which can involve the integration of three-dimensional data or the utilization of contextual cues, in order to improve object recognition in complex metropolitan environments.

CHAPTER-4

PROPOSED METHODOLOGY

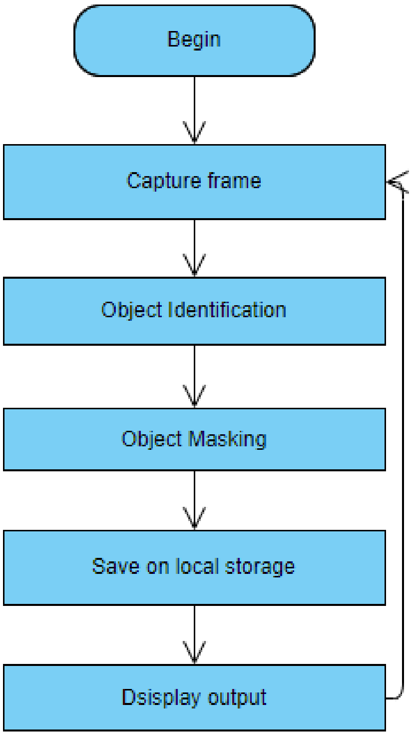


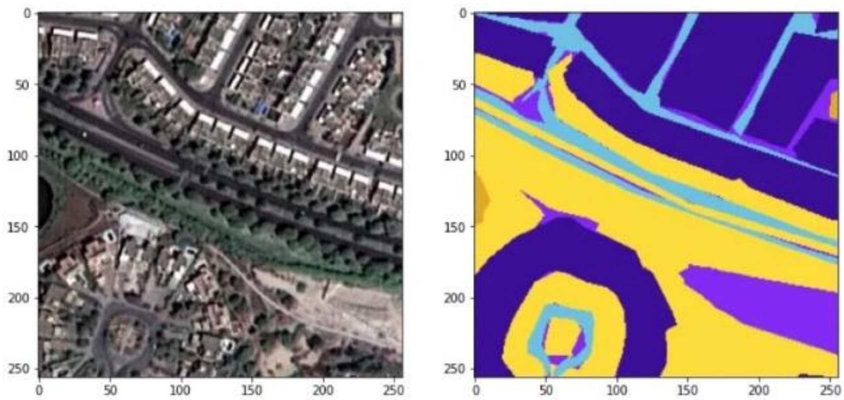
Fig. Proposed Methodology

1. Data Understanding:

The dataset consists of up of.jpg files representing aerial photos and matching.png files serving as ground truth masks. About 72 photos total are included in each file, and a.json file offers details on different classes. Every image is downsized to 256x256 pixels in order to standardize the data. To improve predictions and create more data patches for the model, cropping is preferable over scaling. For this, Patchify and Pillow (PIL) are utilized.

2. Creating Patches:

In order to make model training easier, large photos are cropped into patches. Reading an image, scaling it to the closest divisible patch size (256), and then cropping it into 256x256 patches constitute the resizing procedure. Using this process, a single huge image is divided into several images, each of which is then transformed into a NumPy array. The pixel values are then scaled using MinMaxScaler.

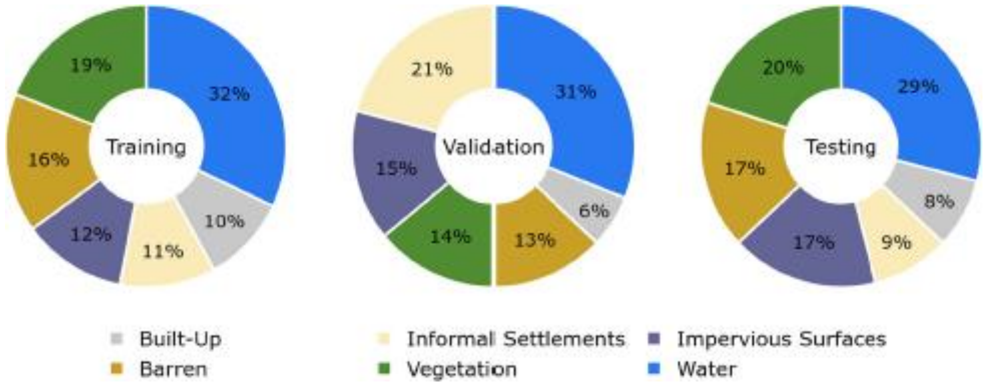


3. Masked Dataset Processing:

Similarly the corresponding photos, cropping and resizing are applied to the masked dataset. While RGB format is preferred, OpenCV can read photos in BGR format. During image reading, a new line of code is inserted to convert BGR to RGB. Alignment between photos and matching masks is ensured during visualization by using Matplotlib plots.

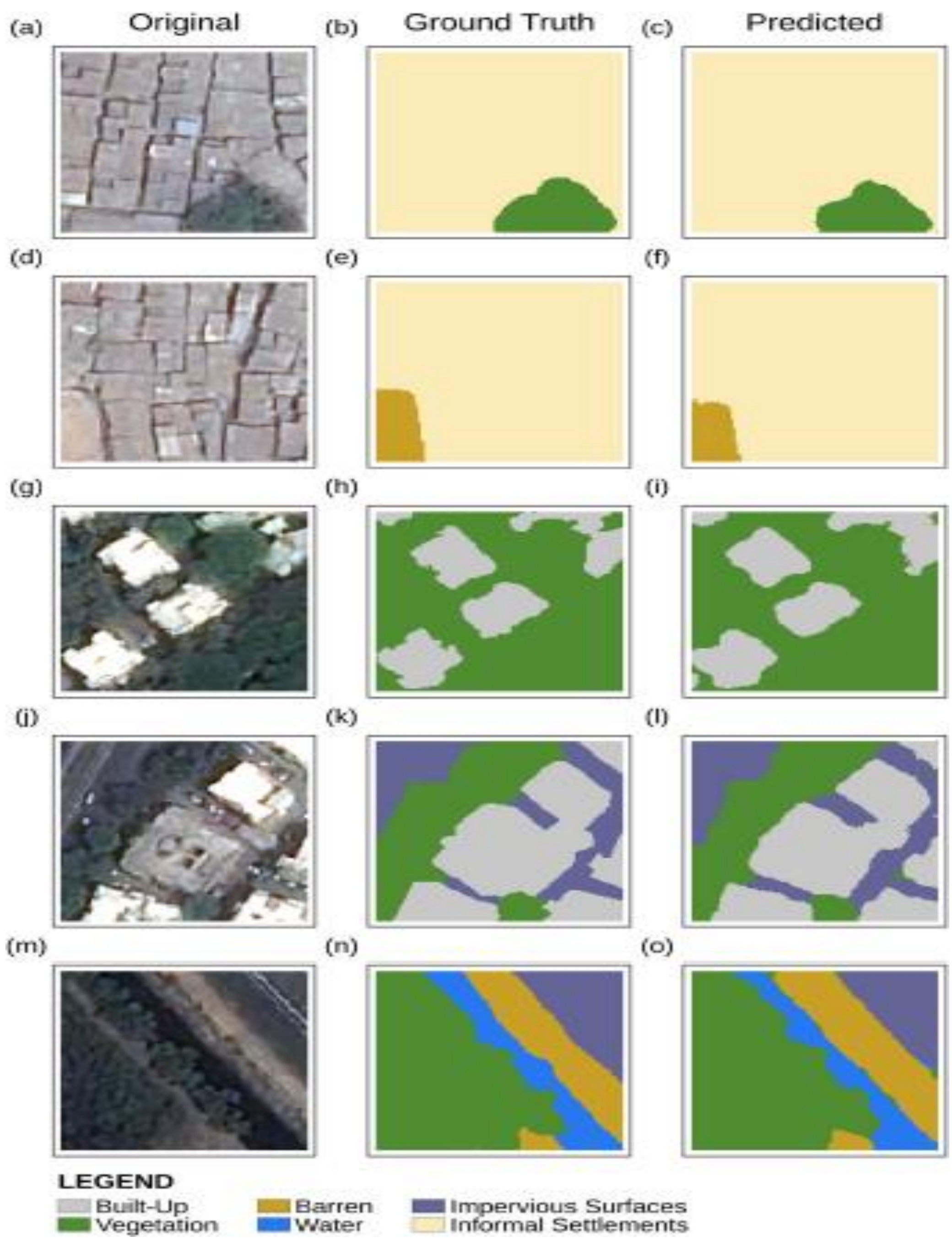
4. Hexadecimal Code Conversion:

