Remittances Review
April 2024,
Volume: 9, No: 2, pp.5593-5633
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

Received: 15 March 2024, Accepted: 25 April 2024
DOI: https://doi.org/10.33282/rr.vx9i2.292

SEMANTIC SEGMENTATION AND CLASSIFICATION USING DEEP LEARNING

Muhammad Abbas[1], Haris Anjum[2], Hamza Anjum[3], Abdul Wahab Paracha[4], Muhammad Romail Imran[5], Muhammad Fasih[6].

1. (Corresponding Author) AI Research Group, Department of Computer Science & Engg. GIK Institute of Engg. Sciences & Tech. Topi, Khyber Pakhtunkhwa, Pakistan
ma_abbas2001@hotmail.com
2. AI Research Group, Department of Computer Science & Engg. GIK Institute of Engg. Sciences & Tech. Topi, Khyber Pakhtunkhwa, Pakistan
harisanjum1061@gmail.com
3. AI Research Group, Department of Computer Science & Engg. GIK Institute of Engg. Sciences & Tech. Topi, Khyber Pakhtunkhwa, Pakistan
hamzaanjum63863@gmail.com
4. AI Research Group, Department of Computer Science & Engg. GIK Institute of Engg. Sciences & Tech. Topi, Khyber Pakhtunkhwa, Pakistan
abdulwahabzahid788@gmail.com
5. AI Research Group, Department of Computer Science & Engg. GIK Institute of Engg. Sciences & Tech. Topi, Khyber Pakhtunkhwa, Pakistan
romailimran26@gmail.com
6. AI Research Group, Department of Computer Science & Engg. GIK Institute of Engg. Sciences & Tech. Topi, Khyber Pakhtunkhwa, Pakistan
muhammadfasih35@gmail.com

Abstract

This project focuses on semantic segmentation, the task of assigning a class label to each pixel in an image, using the Pyramid Scene Parsing Network (PSP Net). It involves training the PSP Net model and applying it to generate accurate predictions on new images. A key goal is to provide a user-friendly solution. To achieve this, a command-line interface (CLI) was developed, allowing users to easily train custom PSP Net models, create segmentation masks, and evaluate performance. The project leverages pre-trained PSP Net models on benchmark datasets like VOC12, ADE20K, and City scapes, making it a valuable tool for a variety of tasks such as object recognition and scene understanding. Its adaptability makes it a useful resource for both researchers and practitioners in the field of computer vision.

Key Words : UNET, CNN, Segmentation, Computer Vision, Image classification.

Introduction

Semantic segmentation is a crucial task in computer vision that involves dividing a digital image into distinct segments, with the goal of simplifying and making the image easier to analyze. In this process, each pixel is assigned a class label, turning the task into a pixel-level classification challenge. Unlike object detection, which identifies and locates objects in an image, semantic segmentation requires a more detailed understanding. It involves accurately outlining the shapes and boundaries of multiple objects or entities within an image, making it a more complex and granular task than standard image classification or object recognition (Liu et al, 2023).

Semantic segmentation holds immense importance in today's tech-driven world due to its wide range of applications. In autonomous driving, it helps systems understand road scenes by classifying elements like pedestrians, vehicles, and signs into separate categories, improving safety and navigation. In the medical field, it enables precise and automated diagnosis by segmenting medical images such as MRI and CT scans, allowing for the clear identification of anatomical structures and diseases. This increases diagnostic accuracy while saving time for medical professionals. Additionally, in remote sensing and geographic information systems, semantic segmentation is essential for land cover classification, urban planning, and environmental monitoring by extracting useful insights from satellite and aerial imagery (Chen et al, 2023).

Lightweight image models have emerged as a more reliable and accurate alternative to conventional classification methods, particularly in tasks that require efficiency without compromising performance. These models are designed to be computationally efficient, with fewer parameters and reduced memory requirements, making them ideal for real-time applications such as mobile devices, embedded systems, and autonomous vehicles. Their superior accuracy comes from their ability to extract essential features from images, despite being compact in size (Neupane et al, 2021).

Unlike traditional classification methods, which often struggle with large datasets and complex tasks, lightweight models can handle intricate tasks with improved speed and precision. They are especially valuable in scenarios where resources are limited, such as in edge computing, where devices need to process data locally rather than relying on powerful cloud servers (Shafaey et al, 2018).

Researchers are actively exploring various lightweight architectures and techniques to enhance their performance further. Some of these studies focus on optimizing existing deep learning models, such as Mobile Net and Squeeze Net, to make them even more efficient. Others explore novel architectures that push the boundaries of what lightweight models can achieve. This ongoing research aims to create models that not only match but often surpass traditional methods in both accuracy and efficiency, making them indispensable for a range of practical applications (Alias et al, 2018).

Land use and land cover maps are crucial tools derived from high-resolution satellite imagery, offering insights into how land is utilized and the various types of cover on Earth's surface. These maps are generated using data from leading satellite missions like Sentinel, Landsat, and Worldview. Sentinel, known for its high-frequency revisits, allows for continuous monitoring of changes, making it ideal for applications requiring up-to-date information, such as disaster management and agricultural monitoring (Drusch et al, 2012).

Landsat, with its long-standing historical data set, provides a unique perspective on land use trends over decades. This archive enables the tracking of long-term environmental changes, helping researchers understand the impact of human activities and natural processes on landscapes. In contrast, the Worldview satellites capture ultra-high-resolution images, offering detailed snapshots of small-scale features, which are especially valuable for urban planning and infrastructure development.(Irons et al, 2012)

The analysis of satellite imagery from these missions allows researchers to classify land into various types forests, water bodies, urban areas—and observe changes over time. Accurate and current land use/cover maps are essential for informed decision-making, supporting initiatives in sustainable development, resource management, and environmental conservation. These maps help policymakers design better strategies for managing land resources effectively and responsibly (Johnson & Koperski, 2017).

A cornerstone of semantic segmentation tasks. These networks excel at processing image data, allowing for the identification of intricate patterns and structures within images. As a result, they enable more precise delineation of objects, leading to improved performance in various applications. The rise of architectures such as the Pyramid Scene Parsing Network (PSP Net) further enhances semantic segmentation capabilities. PSP Net utilizes a pyramid pooling module that captures context at multiple scales, enabling the model to understand global context and local details simultaneously. This approach significantly boosts the accuracy and reliability of segmentation outcomes, making it especially valuable in complex scenes where multiple objects and background elements coexist. Moreover, the integration of transfer learning and pre-trained models has accelerated progress in this field, allowing researchers to leverage existing knowledge from large datasets. This has not only improved the performance of semantic segmentation models but has also made them more accessible, facilitating broader adoption across diverse sectors such as autonomous driving, healthcare, and urban planning. As research continues to advance, the future of semantic segmentation looks promising, with the potential for even greater accuracy and applicability in solving real-world challenges.

Literature Review

Semantic segmentation has experienced extraordinary advancements in recent years, largely fueled by breakthroughs in deep learning and neural networks. This literature review explores foundational works and contemporary studies, highlighting influential models such as U-Net, Multi U-Net, Mobile U-Net, Res Net, and PSP Net. These models represent the cutting edge of semantic segmentation, showcasing innovative architectures that have significantly

Remittances Review
April 2024,
Volume: 9, No: 2, pp.5593-5633
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)
enhanced the accuracy and efficiency of pixel-wise classification tasks. By examining these key developments, we aim to illuminate the trans-formative impact of these models on the field, paving the way for future research and applications across various domains.

Evolution and Current State of Research in Semantic Segmentation

The landscape of semantic segmentation research has undergone a remarkable transformation over the years, driven by groundbreaking architectural innovations and algorithmic improvements. Early methodologies relied heavily on handcrafted features and conventional machine learning techniques, which limited their ability to accurately interpret complex visual data. The advent of deep learning introduced a new era, enabling more sophisticated approaches that could learn directly from raw pixel data, paving the way for significant advancements in performance and efficiency (Scott et al, 2017).

One of the key milestones in this evolution was the introduction of Fully Convolution Networks (FCNs). These networks revolutionized the way semantic segmentation tasks were approached by allowing for end-to-end training. FCNs eliminated the need for traditional classification layers, focusing instead on producing dense pixel-wise predictions directly from the input image. This shift not only enhanced accuracy but also streamlined the training process, making it more efficient (Musaev et al, 2019).

The U-Net architecture emerged as a game-changer in the field, particularly for biomedical image segmentation. Designed with an encoder-decoder structure, U-Net efficiently utilizes limited training data while achieving outstanding performance. Its architecture allows the network to capture context while maintaining precise localization, making it especially suitable for tasks requiring high accuracy in delineating structures. Variants such as Mobile U-Net further extended its capabilities, enabling segmentation tasks to be performed on mobile devices with constrained resources (Mukhamadiyev et al, 2022).

As the field continued to evolve, other notable architectures like VGG Net and Res Net found their way into semantic segmentation applications. Initially designed for image classification, these networks were adapted to handle segmentation tasks by leveraging their deep structures for improved feature extraction. The integration of skip connections in Res Net addressed challenges associated with training deeper networks, while VGG Net's modular design offered an optimal balance between performance and computational efficiency (Valikhujaev et al, 2020).

