Adaptive Semantic Segmentation for Enhanced Crop and Land Use Mapping in India

A report submitted in partial fulfilment of the requirements

for the award of the degree of

B.Tech Artificial Intelligence and Data Science

by

Shivam Mukhedkar
(Roll No: 120AD0028)

Under the Guidance of

Dr. Shounak Chakraborty

Assistant Professor

Dept. of CSE, IIITDM Kurnool

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY DESIGN AND MANUFACTURING KURNOOL

May 2024

Evaluation Sheet

Title of the Project: Adaptive Semantic Segmentation for Enhanced Crop and Land Use Mapping in India

Name of the Student(s): Shivam Mukhedkar

Examiner(s):

Supervisor(s):

Head of the Department:

Date:

Place:

Certificate

I, **Shivam Mukhedkar**, with Roll No: **120AD0028** hereby declare that the material presented in the Project Report titled **Adaptive Semantic Segmentation for Enhanced Crop and Land Use Mapping in India** represents original work carried out by me in the **Department of Computer Science and Engineering** at the **Indian Institute of Information Technology Design and Manufacturing Kurnool** during the years **2023 - 2024**. With my signature, I certify that:

- I have not manipulated any of the data or results.
- I have not committed any plagiarism of intellectual property. I have clearly indicated and referenced the contributions of others.
- I have explicitly acknowledged all collaborative research and discussions.
- I have understood that any false claim will result in severe disciplinary action.
- I have understood that the work may be screened for any form of academic misconduct.

Date: Student's Signature

In my capacity as supervisor of the above-mentioned work, I certify that the work presented in this Report is carried out under my supervision, and is worthy of consideration for the requirements of B.Tech. Project work.

Advisor's Name: Advisor's Signature

Abstract

Semantic segmentation of aerial images plays a pivotal role in various applications, including urban planning, agriculture, and environmental monitoring. However, the performance of segmentation models often deteriorates for images from different domains due to variations in weather conditions, sensor characteristics, and geographical landscapes. This problem is more severe for tasks where acquiring hand-labelled data is extremely hard and tedious. I propose Domain adaptation techniques to address this challenge by enabling models to accurately predict across different domains. In this work, I have used Unet and Siamese networks with adversarial learning for better predictions in crop classes which are affected by seasons.

Acknowledgements

I would like to express my sincere gratitude to the faculty mentor, Dr. Shounak Chakraborty. Their invaluable guidance and support were present throughout the project. Their expertise in the field of my project helped to shape the course of the project. Lastly, I would like to thank all the faculty and staff of the Department of Computer Science and Engineering at IIITDM Kurnool for providing a good academic environment and resources for this project.

Contents

List of Figures

Abbreviations

GIS	**G**geographic **I**nformation **S**ystem
TOP	**T**rue **O**rtho **P**hoto
GT	**G**round **T**ruth
DASS	**D**omain **D**daptive **S**emantic **S**egmentation
DSM	**D**igital **S**urface **M**odel
RGBIR	**R**ed **B**lue **G**reen **I**nfra-**R**ed
RGB	**R**ed **B**lue **G**reen
IRRG	**I**nfra-**R**ed **R**ed **G**reen
GRL	**G**radient **R**eversal **L**ayer

Dedicated to my mentor,friends and family . . .

Chapter 1

Introduction

1.1 Semantic Segmentation

Semantic segmentation is a basic computer vision problem that includes providing a class to each pixel in an image. It is important in a variety of applications, including autonomous driving, land-use classification, and medical picture analysis. Aerial images taken from airborne platforms such as satellites or drones provide significant insights into the Earth's surface and characteristics. These insights are useful in a variety of disciplines, including urban planning, agriculture, environmental monitoring, and disaster response.

1.2 Motivation & Scope

Traditional semantic segmentation approaches typically rely on supervised learning, where a model is trained on a large dataset of labelled images from the target domain. The effective analysis and segmentation of aerial images pose significant challenges due to variations in geographic locations, weather conditions, sensor types, and resolution disparities. However, acquiring large labelled datasets for every new domain is often impractical and expensive. Well-classified crop use area data can help in Precision Agriculture, Environmental Monitoring, Policy Planning, etc.

1.2.1 Domain Adaptation

One of the critical hurdles faced in aerial image analysis is domain shift, as the discrepancy between the characteristics of images captured from different sources or under distinct conditions. This shift impedes the performance of machine learning models trained on specific datasets when applied to new domains, causing a need for robust solutions for domain adaptation. In this report, we delve into domain adaptive semantic segmentation for aerial images and innovate techniques to overcome the challenges posed by domain shifts by leveraging the UNet architecture and gradient reversal layers to enhance the adaptability and generalization of segmentation models of varying domains.

