International Journal of Innovative Research and Scientific Studies, 8(2) 2025, pages: 2047-2064

IJIRSS
International Journal of Innovative Research & Scientific Studies

ISSN: 2617-6548

URL: www.ijirss.com

Graph residual U-Net: Automated deep learning-based road extraction using DeepGlobe dataset, cartosat-2 and sentinel-2 imagery

Avishkar Karnik[1], Shilpa P. Pimpalkar[2], Manjusree Panchagnula[3*]

[1]*Visvesvaraya National Institute of Technology, Nagpur, Maharashtra, India.*
[2]*AISSMS Institute of Information Technology, Pune, Maharashtra, India.*
[3]*National Remote Sensing Centre (NRSC), Indian Space Research Organization (ISRO), Hyderabad, India.*

Corresponding author: Manjusree Panchagnula (*Email: manjusree_p@nrsc.gov.in*)

Abstract

Road detection is crucial for defensive strategies and disaster management since precise and timely mapping of road networks enables efficient operations during critical times. In defence, road detection facilitates strategic troop movement, the identification of secure pathways, and real-time monitoring of potential barriers or dangers, all of which contribute to tactical planning and operational safety. It helps disaster management by finding accessible roadways for first responders, assessing damaged infrastructure, and optimizing evacuation routes. Recent research work using deep learning techniques have significantly enhanced the processing capabilities of remote-sensing images for information extraction. Consequently, these deep-learning methodologies facilitate the automation of road extraction using very high-resolution imagery/satellite imagery. This study presents a novel deep learning approach, Graph Residual U-Net, designed for road extraction from satellite imagery. The Graph Residual U-Net model uses Deep Residual U-Net as the backbone U-Net architecture and incorporates graph neural network (GNN) layers into a residual U-Net framework, enhancing feature representation and improving road extraction accuracy. Findings are presented by implementing two models: i) Graph Residual U-Net Cartosat Model- Model trained using the DeepGlobe and Cartosat-2 Dataset and tested using DeepGlobe and Cartosat-2 Dataset and images acquired using Google Earth Engine, ii) Graph Residual U-Net Sentinel Model- the other model trained and tested using the Sentinel-2 imagery. This work highlights the potential of combining GNNs with convolutional neural networks (CNNs) for remote sensing applications. Graph Residual U-Net Cartosat Model showed an accuracy of 0.97702 and loss of 0.06150 and Graph Residual U-Net Sentinel Model demonstrated an accuracy of 0.99636, a loss of 0.00017, and Graph Residual U-Net Cartosat Model showed superior performance considering overall performance metrics. Graph Residual U-Net Sentinel Model showed low precision and recall but the model demonstrated excellent findings by generating accurate predicted masks when tested on unseen images.

Keywords: Graph Neural Network, Graph Residual U-Net, Residual U-Net, Road extraction.

DOI: 10.53894/ijirss.v8i2.5634
Funding: The authors would like to thank NRSC, ISRO for all the guidance and to ISRO, Bengaluru, & SPPU-STC for the financial support.
History: Received: 31 January 2025 / **Revised:** 3 March 2025 / **Accepted:** 10 March 2025 / **Published:** 24 March 2025

Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.
Transparency: The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
Publisher: Innovative Research Publishing

1. Introduction

Accurate information about road networks is of paramount importance in various sectors, including disaster management, defense strategies, geographic information systems, traffic monitoring, and automated crisis response [1, 2]. Researchers have proposed various deep learning-based methods for automatically extracting road information, which have practical applications across multiple fields. These methods are often demonstrated using open-source datasets, including satellite imagery. Road extraction is a vital task in remote sensing, valued across various domains and has been a key research focus for the past decade. According to the reviewed literature, road extraction is typically framed as a semantic segmentation task, where classes of "road" and "non-road" are identified [3, 4].

The extraction of road information from high-quality remote sensing images is a complex and challenging endeavour influenced by several factors. These factors include intricate and cluttered backgrounds characterized by various buildings, vegetation, and multiple road types. Additionally, the diversity in road shapes, encompassing variations in width and length, poses further challenges. The presence of poor image perspectives, often resulting from occlusions caused by clouds and fog, as well as lighting variations, exacerbates the situation. Furthermore, with the continuous expansion of urban areas, the topological structure of road networks becomes increasingly convoluted, with numerous buildings obstructing significant portions of road visibility. Addressing these challenges necessitates the development of advanced methodologies to improve road information extraction accuracy [2, 5]. With the potential to overcome these challenges, this paper presents the Graph Residual U-Net, a deep learning technique for automated road extraction from Cartosat-2 and Sentinel-2 Imagery. This innovative approach could significantly advance the field of remote sensing, offering a promising solution to the complex and diverse factors that have traditionally hindered road extraction.

The structure of the research presented further is organized as follows: - Section 2 presents review of exiting research work in the field and emphasizes significant findings. Section 3 delineates the dataset, including its source and key characteristics. Section 4 articulates the proposed methodology. Section 5presents the experimental results accompanied by a comprehensive analysis. Section 6 concludes with a synthesis of insights and outlines future research directions.

2. Related Work

Extensive literature survey has been carried out based on 36 relevant articles covering years 2016 to 2024. Distribution of papers referred year-wise is shown in Figure 1 and Figure 2 shows the dataset usage in the papers referred.

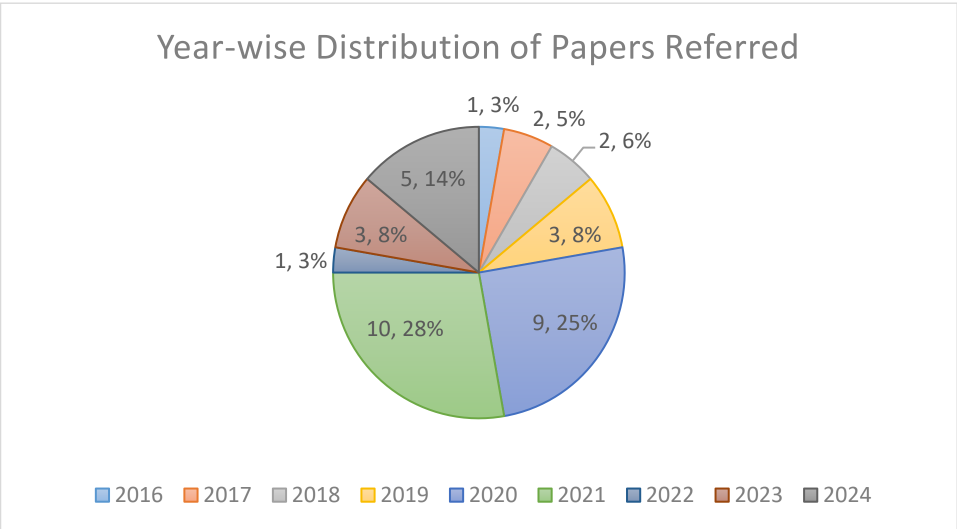


Figure 1.
Year-wise distribution of papers referred.

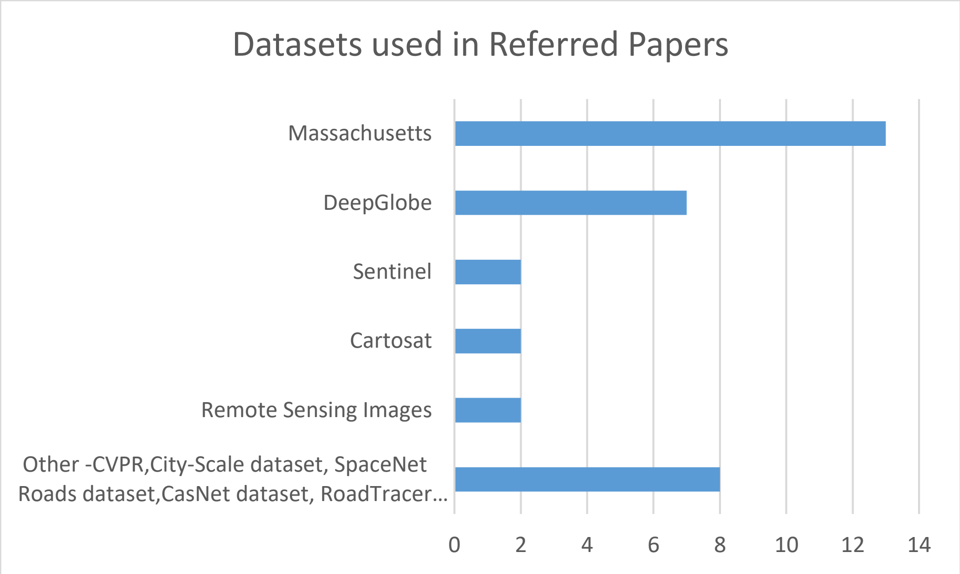


Figure 2.
Usage of Datasets in referred papers.

Table 1. Summarizes the datasets used and methodology followed in each referred paper. These findings are helpful for the future researchers to find the existing implementations and dataset usage.

Table 1.
Literature review of datasets used and methodologies followed in referred papers.