The trend towards increasingly sophisticated architectures reflects a broader goal within the semantic segmentation community: not only to improve accuracy but also to enhance computational efficiency and generalization. Researchers are now focused on creating models that can adapt to various datasets and applications, ensuring that these advanced tools are practical for real-world scenarios. This adaptability is crucial, especially given the diverse nature of tasks that semantic segmentation can address (uchkorov et al, 2010).

Looking ahead, the ongoing research in semantic segmentation is not just about achieving higher accuracy. It aims to make models more accessible and applicable across different domains, from healthcare to autonomous driving. As technology advances, the integration of semantic segmentation into practical applications will continue to grow, ultimately enhancing our ability to analyze and interpret visual data in a variety of contexts. This evolution signifies a promising future for semantic segmentation, where robust and efficient models can transform industries and improve outcomes across numerous fields (Zhu et al, 2023).

Fully Convolutional Networks for Semantic Segmentation (2015)

- **Authors**: Jonathan Long, Evan Shelhamer, Trevor Darrell
- **Model**: Fully Convolution Networks (FCNs)
- **Accuracy/Results**: FCNs achieved state-of-the-art segmentation results on standard datasets like PASCAL VOC.
- **Limitations**: While effective, FCNs struggled with capturing fine details and required further refinement for handling small object segmentation.
- **Contribution**: This study revolutionized semantic segmentation by introducing FCNs, capable of end-to-end pixel-wise segmentation and setting a new benchmark in the field.

Remittances Review
April 2024,
Volume: 9, No: 2, pp.5593-5633
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

U-Net: Convolution Networks for Biomedical Image Segmentation (2015)

- **Authors**: Olaf Ronneberger, Philipp Fischer, Thomas Brox
- **Model**: U-Net
- **Accuracy/Results**: U-Net demonstrated superior performance in biomedical image tasks, particularly in cell segmentation challenges.
- **Limitations**: U-Net requires a substantial amount of memory, which can be a limitation for processing large images.
- **Contribution**: U-Net's introduction marked a significant advance, particularly in medical image analysis, due to its efficient use of data and ability to capture fine-grained details.

Semantic Image Segmentation with Deep Convolution Nets, Atrous Convolution, and Fully Connected CRFs (2016)

- **Authors:** Liang-Chieh Chen, George Papandreou, Alan L. Yuille
- **Model**: Deep Lab
- **Accuracy/Results**: Deep Lab showed improved segmentation accuracy, especially in capturing object boundaries.
- **Limitations:** The model complexity and the computational cost of CRFs were challenges for real-time applications.
- **Contribution**: Deep Lab's introduction of atrocious convolutions and integration of CRFs provided a significant improvement in semantic segmentation, particularly in boundary delineation.

Pyramid Scene Parsing Network (2017)

- **Authors**: Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, Jiaya Jia
- **Model**: PSP Net

- **Accuracy/Results**: PSP Net achieved impressive results on benchmarks like Image Net scene parsing challenge, setting new records.
- **Limitations**: The model's size and computational requirements limit its deployment in resource-constrained environments.
- **Contribution:** PSP Net addressed the challenge of scene parsing and object scale variations by introducing a pyramid pooling module, enhancing global context capture.

Inverted Residuals and Linear Bottlenecks (2018)

- **Authors**: Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, Liang-Chieh Chen
- **Model**: MobileNetV2
- **Accuracy/Results:** While designed for classification, adaptations for segmentation showed competitive performance with reduced computational demands.
- **Limitations:** The trade-off between model size and accuracy needs balancing for specific applications.
- **Contribution:** MobileNetV2's introduction of inverted residuals and linear bottlenecks marked a significant step towards efficient, mobile-friendly architectures.

Deep Residual Learning for Image Recognition (2016)

- **Authors**: Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun
- **Model**: ResNet
- **Accuracy/Results**: Res Net demonstrated exceptional performance in image classification challenges and its adaptations have shown similar success in segmentation tasks.
- **Limitations**: The depth of Res Net can lead to increased computational complexity.
- **Contribution**: Res Net's introduction of residual connections revolutionized deep learning, enabling the training of very deep networks and impacting subsequent segmentation models.

A Deep Neural Network Architecture for Real-Time Semantic Segmentation (2016)

- **Authors**: Adam Paszke, Abhishek Chaurasia, Sangpil Kim, Eugenio Culurciello
- **Model**: E Net
- **Accuracy/Results**: E Net offered a good balance between speed and accuracy, making it suitable for real-time applications.
- **Limitations**: The focus on speed can result in a slight compromise on segmentation accuracy.
- **Contribution:** E Net addressed the need for fast, efficient semantic segmentation, paving the way for real-time applications on edge devices.

Exploiting Encoder Representations for Efficient Semantic Segmentation (2017)

- **Authors**: Alexandre Araujo, Abdel-rahman Mohamed, Gabriel Carneiro
- **Model**: LinkNet
- **Accuracy/Results**: LinkNet showed effective segmentation capabilities while being computationally efficient.

Remittances Review
April 2024,
Volume: 9, No: 2, pp.5593-5633
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

- **Limitations**: There's a trade-off between model complexity and performance, especially for very detailed segmentation tasks.

- **Contribution:** Link Net's efficient use of encoder representations and skip connections offered a new perspective on achieving precise segmentation with reduced computational resources.

Each of these studies has significantly contributed to the advancement of semantic segmentation, addressing different challenges and application requirements. Future research directions include enhancing the efficiency and generalization of these models, reducing their dependency on large annotated datasets, and improving their applicability in diverse real world scenarios.

Gap Analysis

In the context of applying DNN and CNN techniques, such as UNET, Mobile Unet, Multi U net, Res Net, and PSP Net, to the Semantic Drone Data set, several critical research gaps and challenges have emerged. First and foremost, the dataset's limited diversity poses a challenge, as it may not encompass all possible real-world scenarios and environmental conditions encountered by drones. Expanding the data set to include a wider range of challenging scenes, such as extreme weather conditions, diverse terrains, and complex urban environments, is essential to improve model robustness. Furthermore, the resource-intensive nature of models like Res Net and PSP Net raises concerns about their applicability in resource-constrained drone systems. Research efforts should be directed toward optimizing these models for efficient deployment on drones with limited computational resources. Additionally, adapting DNN and CNN models to function seamlessly within the constraints of drones, such as battery life and payload capacity, remains an area requiring further investigation. Addressing these gaps will be pivotal in enhancing the effectiveness of semantic segmentation techniques in the realm of drone-based data analysis.

Remittances Review
April 2024,
Volume: 9, No: 2, pp.5593-5633
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

Research Questions

In the pursuit of advancing the application of deep neural network (DNN) and convolution neural network (CNN) techniques, specifically UNET, Mobile U net, Multi U net, Res Net, and PSP Net, on the Semantic Drone Data set, several fundamental research questions have arisen. Addressing these questions will guide the direction of the study and help unlock valuable insights into the utilization of these techniques in drone-based semantic segmentation tasks:

❖ How can we enhance the generalization capabilities of DNN and CNN models, such as Mobile U net and Res Net, for semantic segmentation on diverse drone imagery? - This question delves into the need for effective transfer learning and fine-tuning strategies that can adapt pre-trained models to the specifics of the Semantic Drone Data set. It also involves exploring data augmentation techniques to improve model robustness.

❖ What methods can be developed to optimize the resource utilization of computationally intensive models like Res Net and PSP Net for deployment on resource-constrained drones? - This question addresses the challenge of making stout-hearted models suitable for real-time processing on drones with limited computational power and memory. Potential solutions may include model compression, quantization, and architecture modifications.

❖ How can we handle class imbalance effectively in drone based semantic segmentation, ensuring that the model performs well on both majority and minority classes? - Dealing with class imbalance is crucial for accurate semantic segmentation. Research should investigate techniques such as class re-weighting, oversampling, or the development of specialized loss functions.

❖ What are the most appropriate evaluation metrics for assessing the performance of semantic segmentation models in the context of drone imagery? - This question focuses on the need for domain-specific evaluation metrics beyond traditional accuracy. Metrics like intersection over union (IoU), mean average precision (mAP), and class-specific performance measures should be explored to provide a more comprehensive assessment of model performance.

