Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.777503

ROAD DETECTION WITH SATELLITE/AERIAL IMAGES

Sowmya Rajnarayanan, Nirmal Kumar Kumaresan, Dr. J. Jayashree* and Yoganarasimha Kollipalli
School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India

Abstract – Satellite and aerial images can play a crucial role in supporting the planning and coordination of global change research by aiding in the development of research strategies and facilitating the implementation of methodologies. Deep learning algorithms have advanced, allowing image sensors to comprehend the scene for the target object with greater accuracy, notably in the domain of segmentation. Due to variable lighting conditions, irregular road geometries, and fuzzy boundaries between the road and other objects, segmenting a road scene from colour photos using a computer vision approach might be difficult. We have used U-Net, Seg-Net, and fully convolutional network (FCN) models to achieve this goal and to clearly separate the road from the non-road component. According to the studies, U-Net outperformed Seg-Net and FCN-32 in terms of mIoU and dice coefficient. Also, the well-known encoder-decoder structure is used by deep convolutional neural networks like SegNet and UNet to segment pictures. These networks' encoders downscale the image gradually while expanding the receptive field in order to encode the features. After recovering the features' spatial dimensions and making the final predictions in full resolution, the decoder utilizes up-sampling techniques to increase the resolution or size of the features.

Keywords – satellite images, road detection, UNET, deep learning, computer vision

1. INTRODUCTION

Satellite and aerial images can play a crucial role in supporting the planning and coordination of global change research by aiding in the development of research strategies and facilitating the implementation of methodologies. The presence of high-resolution satellite images / aerial images and their potential to be used in a wide variety of applications such as road detection and building identification. Road segmentation is crucial to the automation of digital mapping; in recent years, it has received a lot of attention and evolved significantly [1]. However, using the road segmentation approach is difficult due to the background complexity, noise, shadows, and occlusions that go along with it. While some recent work has produced sophisticated neural network topologies, others have significantly enhanced earlier work [2]. The results of all this research have improved segmentation efficiency [1,2,3].

The well-known encoder-decoder structure is used by deep convolutional neural networks like UNet to segment pictures. These networks' encoders downscale the image gradually while expanding the receptive field in order to encode the features. Once the spatial dimensions of the features have been restored and the final predictions are made at full resolution, the decoder employs up-sampling techniques to further increase the size or resolution of the features [4]. To do this, we require Keras with TensorFlow to be installed. In this paper, a method for predicting segmentation masks and object edges simultaneously in difficult situations, specifically highways in satellite pictures, is proposed. It produces road segmentation masks and then creates the road's edges using these segmented masks. Our project aims to detect roads

from satellite images using segmentation models and augmentations libraries on Keras. Image categorization, which is a fundamental problem in computer vision and a key focus in various visual recognition domains, has been a traditional research topic in recent years. Image feature extraction has been a pivotal aspect of traditional methods for image categorization, and has continued to be an active area of research. But around 2015, the use of CNNs for large-scale visual classification applications saw success, and the area of remote sensing image analysis has now fully embraced the technology [5]. Several CNN-based scene categorization techniques have been developed by utilizing various CNN techniques [6]. The images from the repository are three dimensional. This is because coloured pictures have the tricolour RGB channels [7]. For this machine learning model, these are not required. So, passing the dataset images through a grayscale function clears up data other than the important information, also reducing the image to a single dimension. CNNs are those layers that are used for image processing, classification, and segmentation. And before sending images for classification, images are processed with open cB and contours are identified, which are curves joining the continuous paths. Contours are used for object detection. Once it is done, a model will be used to predict the result.

2. LITERATURE SURVEY

This literature survey provides a summary of eight research papers related to road detection from satellite imagery. These papers cover different approaches, including deep learning models, image segmentation, and object-based classification, to address the challenges of detecting roads in underdeveloped regions and low-quality images.