Ref.No.	**Dataset used**	**Methodology**	**Ref. No.**	**Dataset used**	**Methodology**
1	Cartosat2	Line Segment Detector	19	DeepGlobe	Global-aware deep network model
2	Cartosat2	Object-Oriented Analysis	20	Remote Sensing Images	U-net based model
3	Sentinel-2	Spatial Relationship-Informed Network	21	Massachusetts	D-Linknet
4	Sentinel-1 and Sentinel-2	Data augmentation method.	22	Massachusetts	Deep Residual Network based Model
5	Survey Paper	Survey of "Deep Learning based Road Extraction" approaches	23	Findings using Massachusetts, DeepGlobe and CHN6-CUG dataset	DFC-UNet
6	Survey Paper	Survey of" Deep Learning based Road Extraction" approaches	24	CasNet dataset, RoadTracer dataset	Dilated convolutional neural network approach using global context (GC-DCNN)
7	Massachusetts	Deep Residual U-Net	25	Massachusetts Road dataset, GF-2 Road dataset	Deep Residual Convolutional Neural Network
8	Aerial Image Segmentation Dataset	Non-Local Feature Search Network	26	Conghua roads dataset, Massachusetts roads dataset	DenseUNet
9	DeepGlobe	D-LinkNet	27	Massachusetts Roads dataset, GF-2 dataset, UAV dataset, DigitalGlobe Road dataset, SpaceNet dataset	Review Article
10	Computer Vision and Pattern Recognition (CVPR)+Massachusetts	Cascaded Attention DenseUNet	28	GE-Road dataset	Dual-Attention Capsule U-Net
11	DeepGlobe+VHR	Enhanced neural network	29	Findings using Massachusetts, DeepGlobe and CH6-CUG Roads Dataset	Local and global context reasoning
12	CHN6-CUG	Global Context-aware deep learning model	30	Comparisons of various models was done based on training on different	Review Article

Ref.No.	Dataset used	Methodology	Ref. No.	Dataset used	Methodology
				datasets like MNIST, NORB and ImageNet	
13	DeepGlobe+VHR	FCN8s, SegNet, U-Net and D-Linknet	31	Diverse images captured using Google Earth	Generative Adversarial Networks
14	Review Paper	Comprehensive Review	32	DeepGlobe dataset	Attention-Assisted UNet
15	City-Scale dataset, SpaceNet Roads dataset	Graph-Tensor Encoding	33	Mass-Roads, Cheng-Roads, Zimbabwe Roads	Deep Neural Networks
16	LRSNY dataset, Massachusetts Road dataset, Shaoshan dataset	U-net Based Model	34	Massachusetts	Complement UNet
17	Review Paper	Review Article	35	CityScale dataset, SpaceNet dataset	Graph Extraction approach
18	VHR images	Fully convolutional neural network	36	RoadTracer dataset	Point-based Iterative Graph Exploration

2.1. Key Findings Based on Review of 36 Papers is Outlined Below

i. The majority of the findings examined employed a fully supervised learning approach [2].

Liu, et al. [2] thoroughly reviewed 232 papers published from 2012 to 2024, presented details about datasets and methodologies for road extraction. Findings are categorized into three distinct groups: "fully supervised", "semi-supervised", and "unsupervised", based on the varying requirements for annotated datasets. Review also highlighted significant challenges-the complexity of remote sensing images, high data annotation costs, and the need for improved model generalization and robustness, are significant areas for further research. The review notes that most current research work demonstrated findings using "fully supervised learning"

ii. There are very few studies that demonstrate findings using Sentinel-2 imagery.

Jia, et al. [6] presented their findings utilizing the Spatial Relationship-Informed Network, achieving a Producer's Accuracy of 75.9 and a User's Accuracy of 88.1. They indicated that future developments should focus on enhancing feature extraction and improving the accuracy of the model.

Mo, et al. [5] conducted a review of various deep learning approaches for road extraction, referencing the research by Ayala, et al. [7]. The study by Ayala, et al. [7] specifically focused on data augmentation techniques for Sentinel imagery, which Mo, et al. [5] identified as an effective strategy in their analysis.

iii. There are very few research papers that demonstrate findings based on Cartosat-2 imagery.

Vani and Kumar [1] proposed a method that enhances Cartosat images using 'Contrast Limited Adaptive Histogram Equalization (CLAHE)' and also implemented Line Detector approach to identify road segments. Also performed 'connected component analysis (CCA)' on the segmented images to reconnect disjointed objects. The experimental results were demonstrated using images only and quantitative findings are missing.

Leena, et al. [8] conducted a study on 'object-oriented analysis' and 'object-based feature extraction algorithms' to detect road networks from high-resolution Cartosat-2F multispectral data in an Indian city. This analysis addressed various terrain conditions, including densely built-up areas and predominantly vegetated regions. The methodology employed in this work consisted of multi-resolution segmentation (MRS) and spectral difference segmentation (SDS), leading to the extraction of roads through a fuzzy rule-based algorithm. Findings highlighted the need of future work for Segmentation optimization. Future work can be carried out with multi-model approach to address various challenges of road extraction. Multi model approach is suggested to demonstrate the capabilities of handling all complexities of road extraction which is challenging using single module approach.

iv) The key challenges associated with road extraction

- Zhang, et al. [9] proposed a Deep Residual U-Net that utilizes the Massachusetts Dataset for road extraction. This approach effectively handles various types of roads, including narrow and complex ones, by combining the advantages of residual learning and U-Net architecture. However, the model encountered limitations in detecting roads beneath parking slots, as these areas were not labeled in the dataset.
- Ding, et al. [10] developed a model designed to enhance the segmentation accuracy of remote-sensing images of buildings and roads, thereby minimizing the misclassification of buildings and the disconnection of roads during the segmentation process. Nevertheless, the efficacy of this model is hindered in situations where it must process remote-sensing images that contain substantial noise, resulting in a decrease in segmentation accuracy.
- Zhou, et al. [11] have identified several challenges associated with satellite imagery, which often exhibits slender and complex characteristics while covering only a limited portion of the overall scene. In such instances, it is imperative to preserve detailed spatial information. Additionally, roads possess inherent connectivity and extend over long distances; however, many existing models fail to adequately capture this connectivity. In response to these challenges, Zhou, et al. [11] proposed a semantic segmentation model known as D-LinkNet. This model

incorporates an architecture that features skip connections, residual blocks, and an encoder-decoder framework. D-LinkNet still has the wrong recognition and road connectivity problems.

- Li, et al. [12] demonstrated that the Improved U-Net has the potential to reduce training time while delivering commendable results. A key challenge identified is the necessity of preserving the connectivity of the road network to prevent any omissions or interruptions in the roadway Li, et al. [12].

v) Recommendations for enhancing road extraction techniques through the application of deep learning methodologies-

- He, et al. [13] presented approached which combines pixel wise segmentation and graph-based approach. Proposed model trained a simple, non-recurrent, supervised model to predict a rich set of features that capture the graph structure directly from an image. The paper highlights future direction in the area to incorporate the graph-based approach for road extraction.
- Findings by Jia, et al. [6] suggest that the future research work can extend the scope of the work using large dataset to enhance the model's adaptability and accuracy..

2.2. Research Gaps Identified Based on Above Mentioned Key Findings Are as Follows

The existing body of research reveals a significant limitation in the utilization of Sentinel-2 and Cartosat-2 imagery. Notably, there is a lack of experimental findings that employ Graph Neural Networks (GNN) and Residual U-Net with these satellite images, which constitutes an important gap that must be addressed to further the field of road extraction. Additionally, labeled datasets for Sentinel-2 and Cartosat-2 imagery is not available for the research.

Moreover, a considerable number of prior studies have failed to consider the features of the entire spatiotemporal space of the input images. This oversight hinders the ability to interpret relationships among multiple road regions and to generate more thorough road extraction results [14].

Deep learning techniques have demonstrated considerable promise in the field of automatic road detection; however, they frequently yield fragmented road segments, which presents significant challenges in practical applications. Two primary factors contribute to this phenomenon: (1) occlusions and shadows created by trees, buildings, and other obstacles, and (2) the similarity in texture between roads and surrounding objects. Distinctive features of roads differentiate them from other ground-level entities, as roads are interspersed throughout the image and exhibit connectivity [14].

The process of convolutional operations examines one local neighbourhood at a time, resulting in progressive signal propagation. Consequently, capturing long-range dependencies necessitates the repetition of convolution operations, leading to diminished computational efficiency and difficulties in optimization [2].

To achieve substantial improvements in segmentation accuracy and adaptability, it is essential to enhance the extraction of road details significantly [2].

Graph neural network (GNN) methodologies have gained traction in recent years owing to their compelling performance and high interpretability. Within this framework, the road network can be conceptualized as a graph composed of intersections (vertices) and road segments (edges). Subsequently, road extraction utilizing graph learning techniques represents a promising avenue for research advancement [14].

Enhancement in deep learning models required, particularly regarding the challenges of insufficient and excessive semantic segmentation of roadways. These challenges are associated with sideline smoothness, interruptions, and the overall connectivity of the road network. Furthermore, the quality of the labeled dataset is to be improved in future endeavours [5].

Graph neural network (GNN)-based methodologies have recently garnered substantial attention due to their robust performance and high interpretability. The road network can be effectively modeled as a graph, wherein intersections are represented as vertices and road segments as edges. It is proposed that the application of graph learning for road extraction is a promising area for advanced research [15].

In future work, model with enhanced feature extraction capabilities and with training efficiency are required [4].

Following a thorough review of over 40 scholarly articles, it is recommended that future research investigate the integration of multiple models in the extraction of road features. Relying solely on a single type of road feature often results in suboptimal extraction outcomes. By leveraging a combination of models, there is a significant opportunity to enhance feature extraction and improve accuracy in road extraction efforts. This approach has the potential to elevate the effectiveness of deep learning methodologies within this field [13].