Remittances Review
April 2024,
Volume: 9, No: 2, pp.5593-5633
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

Remittances Review
April 2024,
Volume: 9, No: 2, pp.5593-5633
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

TABLE I

SUMMARY OF SIGNIFICANT RESEARCH IN SEMANTIC SEGMENTATION

Year	Author(s)	Paper Title	Method	Dataset	Contribution	Limitation	Results
2015	Jonathan Long et al.	Fully Convolutional Networks for Semantic Segmentation	FCN	PASCAL VOC	Introduced end-to-end pixel-wise segmentation	Struggled with fine details	State-of-the-art on PASCAL VOC
2015	Olaf Ronneberger et al.	U-Net: Convolutional Networks for Biomedical Image Segmentation	U-Net	Medical Images	Efficient use of data in medical imaging	High memory requirement	Superior performance in cell segmentation
2016	Liang-Chieh Chen et al.	Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs	DeepLab	Various	Introduced atrous convolutions and CRFs	High computational cost	Improved boundary segmentation
2017	Hengshuang Zhao et al.	Pyramid Scene Parsing Network	PSPNet	ImageNet	Pyramid pooling module for global context	Large model size	Set records on ImageNet
2018	Mark Sandler et al.	Inverted Residuals and Linear Bottlenecks	MobileNetV2	Various	Efficient architecture for mobile devices	Trade-off between size and accuracy	Competitive performance
2016	Kaiming He et al.	Deep Residual Learning for Image Recognition	ResNet	ImageNet	Introduced residual connections	Increased complexity	Exceptional in classification tasks
2016	Adam Paszke et al.	A Deep Neural Network Architecture for Real-Time Semantic Segmentation	ENet	Various	Compact architecture for real-time segmentation	Compromise on accuracy	Good balance between speed and accuracy

Remittances Review
April 2024,
Volume: 9, No: 2, pp.5593-5633
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

2017	Alexandre Araujo et al.	Exploiting Encoder Representations for Efficient Semantic Segmentation	LinkNet	Various	Efficient use of encoder representations	Complexity-performance trade-off	Effective and efficient segmentation

- How can we improve the contextual understanding of scenes in drone-based semantic segmentation, moving beyond pixel-level segmentation to recognize object relationships and context? - Enhancing semantic segmentation models to grasp spatial and contextual information in complex urban environments is essential. Research should investigate techniques for incorporating contextual cues and understanding object interactions.
- What interdisciplinary collaborations and methodologies can facilitate a more holistic approach to drone-based semantic segmentation, involving experts from the fields of remote sensing and Geo spatial analysis? - Bridging the gap between deep learning and remote sensing requires interdisciplinary collaboration. This question explores how insights from remote sensing experts can be integrated into the development of robust drone-based semantic segmentation solutions.
- What ethical considerations should be addressed when using drones for semantic segmentation, particularly concerning privacy, data security, and responsible data collection practices? - As drones become more prevalent, ethical concerns arise. Researchers should investigate methods for anonymize sensitive information in drone imagery and ensuring responsible data usage.

These research questions provide a framework for exploring the challenges and opportunities in the application of DNN and CNN techniques to the Semantic Drone Data set. Addressing these questions will contribute to the development of more effective, efficient, and ethically sound solutions for drone based semantic segmentation tasks.

Problem Statement

Remittances Review
April 2024,
Volume: 9, No: 2, pp.5593-5633
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

This study focuses on assessing the influence of integrating deep neural network (DNN) and convolution neural net- work (CNN) models, such as UNET, Mobile U net, Multi U net, Res Net, and PSP Net, into existing semantic segmentation models when applied to the Semantic Drone Data set. The primary goal is to evaluate how these advanced models impact the quality of semantic segmentation results in drone imagery analysis. Key challenges include determining the transfer ability of per-trained models to the drone domain, optimizing resource-intensive models for real-time processing on drones, handling class imbalance effectively, defining appropriate evaluation metrics, improving contextual understanding, fostering interdisciplinary collaboration, and addressing ethical considerations related to data privacy and security. Addressing these challenges will yield valuable insights into the effectiveness and ethical implications of deploying advanced models for semantic segmentation tasks in drone-based applications.

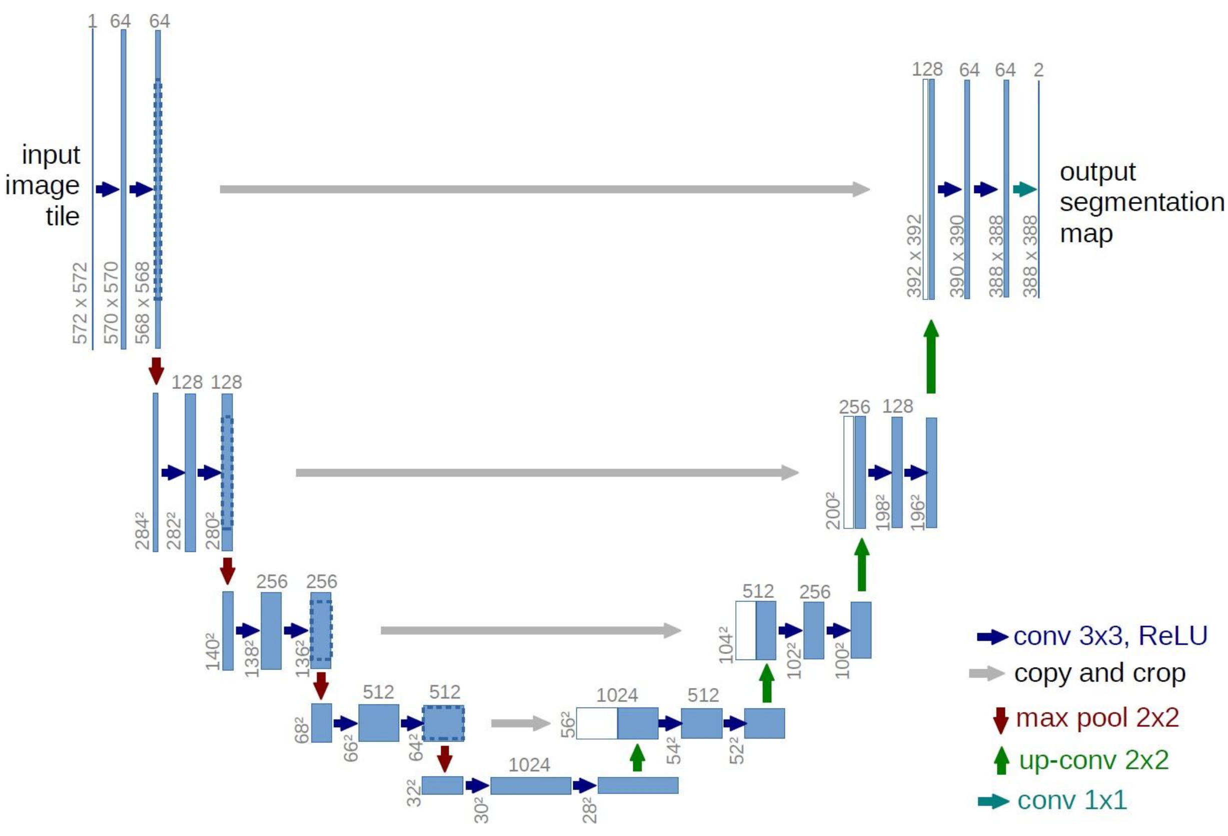

Fig. 1- Architecture of Unet model

Novelty of the Study

Remittances Review
April 2024,
Volume: 9, No: 2, pp.5593-5633
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

This study brings several novel aspects to the field of drone- based semantic segmentation (Wu, et al, 2023):

- **Model Integration Assessment**: This research uniquely focuses on evaluating the impact of integrating advanced DNN and CNN models, including UNET, Mobile U net, Multi U net, Res Net, and PSP Net, into existing semantic segmentation frameworks for drone imagery analysis. Few studies have comprehensively explored the effects of these advanced models in the context of drones.

- **Resource Optimization for Drones**: The study pioneers the investigation of resource optimization techniques for deploying resource-intensive models like Res Net and PSP Net on drones. This addresses a critical gap in the literature, making it more practical to utilize such models in real-time drone applications.

- **Class Imbalance Solutions:** Addressing class imbalance challenges in drone-based semantic segmentation is a novel aspect of this research. It delves into techniques for accurately segmenting both majority and minority object classes in drone imagery, a topic with limited prior exploration.

- **Contextual Understanding**: The study goes beyond pixel-level segmentation by examining how advanced models contribute to contextual understanding in complex drone- captured scenes. This focus on object relationships and context adds a unique dimension to the research.

- **Interdisciplinary Collaboration:** The study emphasizes the importance of interdisciplinary collaboration between deep learning experts and specialists in remote sensing and Geo spatial analysis. Such collaboration is crucial for creating holistic and domain-specific solutions for drone-based semantic segmentation.

Remittances Review
April 2024,
Volume: 9, No: 2, pp.5593-5633
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

- **Ethical Considerations:** The study underscores the ethical dimensions of using drones for data collection, emphasizing responsible data practices and privacy protection. This ethical perspective adds a distinctive and responsible approach to the study.

In summary, the novelty of this study lies in its holistic evaluation of advanced DNN and CNN models within the specific context of drone-based semantic segmentation. It addresses resource optimization, class imbalance, contextual understanding, interdisciplinary collaboration, and ethical considerations, collectively advancing the field and contributing to more effective and responsible drone-based semantic segmentation solutions.

Remittances Review
April 2024,
Volume: 9, No: 2, pp.5593-5633
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

Fig. 2 - Workflow model of the whole report

Significance of Our Work

Our research carries significant implications and contributions to the fields of computer vision, remote sensing, and drone-based applications:

- **Advancing Semantic Segmentation**: This study con- tributes to the advancement of semantic segmentation techniques, a fundamental task in computer vision. By evaluating the impact of state-of-the-art DNN and CNN models on drone imagery, we enhance the accuracy and reliability of object recognition and scene understanding.

- **Efficient Resource Utilization**: Our exploration of re- source optimization techniques for resource-intensive models, such as Res Net and PSP Net, addresses a pressing need in the

Volume: 9, No: 2, pp.5593-5633
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

drone industry. This can lead to the development of more power-efficient and real-time drone systems, expanding their applicability in various domains.