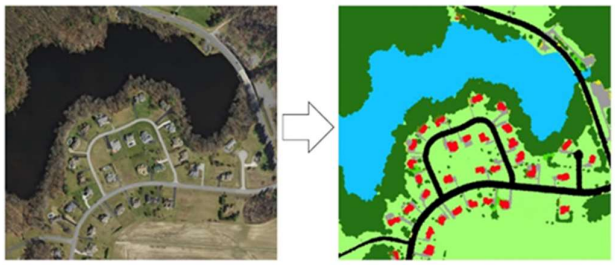
Nachmany and Alemohammad's paper focuses on the need for regionally trained models for road detection in developing regions [8]. They argue that models trained in well-developed regions may not be efficient in detecting roads in underdeveloped/developing regions due to the varying topological differences in these regions. To address this issue, they propose using local training data to build regionally trained models. They use the Raster Vision framework, which uses DeepLab for semantic segmentation.

In Wu et al.'s paper, a novel model called Attention Dilation Linknet is introduced, which extends the D-Linknet34 architecture by incorporating a series parallel combination of diluted convolutions and an attention mechanism to create an enhanced semantic segmentation network [9]. The model is trained on three different datasets, including DeepGlobe's road extraction dataset, DeepGlobe's land classification dataset, and Inner Mongolia's land classification dataset. Comparative evaluations are conducted against existing models to assess the performance of the proposed model. The results show a 0.64% improvement in accuracy over D-Linknet34 in road extraction and 0.35% in land classification.

Oehmcke et al.'s paper[10] proposes two models, one of which is U-Net+, which builds upon the popular U-Net architecture, originally proposed by Ronneberger et al. in 2015. While U-Net+ is good at road detection for high-quality images, they point out that we don't always get good quality images as they are often occluded by clouds. The second architecture is U-Net+Time, which is a model for sequences of low-quality scenes. It employs volumetric convolutions to merge time-series data from twelve different timestamps, combining the imagery per band in a three-dimensional manner.

In Yang et al.'s paper, a novel deep learning model called recurrent convolutional neural network UNet (RCNN-UNet) is introduced. [11]. The model has three major benefits,

including getting rid of propagation errors using an end-to-end deep learning strategy, using a carefully crafted RCNN unit to better take advantage of spatial context and low-level visual data, and using a multi-task learning scheme to train two predictors simultaneously to increase both efficacy and efficiency.



Malpani et al.'s paper proposes using Otsu's method of image segmentation, binarization, and morphological operations to detect roads from satellite images in a more efficient manner[12]. They demonstrate that their proposed method shows the proper edges of the road and the shapes and edges of various objects and buildings.

Zhang et al.'s paper proposes an unsupervised co-segmentation technique that can be used on images with many foreground items present at once and a significantly changing backdrop [13]. For semantic extraction, the RGB space colour edge image is extracted. By recursively modeling the appearance distribution of pixels and regions, this technique effectively differentiates between foreground and background. Experimental results reveal that deep convolutional neural networks exhibit high accuracy and efficiency in performing semantic segmentation and classification of scene images.

Yigit and Uysal's paper considers properties of objects, object-based extraction of information, and object resolution as the main reasons behind the efficiency of their approach[14]. The classification of objects is mostly accomplished by segmentation using the object-based approach method, which replaces the pixel-based technique employed for many years. They use the classification program defines Cognition to classify data more quickly and accurately. The classification result can be transformed to vector format and connected with geographic information systems, and errors or incorrect class assignments can also be swiftly remedied.

In Zhu et al.'s (2021) study, a novel Global Context-aware and Batch-independent Network (GCB-Net) was introduced for road extraction from very high-resolution (VHR) satellite imagery [15]. The GCB-Net utilized Global Context Aggregation (GCA) blocks to enhance the global spatial relationship, resulting in state-of-the-art performance on the DeepGlobe Roads and SpaceNet Roads datasets.