1.3 Problem Statement

- Given labeled training sets of source images S_i with ground truth masks Y_i from a specific regions (source domain) D_i. (where $i = 1, \ldots, n$)
- There exists a separate test set of images T_j (where $j = 1, \ldots, m$) with m total images from other regions(target domain) $\hat{D}_j$.
- The objective is to develop a robust semantic segmentation model that accurately generates segmentation masks $\hat{Y}_j$ for each test image T_j, regardless of its domain $\hat{D}_j$.
- by learning domain-invariant features that are transferable across different regions despite variations in geography, climate, and agricultural practices.

1.4 Challenges

- **Data Variability:** Diverse soil types, weather patterns, and agricultural practices lead to varying spectral and spatial characteristics in X_i across regions.
- **Class Similarities:** Optical similarities between classes like crops and forests significantly affect the results.
- **Class Imbalance:** Various classes have imbalanced occurrences like double/triple crop class and wasteland class.
- **Limited Labeled Data:** Acquiring ground truth data (Y_i) for all crops and land use categories in all regions R_i is resource-intensive.

Chapter 2

Related Work

2.1 Semantic segmentation model for land cover classification from satellite images in Gambella National Park, Ethiopia[1] 2023

- The authors compared the performance of several deep learning algorithms including U-Net, DeepLab v3, MobileNet v2, and EfficientNet B0. But U-Net and DeepLab v3 outperformed the others with overall accuracy exceeding 90%.
- U-Net offered the best balance between accuracy and computational efficiency. DeepLab v3 provided slightly higher accuracy but required longer training times and more computational resources.

2.2 Deep Active Learning in Remote Sensing for data efficient Change Detection[2] 2020

- The authors use a Siamese network layout, where two images from different times are encoded with two convolutional branches with shared weights, and then decoded into a change map. They use ResNet34 as the encoder and Monte Carlo Batch Normalisation to introduce stochasticity.
- Their method can achieve the same performance as a model trained on the full dataset, with only 1.02% of the annotated samples. The choice of uncertainty method and metric is less critical, as all models quickly reach the baseline performance5.

2.3 U-net: Convolutional networks for biomedical image segmentation[3] 2015

- U-Net is a convolutional neural network that was developed for biomedical image segmentation(**O. Ronneberger et al., 2015** [?]). It is made up of a contracting path and a symmetric expanding path that allows for the exact localization of segmented sections while also combining low-level and high-level characteristics using skip connections.
- They outperformed the previous best technique (a sliding-window convolutional network) on the ISBI challenge for segmentation of neural structures in electron microscopic stacks, attaining an error rate of 0.92% compared to 3.77%.

2.4 Optimised U-Net for Land Use–Land Cover Classification Using Aerial Photography[4] 2023

- Compared the performance of the U-Net model with different configurations of hyperparameters (e.g., number of filters, kernel size, learning rate) for land use/land cover (LULC) classification
- The best compromise between accuracy and efficiency was achieved with 56 initial convolutional filters and a kernel size of 5 x 5.

2.5 Image classification using embedded spaces generated by Siamese Networks[5] 2019

- Siamese networks were introduced by researchers at AT&T Bell Laboratories to solve the problem of signature verification (Bromley et al., 1994).
- **One-Shot Siamese Network (Gregory et al.,[7]) and Image classification using embedded spaces (Guillermo et al., 2019 [5])** established a baseline to deal with problems that were difficult to address due to the reduced amount of data available and gave optimisations in the loss function, optimizer, pairing proportion, etc.

2.6 Unsupervised Domain Adaptation by Backpropagation[6] 2015

- The paper introduces a novel gradient reversal layer that reverses the sign of the gradient during the backpropagation, thus encouraging the feature extractor to learn domain-invariant representations. They attempted to match feature space distributions which is accomplished by modifying the feature representation itself.
- The paper demonstrates the effectiveness of the proposed method on several image classification tasks, such as digit recognition (MNIST and SVHN), object recognition (Office-31 and Office-Home), and action recognition (UCF and HMDB).

Chapter 3

Data Preparation

3.1 Data Collection

Identifying the specific area of interest, desired date range, and spectral resolution based on project goals. Raw data from sources like Copernicus are in formats like jp2, and tiff which need specialized software. With these, the raw data is converted into usable formats. A geographic information system (GIS) is a system that creates, manages, analyzes, and maps all types of data. Raw data is imported and processed with this software for geospatial tasks.