Graph-based residual U-net architecture is proposed to improve feature extraction ability and training efficiency. The architecture uses a graph neural network to improve performance and high interpretability. Also, the model was implemented to strengthen feature extraction ability and accuracy. The contribution of the research includes-

- Introducing a novel deep learning approach - Graph Residual U-Net Cartosat Model and Graph Residual U-Net Sentinel Model. These models that combines GNN and Residual U-net architecture, promising to push the boundaries of road extraction research.
- Model has demonstrated a significant improvement in accuracy, paving the way for more effective deep learning techniques in road extraction.
- Sentinel2 labelled dataset of 11640 images along with binary mask is generated for the research work.

2.3. Dataset Details

The model is trained using three different datasets: DeepGlobe, Cartosat and Sentinel-2 satellite imagery.

- VHR Input: Popular DeepGlobe (Resolution 50 cm/pixel) road dataset was chosen. A total of 30,000 samples were acquired from this dataset, each with dimensions 224x224 pixels along with the corresponding road masks of the same dimensions. Out of these 30,000 samples, 24,000 were used for training and 6,000 were used for testing. Dataset Link: DeepGlobe Road Extraction Dataset
- CARTOSAT-2(Resolution 1m), launched by the 'Indian Space Research Organisation (ISRO)', represents an advanced remote sensing satellite capable of providing scene-specific spot imagery. This satellite serves a multitude of purposes, including cartographic applications, urban and rural planning, coastal land use regulation, monitoring road networks and water distribution.
- Sentinel-2 Input: Sentinel-2 is an earth observation mission from the Copernicus programme of the European Space Agency. It provides free imagery acquired by the twin constellation Sentinel-2A and Sentinel-2B at a combined revisit rate of 5 days anywhere on the globe Data (GeoTIFF)The Sentinel-2 imagery has a spatial resolution of 10 meters. Due to unavailability of open-sourced Sentinel-2 Road datasets or any other geospatial road dataset matching the Sentinel 2's spatial resolution, a custom dataset was proposed. Diverse Sentinel 2 scenes were captured using GeoJson and Google Earth Engine. Corresponding road masks were constructed using geospatial tools like QuickOSM and QGIS. A total of 10,640 Sentinel-2 images patches, each with dimensions 256x256 pixels were acquired along with their binary masks. These samples were used for training the Graph Residual U-Net model.

3. Proposed Methodology

The Graph Residual U-Net model aims at utilising global information and preserving the topological connectivity and spatial relationships between pixels that traditional convolutional layers may fail to detect. The Graph Residual U-Net model is specifically designed for the efficient extraction of road features from remote sensing imagery. This model integrates a specialized architecture that combines convolutional neural networks (CNNs) with graph-based methodologies to enhance both feature extraction and segmentation accuracy. The primary components of this model comprise a Residual U-Net backbone, a graph-based convolutional layer, and a feature extraction pipeline that is optimized for image processing. Architecture of the proposed Graph Residual U-Net model is shown in Figure 3. Details about the components of architecture are as follows: -

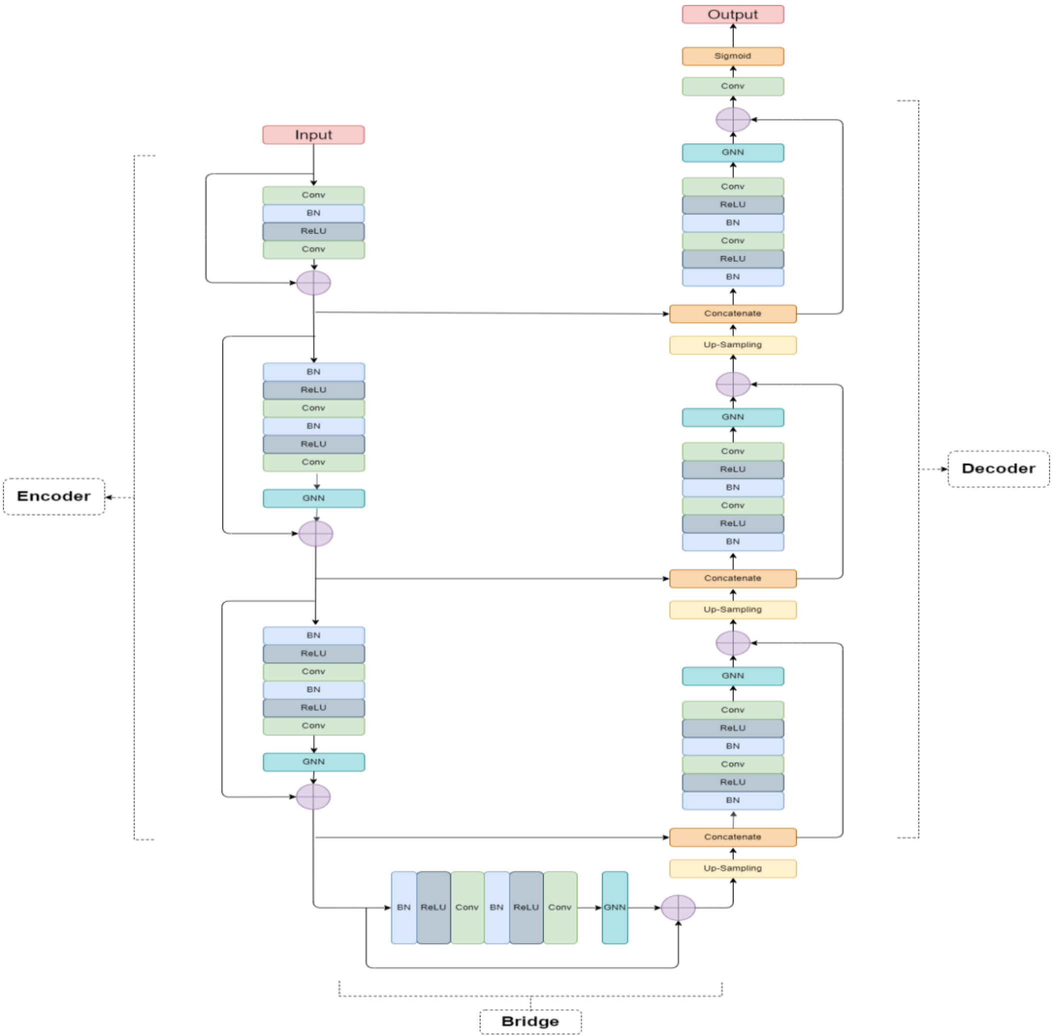


Figure 3.
Architecture of the proposed Graph Residual U-Net model.

3.1. Residual U-Net Backbone

The Residual U-Net backbone is an enhanced version of the conventional U-Net architecture, commonly used for image segmentation. It features a U-shaped encoder-decoder structure that incorporates residual connections by utilizing residual units instead of traditional ones. This adaptation aids in the training of deep neural networks. Additionally, the enhanced skip connections within the network improve information propagation, enabling the development of networks with fewer parameters that achieve superior performance [9, 16, 17]. The organization of the encoder-decoder structure is as follows:

Encoder Path: The encoder path consists of convolutional layers and batch normalization processes within each residual block. This structure reduces the spatial dimensions of the input image while increasing the number of channels. Instead of conventional pooling, a stride of two in the first convolutional block of each residual unit halves the feature map size. Skip connections mitigate the vanishing gradient problem, aiding the training of deeper neural networks and preserving essential spatial information while lowering data dimensionality.

Decoder Path: The decoder path mirrors the encoder by executing up-sampling operations to restore the spatial dimensions of the image. It combines high-level features from the encoder by concatenating each up-sampling layer with corresponding feature maps. This approach yields more precise segmentation by utilizing low-level details and high-level semantic information, crucial for accurate road segmentation.

Additionally, this architecture enhances computational efficiency and performance by eliminating traditional cropping operations in U-Net, thus streamlining network design.

3.2. Graph-Based Convolutional Layer

The Graph-Based Convolutional Layer (GraphConvLayer) introduces a novel methodology by integrating graph-based techniques into the Residual U-Net architecture. This layer is specifically engineered to effectively capture spatial relationships and dependencies among pixels by representing the image as a graph [13, 18, 19]. It encompasses two fundamental aspects:

- GraphSAGE Convolution: GraphSAGE (Graph Sample and Aggregate) is an advanced graph convolutional method. Unlike traditional Graph Convolutional Networks (GCNs) that aggregate features from all neighboring nodes, GraphSAGE samples a fixed-size set of neighbors for feature aggregation. This enables scalable training on large graphs. By combining the aggregated features with the node's own features, the model effectively understands the spatial structure of roads and their surroundings**.**
- Sparse Adjacency Matrix: The sparse adjacency matrix represents the connections between nodes. Here, each pixel is treated as a node. The connections or edges connect each pixel to its right and bottom neighbours, forming a matrix. Thus, this matrix encodes relationship between neighbouring pixels, allowing the model to consider local spatial relationships. The matrix is sparse because each pixel only connects to a few others. As a result, sparse adjacency matrix was opted for this purpose instead of a traditional adjacency matrix. This is because the traditional adjacency matrix represents every possible pair of nodes, leading to a very large and mostly empty matrix for large graphs. A sparse adjacency matrix only stores the non-zero elements thus focusing only on the relevant connections. The makes operations like GraphSAGE convolution more efficient in terms of memory and computation, especially for large graphs like images where each node (pixel) is connected only to a few neighbours.

The GraphConvLayer combines the above two aspects of Sparse adjacency matrix and GraphSAGE convolution in the following way:

- Graph Convolution:
- Reshape: The input image tensor is reshaped from [batch size, height, width, channels] to [batch size * height * width, channels]. This basically flattens in input tensor spatially but retains channel information. This allows each pixel to be treated as a node.
- GraphSAGE convolution: Applies the GraphSAGE convolution using the sparse adjacency matrix, aggregating features from neighbouring pixels and adding these aggregated features to the current node’s features, thereby capturing structure and spatial relationships between pixels beyond the local receptive field of standard convolutions.
- Reshape Back: The output is reshaped back to the original image shape [batch size, height, width, channels].
- 1x1 Convolution:
- A 1x1 convolution is applied after the GraphSAGE convolution. This convolution uses kernels (filters) of dimensions 1x1 to process each pixel independently in the spatial dimension but across all channels. Thus, it combines information across all the channels for each pixel which helps in refining the features and improve the learning capacity of the model.