The study by Ayala et al. (2021) presented a deep learning-based approach in their study for improved building footprint and road detection in high-resolution satellite imagery [16]. They used Sentinel 1 and Sentinel 2 images and showed that fusing these two images improves the model's performance compared to using only Sentinel 2 images. They also suggested including more zones and continents to improve the dataset's variety for training and testing.

In Ghandorh et al.'s (2022) study, a hybrid encoder with two sections was proposed for road detection in very high-resolution (VHR) satellite images[17]. The first section extracted features comprehensively, while the second section utilized max-pooling layers to enhance the network's receptive field and produce high-resolution feature encoding. The study also showed that a combination of weighted cross-entropy loss and Tversky loss functions can effectively

improve training performance in scenarios where class imbalance is prominent, such as road extraction.

3. BACKGROUND STUDY

3.1 Satellite and Aerial images

Making an effective GIS system involves a lot of work in the field of road detection using satellite photos. The development of a GPS system that can recognise its surroundings is also aided by this. Prior researchers mapped the routes using satellite imagery. However, contemporary researchers have combined satellite and aerial pictures to better interpret the images. This is because the forecasts are significantly more accurate when made using aerial photos, which offer superior image resolution. All imagery captured by an airborne craft is referred to as aerial imagery. It is divided into groups based on the camera axis, scale, and sensor. While aerial photography has more small-scale commercial applications, satellite photos have more large-scale scientific applications.

3.2 Computer vision

Computer vision is the field of using computers and advanced AI algorithms to enable computers to gain an understanding of images. There are different granularities at which computers can comprehend images. There is an issue in the Computer Vision area that has been defined for each of these levels starting with a coarser knowledge and working our way up to a finer understanding. The area of computer vision known as image classification focuses on grouping the provided image into a specific cluster. Extraction of information classes from a multiband raster image is what this process entails. Thematic maps can be made using the raster that is produced after picture categorization.

To locate instances of things in images or movies, a computer vision technique known as object detection and localization is used. Object detection algorithms usually make use of deep learning or machine learning to get findings that are helpful. Object detection, also known as object recognition, is used in a variety of applications, such as security systems, self-driving cars, pedestrian counting, vehicle recognition, and many more.

Another methodology that improves on sematic segmentation is instance segmentation. This model finds each instance of the class in addition to the appropriate pixels. For instance, when segmenting a picture of five people, each person is assigned to a separate class and will be represented by a different colour. In contrast, all humans will be placed into the same class in semantic segmentation.

3.3 Sematic segmentation

We used sematic segmentation model for the detection of roads. The goal of semantic segmentation is to label each pixel of an image with an appropriate class. Because this process classifies every pixel in an image it is also called dense prediction.

To categorize each pixel of a picture (such as a tree, animal, nest, sky, lake, or person), a procedure known as semantic segmentation is used. A few applications for semantic segmentation include autonomous driving, industrial inspection, and the classification of terrain visible in satellite images.

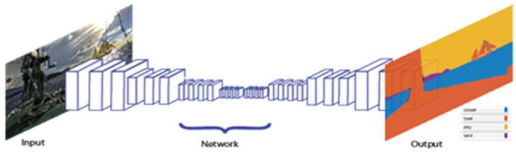


Figure 2 Segmentation process

4. PROPOSED MODEL

4.1 UNET architecture

mage segmentation is a highly intricate task in computer vision that involves the partitioning of an image into meaningful regions or segments, each representing a distinct object or region of interest. We have experimented with image classification a lot; in fact, convolutional neural networks and their application to image classification come to mind when we hear of computer vision applications. But there's a brand-new issue in town that goes by the name of "semantic segmentation." Image segmentation has two tasks to complete: localization and classification, in contrast to image classification, which seeks to predict a single class for the entire input image. Finding a specific object's location (in pixels) inside a much bigger image is referred to as localization. Next is classification, which means categorising the object that has been located within the image.