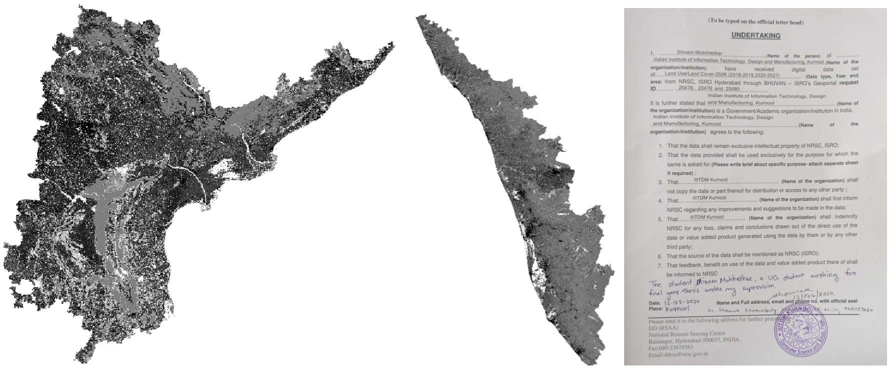
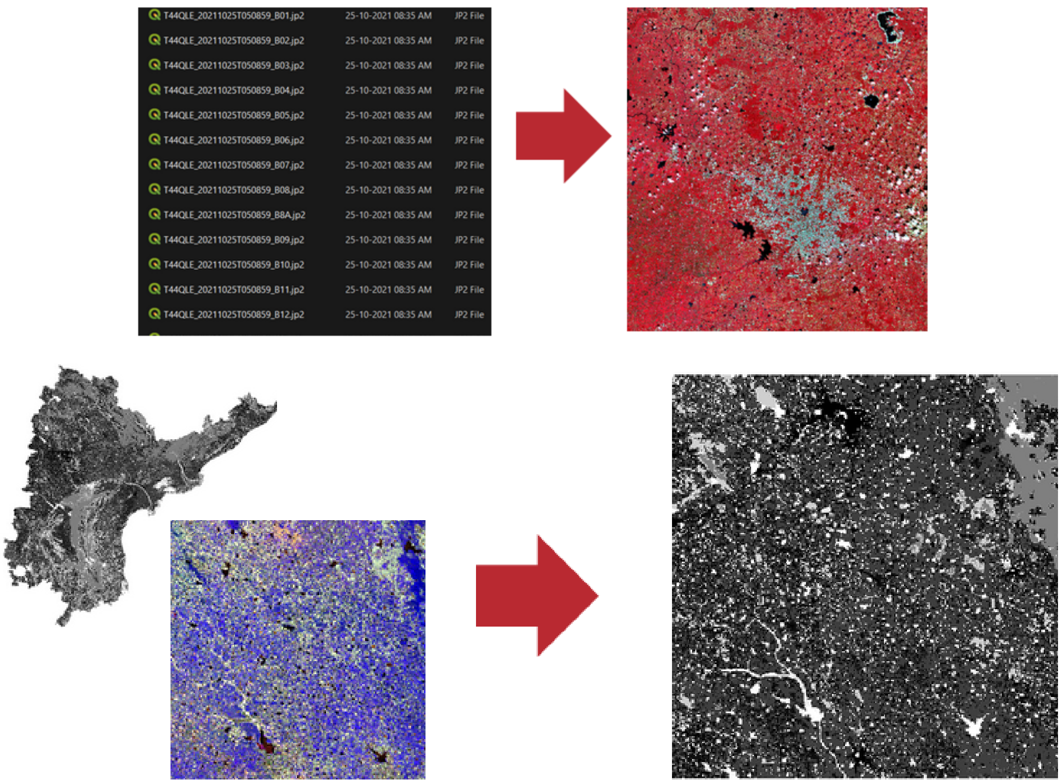


FIGURE 3.1: (a)AP & Telangana mask (b) Kerala Mask (c)request letter

[1] https://bhuvan-app1.nrsc.gov.in/thematic/thematic/index.php
[2] https://dataspace.copernicus.eu

3.1.1 Data Processing

- Coordinating images with different sensors, pre-processing, Sunlight reflectance, etc. configured by source.
- Corrections like Geometric Corrections, CRS conversion, etc are done manually in QGIS software.
- Satellite images from monsoon or subsequent months are always filled with clouds and cloud shadows which are removed using Google Earth Engine API.
- Images and Masks pairs were created by modifying the extent and CRS information present in their tiff files to exactly align the satellite images with masks. Rescaling, Standardization, etc are done before patching image and mask pairs.



3.1.2 Bands

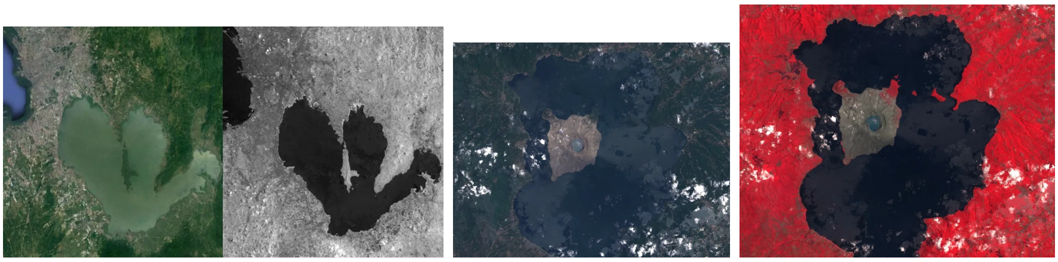


FIGURE 3.2: 1.RGB and IR 2.Comparison of RGB and IRRG

TOP images extracted have NIR(near-IR), Red and Green bands (false color composite) instead of RGB bands because IRRG gives all vegetation a distinct red color, making it easier to discern from its surroundings. This is conceivable because plants are highly reflective in the near-infrared spectrum. Additionally, the NIR, Red, and Green scheme aids in distinguishing clear water (darker shade of blue) from turbid water (cyan) in a false color image.[8][9]

3.1.3 Dataset

This dataset has 18 classes as shown in Figure 3.5.
Contains 1462 TOP (TrueOrtho-photo) images and 731 masks (GT).
Half of the TOP images are from March-April months and others are from sept-oct months of years 2019 and 2021.