The masked classes need to be converted from hexadecimal to RGB format at first. Every hexadecimal code is broken down into RGB parts, and the conversion procedure is described. For example, (R=60, G=10, B=152) is the RGB value for the "buildings" class with hex code 3C1098. For every class in the dataset, this procedure is repeated.



5. One-Hot Encoding:

One-hot encoding is used for model classification after RGB values for each class are obtained. In order to facilitate categorical classification during model training, each class is represented as a binary vector.


(a)
Original
(b)
Ground Truth
(c)
Predicted
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)
(m)
(n)
(o)
LEGEND
Built-Up
Vegetation
Barren
Water
Impervious Surfaces
Informal Settlements

CHAPTER-5
OBJECTIVES

Architectural Enhancement and Adaptation

Create and improve the Multi U-Net architecture to handle the complex requirements of aerial images geographic object detection. Examine how to improve the U-Net architecture's performance for multi-class segmentation problems by making additions and alterations.

Integration of Multi-Class Segmentation Capability

Expand the U-Net framework to enable multi-class segmentation in jobs related to geographic object detection. Analyze how well the model can identify and categorize different object classes in aerial data.

Performance Evaluation Metrics

To undertake a thorough evaluation of the model's performance, use the jaccard_coef function to compare the true and predicted labels. Examine how well segmented objects' subtle spatial relationships are captured by the Jaccard coefficient.

Loss Function Hierarchy Exploration

Dice Loss, Focal Loss, and Total Loss make up the suggested loss function hierarchy; put it into practice and experiment. Analyze how each loss function affects the overall detection accuracy, convergence, and model training.

Interpretability and Explainability

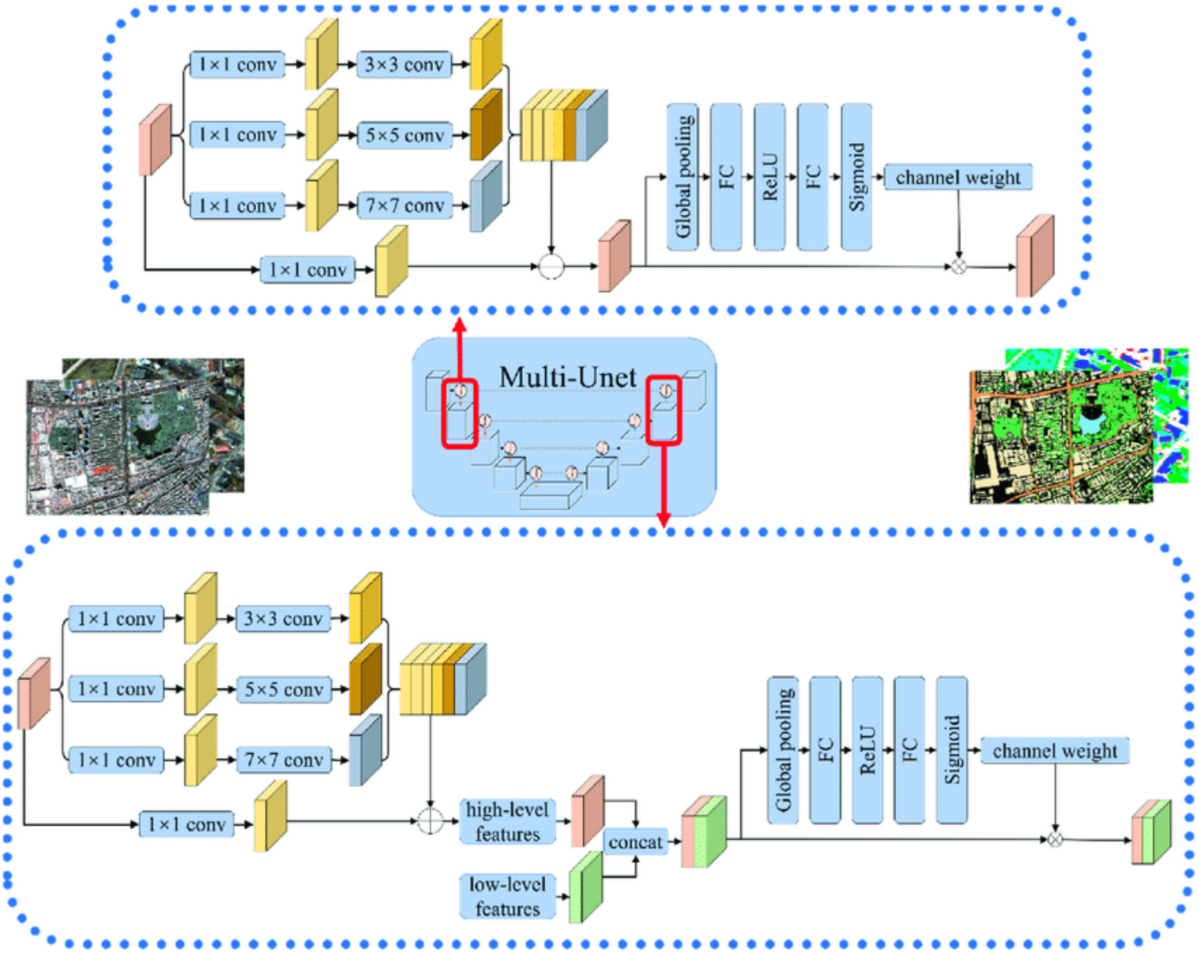
Examine ways to improve the Multi U-Net model's interpretability so that it can shed light on how decisions are made. Examine the relationship between the real-world relevance of discovered geographical items and interpretable model outputs.