3.3. Feature Extraction and Processing Pipeline

The feature extraction pipeline in the Graph Residual U-Net model is designed to process satellite images effectively. The model integrates the Graph-Based Convolutional Layer in each of the residual units to help refine the extracted features and prepare them for segmentation. For tasks like road extraction, where continuity and connectivity of features are important, the GNN layer helps the model to better understand and predict these connections, leading to more accurate segmentation. The skip connections throughout the pipeline help maintain the flow of gradients during training, thus improving the model’s ability to learn and generalise from complex image features.

3.4. Road Segmentation

Once the feature maps are processed, the model outputs the segmentation maps highlighting the detected road regions in satellite images in the form of a binary mask. The output is a probability map where each pixel is classified as either road or non-road**.**

Based on the architecture of the model shown in Fig.3. overall flow of execution is presented in Figure 4.

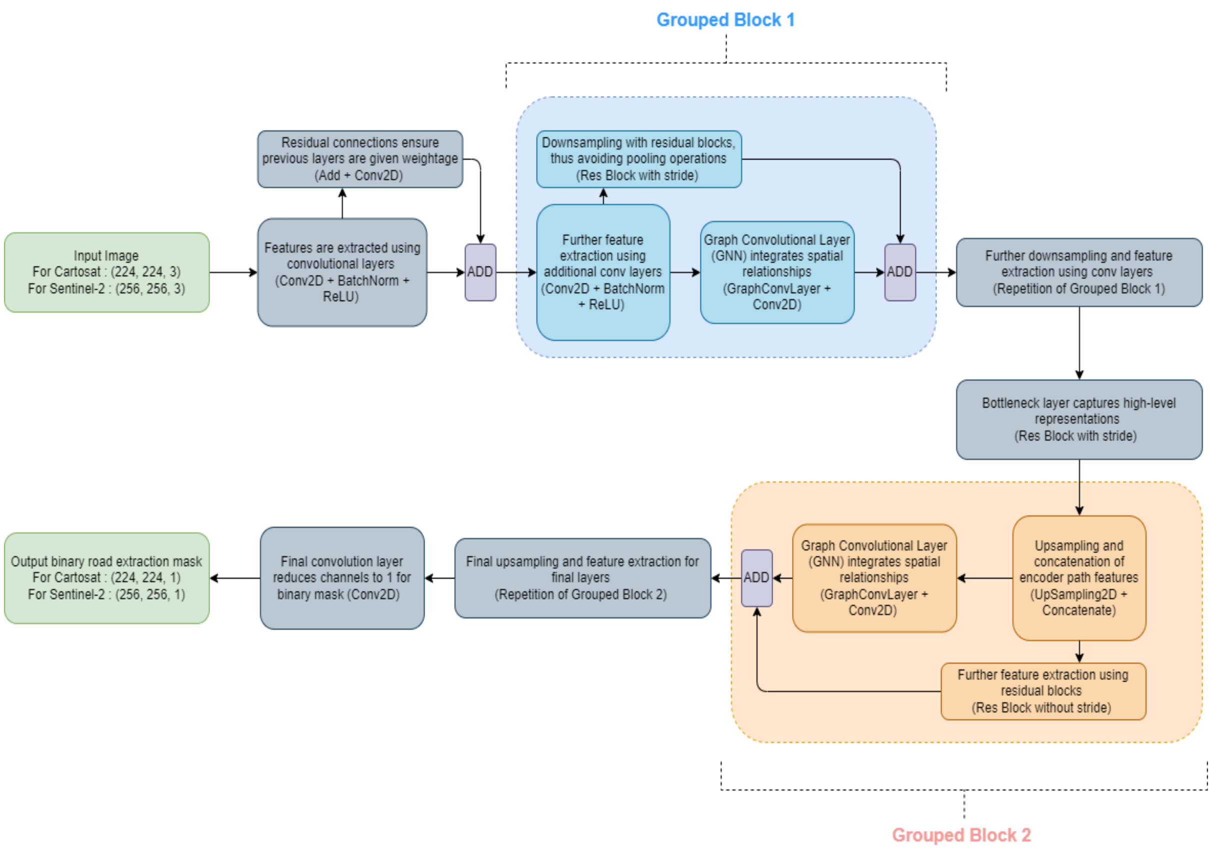


Figure 4.
Operations flow diagram of Graph Residual U-Net model.

There has been extensive research into several deep learning methods for fully supervised road extraction. Among these methods, the graph-based methods are becoming increasingly popular due to their ability to preserve the topological connectivity of roads, enhancing feature extraction and spatial relationship understanding. For addressing these needs, the Graph Residual U-Net model, a sophisticated deep learning model is proposed, which integrates the strengths of both, Residual Networks (ResNet) as well as Graph Neural Networks (GNNs) within a U-Net Architecture.

Foundational components and mechanisms of Graph Residual U-Net model is as follows:

- Input Preprocessing:
- Patch Extraction: Images are converted to patches of size 224x224 pixels (for Cartosat Model) or 256x256 pixels (for Sentinel-2 Model). Padding is applied if necessary.
- Processing of Patches: Pre-processing techniques like gamma correction, CLAHE and normalisation are applied to each image patch to improve contrast and enhance features.
- Model Initialization:
- The model is based on traditional U-Net architecture, with two key integrations; a Residual Block (Res_Block) and a Graph Neural Network Layer (GNN Layer). The model is trained separately for high resolution VHR imagery and for low resolution Sentinel-2 imagery. For Cartosat model, training is conducted using DeepGlobe road dataset, while Sentinel-2 training uses a custom dataset created using geospatial tools.
- Inference Process:
- The model processes image patches by converting them into tensors and then conducts inference to determine if each pixel represents road or non-road class.
- Post Processing:
- To refine the predictions and enhancing finer road features, a custom function defined “Join Lines”. It combines two output predictions of the same input image, where morphological operations like closing and binary thresholding are applied to one of the predictions. The ensures all the roads are visible while eliminating as much noise as possible and improving visibility.

```
# Function to join disconnected white lines
def join_white_lines(image, kernel_size=(5, 5), iterations=2):
    # Convert the image to grayscale
    gray = cv2.cvtColor(image, cv2.COLOR_RGB2GRAY)

    # Apply binary thresholding
    _, binary = cv2.threshold(gray, 128, 255, cv2.THRESH_BINARY)

    # Apply morphological closing
    kernel = np.ones(kernel_size, np.uint8)
    closed = cv2.morphologyEx(binary, cv2.MORPH_CLOSE, kernel, iterations=iterations)

    # Convert back to 3-channel image
    closed_color = cv2.cvtColor(closed, cv2.COLOR_GRAY2RGB)

    return closed_color
```

- Output Visualization:

Individual road patches are then restitched together and cropped to match the dimensions of the original input image. During this process, the Coordinate Reference System (CRS) information of the original input image is preserved to ensure the output aligns geographically with the input. Thus, the final output is available in both .PNG and .TIFF formats.

- Evaluation of the model:

Graph Residual U-Net Cartosat Model -The model's performance is assessed using a dataset of 6,000 unseen DeepGlobe images, images acquired using google earth engine and also Cartosat-2 images. The model exhibited an accuracy of 0.97702, a precision of 0.78589, an F1-score of 0.68786, a recall of 0.61158 and a loss of 0.06150.

Graph Residual U-Net Sentinel Model - Model is created and exclusively trained and tested usingSentinel-2 dataset. 950 images used for training and 50 images used for the testing. Model demonstrated an accuracy of 0.99636, a loss of 0.00017.

4. Results

This section presents in details about findings of both models - Graph Residual U-Net Cartosat model and Graph Residual U-Net Sentinel model.

4.1. Graph Residual U-Net Cartosat model –Trained using DeepGlobe and Cartosat-2

This model is tested on three different types of images: 6000 unseen VHR images from the DeepGlobe dataset ,Google Earth Engine images, a Cartosat-2 image .This section further presents the findings using each dataset.

4.1.1. DeepGlobe Dataset Testing

To calculate the metrics, the DeepGlobe roads dataset, consisting of a total of 30,000 samples was split into two parts, with 6,000 unseen samples used for calculating the metrics. Below are the model metrics derived from the testing. As a comparison, the model was compared with its backbone, Residual U-Net. The model metrics were computed twice for each model: once when the models were trained on 8000 samples and once when the models were trained for all 24,000 samples. The Table 2 shows how Graph Residual U-Net can show better results and learn faster with fewer training samples as compared to its Residual U-Net backbone. Figure 5 shows the confusion matrix of the models indicating the performance improvement of the Graph Residual U-net over Residual U-net model.

Table 2.
Comparative of the performance evaluation of the model.