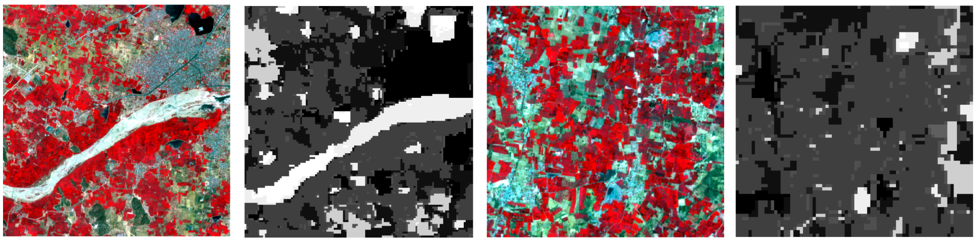


FIGURE 3.3: Image and Masks pairs

	Training		Testing		Total	
Domain	Images	Masks	Images	Masks	Total Images	Total Masks
AP & TEL	1156	578	20	10	1176	588
KERALA	256	128	10	5	266	133

FIGURE 3.4: No. of Images/Masks

Value	Description
1	Built-up
2	Kharif Crop
3	Rabi Crop
4	Zaid Crop
5	Double/Triple Crop
6	Current Fallow
7	Plantation
8	Evergreen Forest
9	Deciduous Forest
10	Degraded/Scrub Forest
11	Littoral Swamp
12	Grassland
13	Shifting Cultivation
14	Wasteland
15	Rann
16	Waterbodies max
17	Waterbodies min
18	Snow Cover

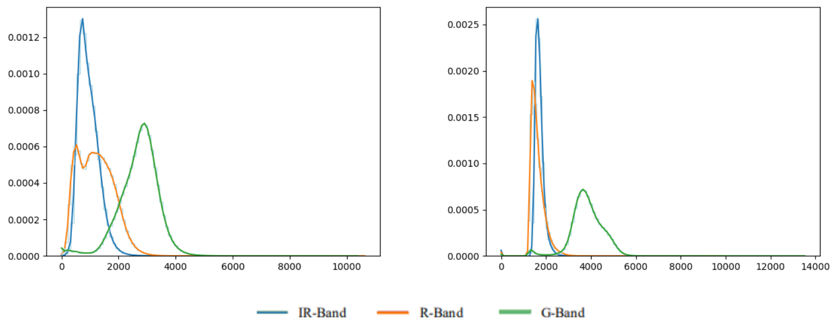
FIGURE 3.5: Image and Masks pairs

Chapter 4

Background

4.1 Domain Shift

- The Indian subcontinent's land cover is unique due to diverse geography, climate, agriculture, population, monsoon dependence, and traditional practices. Regions(states) like Northeast India and Kerala stay green and Rajasthan stays barren all year while other states like AP, Telangana, etc change according to seasons.
- The domain shift between both datasets is generated from factors like changes in imaging sensors, Position of the sun and satellite observation direction(and reflectance), Atmospheric effects and class distribution. Classes with vegetation and clutter are highly affected by the sensor factor and the class distribution[10].



.

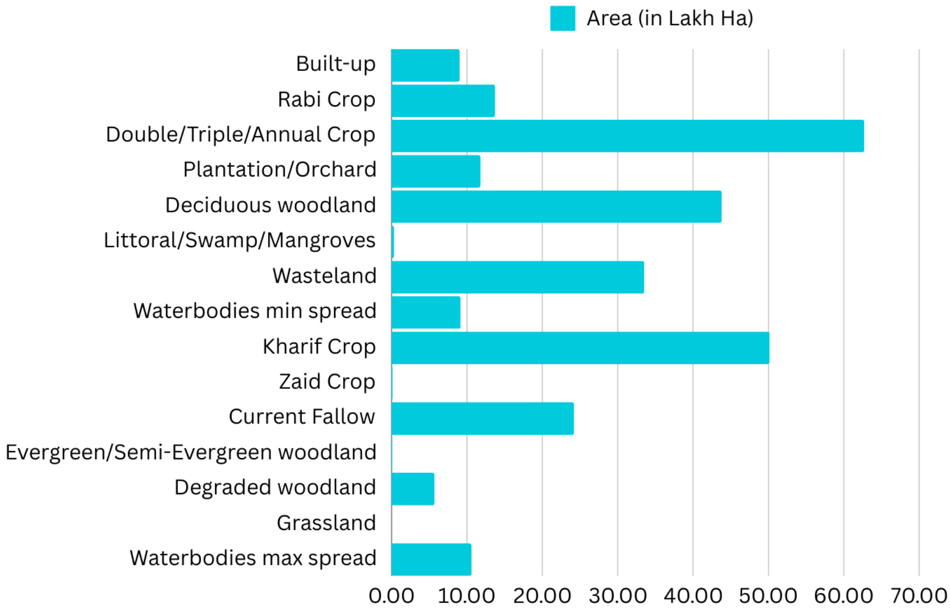


FIGURE 4.1: AP & Telangana (275.12 La Ha)