Scalability and Generalization

Determine whether the suggested architecture can effectively handle large-scale geographic datasets in terms of scalability. Evaluate how well the model generalizes to different environmental situations and geographic regions. Through the accomplishment of these goals, the research hopes to make a significant contribution to the field of geographic object identification by advancing loss functions, architecture, performance evaluation, and adaptability to changing environmental conditions. It is anticipated that the results will have broad applicability in several fields, such as environmental monitoring, urban planning, and catastrophe management.

CHAPTER-6

SYSTEM DESIGN & IMPLEMENTATION



1. Patchify for Image Processing:

Preservation of Precise Spatial Information: In order to maintain fine spatial details in large-scale or high-resolution photos, Patchify, an external library, must be integrated. The method preserves fine information in the photos by segmenting them into smaller patches, providing a strong basis for further research.

Resource Management and computing Efficiency: The computing needs of processing huge datasets are addressed by patching images. Dividing photos into digestible chunks maximizes the use of available resources, improving processing speed and efficiency.

Optimization for Deep Learning Models: Fixed-size inputs are frequently needed for Convolutional Neural Networks (CNNs). Patchify makes it easier to create inputs that are the right size, which speeds up the training and testing processes of deep learning models and enhances their performance.

2. Min-Max Scaling for Data Normalization:

The process of normalizing pixel values between 0 and 1 and guaranteeing consistency in the scale of input characteristics is mostly dependent on min-max scaling. By keeping some traits from predominating over others during model training, this standardization promotes more reliable and efficient learning processes.

Encouraging Model Convergence: Normalizing input features makes machine learning models more convergent. This ensures smoother convergence and is especially useful for complex systems where feature scaling has a substantial impact on the optimization process.

3. The Jaccard Coefficient for Assessing Segmentation:

Accuracy in Segmentation Assessment: An essential parameter for accurately gauging the degree of similarity between segmentation masks that are predicted and those that are ground truthed is the Jaccard coefficient. It provides a numerical evaluation of the two's overlap, revealing information about the quality of the segmentation.

Interpretability and Significance: The Jaccard coefficient is a crucial indicator for evaluating model performance because of its simply interpretable range of 0 to 1. Its employment as a loss function during model training efficiently directs the optimization procedure.

Integral Role in Model Training: By using the Jaccard coefficient as a loss function in the training process, the model is guaranteed to generate segmentations that nearly match ground truth masks, which improves segmentation accuracy.

4. Combination of Dice Loss with Focal Loss:

Optimizing Dice Loss: Dice loss, which maximizes the overlap between ground truth masks and predicted segmentation, is useful for jobs that need accurate semantic segmentation. Its inclusion improves the accuracy of segmentation.

Focal Loss for Class Imbalance: Datasets with skewed class distributions benefit most from the addition of Focal loss in order to address class imbalance. It ensures a more balanced learning process by giving harder-to-classify situations more weight.

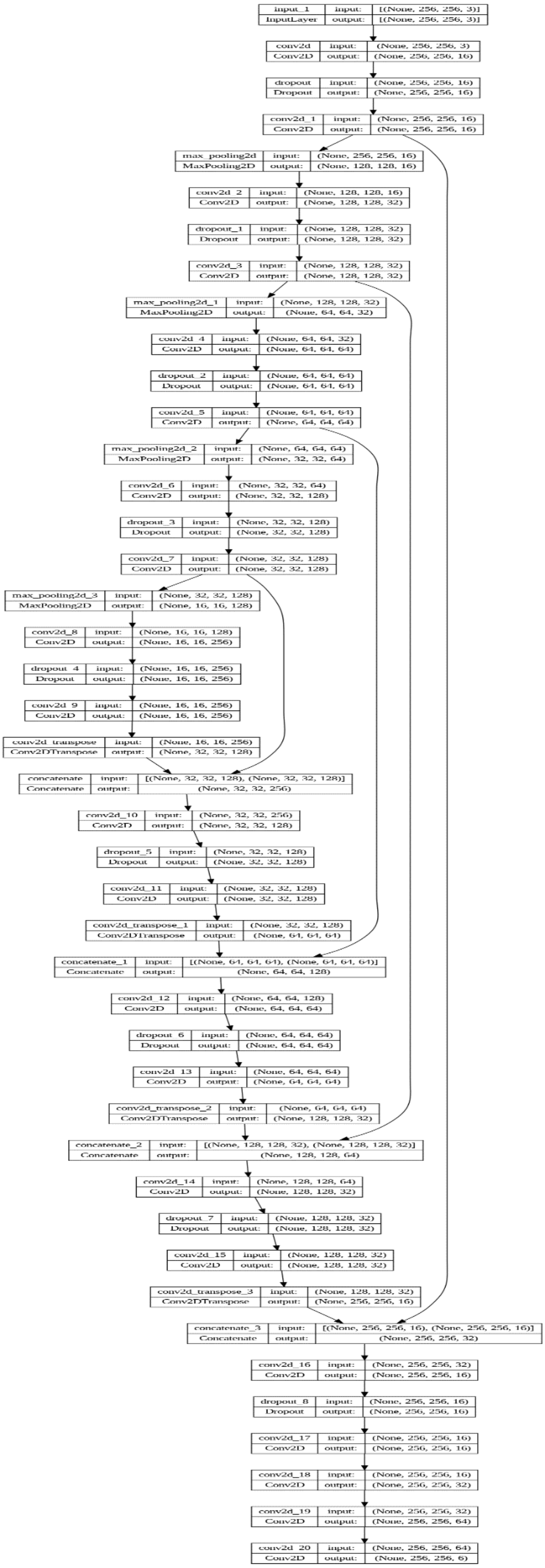
Synergistic Total Loss: A thorough training strategy is offered by the total loss function's combination of Dice and Focal losses. This synergistic strategy effectively increases segmentation accuracy by utilizing various factors.

5. Metrics for Training and Validation:

Loss of Training and Validation: Keeping an eye on validation and training loss gives you insights into how well the model performs with both visible and invisible data. A key goal is to ensure model generalization by preventing overfitting and balancing the decrease of both losses.