Model	**Number of Training Samples**	**Binary Accuracy**	**Precision**	**Recall**	**F1 Score**	**Loss**
Graph Residual U-Net	8000 Deepglobe	0.97490	0.75692	0.58001	0.65674	0.16953
Residual U-Net	8000 Deepglobe	0.81580	0.11306	0.50404	0.18469	0.38371
Graph Residual U-Net	24000 Deepglobe	0.97702	0.78589	0.61158	0.68786	0.06150

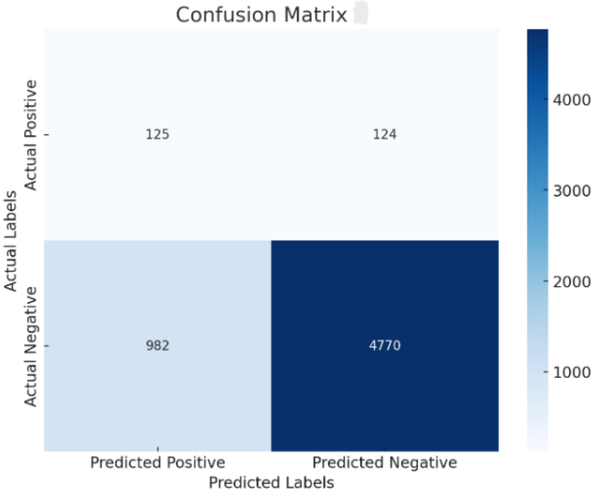


Figure 5.a.

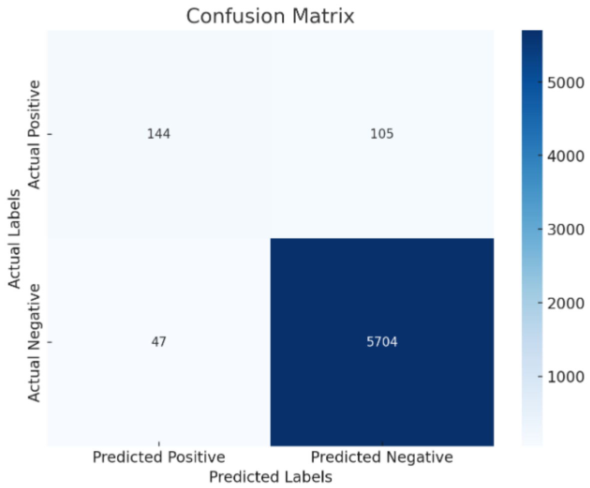


Fig 5.b.

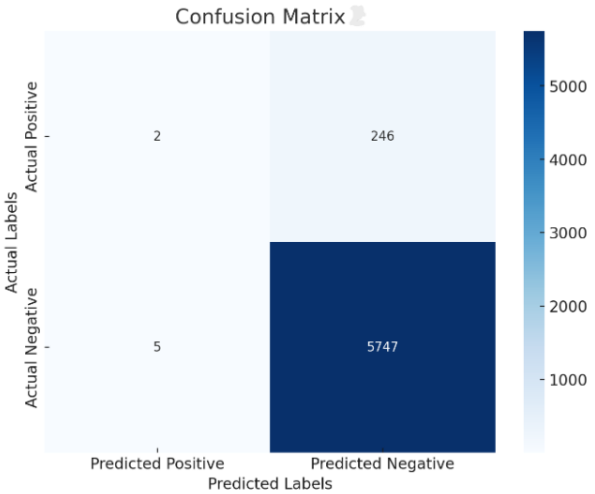
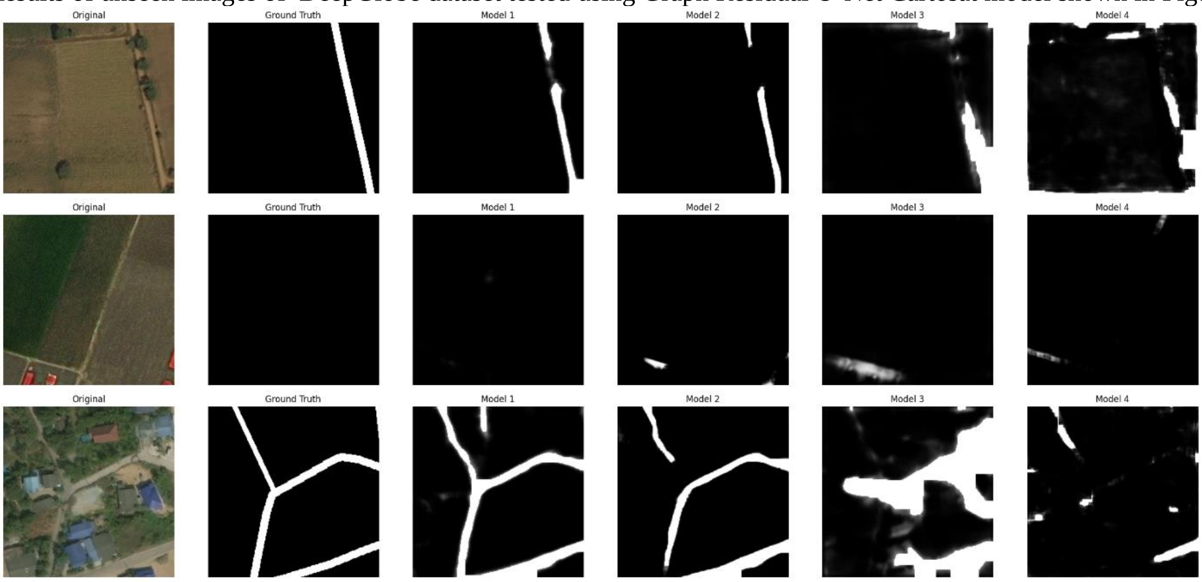


Fig 5.c.

Figure 5.
Confusion matrix of showing performance improvement a) Residual U-Net model trained on 8,000 samples, , b) Graph Residual U-Net model trained on 8,000 samples c) Residual U-Net model trained on 24,000.

Results of unseen images of DeepGlobe dataset tested using Graph Residual U-Net Cartosat model shown in Figure 6.



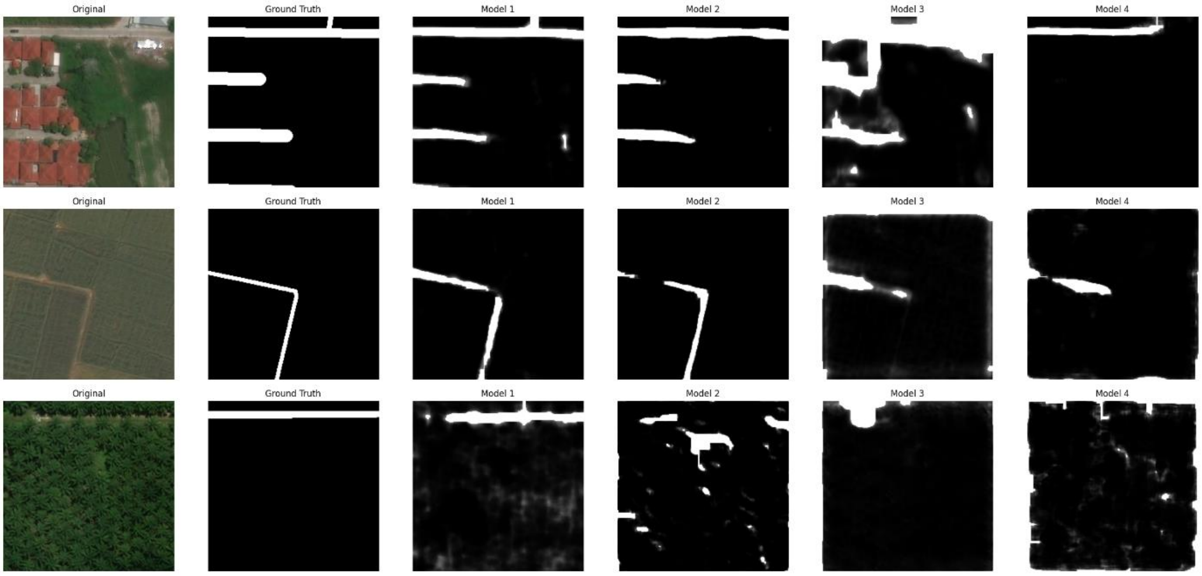


Figure 6. Here, Model 1 and 2 represent Graph Residual U-Net models trained on 24,000 and 8,000 samples respectively, and Models 3 and 4 represent Residual U-Net models trained on 24,000 and 8,000 samples respectively.

4.2. Google Earth Images

In order to test whether the model can handle different image scenarios and varied input sizes, it was also tested on google earth imagery. The images were acquired using Google Earth Pro software. The testing was successful in proving that the model gives satisfactory output for images with different resolutions than the training data. All the images were from Hyderabad city and its surrounding towns. Findings are shown in Figure 7.