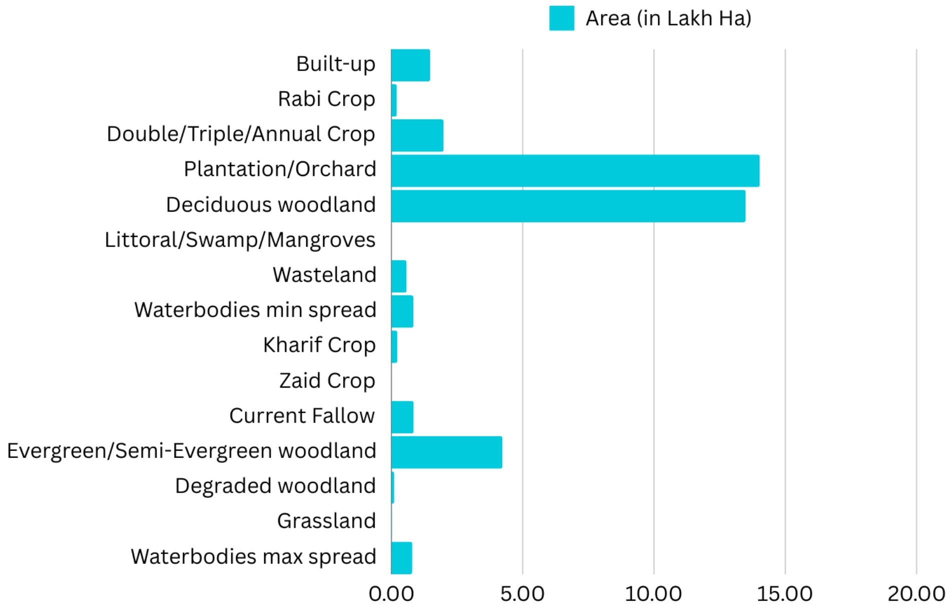


FIGURE 4.2: Kerala (38.86 La Ha)

- A single model trained on data from one region might perform poorly when applied to data from regions with vastly different characteristics

  - The plots below are completely different from each other. Substantial changes in relative distances, particularly for clusters representing the same classes, we can infer the domain shift. By the cluster pattern, some classes have dense clusters while others like the crop-(2) overlapping other classes and also the amount shows a class imbalance.

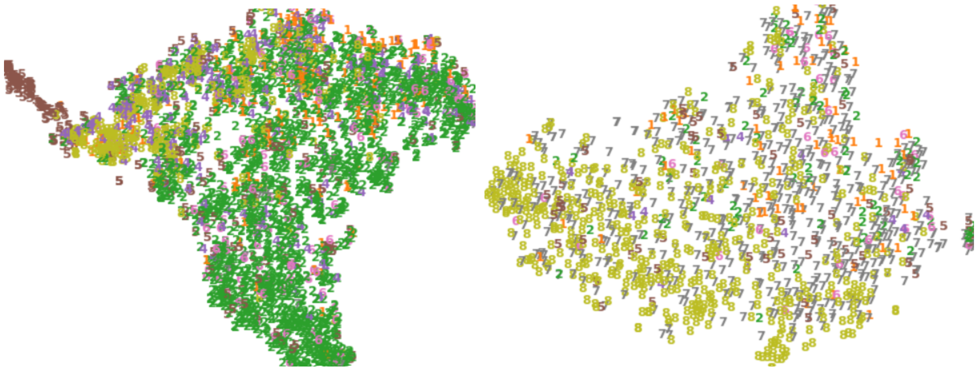


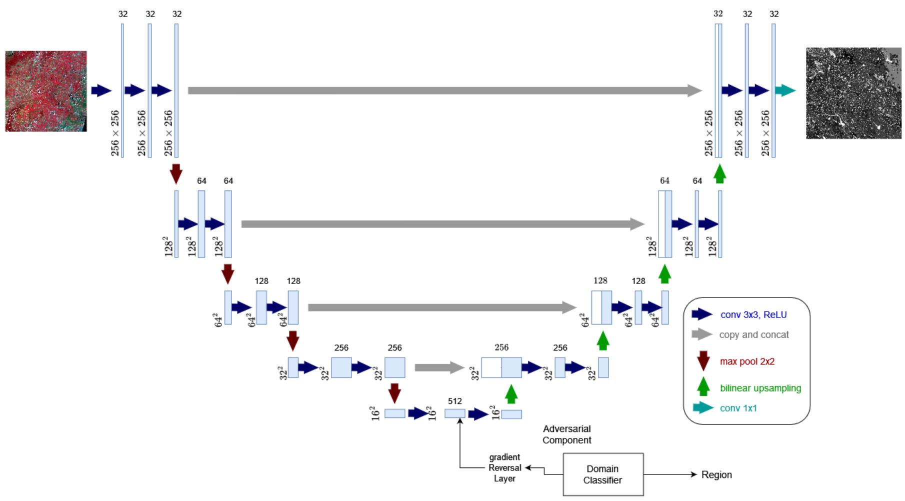
FIGURE 4.3: T-SNE plots - 1. Source Domain 2. Target Domain