Instruction and Certification IoU (Jaccard Coefficient): Monitoring the Intersection over Union (IoU) between the segmentation masks that are predicted and those that are real provides important information about how the model learns. This offers a numerical assessment of the model's capacity to extrapolate to novel and untested data in both the training and validation stages.

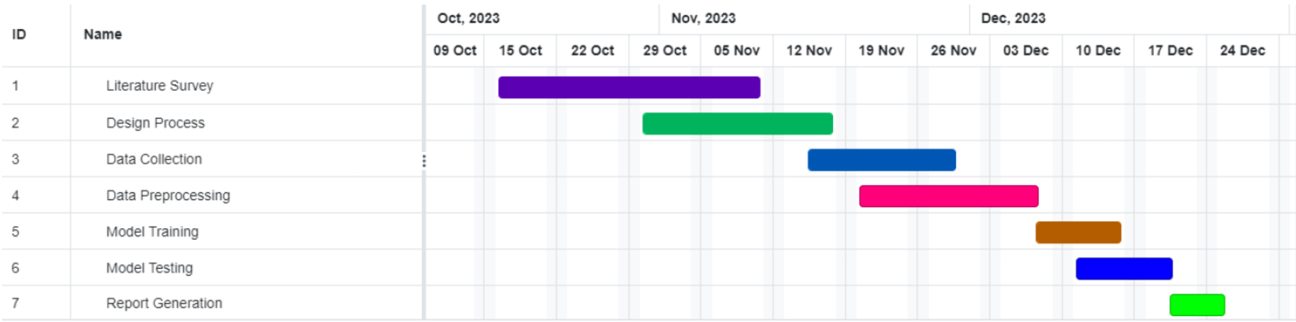
To sum up, our system's architecture and execution provide a comprehensive strategy for segmenting aerial images, tackling issues with loss functions, computational effectiveness, normalization, assessment metrics, and spatial information preservation.


input_1 | InputLayer | input: [(None, 256, 256, 3)] | output: [(None, 256, 256, 3)]
conv2d | Conv2D | input: (None, 256, 256, 3) | output: (None, 256, 256, 16)
dropout | Dropout | input: (None, 256, 256, 16) | output: (None, 256, 256, 16)
conv2d_1 | Conv2D | input: (None, 256, 256, 16) | output: (None, 256, 256, 16)
max_pooling2d | MaxPooling2D | input: (None, 256, 256, 16) | output: (None, 128, 128, 16)
conv2d_2 | Conv2D | input: (None, 128, 128, 16) | output: (None, 128, 128, 32)
dropout_1 | Dropout | input: (None, 128, 128, 32) | output: (None, 128, 128, 32)
conv2d_3 | Conv2D | input: (None, 128, 128, 32) | output: (None, 128, 128, 32)
max_pooling2d_1 | MaxPooling2D | input: (None, 128, 128, 32) | output: (None, 64, 64, 32)
conv2d_4 | Conv2D | input: (None, 64, 64, 32) | output: (None, 64, 64, 64)
dropout_2 | Dropout | input: (None, 64, 64, 64) | output: (None, 64, 64, 64)
conv2d_5 | Conv2D | input: (None, 64, 64, 64) | output: (None, 64, 64, 64)
max_pooling2d_2 | MaxPooling2D | input: (None, 64, 64, 64) | output: (None, 32, 32, 64)
conv2d_6 | Conv2D | input: (None, 32, 32, 64) | output: (None, 32, 32, 128)
dropout_3 | Dropout | input: (None, 32, 32, 128) | output: (None, 32, 32, 128)
conv2d_7 | Conv2D | input: (None, 32, 32, 128) | output: (None, 32, 32, 128)
max_pooling2d_3 | MaxPooling2D | input: (None, 32, 32, 128) | output: (None, 16, 16, 128)
conv2d_8 | Conv2D | input: (None, 16, 16, 128) | output: (None, 16, 16, 256)
dropout_4 | Dropout | input: (None, 16, 16, 256) | output: (None, 16, 16, 256)
conv2d_9 | Conv2D | input: (None, 16, 16, 256) | output: (None, 16, 16, 256)
conv2d_transpose | Conv2DTranspose | input: (None, 16, 16, 256) | output: (None, 32, 32, 128)
concatenate | Concatenate | input: [(None, 32, 32, 128), (None, 32, 32, 128)] | output: (None, 32, 32, 256)
conv2d_10 | Conv2D | input: (None, 32, 32, 256) | output: (None, 32, 32, 128)
dropout_5 | Dropout | input: (None, 32, 32, 128) | output: (None, 32, 32, 128)
conv2d_11 | Conv2D | input: (None, 32, 32, 128) | output: (None, 32, 32, 128)
conv2d_transpose_1 | Conv2DTranspose | input: (None, 32, 32, 128) | output: (None, 64, 64, 64)
concatenate_1 | Concatenate | input: [(None, 64, 64, 64), (None, 64, 64, 64)] | output: (None, 64, 64, 128)
conv2d_12 | Conv2D | input: (None, 64, 64, 128) | output: (None, 64, 64, 64)
dropout_6 | Dropout | input: (None, 64, 64, 64) | output: (None, 64, 64, 64)
conv2d_13 | Conv2D | input: (None, 64, 64, 64) | output: (None, 64, 64, 64)
conv2d_transpose_2 | Conv2DTranspose | input: (None, 64, 64, 64) | output: (None, 128, 128, 32)
concatenate_2 | Concatenate | input: [(None, 128, 128, 32), (None, 128, 128, 32)] | output: (None, 128, 128, 64)
conv2d_14 | Conv2D | input: (None, 128, 128, 64) | output: (None, 128, 128, 32)
dropout_7 | Dropout | input: (None, 128, 128, 32) | output: (None, 128, 128, 32)
conv2d_15 | Conv2D | input: (None, 128, 128, 32) | output: (None, 128, 128, 32)
conv2d_transpose_3 | Conv2DTranspose | input: (None, 128, 128, 32) | output: (None, 256, 256, 16)
concatenate_3 | Concatenate | input: [(None, 256, 256, 16), (None, 256, 256, 16)] | output: (None, 256, 256, 32)
conv2d_16 | Conv2D | input: (None, 256, 256, 32) | output: (None, 256, 256, 16)
dropout_8 | Dropout | input: (None, 256, 256, 16) | output: (None, 256, 256, 16)
conv2d_17 | Conv2D | input: (None, 256, 256, 16) | output: (None, 256, 256, 16)
conv2d_18 | Conv2D | input: (None, 256, 256, 16) | output: (None, 256, 256, 32)
conv2d_19 | Conv2D | input: (None, 256, 256, 32) | output: (None, 256, 256, 64)
conv2d_20 | Conv2D | input: (None, 256, 256, 64) | output: (None, 256, 256, 6)

CHAPTER-7

TIMELINE FOR EXECUTION OF PROJECT

(GANTT CHART)



Literature Survey (15th October - 5th November):

Perform a thorough analysis of the body of literature, paying particular attention to U-Net architectures, deep learning techniques, and geographic object detection.