For the purpose of segmenting images, UNET is a fully convolutional neural network. It is one of the most widely used approaches in semantic segmentation today and is built to learn from fewer training datasets. The U-shaped encoder-decoder arrangement known as UNET consists of four encoder and decoder blocks connected by a bridge. The feature extractor is an encoder network, which learns the input image's abstract representation through a series of encoder blocks.

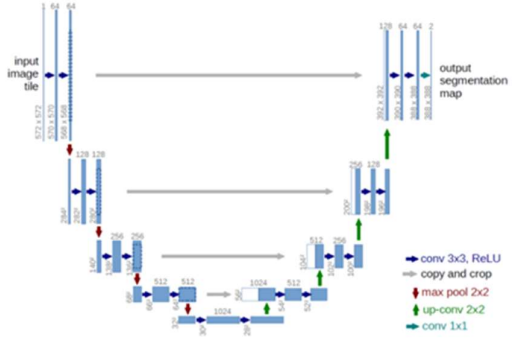


Figure 3 Unet Architecture

Three 3X3 convolutions make up each encoder block, and a ReLU (Rectified Linear Unit) activation function follows each convolution.

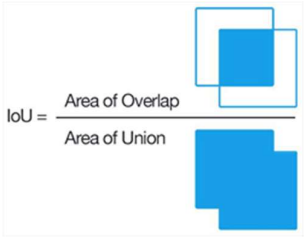
ReLU's output is a skip connection for the decoder block after it. Skip connections make better semantic features possible. The bridge, which has 3X3 convolution and ReLU activation function in each convolution, joins the encoder and decoder networks and completes the information flow. The decoder network receives the abstract representation and produces a semantic segmentation mask. Beginning with 2X2 transpose convolution, the decoder block is joined by the encoder block's skip connection function. The output of the final decoder is subjected to a 1X1 convolution with sigmoid activation to create a segmentation mask that represents the categorization of pixels.

The first path, or contracting path, uses a reduction convolution to process the input image. This multiplies the quantity of feature vectors generated from the source data (input image). As a result, there are no feature vectors present at the beginning of the operation. The data is then max pooled to speed up computing and continue the model's architecture after a sequence of reductive convolutions. This route consists of a sequence of max pool operations and reductive convolutions. We are left with a very low dimensional piece of data (a highly max pooled image) at the end of the journey, along with a large number of feature vectors that characterise the key characteristics of the data. The subsequent component of the model's design, the extensive path.

The second path, or expanding path, applies a series of beneficial convolutions to the low dimensional, high feature vector data from the contracting path. At the conclusion of each step of sets of reductive convolutions, the data from the contracting path is transmitted forward to adjoin the data in each subsequent step in the expanding path. The data "expands" in terms of dimensionality in this way, regaining its original dimensions. At this point, the data has been roughly extended to its original and the only features left are the determinant features that provide the solution to the problem at hand.

4.2 IOU metric or Jaccard Loss

An evaluation metric called Intersection over Union (IOU) is used to evaluate how accurately an object detector performs on a specific dataset.



IoU

$$IoU = \frac{target \cap prediction}{target \cup prediction}$$

We compare the forecast from the model with the ground truth by determining areas of union and overlap. To obtain the final result for your test set, take the mathematical average of the IoU scores.

4.3 Binary Cross Entropy:

The Binary Cross Entropy is calculated using the following formula:

$$Loss = -\frac{1}{Output\ size} \sum_{i=1}^{Output\ size} y_i . log\hat{y}_i + (1 - y_i) . \log(1 - \hat{y}_i)$$

where y^_i is the i-th scalar value in the model output, yi is the corresponding target value, and output size is the number of scalar values in the model output.