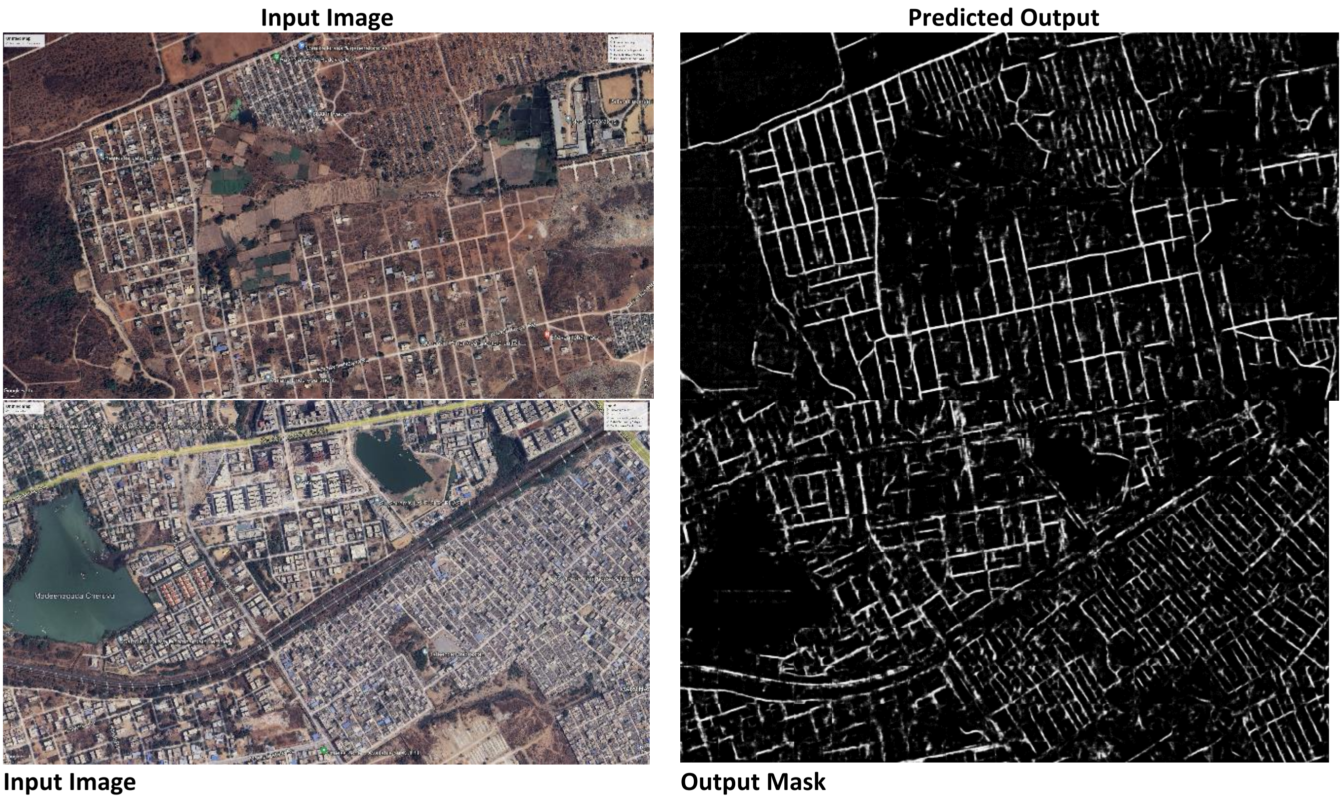


Figure 7.
Google Earth Image inputs and their corresponding results.

4.3. Cartosat-2 Input

Due to limitations in availability of open-sourced Cartosat-2 data, the model was tested on few Cartosat-2 images. Result of model with Cartosat-2 Image representing complete scene of spatial dimension 10 km x 10 km. The model successfully extracted road and efficiently ignored railway tracks and water bodies such as rivers. Result shown in Figure 8.

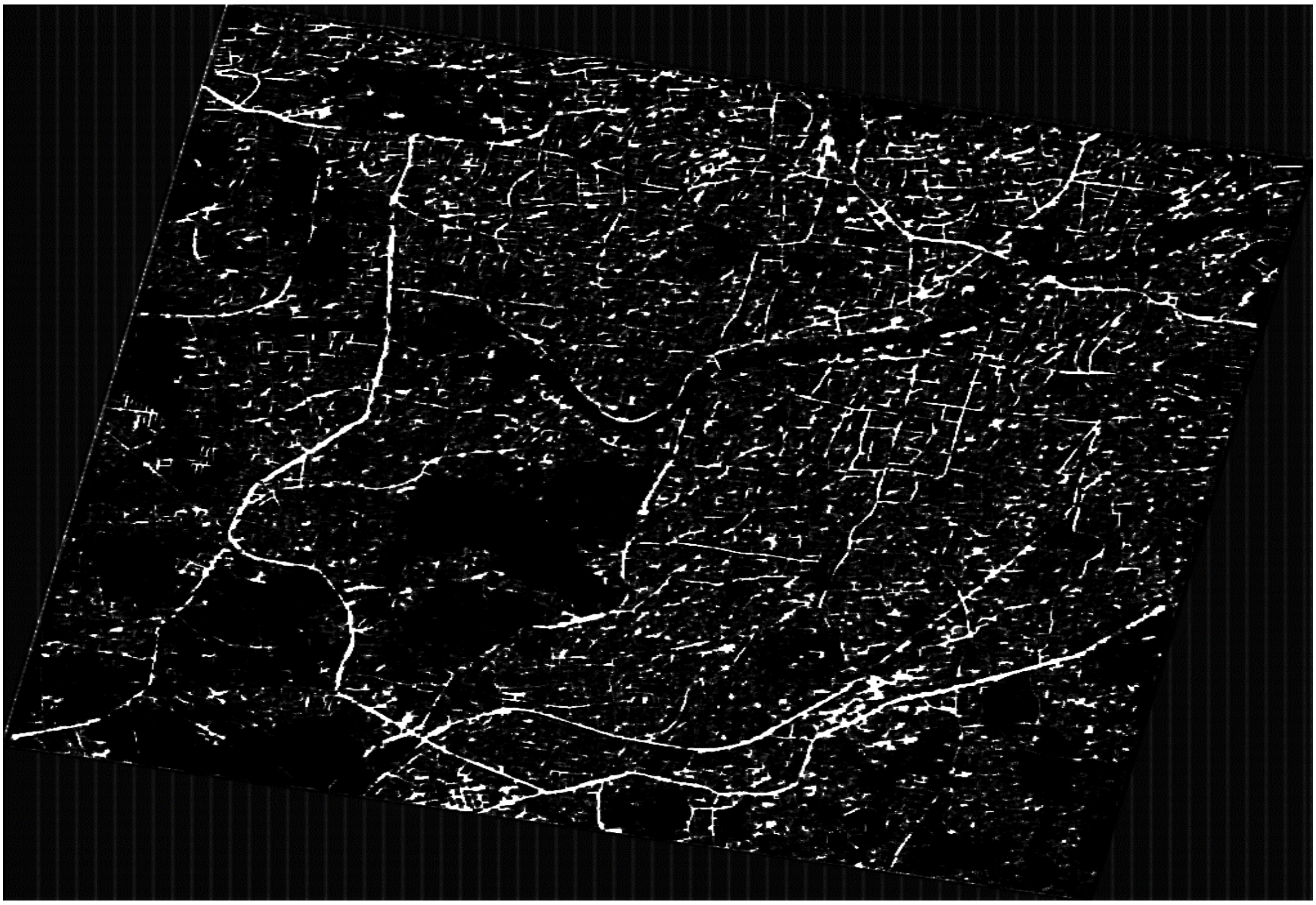


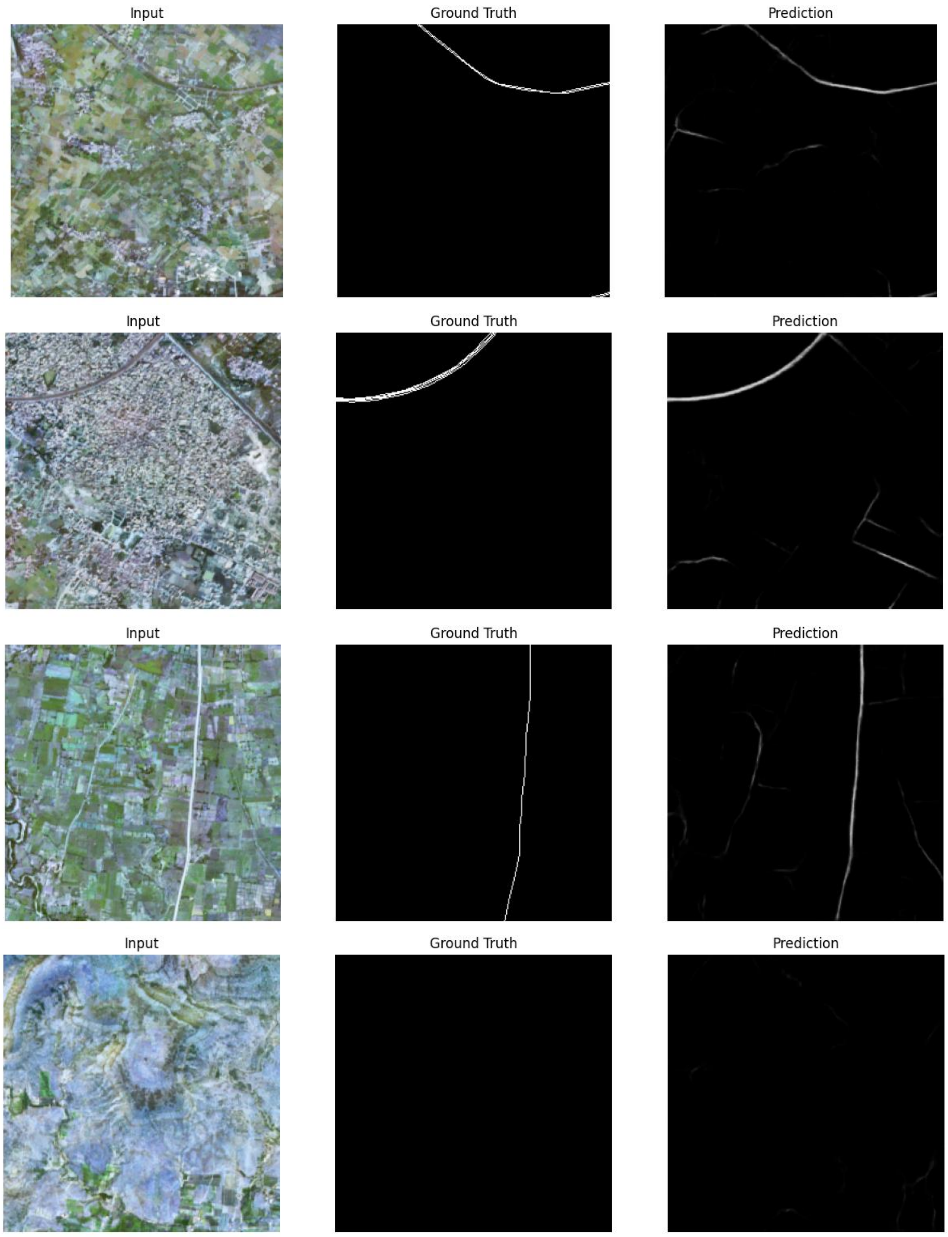
Figure 8.
The output mask of a Cartosat-2 Image of size 10kmx10km.