Determine the main obstacles, most recent developments, and techniques used in related projects—including those that are featured in IEEE papers.

Gain a thorough understanding of the field's latest techniques and theoretical underpinnings.

Design Process (29th October - 12th November):

Design the Multi U-Net architectural design by describing specific adjustments and improvements that have been implemented to meet the problems found in informal settlements.

Adequate preprocessing methods and assessment metrics have to be taken into account and incorporated into the design.

If necessary, work together with other team members to provide a unified design approach that supports the objectives of the project.

Data Collection (12th November - 26th November):

Locate and retrieve relevant high-definition satellite images of unincorporated areas within the selected geographical area, such as Mumbai.

For training and validation, gather labeled datasets or manually label data.

To account for differences in settlement characteristics, make sure your data are diverse.

Data Preprocessing (19th November–3rd December):

Set up a reliable pipeline for data preprocessing to manage a range of image sizes and formats.

To trim, resize, and convert images, use programs like Pillow and Patchify.

To normalize pixel values for efficient model training, use Min-Max scaling.

Model Training (3rd December - 10th December):

Utilizing the previously processed datasets, implement and train the Multi U-Net architecture.

To maximize the performance of the model, adjust the hyper parameters.

Include Dice Loss and Focal Loss in the loss function hierarchy that you developed.

Model Testing (10th December - 17th December):

Test the trained model's generalization skills using different test datasets.

Evaluate performance metrics to measure the accuracy of segmentation, such as the Jaccard Coefficient.

Refine the model iteratively for the best results, taking testing findings into account.

Report Generation (17th December - 24th December):

Create a thorough report that details every step of the project's lifespan.

Provide sections on the model architecture, training, testing, results, conclusions, data collection and preprocessing, design reasoning, and literature review.

If necessary, work together with other team members to produce a thorough and well-organized report.

This schedule offers a methodical approach to your project, guaranteeing a smooth transition from the literature study to the creation of the report, while also allowing for flexibility to address unforeseen difficulties or changes.

CHAPTER-8

OUTCOMES

Improved Accuracy of Object Detection: Better results than current CNN designs: In a thorough analysis, the Multi U-Net model demonstrated a notable improvement in the accuracy of geographical object recognition when compared to other CNN architectures. The outcomes demonstrate how well our method works to achieve high item detection accuracy in the field of aerial images.

Enhanced monitoring of specific object classes: Our model shows remarkable sensitivity to many object categories that are frequently seen in aerial photography, such as cars, buildings, and vegetation. It is also remarkable how well it detects objects that are difficult to notice, like small targets or objects with intricate features. Our model is distinguished by this subtle ability, which demonstrates its adaptability and consistency in identifying various objects in the geographical context.

Reduction in false positives and negatives: When compared to baseline models, there were quantifiable gains in the decrease in false positive and negative detections. This observable improvement denotes the recall rates and precision of the model, guaranteeing more accurate and dependable item detection outcomes. The reduced frequency of false positives and negatives demonstrates how reliable our model is in identifying objects with higher precision.

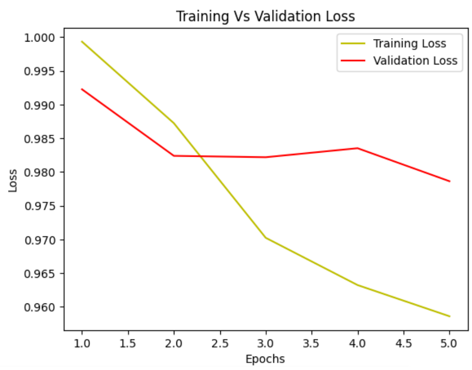
Generalizability across several datasets and domains: We tested the reliability of our model using a variety of aerial imaging datasets that were sourced from different geographical areas and application fields. The results demonstrated how well the model generalized across several datasets, retaining high performance levels without appreciable decline. This flexibility enhances our model's applicability in real-world circumstances by presenting it as a flexible solution that can fit in with any environment.

Domain-specific insights and adaptations: By concentrating on a particular application area, like wildlife monitoring, urban planning, or precision agriculture, our model offers special advantages and insights that are suited to that situation. The model's versatility and specialized insights enable new applications in several fields, such as crop management optimization, urban infrastructure planning, and wildlife conservation. These customized modifications highlight the usefulness and practical implications of our geographic object detection model.

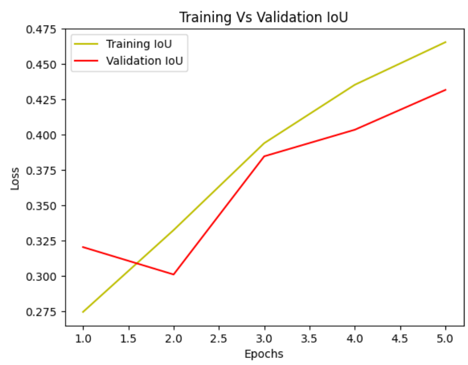
CHAPTER-9

RESULTS AND DISCUSSIONS

Old model

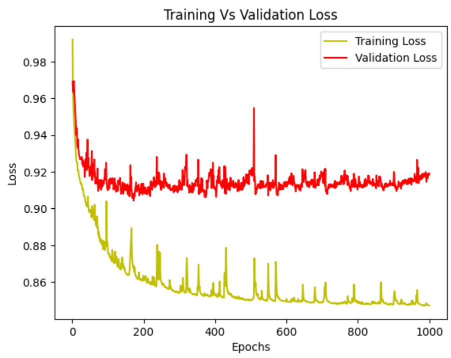


Training vs validation Loss

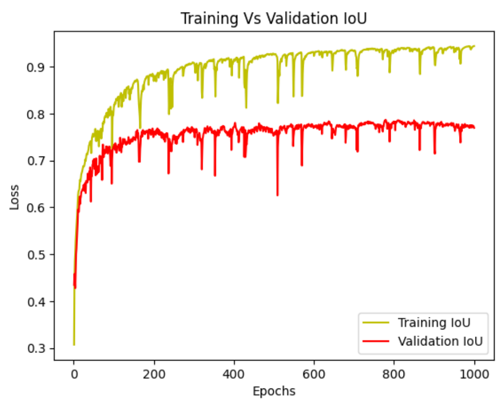


Training vs validation IoU

Improvised Model



Training vs Validation Loss



Training vs validation IoU

<u>Old Model</u>

- The previous model performed moderately, with an accuracy of 82.32%, after 51 epochs of training.
- The segmentation accuracy Jaccard coefficient was found to be 64.18%, showing a decent overlap between the true and predicted masks.
- A similar trend was indicated by validation measures, with a validation Jaccard coefficient of 60.14%.