- **Improved Object Detection**: By addressing class imbalance challenges, we enhance object detection and segmentation, particularly for minority object classes in drone imagery. This is crucial for applications like precision agriculture, infrastructure inspection, and disaster response.
- **Enhanced Contextual Understanding**: Our study emphasizes contextual understanding, extending beyond basic pixel-level segmentation to enable drones to comprehend object relationships and the broader scene context. This capability is crucial for applications in complex urban environments and large-scale monitoring, where understanding the interplay between objects and their surroundings is essential for effective decision-making and navigation.
- **Interdisciplinary Collaboration**: The focus on interdisciplinary collaboration effectively bridges the divide between deep learning experts and remote sensing specialists. This synergy not only leads to the development of more robust, domain-specific solutions but also nurtures a collaborative spirit within research communities. By combining diverse expertise and perspectives, we can tackle complex challenges more effectively, ultimately driving innovation and enhancing the impact of our work.
- **Ethical Data Practices**: Our research highlights the critical importance of ethical data practices, emphasizing the need to maintain data privacy and security when employing drones for data collection. By prioritizing responsible data usage, we establish a benchmark for ethical considerations in drone-based applications. This commitment not only safeguards individuals' privacy but also fosters trust in the technologies we develop, encouraging broader acceptance and adoption of drone solutions in various fields.
- **Industry Applications**: The results of our work have significant practical applications across multiple industries, including agriculture, environmental monitoring, disaster management, and infrastructure inspection. By enhancing semantic segmentation

Remittances Review
April 2024,
Volume: 9, No: 2, pp.5593-5633
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)
capabilities on drones, we can achieve more precise decision-making and optimized resource allocation. This advancement allows for better analysis of agricultural land, more effective monitoring of environmental changes, timely responses to disasters, and thorough inspections of infrastructure, ultimately leading to improved efficiency and outcomes in these critical sectors.

Academic Advancement: The findings from this research contribute significantly to the academic discourse by addressing essential research gaps at the intersection of deep learning and drone technology. By exploring these uncharted territories, we lay a solid foundation for future studies and innovations in the field. This work not only enhances understanding but also inspires new avenues of research that can lead to groundbreaking advancements in both deep learning and drone applications.

In conclusion, the significance of our work lies in its ability to elevate the capabilities of drone-based semantic segmentation, facilitating more accurate and efficient data analysis across a variety of applications. This research not only enriches academic knowledge but also drives practical advancements, ultimately benefiting various industries and society as a whole. By improving how drones analyze and interpret data, we pave the way for smarter decision-making and resource management, showcasing the profound impact of our findings on real-world challenges.

METHODOLOGY

Data set Description

The study employs the Semantic Drone Data set, which is curated from Kaggle and tailored for semantic segmentation tasks in drone applications. This data set features high-resolution aerial imagery, showcasing a wide range of landscapes, urban scenes, and diverse objects captured from above. Each image in the data set is meticulously labeled, with every pixel annotated to represent various objects and terrain types. This comprehensive data set plays a crucial role in training and assessing the performance of semantic segmentation models in real-

Remittances Review
April 2024,
Volume: 9, No: 2, pp.5593-5633
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)
world drone scenarios, ensuring that the models can effectively interpret complex aerial environments.

Image Preprocessing and Augmentation: The initial phase of the project involves rigorous preprocessing and augmentation of the images from the Semantic Drone Data set. These steps are crucial to standardize the input data format, enhance the dataset's variability, and improve the robustness of the model against real-world scenarios.

- **Prepossessing**: To prepare the images for the deep learning model, we implemented a series of prepossessing steps:

  - **Re sizing**: All images were re-sized to a standard dimension to maintain consistency across the data set. This step ensures that the input to the neural network is of a uniform size.

  - **Normalization**: Pixel values were normalized to have a mean of 0 and a standard deviation of 1. This normalization helps in speeding up the training by ensuring that the optimization landscape is smoother.

  - **Color Space Conversion**: Images were converted from RGB to a more uniform color space, such as LAB, to reduce the effects of lighting variations and improve the model's performance on different color schemes.

- **Augmentation**

To increase the robustness and performance of the model, we employed the Augmentations library, a fast and flexible image augmentation tool. The augmentation strategy included a variety of transformations designed to expose the model to various aspects of real-world data (Sun et al, 2023):

  - **Horizontal and Vertical Flip (p=0.5):** Images were flipped horizontally or vertically with a 50

- **Rotation (limit=45 degrees, p=0.5):** The images were randomly rotated within ±45 degrees. This increase in range allows the model to become invariant to more drastic rotational changes, simulating a wider array of possible drone camera angles.
- **Random Brightness and Contrast (p=0.5)**: To simulate different lighting conditions, the brightness and contrast of the images were randomly altered, making the model more robust to changes in illumination.
- **Random Gamma Correction (p=0.5):** Gamma correction was applied to adjust the image's gamma values, thus altering the illumination of images and making the model adaptable to various lighting conditions.
- **Random Crop and Scale (p=0.5):** Random cropping and scaling of the images were performed to mimic the effect of varying distances and sizes of objects as seen from a drone. This technique helps in recognizing objects of different scales and at varying positions.
- **Elastic Transformations (p=0.1)**: To introduce geometric distortions in the images and make the model resilient to shape variations, elastic transformations were applied with a small probability. This mimics real-world scenarios where objects might be partially obscured or deformed.
- **Noise Injection (p=0.25):** Random noise was injected into the images to simulate the effect of a noisy sensor or environmental perturbations that might affect image quality.

These augmentation techniques collectively expand the variability within the training data set, thereby enhancing the generalization capability of our semantic segmentation model. They simulate a wide range of possible environmental conditions and object appearances, preparing the model for diverse real- world scenarios encountered in drone imagery.

- **Data Splitting**: The data set is divided into training, validation, and test sets. This separation is crucial for an unbiased evaluation of the models, with the training set used to train the

Remittances Review
April 2024,
Volume: 9, No: 2, pp.5593-5633
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

models, the validation set for hyper parameter tuning, and the test set for the final evaluation (Zhu et al, 2023).

- **Model Development**: For tackling the intricate task of semantic segmentation in drone imagery, our study incorporates a diverse range of models. Each model is chosen fortis unique architectural strengths and suitability for analyzing aerial views, particularly for detecting and segmenting people within the images.

Volume: 9, No: 2, pp.5593-5633
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

Fig. 3 - How does our images data set look like

- ❖ **UNET:** Architecture and Application: UNET, initially designed for biomedical image segmentation, boasts a symmetric encoder-decoder architecture. The encoder progressively compresses the spatial dimensions while increasing the feature depth, capturing comprehensive contextual information. The decoder then reconstructs the detailed segmentation map by up sampling and combining features from the encoder through skip connections. These connections are pivotal for retaining critical boundary details, making UNET highly effective in delineating precise object boundaries, such as the outline of a person in aerial images (Wang et al, 2023). Adaptation for People Detection: For our specific task of identifying people in drone imagery, we fine-tuned the UNET architecture to focus more on the human shape and size variations from aerial perspectives. The network was trained extensively on annotated images where people are marked, learning to segment these regions accurately from the surrounding environment.
- ❖ **Mobile UNET:** Architecture and Application: Mobile- UNET adapts the robust structure of UNET for environments constrained by computational resources. By incorporating depth wise separable convolutions, Mobile UNET significantly reduces the computational load, making real-time semantic segmentation feasible on mobile or

Remittances Review
April 2024,
Volume: 9, No: 2, pp.5593-5633
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

embedded devices like drones. Despite its lightweight nature, it retains considerable accuracy in segmenting images, crucial for timely and reliable person detection in aerial surveillance.

- **Use in Drone Imagery:** In the context of drones, Mobile-U Net's efficient architecture ensures that the model can run on the limited computational resources available onboard. This ability makes it particularly valuable for scenarios requiring immediate analysis and decision-making based on the drone’s footage, such as search and rescue operations or crowd monitoring.

- **Multi UNET:** Architecture and Application: Multi UNET extends the capabilities of the original UNET by integrating multiple UNET structures either sequentially or in parallel. This amalgamation allows the model to capture and combine a rich set of features from different levels of abstraction, enhancing its ability to understand and segment complex scenes. The diversity in features is particularly beneficial for aerial images, which may contain varied textures, patterns, and objects (Zhang et al, 2023).

- **Enhancement for Semantic Segmentation:** For semantic segmentation, especially in discerning people within diverse urban or rural landscapes, Multi UNET’s comprehensive feature extraction proves invaluable. It provides a more nuanced understanding of human figures against varying backgrounds, improving the model’s accuracy in person segmentation.

- **RESNET:** Architecture and Application: Res Net introduces a revolutionary design with residual connections to facilitate the training of very deep networks. These connections enable the network to bypass certain layers, effectively addressing the vanishing gradient problem and promoting feature reuse. This architectural innovation allows Res Net-based models to capture an extensive range of features, from basic to complex, enhancing their performance in detailed tasks like semantic segmentation.