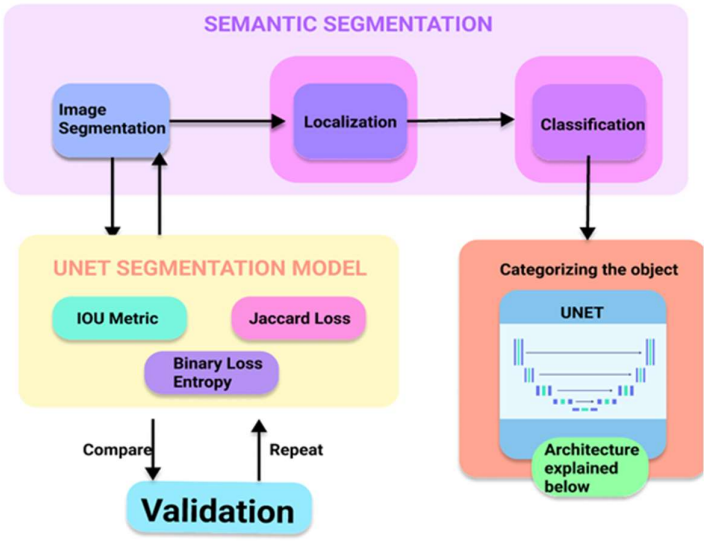


Figure 4 Framework Diagram

5. METHODOLOGY

5.1 Dataset generation

Road detection from satellite images using segmentation models is an arduous process. This is because most often roads are obscured by shadows of nearby buildings, trees and varying colours and textures of the road and soil. A high-quality aerial imagery dataset is therefore really useful. We are using the Massachusetts Road Dataset [18].

The Massachusetts Road Dataset comprises 1171 aerial images of the state of Massachusetts, with each image measuring 1500 x 1500 pixels and covering an area of 2.25 square kilometres. The training data was randomly split into 1108 images for training, 14 images for validation, and 49 images for testing.

Figure 5 Sample image from dataset

It has a variety of rural, urban regions and it covered a wide area of 2600 square kilometres. Segmentation of the images consisting of streets, buildings, and ignoring the images which didn't had buildings like water, plains, is necessary. To train the model, extracted images was

used, and before that pre-processing of the data, which involves filtering the images labelled with “roads”, segmentation of data, resizing of images to required size is done. Masking was done for every image using binary masks, 0 intensity value was for non-road class and intensity value of 1 was used for building classed

5.2 Dataset Augmentation

Deep neural networks face challenges such as limited training data and class imbalance in datasets. To address these challenges, data augmentation techniques are commonly used. In this study, various image alterations such as cropping, zooming, rotating, histogram-based techniques, finishing, style transfer, and generative adversarial networks were applied for data augmentation. These newly generated images were then used to pretrain the neural networks, such as Unet, Keras, and SegNet, to enhance the training process efficiency. Satellite images were utilized to increase the density of the training dataset by performing image rotation, cropping, and zooming using Keras pre-processing layers such as tf.keras.resizing, tf.keras.layers.rescaling, and tf.keras.layers.randomRotation.

5.3 Training

After the augmentation of images, we start working on the model. UNET architecture is available directly from some python packages, and we are going to use the segmentation_models library. There are a few parameters to tune and some functions to initialize to result in a better model. We are using EfficentNet as backbone, which has faster inference and fewer training parameters. We use the ImageNet weights as encoder weights which speeds up the training process. We also use the a few call-backs from the Keras library. Model Checkpoint save weights of the model while training, ReduceLROnPlateau reduces training if a plateau is found, Early Stopping stops training once a metric stop changing after several epochs. We also use TensorBoard to monitor the training process. After initializing and compiling the Unet model, we set IOU as a metric to monitor and bce_jaccard_loss as the loss to optimize.

6. RESULTS

Below are the results for the implemented method.