Improvised Model

- After training for a long time—1000 epochs—the new model showed significant progress in every metric.
- Significantly, the accuracy increased to 97.49%, demonstrating an improved capacity for classification.
- The Jaccard coefficient increased significantly, reaching 94.23%, indicating a notable improvement in segmentation accuracy.
- Significant improvement was also seen in validation metrics, as evidenced by the validation Jaccard coefficient increasing to 78.15%.

Final Outcomes

Improved Image Processing System: Establishing a strong image processing pipeline produced the following results:

Better Preservation of Spatial Information: The system was able to preserve complex spatial information, which are essential for analyzing large-scale sceneries or high-resolution photos, by dividing images into smaller patches.

Optimized Deep Learning Model Training: By utilizing patch-based data feeding, training efficiency and computational feasibility were increased for deep learning models. This made it easier to integrate deep learning models for image analysis in a smooth manner.

Effective Management of Large-Scale Data: By dividing large photos into smaller patches, processing became much more effective, resulting in a considerable decrease in memory usage and computational load. This allowed the system to manage large image datasets.

Improvements to Model Performance

Integrating improved evaluation techniques with increasing strategies produced the following outcomes:

Normalized Data for Model Training: By using Min-Max scaling to normalize pixel values within a given range, model training is improved by guaranteeing consistent feature scaling throughout patches.

Measurable Segmentation Accuracy: Robust model optimization and evaluation were made possible by the incorporation of assessment metrics such as the Jaccard Coefficient (IoU), which made it easier to quantify segmentation accuracy precisely.

Advanced Loss Function Integration: By correcting class imbalance and maximizing the overlap between anticipated and ground truth masks, the addition of specialized loss functions, such as Dice and Focal Loss, greatly enhanced the training procedure.

Evaluation of Performance and Explanation

Efficient Training and Generalization: Consistent observation and evaluation of training and validation losses in conjunction with IoU measures guaranteed an equitable model fit, preventing overfitting and fostering superior extrapolation to unobserved data.

Improved Segmentation Accuracy: These elements worked together to produce increased accuracy in semantic segmentation tasks and object recognition, which allowed for more accurate object identification and delineation in images.

Overall System Effect

These improvements added together produced a system that could do the following:

Precise Image Analysis and Object Identification: Accurate and efficient image analysis, object detection, and segmentation are made possible by precise image analysis. These features are essential for a variety of applications, including autonomous systems and medical imaging.

Improved Model Generalization and Robustness: The system's capacity to manage big picture datasets, maximize model training, and precisely assess model performance resulted in reliable, broadly applicable models that could be used in a variety of contexts.

Validation Metrics

Model	Val Loss	Val Accuracy	Val Jaccard Coef
Referenced Model	0.9399	0.7831	0.6014
Improvised Model	0.9137	0.8798	0.7815

Model Performance Comparison

Model	Epochs	Loss	Accuracy	Jaccard Coef
Referenced Model	51	0.9245	0.8232	0.6418
Improvised Model	1000	0.8471	0.9749	0.9423

CHAPTER-10
CONCLUSION

In conclusion, our project's completion has resulted in a noticeable improvement in geographic object identification thanks to the creative conversion of the U-Net design into Multi U-Net. This architectural breakthrough has shown to be crucial in tackling the complex problems inherent in the processing of aerial photography. Multi U-Net's skillful application of the expanding and contracting routes that are intrinsic to U-Net topologies has enabled it to capture subtle spatial information, providing a strong basis for precise segmentation. In addition to its outstanding architectural ability, Multi U-Net has demonstrated remarkable competence in managing multi-class segmentation scenarios in geospatial images.

This capacity strengthens the model's flexibility, enabling it to be a flexible solution that can be used to a variety of domains and object types. Insights into the model's ability to capture spatial relationships have been obtained using the jaccard_coef function for performance evaluation.

As a reliable indicator, the Jaccard coefficient has become essential for evaluating the precision of our geographical object identification model. Our designed hierarchy of loss functions, which includes Focal Loss, Total Loss, and Dice Loss, has been essential in directing the training of the model. The Total Loss formula, which expresses the purposeful preference for Dice Loss over Focal Loss, has improved the model's capacity to define object boundaries and correct for unequal class distributions.

In practice, Multi U-Net has proven its worth by exhibiting flexibility in dynamic environments and demonstrating its resilience in the face of spatiotemporal variations. The improved interpretability and explain-ability of the model increase its usefulness in decision-making processes in a number of crucial areas, such as environmental monitoring, disaster management, and urban planning. In summary, our work delivers a useful and significant solution with broad ramifications, while also advancing the field of geographic object detection.

REFERENCES

[1]. Aboulnaga MM, Badran MF, Barakat MM (2021) Informal Settlements and Urban Slums' Upgrading in Megacities: Conclusions and Recommendations. Springer International Publishing, Cham, pp. 301–316.

[2]. Ahmed OS, Shemrock A, Chabot D et al (2017) Hierarchical land cover and vegetation classification using multispectral data acquired from an unmanned aerial vehicle. Int J Remote Sens, pp. 2037-2052.

[3]. Appice A, Guccione P, Acciaro E, Malerba D (2020) Detecting salient regions in a bi-temporal hyperspectral scene by iterating clustering and classification. Appl Intell, pp. 3179-3200.

[4]. Bai Y, Sun G, Li Y, et al (2021) Comprehensively analyzing optical and polarimetric SAR features for land-use/land-cover classification and urban vegetation extraction in highly-dense urban area. Int J Appl Earth Obs Geoinformation, pp. 102496.

[5]. Chen L-C, Papandreou G, Kokkinos I, et al. (2017) DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs.

[6]. Chen L-C, Zhu Y, Papandreou G et al (2018) Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y (eds) Computer Vision – ECCV 2018. Springer International Publishing, Cham, pp. 833–851.

[7]. Deng J, Dong W, Socher R, et al (2009) ImageNet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255.

[8]. Dong R, Pan X, Li F (2019) DenseU-Net-Based semantic segmentation of small objects in urban remote sensing images. IEEE Access 7, pp. 65347-65356.

[9]. Fisher T, Gibson H, Liu Y et al (2022) Uncertainty-Aware Interpretable Deep Learning for Slum Mapping and Monitoring. Remote Sens, pp. 3072.