- Implementation of the U-Net model to perform semantic segmentation on the dataset with 15 classes with thorough hyper-parameter tuning.

Domain	Setting (Train-Test)	No. of Images	Jaccard Coef (train)	Jaccard Coef (test)	Accuracy (test)	Precision (test)	Recall (test)	F1 score (test)
Same	AP & TEL	588	0.5989	0.4607	0.5892	0.6181	0.5546	0.5847
	KERALA	133	0.5115	0.3581	0.5562	0.5709	0.5377	0.5538
Different	APT - KERALA	721	-	0.02405	0.04522	0.04308	0.0401	0.04154
	KERALA - APT	721	-	0.07212	0.1111	0.1023	0.0835	0.092

FIGURE 4.4: UNET

- In the above implementation all results with different domains perform poorly which shows domain shift and the need for domain adaptation.
  To achieve this, we explore domain adaptation techniques such as Adversarial Domain Adaptation (ADA) or Gradient Reversal Layer (GRL), which bridge the gap between source and target domains and promote feature transferability.



Domain	No. of Images	Jaccard Coef (train)	Jaccard Coef (test)	Accurac y (test)	Precision (test)	Recall (test)	F1 score (test)
Different	721	0.6092	0.4828	0.6770	0.6949	0.6624	0.6783

FIGURE 4.5: UNET with Domain Adaptation

- Above implementation is of a model trained gave good results on both datasets. The domain-adapted model performed better than all other counterparts. Further, the proposed model will be highly affected by seasons and our target domain and source domain act differently to different seasons. Which implies a need for a domain adaptive model.

4.2 Gradient Reversal Layer

Applying transformations to embeddings of both the source and target domains, bringing their feature space distributions into closer alignment. Subsequently, train the classifier on the transformed source distribution. Given the increased similarity between the transformed distributions, the model is expected to attain higher accuracy when applied to the target domain. The Gradient Reversal Layer functions as an identity during forward propagation, producing outputs identical to its inputs. However, during backpropagation, it inversely multiplies its input by -1. In essence, during backpropagation, the GRL's output effectively steers towards a reverse effect of gradient descent. Here the feature extractor is trying to confuse the domain classifier by thus bringing the two distributions closer. The feature extractor(Unet encoder) will therefore be trained to minimize the classification loss of the mask generator(Unet decoder) and maximize the classification loss of the domain predictor. the mask generator(Unet decoder) and domain predictor will be trained to minimize their respective classification loss.

4.2.1 Pair Network

These subnetworks take input data and transform it into an embedding space, where the distance between embeddings shows the similarity between the input data. Based on the embedding similarity the target input data will be classified. Siamese networks are robust to variations in data, as they are trained on the relative similarities between pairs rather than absolute feature values. These embeddings won't hold data belonging to a particular domain and instead similarities, shared and discriminative features thus also making it domain invariant. The distance mentioned here will be calculated by the Euclidean distance function as stated optimal in [5], with the loss function.