- **Specifics for Person Detection:** In the scope of detecting people in drone imagery, a Res Net-based model can learn detailed and nuanced patterns associated with human figures, even from high altitudes or in crowded scenes. The depth of the network, combined with residual learning, ensures that subtle distinctions between people and other objects are captured, leading to more accurate segmentation and classification.
- **PSPNET:** Architecture and Application: PSP Net stands out with its pyramid pooling module, designed to collect contextual information at various scales. This feature is incredibly beneficial for semantic segmentation, where understanding both the finer details and the broader context is essential. By aggregating features under different sub-region scales, PSP Net effectively captures the essence of objects and their surroundings, leading to a robust segmentation performance (Li et al, 2023).
- **Application in Drone Imagery**: For drone-based applications, PSP Net's ability to handle variations in object size and appearance is vital. As drones capture images from varying altitudes and angles, objects like people can appear significantly different. PSP Net's pyramid pooling ensures that the model remains sensitive to these variations, accurately segmenting and classifying people irrespective of the drone's altitude or the scene's complexity.

In summary, each model employed in our study offers unique benefits for the task of semantic segmentation in drone imagery, with a particular emphasis on accurately detecting and classifying people within the scenes. By leveraging these models' strengths and adapting them to the specific nuances of aerial images, we aim to achieve a robust and reliable system capable of assisting in various applications, from urban planning and crowd monitoring to search and rescue operations.

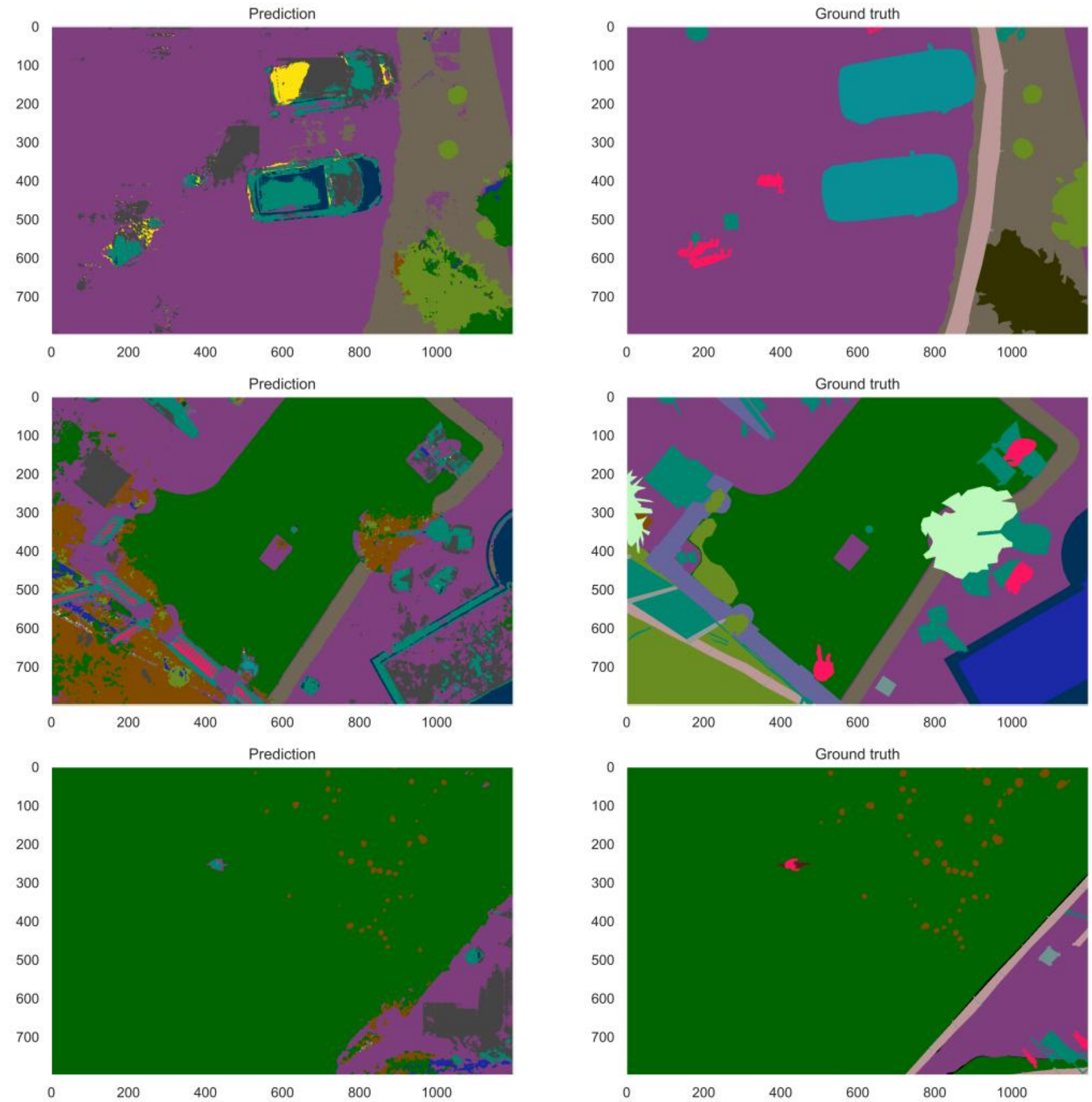

Fig. 4 - Training loss and accuracy of the Mobile Unet model.

Evaluation: The evaluation of these models is conducted rigorously on the test set using a suite of metrics to assess their segmentation performanc**e.**

- **Evaluation of UNET:** The performance of UNET is evaluated based on its ability to accurately segment fine details in the images. Given its architecture, the focus is on

Volume: 9, No: 2, pp.5593-5633
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

assessing how well the skip connections help in recovering detailed information in the segmented output.

- **Evaluation of Mobile UNET:** For Mobile UNET, the evaluation emphasizes not only on segmentation accuracy but also on the model's efficiency. Metrics such as model size and inference time are crucial, alongside traditional segmentation accuracy metrics, to determine its suitability for real-time applications on drones (Gulzar, 2023).
- ,**Evaluation of Multi UNET:** The effectiveness of Multi UNET is measured by its ability to handle complex scenes with diverse features. The evaluation focuses on the model's capacity to leverage multiple networks for enhanced segmentation performance.
- **Evaluation of RESNET:** Res Net-based models are evaluated for their depth and ability to learn complex patterns. The emphasis is on how effectively the residual connections aid in training deeper networks and their impact on the segmentation accuracy, especially for intricate patterns in aerial images. Here we would be working on classification of images where we classify based on if a person is found in the drone image or not. This was done to check how segmentation works as a classification model. We used confusion matrix, f1-score, recall, precision and support. Each model's performance is assessed using metrics like ac- curacy, precision, recall, and F1-score, or image segmentation providing a comprehensive understanding of their strengths and limitations in the context of semantic segmentation for drone imagery.

Further Evaluation: Our comprehensive evaluation process rigorously assesses the performance of each model on the test set using a suite of metrics designed to measure segmentation and classification effectiveness. The evaluation strategy is with specific measures of boundary accuracy, are used to understand UNET's performance in this nuanced task.

- **Evaluation of Mobile UNET:** Performance and Efficiency Metrics: Mobile UNET's evaluation balances segmentation accuracy with model efficiency, critical for real-time applications. We assess not only the traditional accuracy metrics but also compute the model's size, inference time, and power consumption to gauge its viability for on-drone deployment. The focus is on maintaining high precision and recall while ensuring the model remains lightweight and fast.

- **Evaluation of Multi UNET:** Complex Scene Handling: Multi UNET's ability to handle complex scenes with diverse features is under scrutiny. We evaluate how effectively the ensemble of networks within Multi UNET enhances performance, particularly in segmenting and classifying images with multiple people or challenging environmental conditions. The evaluation focuses on the model's capacity to leverage its comprehensive feature set for improved classification accuracy (Le at al, 2023).

- **Evaluation of RESNET:** Depth and Pattern Recognition: The evaluation for Res Net-based models centers on their depth and ability to learn and generalize complex patterns necessary for distinguishing people in varied drone images. We pay special attention to the models' ability to avoid over fitting while maintaining high sensitivity and specificity in person detection. The impact of residual connections on maintaining robustness across different scenes is also a key focus (Singh &Rani, 2020).

- **Classification as a Semantic Segmentation Task:** In all models, we re conceptualize semantic segmentation as a binary classification task - the presence or absence of a person in the image. This approach allows us to use segmentation models to provide valuable insights into scene content, which is crucial for applications like surveillance and search and rescue operations (Zhang, 2020).

Evaluation Metrics, The performance of the semantic segmentation models is assessed using a variety of key metrics:

- **Confusion Matrix:** The confusion matrix visualizes the performance of the models. It shows the number of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN).

- **Recall (Sensitivity):** Recall measures the model's ability to correctly identify all relevant instances (TP) of each particularly focused on understanding how well each model can distinguish between the presence and absence of a person in drone imagery,

Volume: 9, No: 2, pp.5593-5633
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

effectively turning the segmentation task class.**Recall = TP / TP + FN,** into a classification problem.

- **Evaluation of UNET:** Performance Metrics: The evaluation of UNET is based on its precision in segmenting fine details, which is vital for identifying smaller or partially obscured individuals in the images. We examine the effectiveness of skip connections in preserving boundary details and enhancing the model's ability to accurately classify images based on the presence of a person. Standard metrics, along In semantic segmentation, high recall indicates that the model correctly identifies most of the relevant pixels for each class, reducing the chance of missing important features.