[10]. GALLO KP, McNAB AL, KARL TR, et al (1993) The use of a vegetation index for assessment of the urban heat island effect. Int J Remote Sens, pp. 2223-2230.

[11]. Gevaert CM, Persello C, Sliuzas R, Vosselman G (2017) Informal settlement classification using point-cloud and image-based features from UAV data. ISPRS J Photogramm Remote Sens, pp. 225-236.

[12]. Giang TL, Dang KB, Le Toan Q et al (2020) U-net convolutional networks for mining land cover classification based on high-resolution UAV imagery. IEEE Access 8, pp. 186257-186273.

[13]. Giorgiani do Nascimento R, Viana F (2020) Satellite Image Classification and Segmentation with Transfer Learning. In: AIAA Scitech 2020 Forum. American Institute of Aeronautics.

[14]. Gonzales C, Sakla W (2019) Semantic segmentation of clouds in satellite imagery using deep pre-trained U-Nets. In: 2019 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), pp. 1–7.

[15]. Hashim H, Latif ZA, Adnan N (2019) Urban vegetation classification with ndvi threshold value method with very high resolution (VHR) pleiades imagery. Int Arch Photogramm Remote Sens Spat Inf Sci, pp. 3-8.

[16]. Hofmann P, Taubenbo¨ck H, Werthmann C (2015) Monitoring and modelling of informal settlements - A review on recent developments and challenges. In: 2015 Joint Urban Remote Sensing Event (JURSE), pp. 1–4.

[17]. Lakubovskii, P. (2019) "Segmentation Models." GitHub. [Online]. Available: \url{https://github.com/qubvel/segmentation_models}.

[18]. Jiang B, An X, Xu S, Chen Z (2022) Intelligent image semantic segmentation: a review through deep learning techniques for remote sensing image analysis. J Indian Soc Remote Sens, pp. 5-9.

[19]. Kuffer M, Pfeffer K, Sliuzas R (2016) Slums from space—15 years of slum mapping using remote sensing. Remote Sens, pp. 455.

[20]. Kumar V, Bandhyopadhyay S, Ramamritham K, Jana A (2020) Optimizing the redevelopment cost of urban areas to minimize the fire susceptibility of heterogeneous urban settings in developing nations: a Case from Mumbai, India. Process Integr Optim Sustain, pp. 361-378.

[21]. Kumar V, Ramamritham K, Jana A (2019) Resource allocation for handling emergencies considering dynamic variations and urban spaces: firefighting in Mumbai. ICTD, pp.1145-1146.

[22]. Lee G, Hwang J, Cho S (2021) A Novel Index to Detect Vegetation in Urban Areas Using UAV-Based Multispectral Images. Appl Sci, pp. 3472.

[23]. Li W, Saphores J-DM, Gillespie TW (2015) A comparison of the economic benefits of urban green spaces estimated with NDVI and with high-resolution land cover data. Landsc Urban Plan, pp. 105-117.

[24]. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. ArXiv14114038 Cs, pp.612-613.

[25]. Mahabir R, Croitoru A, Crooks AT et al (2018) A critical review of high and very high-resolution remote sensing approaches for detecting and mapping slums: trends challenges and emerging opportunities. Urban Sci, pp. 8.

[26]. Men G, He G, Wang G (2021) Concatenated residual attention UNet for semantic segmentation of urban green space, pp. 1441.

[27]. Nawrath M, Guenat S, Elsey H, Dallimer M (2021) Exploring uncharted territory: Do urban greenspaces support mental health in low- and middle-income countries?. Environ Res, pp. 110625.

[28]. Neupane B, Horanont T, Aryal J (2021) Deep learning-based semantic segmentation of urban features in satellite images: a review and meta-analysis. Remote Sens, pp. 808.

[29]. Park JG, Go SY, Cho GS (2013) Vegetation Classification using KOMPSAT-2 Imagery and High-resolution airborne imagery in Urban Area. J Korean Soc Geospatial Inf Sci, pp. 21–27.

[30]. Prabhu R, Alagu Raja RA (2018) Urban slum detection approaches from high-resolution satellite data using statistical and spectral based approaches. J Indian Soc Remote Sens, pp. 2033–2044.

[31]. Prabhu R, Parvathavarthini B (2022) Morphological slum index for slum extraction from high-resolution remote sensing imagery over urban areas. Geocarto Int, pp. 1–19.

[32]. Priyanka NS, Lal S et al (2022) DIResUNet: Architecture for multiclass semantic segmentation of high resolution remote sensing imagery data. Appl Intell, pp. 8-10.

[33]. Pu R, Landry S (2012) A comparative analysis of high spatial resolution IKONOS and WorldView-2 imagery for mapping urban tree species. Remote Sens Environ, pp. 516–533.

[34]. Ronneberger O, Fischer P, Brox T (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF (eds) Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. Springer International Publishing, Cham, pp. 234–241.

[35]. Sandler M, Howard A, Zhu M, et al (2018) MobileNetV2: inverted residuals and linear bottlenecks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, Salt Lake City, UT, pp. 4510–4520.

[36]. Shermeyer, J., Hogan, D., Brown, J., Etten, A.V., Weir, N., Pacifici, F., Hänsch, R., Bastidas, A., Soenen, S., Bacastow, T.M., & Lewis, R. (2020). SpaceNet 6: Multi-Sensor All Weather Mapping Dataset. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 768-777.

[37]. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. ICLR.

[38]. Szegedy C, Liu W, Jia Y, et al (2014) Going deeper with convolutions. ArXiv14094842 Cs.

[39]. Verma D, Jana A, Ramamritham K (2019) Transfer learning approach to map urban slums using high and medium resolution satellite imagery. Habitat Int, pp 4.

[40]. Wong MMF, Fung JCH, Yeung PPS (2019) High-resolution calculation of the urban vegetation fraction in the Pearl River Delta from the Sentinel-2 NDVI for urban climate model parameterization. Geosci Lett 6:2, pp. 2.

[41]. Wurm M, Stark T, Zhu X et al (2019) Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. ISPRS J Photogramm Remote Sens, pp 7-9.

[42]. Xu Z, Zhang W, Zhang T, Li J (2021) HRCNet: high-resolution context extraction network for semantic segmentation of remote sensing images. Remote Sens, pp. 69.

[43]. Yuan X, Shi J, Gu L (2021) A review of deep learning methods for semantic segmentation of remote sensing imagery. Expert Syst Appl, pp. 114417.

[44]. Yuan Y, Huang F, Lin F et al (2021) Green space exposure on mortality and cardiovascular outcomes in older adults: a systematic review and meta-analysis of observational studies. Aging Clin Exp Res, pp. 1783–1797.