Chapter 5

Implementation

5.1 Proposed model Implementation

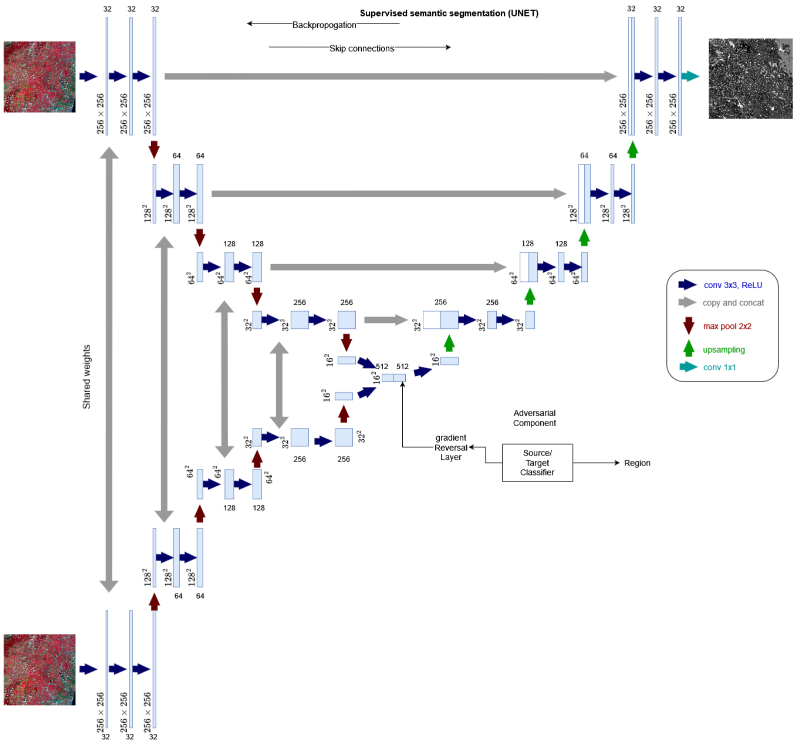


FIGURE 5.1: Proposed Solution Architechture

Here, Unet Encoder, Unet decoder, and half-siamese act as feature extractors, mask generators and domain classifiers respectively. The architecture is shown in Figure 5.1.

5.1.1 Hypothesis

In Deep Active Learning in Remote Sensing for data-efficient Change Detection[2], where the authors use a Siamese network layout and two images from different times, are encoded with two convolutional branches with shared weights and then decoded into a change map. This gives us the idea of changes in the embeddings of the model and the gradient flow, the embeddings of both images get compared and retained till the output of the model. A slight difference in our model is that the mask is not of change detections, but the mask contains classes which can be determined by the change data. These classes are crops(all), forests, waterbodies, etc. By this, we can infer that some of the change data will retained and can help predict those classes more accurately.

5.1.2 Training and Results

Model	Classes	Domain	Jaccard Coef (train)	Jaccard Coef (test)	Accuracy (test)	Precision (test)	Recall (test)	F1 score (test)
UNET	18	Same	0.5078	0.2484	0.4293	0.4900	0.4174	0.4508
UNET	15	Same	0.5989	0.4607	0.5892	0.6181	0.5546	0.5847

Figure 5.2: Old model results

Model	Classes	Domain	Jaccard Coef (train)	Jaccard Coef (val)	Accuracy (test)	Precision (test)	Recall (test)	F1 score (test)
SUNET (DA)	18	Different	0.5281	0.4601	0.5685	0.6391	0.5063	0.5650

Figure 5.3: Proposed model results

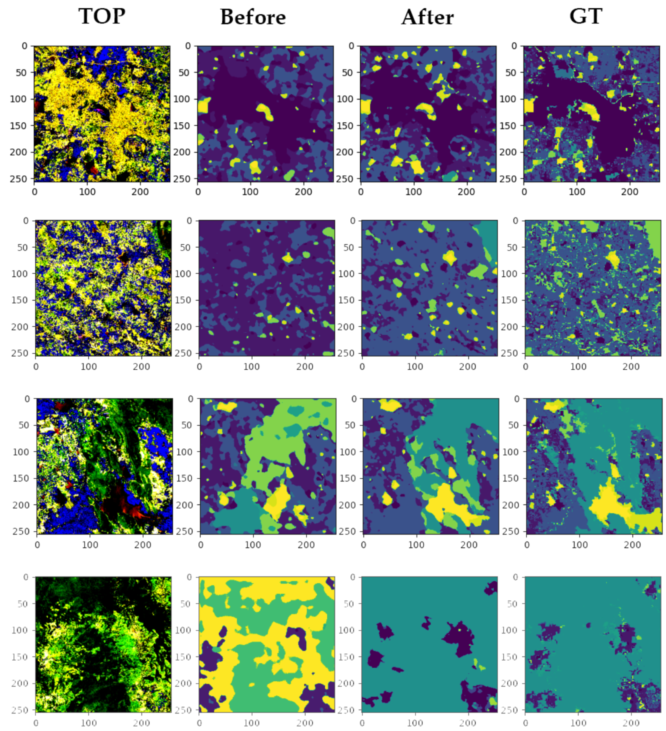


Figure 5.4: Results

The results demonstrate the effectiveness of the proposed domain adaptation network compared to a standard U-Net architecture. Furthermore, comparing novel UNET with the results of the proposed model we can determine the proposed model is superior in predicting season-influenced classes. We can also observe the same in the below diagrams. The season-influenced classes are getting better predicted in the new model.

	precision	recall	f1-score	precision	recall	f1-score
1	0.69	0.55	0.62	0.72	0.63	0.67
2	0.23	0.53	0.32	0.47	0.51	0.49
3	0.03	0.00	0.00	0.25	0.00	0.00
4	1.00	0.10	0.18	1.00	0.10	0.18
5	0.49	0.40	0.44	0.65	0.89	0.75
6	0.07	0.00	0.00	0.10	0.00	0.01
7	0.00	0.00	0.00	0.02	0.00	0.00
8	1.00	1.00	1.00	1.00	1.00	1.00
9	0.04	0.00	0.00	0.87	0.89	0.88
10	0.02	0.10	0.03	0.29	0.04	0.07
11	1.00	1.00	1.00	1.00	1.00	1.00
12	1.00	1.00	1.00	1.00	1.00	1.00
13	1.00	1.00	1.00	1.00	1.00	1.00
14	0.06	0.10	0.08	0.32	0.13	0.18
15	1.00	1.00	1.00	1.00	1.00	1.00
16	0.39	0.19	0.26	0.41	0.27	0.33
17	0.86	0.31	0.45	0.72	0.70	0.71
18	1.00	1.00	1.00	1.00	1.00	1.00
accuracy			0.29			0.64
macro avg	0.55	0.46	0.47	0.66	0.56	0.57
weighted avg	0.31	0.29	0.28	0.59	0.64	0.60