- **Precision:** Precision assesses the accuracy of positive predictions. **Precision= TP / TP + FP,** High precision in semantic segmentation means that the model accurately labels pixels, ensuring that different classes are correctly distinguished.

- **Accuracy:** This metric evaluates the overall correctness of the models. **Accuracy = TP + TN/ TP + TN + FP + FN,** While accuracy is a general indicator of performance, it might not always be the best metric for imbalanced datasets common in semantic segmentation.

- **F1-Score:** The F1-score is a harmonic mean of precision and recall. **F 1-Score = 2x Precision × Recall/ Precision + Recall,** The F1-score is particularly useful in semantic segmentation when balancing the precision-recall trade-off. It ensures that both false positives and false negatives are taken into account, which is crucial for accurately segmented images.

These evaluation metrics collectively offer a comprehensive view of the models' performance, guiding the fine-tuning process and ensuring the development of highly accurate and reliable semantic segmentation models for drone imagery.

RESULTS

- **Overview.** In this study, we compared the performance of Multi Unet, Mobile Net, Res Net, and UNET models for our specific task. Each model was trained over a set of

Volume: 9, No: 2, pp.5593-5633
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

epochs tailored to its computational demands, with Res Net models trained on lower epochs due to higher computational costs. The performance of all models was evaluated using metrics such as accuracy, precision, and recall. It's worth noting that while Res Net focused on classification post-segmentation, the other models were evaluated purely on their segmentation capabilities.

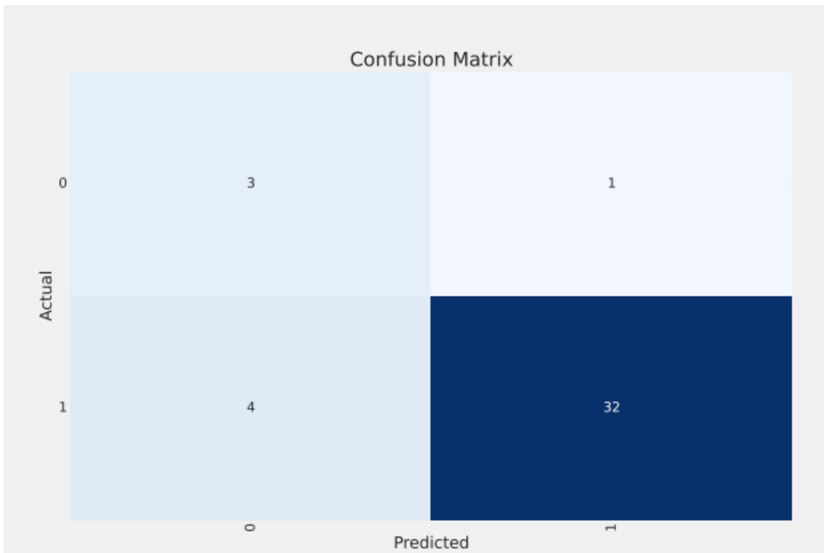

Fig. 5- This figure shows the confusion matrix for RESNET

- **Training Results.** The training process for all models varied in epochs, reflecting their complexity and computational demands. The training's progress is documented in terms of loss and accuracy and depicted in the subsequent figures.

  - **Multi Unet:** The Multi Unet model demonstrated a consistent decrease in loss with a corresponding increase in accuracy over the epochs. The training loss and accuracy for the Multi Unet model are shown below.

  - **UNET:** The UNET model showed steady improvements in both training loss and accuracy as the training progressed. The relevant training metrics are presented below (Singh & Nongmeikapam, 2023).

- **Mobile Net:** The Mobile Net model's training process also reflected a consistent improvement in loss and accuracy. The performance during training is depicted in the following figure.
- **RESNET:** The Res Net model, primarily used for classification after segmentation, exhibited a steady increase in accuracy and decrease in loss over its training period. The training performance is illustrated below.

❖ **Evaluation Metrics.** After training, the models were evaluated on a separate test set to assess their performance in real-world conditions. The key metrics used for evaluation were accuracy, precision, and recall. The results are summarized in the following tables and provide a comprehensive understanding of each model's strengths and limitations.

- **Multi Unet:** The performance metrics for the Multi Unet model are detailed in the corresponding table.
- **UNET:** The performance metrics for the UNET model are detailed in the corresponding table.
- **RESNET:** The Res Net model's classification performance post-segmentation is detailed in Table II. Notably, the Res Net model achieved high accuracy due to its classification focus after segmentation.

Metric	Accuracy	Precision	Recall
Value	87.50%	91.56%	87.50%

TABLE II
EVALUATION METRICS FOR THE RESNET MODEL.

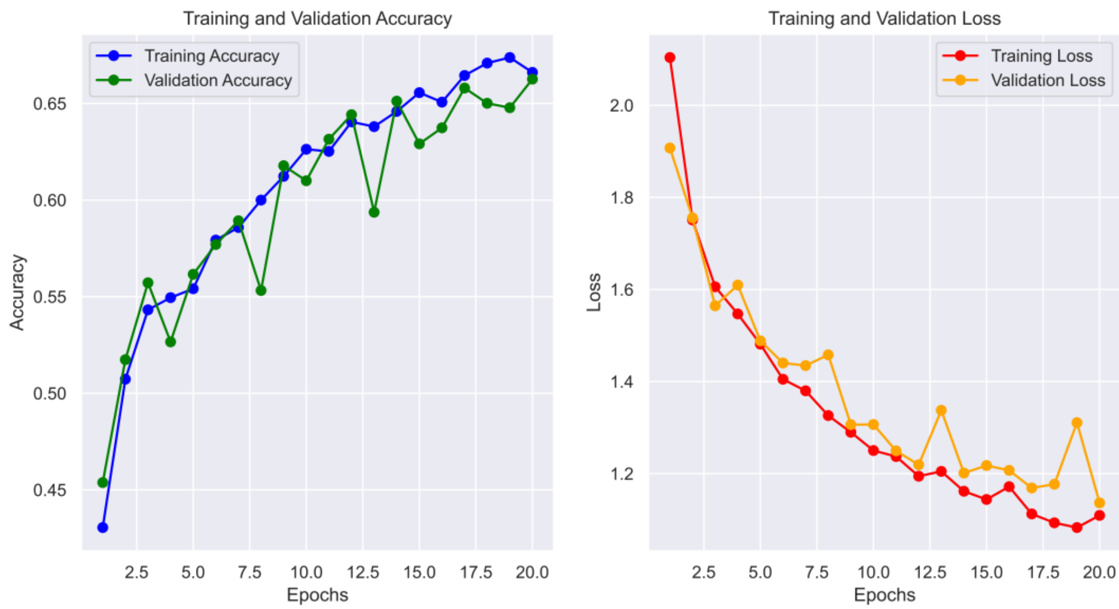

Fig. 6. Training loss and accuracy of the Multi Unet model.

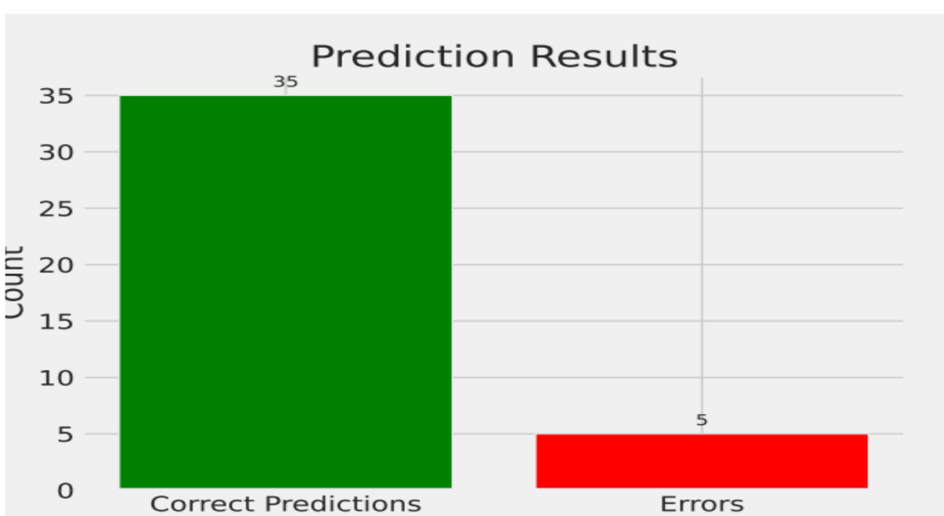

Fig. 7. Training loss and accuracy of the RESNET model.

Volume: 9, No: 2, pp.5593-5633
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

- **Discussion.** The comparative analysis of the Multi Unet, UNET, Mobile Net, and RESNET models provides insightful findings into their respective capabilities and limitations for semantic segmentation, particularly in identifying individuals in drone imagery.