4.4. Graph Residual U-Net Sentinel Model- Trained and Tested Using Sentinel-2 Imagery

Due to unavailability of labelled Sentinel-2 Road dataset, a custom dataset was constructed using Sentinel 2 imagery. Customized dataset with 11,640 Sentinel-2 images is created for the research work. Total 1000sample used for testing purpose. Sentinel2 dataset resolution is 10m which is low as compared to Cartosat and DeepGlobe dataset. Road extraction using lower resolution satellite images is a challenging task, but the model was successful in capturing the major roads. Findings are consistent with the visual results obtained as shown in Figure 9 and Figure10.

The metrics for the Sentinel model were evaluated on 1000 unseen Sentinel-2 samples. The model exhibited an accuracy of 0.99636, a loss of 0.00017 and a precision and recall of ∽0.0. While these results seem like the model is under performing, the issue stems from the nature of the data, where roads occupy a very small portion of the image compared to the background. This extreme class imbalance can hinder the performance metrics, even though the model is learning to identify roads reasonably well. Metrics like precision and recall are heavily influenced by the distribution of classes. When one of the classes, in this case, the background, is overwhelmingly dominant, even a small number of false positives on the background can significantly impact these metrics, making the model appear worse than it is.

Thus, the low precision and recall are not necessarily an indication that the model is wrong. To emphasis on this finding, some of the results from the testing samples are shown below:



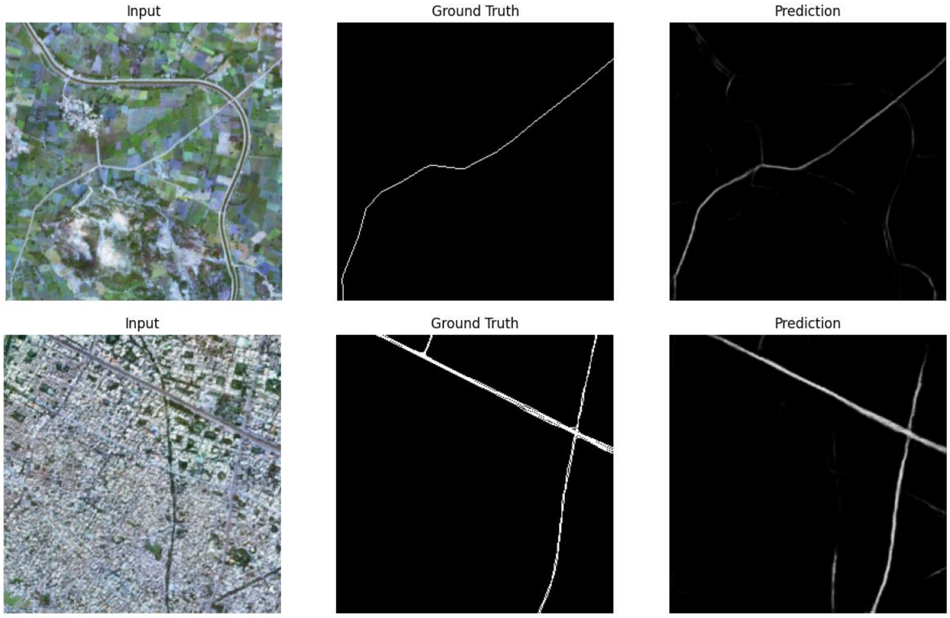
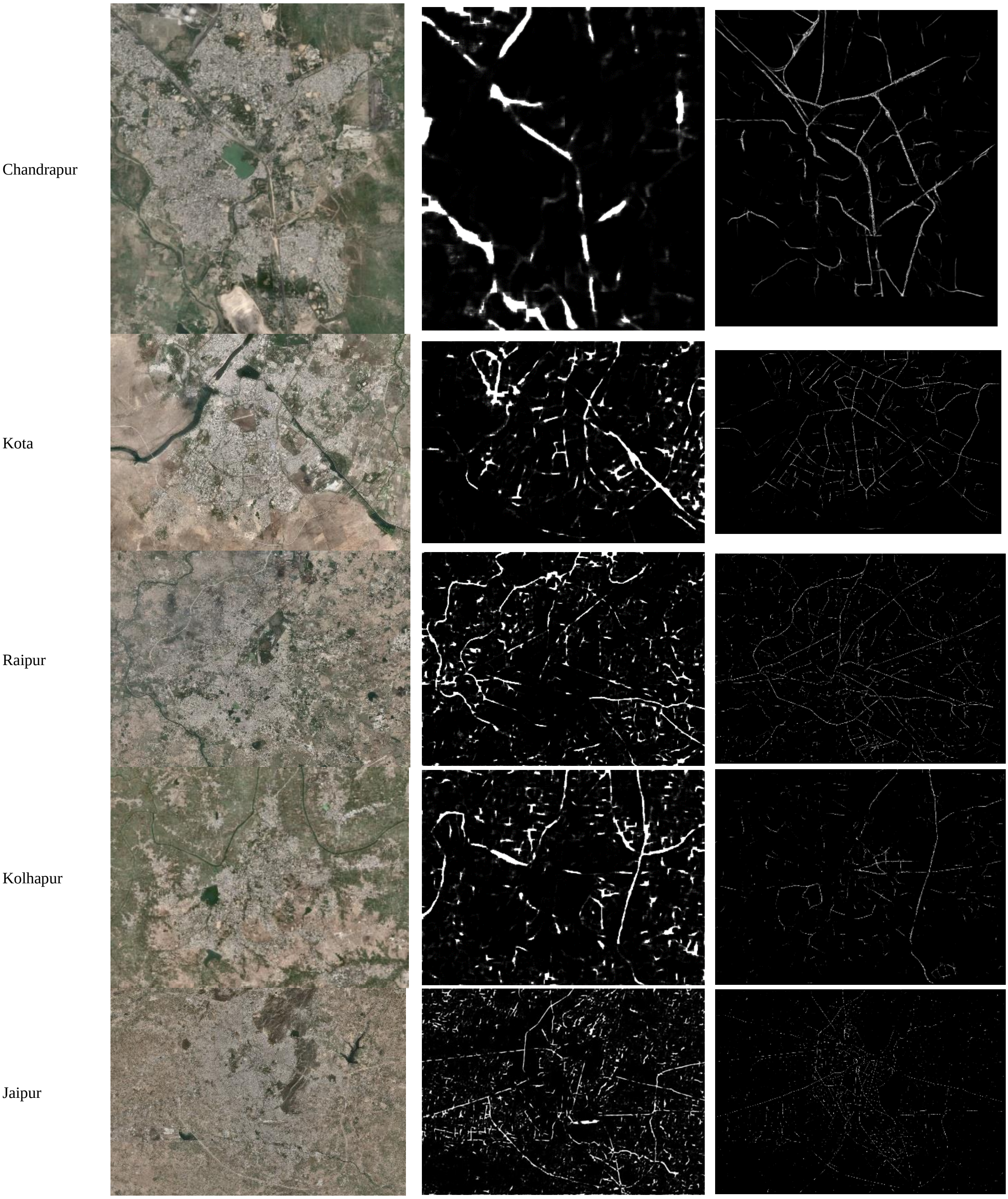


Figure 9.
The ground truths support the conclusion that roads are indeed very thin due to Sentinel-2's lower spatial resolution of 10m.

4.5. Findings With Various Images Acquired from India

To show the effect of changing the training dataset, inputs were also compared with the Cartosat Model. These inputs were scenes representing entire cities, much larger than the inputs mentioned above. The Sentinel Model output gives thinner, more refined roads as compared to the Cartosat Model. The results are given below:

Region	Input Image	Output Mask for Cartosat Model (model 1)	Output Mask for Sentinel Model (model 2)
Indore			
Surat			


Chandrapur
Kota
Raipur
Kolhapur
Jaipur

Madurai

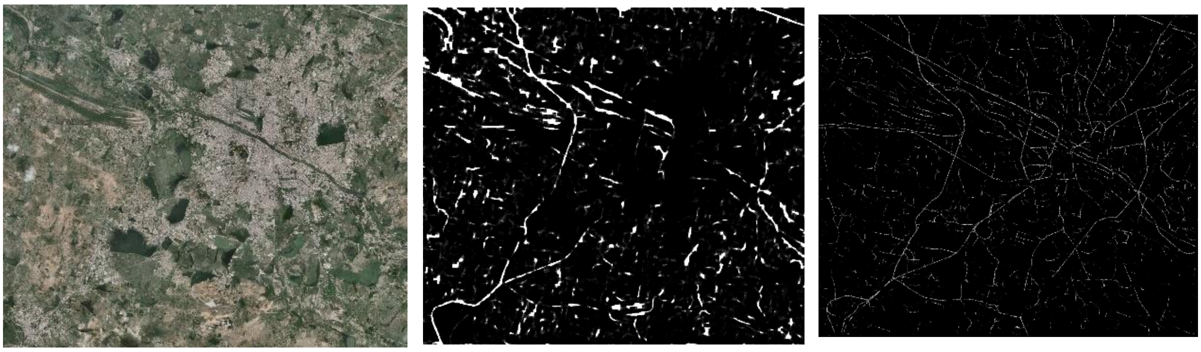


Figure 10.
These outputs demonstrate that the Sentinel-2 model was able to capture all the major roads while preserving the CRS and geospatial information of the input, so that the output was available in .png as well as .tiff format and could be visualised in geospatial software like QGIS and merged seamlessly with other outputs based on their geospatial information.