FIGURE 5.5: 1.Old Model 2.Proposed model

Chapter 6

Conclusions and Future Objectives

6.1 Conclusions

This study presents a strategy for overcoming the restrictions of the traditional semantic segmentation model. We presented a new strategy for domain adaptation along with better crop classification. Outlined new results comparing the performance of these networks to a traditional semantic segmentation model developed for the data set. The Experimentation results show the improvement in the network's ability to adapt to new domains and achieve superior segmentation accuracy. The architecture demonstrates promising results in classes sensitive to seasonal changes.

6.2 Future Work

- Using better Domain adaptation methods.
- Integration of different models such as resnet,deeplabv3, etc.
- Improvement for domain invariant classification for different multiple regions. Experimentation with the use of DSM as depth can give different and useful insights.
- Feasibility for inclusion of Sentinal-1 data which contains surface roughness and moisture content, which is not directly captured by Sentinel-2's spectral bands.

Bibliography

[1] G. D. T. Mulugeta Yikuno Lilay1, "Semantic segmentation model for land cover classification from satellite images in gambella national park, ethiopia," *SN Appl. Sci. 5, 76*, 2023. [Online]. Available: https://doi.org/10.1007/s42452-023-05280-4

[2] J. D. W. K. S. Vıt Ruzicka, Stefano D'Aronco, "Deep active learning in remote sensing for data efficient change detection," *arXiv:2008.11201*, 2020. [Online]. Available: https://www.researchgate.net/publication/343877160_Deep_Active_Learning_in_Remote_Sensing_for_data_efficient_Change_Detection

[3] T. B. O. Ronneberger, P. Fischer, "U-net: Convolutional networks for biomedical image segmentation," *Medical Image Computing and Computer-Assisted Intervention–MICCAI*, 2015. [Online]. Available: https://doi.org/10.1007/978-3-319-24574-4_28

[4] P. S. . S. P. Clark, A., "Optimised u-net for land use–land cover classification using aerial photography," *PFG 91, 125–147*, 2023. [Online]. Available: https://doi.org/10.1007/s41064-023-00233-3

[5] G. Lloret-Talavera, "Image classification using embedded spaces generated by siamese networks," *University of Alicante, Spain*, 2019.

[6] V. L. Yaroslav Ganin, "Unsupervised domain adaptation by backpropagation," 2014.

[7] R. S. Gregory Koch, Richard Zemel, "Siamese neural networks for one-shot image recognition," *ICML deep learning workshop*, vol. 2, 2015.

[8] E. Imam, "Remote sensing and gis module: Colour composite images and visual image interpretation," *University Grand Commission (UGC), MHRD, Govt of India.*, 2019.

[9] C. Gutierrez, "A comparison of false color composites in mapping and discriminating between salt-affected soils in kings county, california," *THE GEOSCIENCES DEPARTMENT OREGON STATE UNIVERSITY*, 2004.

[10] A. K. Bilel Benjdira, Yakoub Bazi and K. Ouni, "Unsupervised domain adaptation using generative adversarial networks for semantic segmentation of aerial images," 2019. [Online]. Available: https://doi.org/10.3390/rs11111369

[11] K. H. A. Sarah Kamel Hussein, "Enhanced semantic segmentation of aerial images with spatial smoothness using crf model," *MICST*, 2022.

[12] A. J. S. L. R. C. S. Sankaranarayanan, Y. Balaji, "Learning from synthetic data: Addressing domain shift for semantic segmentation," *IEEE conference on computer vision and pattern recognition*, pp. 3752–3761, 2018.

[13] C. Z. S. F. C. D. X. M. P. M. A. Libo Wang, Rui Li, "Unet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery," *arXiv:2109.08937*, 2021.

[14] Z. S. W. L. Z. M. R. Gao1, "An improved u-net method for the semantic segmentation of remote sensing images," 2021.

[15] A. C. I. F. a. M. B. Boris Shirokikh, Ivan Zakazov, "First u-net layers contain more domain specific information than the last ones," *arXiv:2008.07357v1 [eess.IV]*, 2020. [Online]. Available: https://doi.org/10.1007/978-3-030-60548-3_12

[16] G. A. R. M. J. K. H.-P. S. B. Borgwardt, Karsten M. and A. J. Smola, "Integrating structured biological data by kernel maximum mean discrepancy," *ISMB*, pp. 49–57, 2006.