- **Performance Insights:** The UNET model demonstrated impressive accuracy, indicating its robust feature extraction and segmentation capabilities, especially with fine details. This makes it a strong candidate for tasks requiring high precision, such as detailed environmental monitoring or specific object identification in complex scenes. On the other hand, Multi Unet, with its ensemble approach, showed promise in handling diverse features and complex scenes, though its accuracy was lower compared to the standalone UNET model. This suggests that while Multi Unet is potentially more versatile, it may require further optimization or a larger training data set to reach its full potential. Mobile Net, tailored for efficiency, showed reasonable accuracy and could be highly beneficial for real-time or on device processing applications. Its lightweight nature makes it particularly suitable for drones with limited computational resources, where maintaining a balance between performance and speed is critical. RESNET, employed for classification post-segmentation, showed high accuracy in determining the presence of a person in the image. Its deep architecture and residual learning make it adept at capturing complex patterns, crucial for accurate classification in varied and challenging environments. Computational Constraints: One of the critical considerations in this study is the computational constraints, which limited the extent of training for especially computation-intensive models like RESNET. Despite these limitations, all models exhibited positive growth in performance metrics with each epoch. This trend is indicative of the potential for significant improvements in model performance with access to more powerful computational resources. It is expected that with extended training on a high-performance GPU, the models would achieve even higher accuracy and generalization capabilities, making them more effective for practical applications.

- **Choosing the Right Model**: Selecting the appropriate model for a specific task in drone-based applications depends on several factors, including the required accuracy,

Remittances Review
April 2024,
Volume: 9, No: 2, pp.5593-5633
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

computational limitations, and the nature of the task. For instance, UNET’s high precision might be preferred for tasks requiring detailed segmentation, while Mobile Net’s efficiency would be advantageous for real-time processing. Multi Unet could be beneficial for varied and complex scenes, offering a blend of features and perspectives. Meanwhile, RES NET’s classification capabilities make it ideal for tasks where quick and accurate identification of specific elements, such as people, is critical.

In conclusion, despite the challenges posed by computational constraints, the encouraging trends observed in the models’ training trajectories indicate that, with sufficient resources, these models can be refined to achieve even greater performance. Each model possesses unique strengths that contribute valuable capabilities to semantic segmentation tasks. Future research should aim to leverage these strengths in an optimized manner, potentially by integrating the advantages of different models or further customizing them to meet specific application needs. Ongoing advancements in computational resources and model optimization techniques will undoubtedly facilitate the realization of these models' full potential, enhancing their effectiveness for practical, real-world applications in drone imagery analysis.

Remittances Review
April 2024,
Volume: 9, No: 2, pp.5593-5633
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

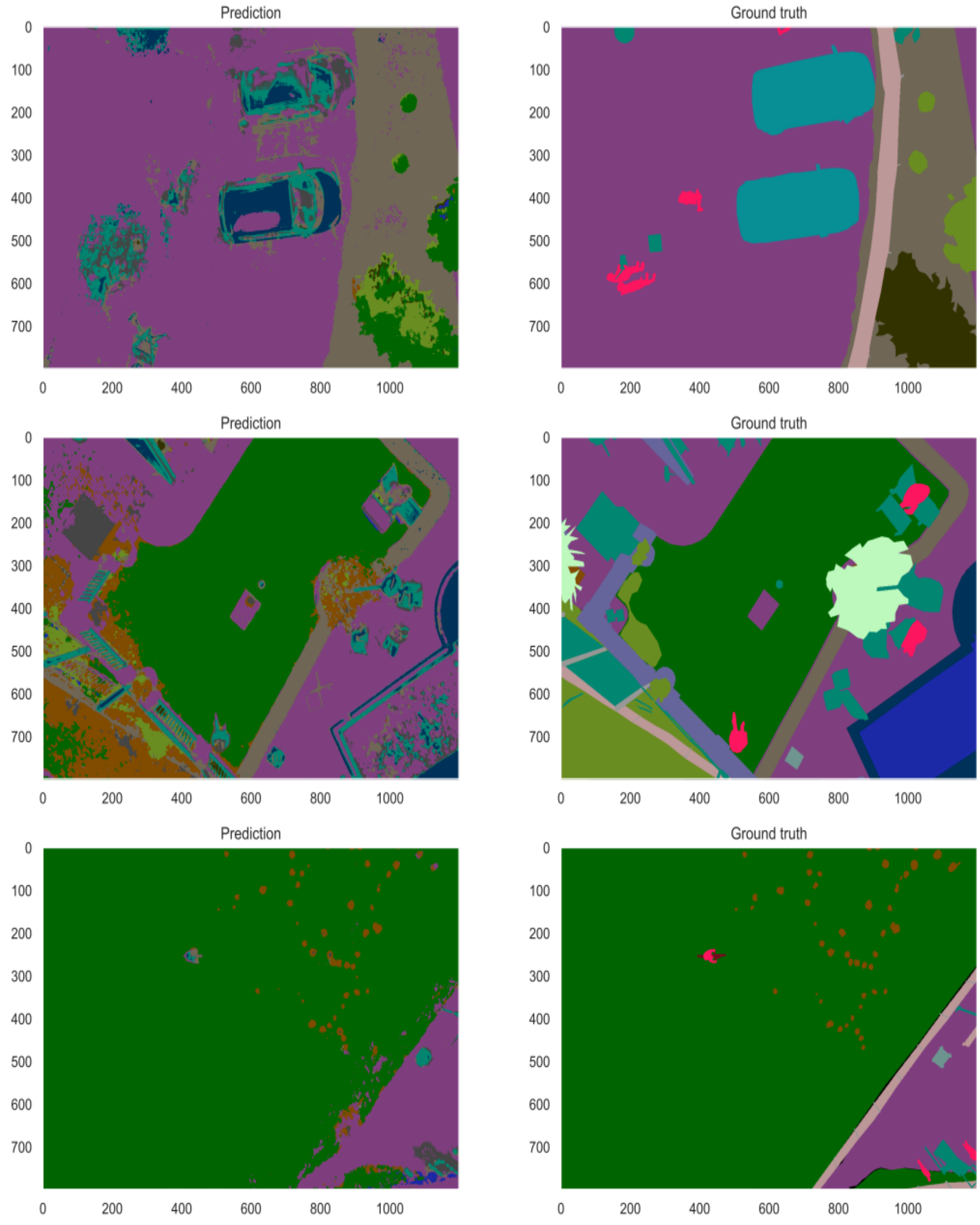

Fig. 8. Prediction and ground truth of the Multi Unet model.

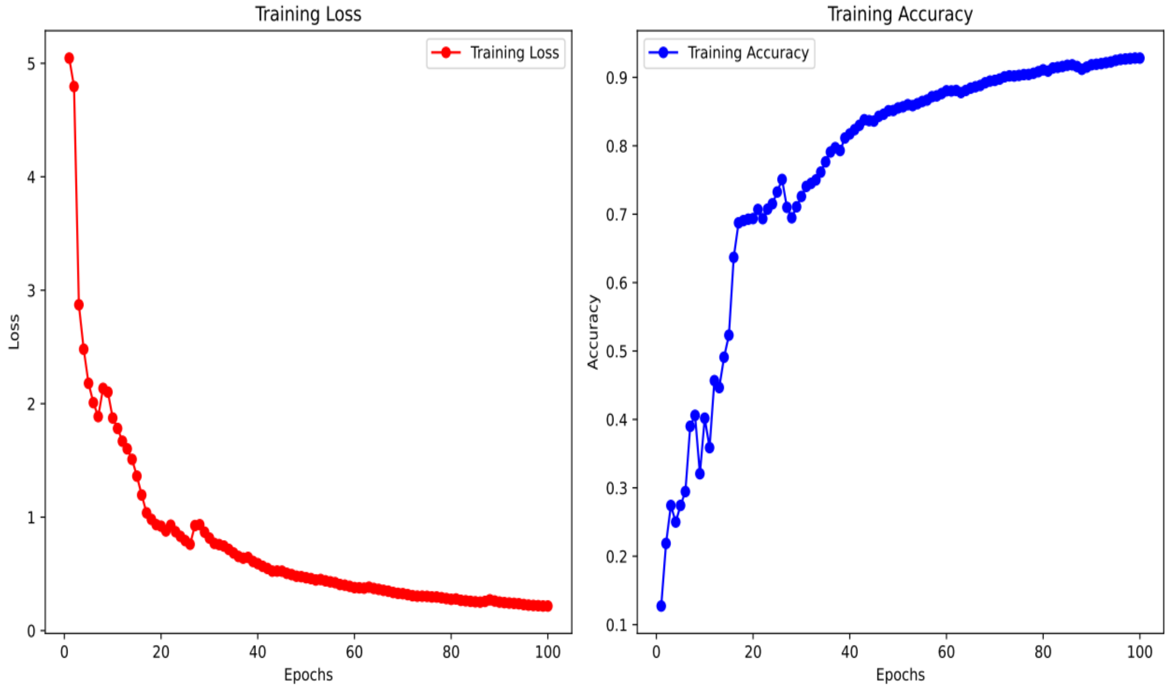

Fig. 9. Training loss and accuracy of the UNET model.

Fig. 10. Training loss and accuracy of the MobileNet model.