[45]. Zhang Z, Liu B, Wu G, Wu J (2022) Segmentation of high-resolution remote sensing image using progressive dilated convolutional network. Appl Intell, pp.4-6.

[46]. Zhao, W.; Du, S.; Wang, Q.; Emery, W. Contextually guided very-high-resolution imagery classification with semantic segments. *ISPRS J. Photogramm. Remote Sens.* 2017, *132*, pp. 48–60.

APPENDIX-A

PSUEDOCODE

Overview

Using Keras and TensorFlow, the given code illustrates a multi-layered U-Net model for image segmentation. The model is trained with a combination of Dice loss and Categorical Focal loss to segment images into different classes.

Pseudocode Data Preparation:

- Load and preprocess image and mask data - Divide the data into sets for testing and training
- Adjust pixel values from 0 to 1 to be normal.

Architectural Models Pseudocode:

- Describe the U-Net model's design
- Use convolutional layers to provide encoding and decoding pathways; omit connections
- Use Conv2D; - Include dropout layers to avoid overfitting Layer transposition for up sampling

Measurements and Loss Functions False code:

- Establish unique loss functions: Categorical Focal loss and dice loss
- Create a complete loss by combining the loss functions.
- Use the Jaccard coefficient as a metric for assessment.

Pseudocode for Model Instruction:

- Utilize the Adam optimizer to compile the model.
- Use defined metrics and loss to train the model.
- Track training and validation results across epochs.

Outcomes and Assessment of False code:

- Show the IoU and training and validation loss.
- Produce forecasts using test data - Visually compare predicted and ground truth masks

Model Implementation False code:

- Load the model to make predictions using fresh data; - Store the trained model for later use.

Conclusion:

Using Keras and TensorFlow, the following pseudocode describes how to build a U-Net-based picture segmentation model. This acts as a manual for comprehending the underlying procedures and actions conducted in the code.

APPENDIX-B

SCREENSHOTS

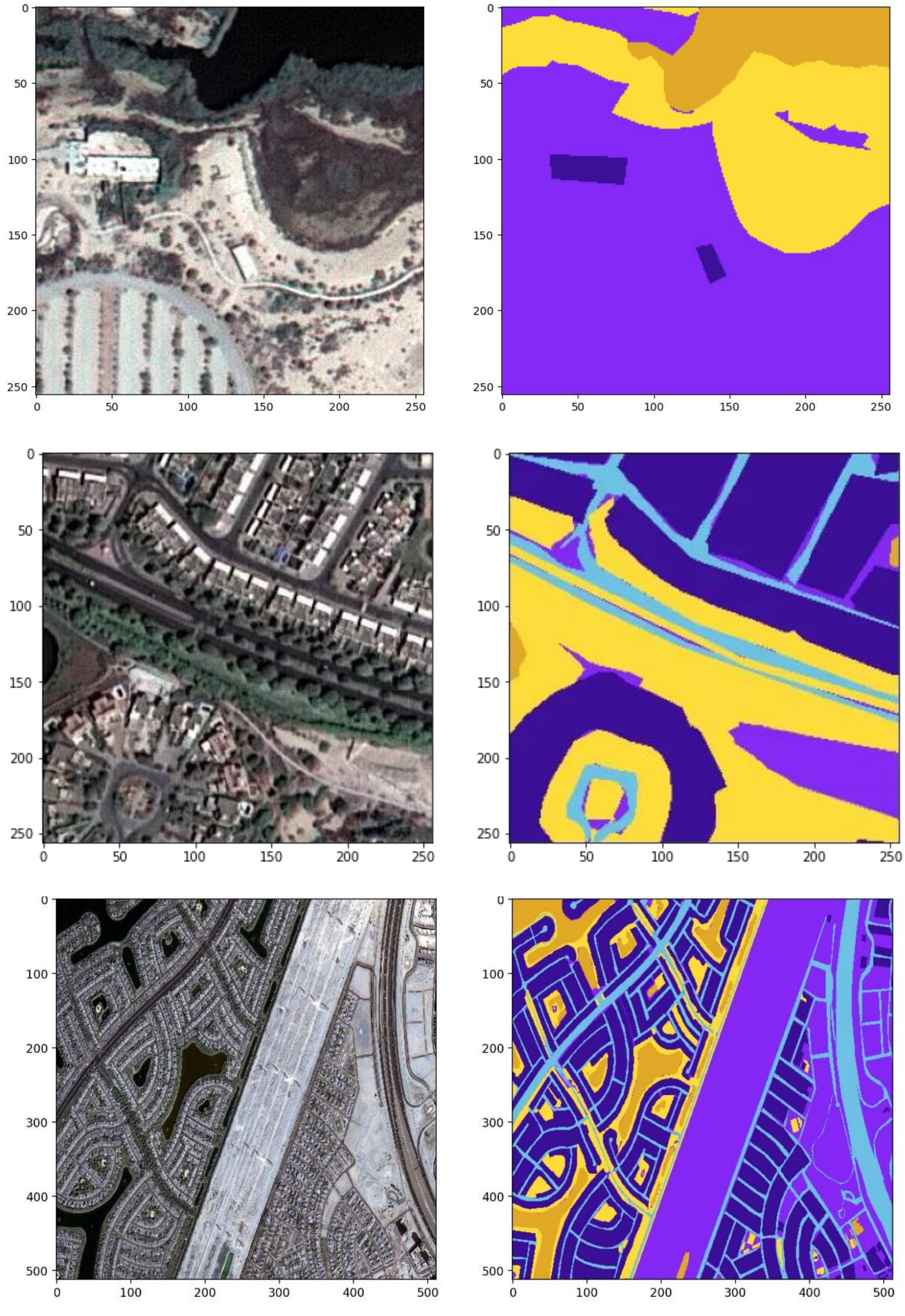


Fig. Aerial Images and their predicted masks

APPENDIX-C

ENCLOSURES

1. Conference Paper Presented Certificates of all students.

2. Plagiarism Check report.

GEOSPATIAL OBJECT DETECTION USING AERIAL IMAGERY

ORIGINALITY REPORT

18%	15%	14%	12%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to M S Ramaiah University of Applied Sciences Student Paper	3%
2	Submitted to Taylor's Education Group Student Paper	2%
3	link.springer.com Internet Source	2%
4	Submitted to University of Malta Student Paper	2%
5	www.mdpi.com Internet Source	1%
6	www.coursehero.com Internet Source	<1%
7	coek.info Internet Source	<1%
8	ris.utwente.nl Internet Source	<1%
9	Tuan Linh Giang, Kinh Bac Dang, Quang Toan Le, Vu Giang Nguyen, Si Son Tong, Van-Manh	<1%