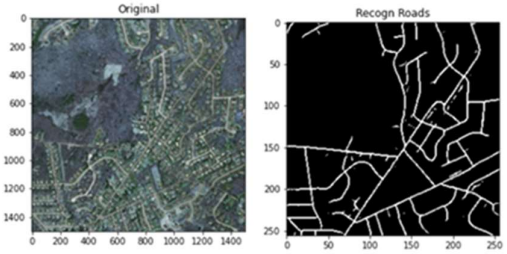


Figure 6

Figure 7

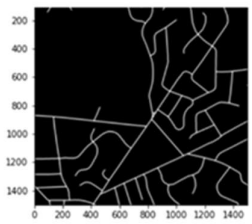


Figure 8

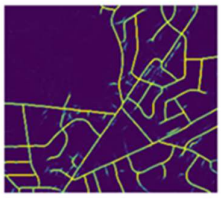


Figure 9

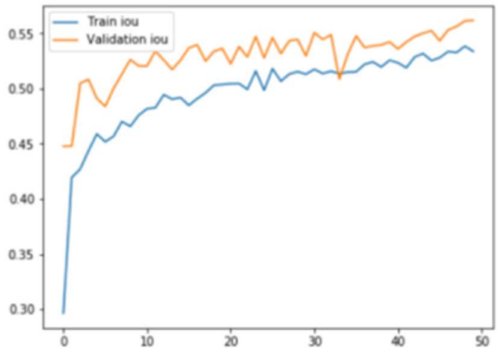
Figure 6 is the original image that is the input of the model, and Figure 7 is the original mask or road extracted from Figure 6. Figure 8 is the binary image which is obtained using a binary threshold function from Figure 9 that is the actual output of the UNet model. Table 1 compares the IoU, correlation and F1 scores of the U-Net model against standard fully convoluted network models

Table 1

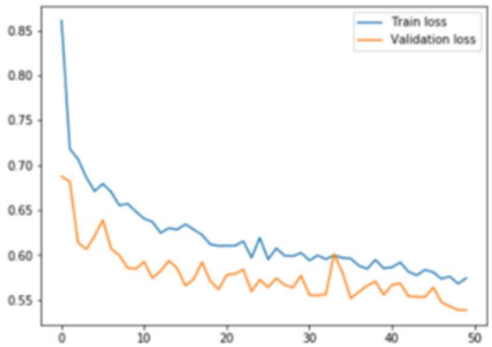
Method's name	IoU	Correlation	F1
FCN	0.52	0.9098±0.0407	0.8178±0.0392
U-Net	0.59	0.9285±0.0038	0.9656±0.004

The mask and the results of our model matched for the majority of the dataset. The model exhibits good performance in both dense and sparse environments, as seen. As we can observe the validation IOU almost matches the training IOU.

The slight difference indicates that the Intersection over Union (IoU) score achieved on the validation dataset was slightly higher compared to the score obtained on the training dataset.



The above graph show that the validation loss and training loss are nearly comparable. The model exhibits good performance on the validation dataset since it is somewhat underfitted. This is preferable to overfitting the model because that would make it overly dependent on the training data. The validation IOU almost exactly matches the training IOU, as can be shown. The subtle variation in scores suggests that the Intersection over Union (IoU) score for the validation dataset was slightly higher compared to the IoU score obtained for the training dataset. This implies that the model is likely to perform better on new data.



The IOU measure, or Intersection Over Union, was used to determine how similar the two images were. With the suggested model, we have a validation score of 0.55 IOU, but we count it as a mask for any pixel prediction that is higher than 0. By selecting an appropriate threshold, we can improve our outcome by 0.047.

Loss and IoU metric history

7. Conclusion

The work done in this paper will assist in creating a better GIS system by mapping roads using aerial and satellite images. As we are working to detect the roads, clearer aerial images are much more valuable than satellite images. Among other methods, we chose semantic segmentation, a dense prediction model that classifies each pixel into respective category. U-NET architecture, a fully convolutional neural network is used to perform our semantic segmentation. In order to produce outputs with higher resolution on the input images, UNET uses up-sampling operators immediately after successive contracting layers. As for evaluation, we used IOU index and binary cross entropy. These provide the similarity between the mask and the obtained output. Our model obtained an IOU score of 0.55, which is a fair score. By picking suitable threshold values, we can still improve the results. This research on road detection can be expanded to include the detection of buildings, water resources, and arable fields, which can aid in creating accurate maps of the locations.