Remittances Review
April 2024,
Volume: 9, No: 2, pp.5593-5633
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

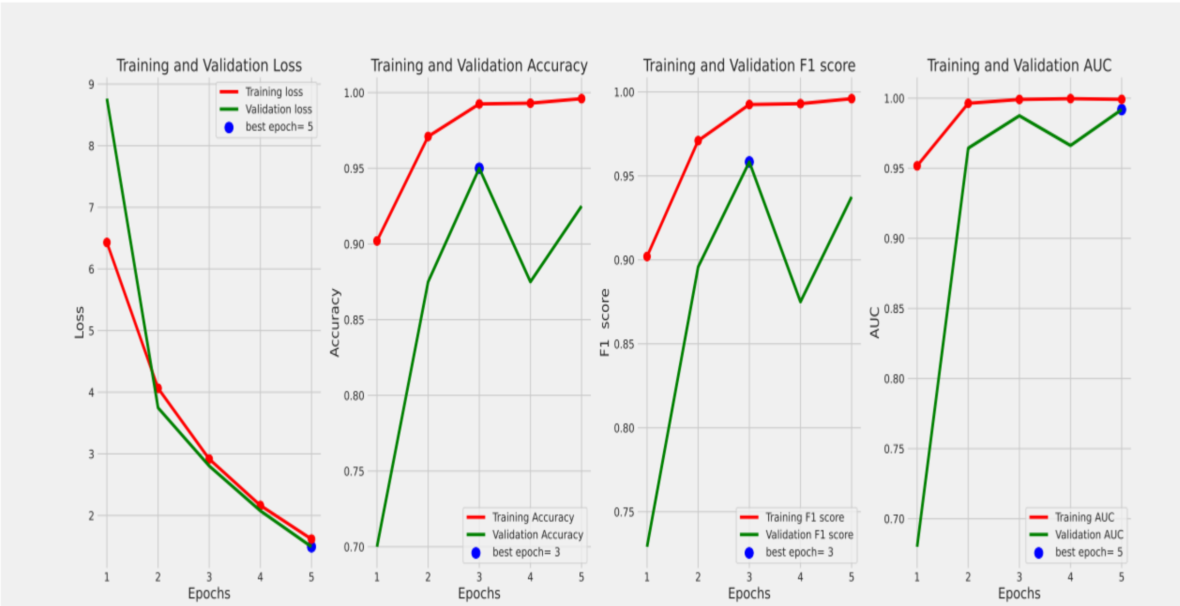

Fig. 11. Training loss and accuracy of the ResNet model.

Conclusion

In this project, we have delved into the application of the Pyramid Scene Parsing Network (PSP Net) for semantic segmentation, a core task in computer vision. Our efforts focus on two key components: training the PSP Net model and employing it for various semantic segmentation tasks. User-friendliness and adaptability are at the forefront of our design. To this end, we have developed a comprehensive command-line interface (CLI) that enables users to effortlessly train custom PSP Net models, perform pixel-wise predictions on new images, and assess model performance. The project's versatility is further highlighted by its capacity to leverage per-trained PSP Net models on benchmark datasets, including the Aerial Drone data set from Kaggle and Cityscape. The implications of our work are substantial, as semantic segmentation is crucial across various fields, such as object recognition, scene understanding, and robotics. Our project provides an efficient and flexible solution for researchers and practitioners engaged in these domains. Looking ahead, we plan to keep our project updated with the latest advancements in semantic segmentation models and datasets, ensuring it remains a state-of-the-art tool for the community. Fine-tuning the training process and exploring transfer learning techniques will enhance the performance of custom PSP Net models. Additionally,

Volume: 9, No: 2, pp.5593-5633
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

optimizing the CLI and expanding its capabilities to support more datasets and evaluation metrics will further improve usability. This project not only advances the field of semantic segmentation but also prioritizes accessibility and versatility through its intuitive interface. As we continue refining and expanding our work, we aspire for it to become a valuable resource for the computer vision community and beyond.

References

❖ Liu, S., Wang, Y., Chen, Q., Cheng, J., Wang, Y., & Wang, T. (2023). Robustness of semantic segmentation models for medical images: Evaluation, augmentation, and regularization. *IEEE Transactions on Medical Imaging, 42*(6), 1546–1558.

❖ Chen, Y., Fan, R., Yang, X., Wang, J., & Latif, A. (2023). Multi-scale semantic segmentation of remote sensing images with attention-guided dense dilated convolutions and efficient domain-specific image mixing. *Remote Sensing*, *15*(15), 3546.

❖ Neupane, B., Horanont, T., & Aryal, J. (2021). Deep learning-based semantic segmentation of urban features in satellite images: A review and meta-analysis. *Remote Sensing, 13*(4), 808.

❖ Shafaey, M. A., Salem, M. A.-M., Ebied, H. M., Al-Berry, M. N., & Tolba, M. F. (2018). Deep learning for satellite image classification. *International Conference on Advanced Intelligent Systems and Informatics* (pp. 383–391). Springer.

❖ Alias, B., Karthika, R., & Parameswaran, L. (2018). Classification of high-resolution remote sensing images using deep learning techniques. In *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (pp. 1196–1202). IEEE.

❖ Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., et al. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, *120*, 25–36.

- Irons, J. R., Dwyer, J. L., & Barsi, J. A. (2012). The next Landsat satellite: The Landsat Data Continuity Mission. *Remote Sensing of Environment*, *122*, 11–21.

- Johnson, K., & Koperski, K. (2017). WorldView-3 SWIR land use-land cover mineral classification: Cuprite, Nevada. Remote Sensing and GIS.

- Scott, G. J., England, M. R., Starms, W. A., Marcum, R. A., & Davis, C. H. (2017). Training deep convolutional neural networks for land–cover classification of high-resolution imagery. *IEEE Geoscience and Remote Sensing Letters*, *14*(4), 549–553.

- Musaev, M., Khujayorov, I., & Ochilov, M. (2019). Image approach to speech recognition on CNN. In *Proceedings of the 2019 3rd International Symposium on Computer Science and Intelligent Control(*pp. 1–6).

- Mukhamadiyev, A., Khujayarov, I., Djuraev, O., & Cho, J. (2022). Automatic speech recognition method based on deep learning approaches for Uzbek language. *Sensors*, *22*(10), 3683.

- Valikhujaev, Y., Abdusalomov, A., & Cho, Y. I. (2020). Automatic fire and smoke detection method for surveillance systems based on dilated CNNs. *Atmosphere*, *11*(11), 1241.

- Kuchkorov, T., Urmanov, S. N., Nosirov, K. K., & Kyamakya, K. (2020). Perspectives of deep learning based satellite imagery analysis and efficient training of the U-Net architecture for land-use classification. In *Developments of Artificial Intelligence Technologies in Computation and Robotics: Proceedings of the 14th International FLINS Conference (FLINS 2020)* (pp. 1041–1048). World Scientific.

- Zhu, Y., Yang, Y., Xu, Z., & Feng, W. (2023). Semantic segmentation of high-resolution aerial images with edge-enhanced ASPP module. *Remote Sensing*, *15*(20), 5005.

- Wu, J., Wu, Y., Liu, W., & Lu, J. (2023). A novel attention-guided multi-scale semantic segmentation network based on residual U-Net. *Applied Soft Computing*, *122*, 109404.

Volume: 9, No: 2, pp.5593-5633
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

❖ Sun, J., Ke, R., Huang, Z., & Huang, G. (2023). AugNet: Dynamic data augmentation for semantic segmentation with multi-scale deep network. *IEEE Transactions on Image Processing, 32,* 4665–4678.

❖ Zhu, J.-J., Yang, M., & Ren, Z. J. (2023). Machine learning in environmental research: Common pitfalls and best practices. *Environmental Science & Technology*.

❖ Wang, R., Zhang, F., Wang, F., & Song, W. (2023). Residual U-Net++ with deep supervision for semantic segmentation of aerial images. *Remote Sensing, 15*(12), 2805.

❖ Zhang, K., Liu, W., Chen, X., Wang, Q., Tian, Q., Tang, X., & Wang, J. (2023). Enhanced multi-scale U-Net with image augmentation for semantic segmentation of brain tumors. *Frontiers in Neuroscience, 17,* 1015465.

❖ Li, X., He, W., Li, M., & Du, Q. (2023). PSPNet-based semantic segmentation with adaptive pyramid pooling for high-resolution remote sensing imagery. *IEEE Geoscience and Remote Sensing Letters, 20*(4), 1–5.

❖ Gulzar, Y. (2023). Fruit image classification model based on MobileNetV2 with deep transfer learning technique. *Sustainability, 15*(3), 1906.

❖ Li, W., Guo, Z., Yin, W., & Deng, Y. (2023). Multi-scale attention U-Net with hybrid loss function for medical image segmentation. *Applied Sciences, 13*(15), 7704.

❖ Singh, R., & Rani, R. (2020). Semantic segmentation using deep convolutional neural network: A review. In *Proceedings of the International Conference on Innovative Computing & Communications (ICICC)*.

❖ Zhang, R., Du, L., Xiao, Q., & Liu, J. (2020). Comparison of backbones for semantic segmentation network. In *Journal of Physics: Conference Series* (Vol. 1544, No. 1, p. 012196). IOP Publishing.

❖ Singh, N. J., & Nongmeikapam, K. (2023). Semantic segmentation of satellite images using Deep-U-Net. *Arabian Journal for Science and Engineering, 48*(2), 1193–1205.