There are very few research papers that demonstrate findings based on Sentinel-2 and Cartosat-2 imagery. Comparison of proposed methodology with existing findings is shown in Table 3.

Table 3.
Comparative evaluation of the Proposed Graph Residual U-net with exiting findings.

Findings from the exiting research using Sentinel2 and Cartosat imagery				**Findings based on Proposed Model Graph Residual U-net**
Author name, Year and Ref. No.	Methodology followed.	Dataset Details	Accuracy	
Jia, et al. [6]	A "super-resolution road mapping network model" that effectively incorporates the spatial relationships between "roads" and "rural settlements".	Sentinel2 dataset images- Training:850 Testing: 31	Producer's Accuracy of 75.9 and a User's Accuracy of 88.1.	Total 10640 Sentinel-2 samples for training and 1000 used for testing. Accuracy: 0.99636, and loss 0.00017
Vani and Kumar [1]	Line detection and image enhancement techniques	Cartosat-2	Results are demonstrated using visual comparison of the output masks generated. Quantitative findings not mentioned.	Evaluation of Proposed model is as follows: Accuracy0.97702 Precision0.78589 Recall0.61158 F1Score0.68786 Loss0.06150
Leena, et al. [8]	Object oriented approach	Cartosat-2.	The overall accuracy results have demonstrated a range from 77.46% to 92%.	Evaluation of Proposed model is as follows: Accuracy0.97702 Precision0.78589 Recall0.61158 F1Score0.68786 Loss0.06150

5. Conclusion and Future Scope

In this study, extensive literature survey is done to identify the exiting findings related to dataset usage and methodologies proposed for road extraction using deep learning. There were limited findings with Cartosat and Sentinel2 dataset. We proposed and evaluated a novel Graph Residual U-Net Cartosat model and Graph Residual U-Net Sentinel model for road extraction, leveraging the strengths of convolutional neural networks (CNNs) and graph-based methods. The model was trained on different spatial resolution imagery as well as tested on various datasets, including DeepGlobe dataset, Google Earth imagery, Cartosat-2 and Sentinel-2 images. Our results demonstrate that integrating graph-based methods proves to be a promising approach in tackling deep learning tasks like road extraction. All the visual images generated confirm the findings of the results. The proposed research not only contributed in achieving excellent accuracy but also developed Sentinel-2 dataset with 11640 image patches with its binary mask The proposed research demonstrated following key findings i) Generalization Capability of the model: The model's ability to handle different image scenarios as evidenced by its successful testing on Google Earth imagery, highlights its robustness and adaptability to diverse real-world conditions. ii)Road Extraction from Low-Resolution Imagery**:** The successful extraction of major roads from Sentinel-2 imagery, despite its lower resolution, underscores the model's potential for applications in regions where high-resolution data is unavailable. The model can be further improved and evaluated by incorporating the attention mechanism. Also, there are following areas which can be explored for valuable contribution in the proposed research work-

i) Challenges Faced by Current Road Extraction Methodologies-Numerous Road extraction methodologies, including ‘graph-based’ and ‘transformer-based’ approaches, are encountering significant challenges due to their substantial computational demands. Understanding these challenges is crucial in the development of lightweight networks, which has become a critical necessity.

ii) Unlocking the Potential of Multimodal Data Fusion-The investigation of ‘semi-supervised’ and ‘unsupervised’ methods continues to focus on interpretation of the data and enabling adaptive model training without dependence on manual annotation. The potential of multimodal data fusion in this area is vast and inspiring, making it a prominent topic of research.

iii) Urgent Need of ‘Adaptive Learning’: The capability of road extraction models to adapt to remote sensing imagery is paramount when addressing complex scenarios. This adaptability is not just a feature, but a necessity that ensures models can proficiently extract road information considering the challenges of building occlusions, tree cover, and variable lighting conditions.

There are dataset specific challenges which need attention in further research using Cartosat-2 and Sentinel-2. Both Cartosat-2 and Sentinel 2 imagery present its own set of challenges. In future model can be improved to handle following challenges related to dataset-i) Cartosat-2 Imagery: With spatial resolution of 1m, Cartosat-2 imagery aligns well with very high-resolution imagery. Though this might give detailed context and representation of roads, it gives rise to problems such as road occlusions caused by shadows and other peripheral objects like trees, buildings, and vehicles. ii) Sentinel 2 Imagery: Offering a maximum spatial resolution of 10m, Sentinel 2 imagery lacks the finer details in the image and thereby demands the ability to capture thinner roads.

References

[1] K. S. Vani and K. P. Kumar, "Detection and extraction of roads using cartosat-2 high resolution satellite imagery," in *Proceedings of the 2020 3rd International Conference on Image and Graphics Processing*, 2020, pp. 7-11.

[2] R. Liu *et al.*, "A review of deep learning-based methods for road extraction from high-resolution remote sensing images," *Remote Sensing*, vol. 16, no. 12, p. 2056, 2024. https://doi.org/10.3390/rs16122056

[3] X. Bai, L. Guo, H. Huo, J. Zhang, Y. Zhang, and Z.-L. Li, "Rse-net: Road-shape enhanced neural network for Road extraction in high resolution remote sensing image," *International Journal of Remote Sensing*, vol. 45, no. 19-20, pp. 7339-7360, 2024. https://doi.org/10.1080/01431161.2023.2214277

[4] D. Feng, X. Shen, Y. Xie, Y. Liu, and J. Wang, "Efficient occluded road extraction from high-resolution remote sensing imagery," *Remote Sensing*, vol. 13, no. 24, p. 4974, 2021. https://doi.org/10.3390/rs13244974

[5] S. Mo, Y. Shi, Q. Yuan, and M. Li, "A survey of deep learning road extraction algorithms using high-resolution remote sensing images," *Sensors*, vol. 24, no. 5, p. 1708, 2024. https://doi.org/10.3390/s24051708

[6] Y. Jia, X. Zhang, R. Xiang, and Y. Ge, "Super-resolution rural road extraction from Sentinel-2 imagery using a spatial relationship-informed network," *Remote Sensing*, vol. 15, no. 17, p. 4193, 2023. https://doi.org/10.3390/rs15174193

[7] C. Ayala, C. Aranda, and M. Galar, "Multi-temporal data augmentation for high frequency satellite imagery: A case study in sentinel-1 and sentinel-2 building and road segmentation," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 43, pp. 25–32, 2022. https://doi.org/10.5194/isprs-archives-XLIII-B3-2022-25-2022

[8] C. Leena, M. R. Saxena, and R. S. Dwivedi, "Road extraction from Cartosat-2F multispectral data with object-oriented analysis," *Photogrammetric Engineering & Remote Sensing*, vol. 87, no. 3, pp. 171-180, 2021. https://doi.org/10.14358/PERS.87.3.171

[9] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual u-net," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 749-753, 2018. https://doi.org/10.1109/LGRS.2018.2802944

[10] C. Ding, L. Weng, M. Xia, and H. Lin, "Non-local feature search network for building and road segmentation of remote sensing image," *ISPRS International Journal of Geo-Information*, vol. 10, no. 4, p. 245, 2021. https://doi.org/10.3390/ijgi10040245

[11] L. Zhou, C. Zhang, and M. Wu, "D-link net: Link net with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction," presented at the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA: IEEE, 2018.

[12] J. Li, Y. Liu, Y. Zhang, and Y. Zhang, "Cascaded attention DenseUNet (CADUNet) for road extraction from very-high-resolution images," *ISPRS International Journal of Geo-Information*, vol. 10, no. 5, p. 329, 2021. https://doi.org/10.3390/ijgi10050329

[13] S. He *et al.*, "Sat2graph: Road graph extraction through graph-tensor encoding," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, Proceedings, Part XXIV, Springer International Publishing*, 2020, vol. 16, pp. 51-67.

[14] Q. Zhu *et al.*, "A global context-aware and batch-independent network for road extraction from VHR satellite imagery," *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 175, pp. 353-365, 2021. https://doi.org/10.1016/j.isprsjprs.2021.03.016

[15] R. Lian, W. Wang, N. Mustafa, and L. Huang, "Road extraction methods in high-resolution remote sensing images: A comprehensive review," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* vol. 13, pp. 5489-5507, 2020. https://doi.org/10.1109/JSTARS.2020.3023549

[16] J. Fan and Z. Yang, "Deep residual network based road detection algorithm for remote sensing images," presented at the 2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE), Harbin, China: IEEE, 2020.

[17] L. Gao, W. Song, J. Dai, and Y. Chen, "Road extraction from high-resolution remote sensing imagery using refined deep residual convolutional neural network," *Remote Sensing,* vol. 11, no. 5, p. 552, 2019. https://doi.org/10.3390/rs11050552

[18] Y. Zao, Z. Zou, and Z. Shi, "Topology-guided road graph extraction from remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 62, pp. 1-14, 2023. https://doi.org/10.1109/TGRS.2023.3344150

[19] Y.-Q. Tan, S.-H. Gao, X.-Y. Li, M.-M. Cheng, and B. Ren, "VecRoad: Point-based iterative graph exploration for road graphs extraction," presented at the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA: IEEE, 2020.