8. References

- Malpani, S., Kamble, S., Chavan, M., & Nagpure, R. (2020). Road Detection from Satellite Images.
- Mokhtarzade, M., & Zoej, M. V. (2007). Road detection from high-resolution satellite images using artificial neural networks. International journal of applied earth observation and geoinformation, 9(1), 32-40.
- Ayala, C., Sesma, R., Aranda, C., & Galar, M. (2021). A deep learning approach to an enhanced building footprint and road detection in high-resolution satellite imagery. Remote Sensing, 13(16), 3135.
- Dewangan, D. K., & Sahu, S. P. (2021). Road detection using semantic segmentation-based convolutional neural network for intelligent vehicle system. In Data Engineering and Communication Technology: Proceedings of ICDECT 2020 (pp. 629-637). Springer Singapore.
- Massoumi, F., & Afzali, M. (2015). Road Detection from Satellite Images Using Image Mining.

- Ghandorh, H., Boulila, W., Masood, S., Koubaa, A., Ahmed, F., & Ahmad, J. (2022). Semantic segmentation and edge detection—Approach to road detection in very high resolution satellite images. Remote Sensing, 14(3), 613.
- Fakhri, S. A., & Shah-Hosseini, R. (2022). Improved road detection algorithm based on fusion of deep convolutional neural networks and random forest classifier on VHR remotely-sensed images. Journal of the Indian Society of Remote Sensing, 50(8), 1409-1421.
- Nachmany, Y., & Alemohammad, H. (2019). Detecting roads from satellite imagery in the developing world. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (pp. 83-89).
- Wu, M., Zhang, C., Liu, J., Zhou, L., & Li, X. (2019). Towards accurate high resolution satellite image semantic segmentation. IEEE Access, 7, 55609-55619.
- Oehmcke, S., Thrysøe, C., Borgstad, A., Salles, M. A. V., Brandt, M., & Gieseke, F. (2019, December). Detecting hardly visible roads in low-resolution satellite time series data. In 2019 IEEE International Conference on Big Data (Big Data) (pp. 2403-2412). IEEE.
- Yang, X., Li, X., Ye, Y., Lau, R. Y., Zhang, X., & Huang, X. (2019). Road detection and centerline extraction via deep recurrent convolutional neural network U-Net. IEEE Transactions on Geoscience and Remote Sensing, 57(9), 7209-7220.
- Malpani, S., Kamble, S., Chavan, M., & Nagpure, R. (2020). Road Detection from Satellite Images
- Zhang, L., Sheng, Z., Li, Y., Sun, Q., Zhao, Y., & Feng, D. (2020). Image object detection and semantic segmentation based on convolutional neural network. Neural Computing and Applications, 32, 1949-1958
- Yiğit, A. Y., & Uysal, M. (2020). Automatic road detection from orthophoto images. Mersin Photogrammetry Journal, 2(1), 10-17.
- Zhu, Q., Zhang, Y., Wang, L., Zhong, Y., Guan, Q., Lu, X., ... & Li, D. (2021). A global context-aware and batch-independent network for road extraction from VHR satellite imagery. ISPRS Journal of Photogrammetry and Remote Sensing, 175, 353-365.
- Ayala, C., Sesma, R., Aranda, C., & Galar, M. (2021). A deep learning approach to an enhanced building footprint and road detection in high-resolution satellite imagery. Remote Sensing, 13(16), 3135.
- Ghandorh, H., Boulila, W., Masood, S., Koubaa, A., Ahmed, F., & Ahmad, J. (2022). Semantic segmentation and edge detection—Approach to road detection in very high resolution satellite images. Remote Sensing, 14(3), 613.
- https://www.kaggle.com/datasets/balraj98/massachusetts-roads-dataset