1. Hello My name is Mansi uniyal. I'm now going to present my presentation on the topic of Deep Learning-based Semantic Segmentation in Remote Sensing.

2. So let's go through the contents that are going to be covered. They are:
    a. DEEP LEARNING
    b. SEMANTIC SEGMENTATION
    c. REMOTE SENSING
    d. OVERVIEW OF PREVIOUS WORK
    e. WORKING ON DEEP LEARNING MODELS
    f. INFERENCES AND PERFORMANCES
    g. SUMMARY
    h. FUTURE SCOPE
    i. REFERENCES

3. First, let’s define Deep Learning. Here we will be specifically addressing deep learning by the use of a deep neural network. Secondly, let’s define satellite imagery.

4. Then we ponder on what Semantic segmentation is. More specifically, the goal of semantic image segmentation is to label each pixel of an image with a corresponding class of what is being represented. Because we're predicting for every pixel in the image, this task is commonly referred to as dense prediction. One important thing to note is that we're not separating instances of the same class; we only care about the category of each pixel. In other words, if you have two objects of the same category in your input image, the segmentation map does not inherently distinguish these as separate objects.

5. Segmentation models are useful for a variety of tasks, including:
    a. Autonomous vehicles
       We need to equip cars with the necessary perception to understanding their environment so that self-driving cars can safely integrate into our existing roads.
    b. Medical image diagnostics
       Machines can augment analysis performed by radiologists, greatly reducing the time required to run diagnostic tests.
    c. GeoSpatial Remote Sensing
       We will be looking into deeper depths on this in the coming up slides.

6. Semantic Segmentation of images task can be obtained by using various Deep Learning methods. As shown in the example figure below is the model architecture of SegNet. This core trainable segmentation architecture consists of an encoder network, a corresponding decoder network followed by a pixel-wise classification layer.

7. Right now, there are satellites orbiting around the Earth, at more than 600 km above its surface. Each satellite has its own sensor, which gives you specific data for each satellite. They are taking pictures of the Earth, at a given place and at a given time.

8. Among the various examples of remote sensing, let's look into the reasons why analyzing such obtained data is necessary.
    a. Large forest fires can be mapped from space, allowing rangers to see a much larger area than from the ground.
    b. Tracking clouds to help predict the weather or watching erupting volcanoes, and help watch for dust storms.
    c. Tracking the growth of a city and changes in farmland or forests over several years or decades.
    d. Discovery and mapping of the rugged topography of the ocean floor (e.g., huge mountain ranges, deep canyons, and the “magnetic striping” on the ocean floor).

9. Let’s start with why satellite images are specific data.
    a. First, if you compare photography and satellite image, you will see that usually, a satellite image has much more pixels than a classical photograph. And that, considering that a photograph can already have thousands of pixels.
    b. Secondly, if you compare a photography and satellite image, you will see that you don’t have the same number of channels. In a classical image, you will have three channels: one for the red color, one for the blue color, and one for the green color. You will have all these channels in a satellite image, but you may also have others and dozens of others.
    c. Thirdly, when a satellite is taking an image of the Earth, you may want to know where the place is. So, you will have to consider geo-referenced images. These images will first be recorded in an elliptical frame. And from this elliptical frame, you will have to take your coordinates into a flat Earth projection. And this allows us to introduce the ground resolution. When a satellite is looking down at the Earth's surface, each of its sensor’s pixels is able to record a certain distance on the surface. All this information is recorded in the metadata of the image.
    d. But in this metadata, you may also find the sun angle, atmosphere conditions, and the satellite position for a few examples.
    e. A car in a satellite image is just a few pixels. These small objects are for instance cars, while cars are supposed to be big objects for us. This causes a decreased resolution of the objects from such a high altitude.
    f. Now, when a satellite records an image of a specific area, at a specific time, it only has one shot. It means that if the weather is bad if you have clouds if you have snow or haze, you will have to deal with it, and your vehicle detector will have to see the car in the haze or in the snow. Another important thing is the position of the sun, because your image may be overexposed, or on the contrary, you may have huge shadows in your image, and your vehicle detector will have to be able to detect vehicles even in the shadows.
    g. Lastly, when a satellite looks down at the Earth and when the line between the sensor of the satellite and the center of the image makes a right angle with the ground surface. This is the position of the sensor where you will get the best resolution. Else it will end up having buildings to hide objects such as vehicles.

10. First, as we said before, a vehicle in a satellite image will be represented by only a few pixels and only a few features. So you might not have enough information to use classical computer vision methods. Secondly, the context around all these vehicles will always change. And this is where deep learning has proven to be a great solution for such use cases.
    With satellite images, you will have to keep in mind three things: The first one, as we said before, is that our images are really big. This means that your network won't be able to process all of it at once, and you will have to divide it into smaller tiles before putting it in your network. The second thing is that your image may have more than three color channels, not just red, blue, and green colors. Your whole pipeline will have to take care of all of these channels and take them into account, or not. And the last thing is that you have to consider that this data is geo-referenced.

11. With the development of various algorithms to deal with different objects in such big satellite imagery the detection of aircraft, roads, buildings, ships, and different types of land use such as water, green, space, and so on, can make impacts in different scenarios. I would now like to show some specific cases of the same.
    a. Land cover mapping is a basic process to categorize and describe the surface of the earth. It provides fundamental data from various management and research applications such as food production forecast, urban planning, flood control disaster prevention, and other earth system studies like estimating the fire area in Australia.
    b. To find out what happened in a forest if you stay inside sometimes it takes lots of time to find out if there is deforestation. However, without forest change detection we can help to find the deforestation very easily and make efficient responses. We can take two images collected at different times and see the forest change with high accuracy and speed with the segmentation of the images.

12. We can also focus on a city inside, for example, damage selection after earthquake-like building extraction. Let's take building extraction as an example to introduce how we extract and understand urban areas' building extraction from remote sensing images. In recent years more building extraction studies are based on pixel-wise semantic labeling methods semantic segmentation models and instance segmentation models have been widely explored for building extraction tasks. Buildings are another important indicator of urban growth and change, and accurately annotated labels for most regions are already publicly available.

13. It's easy to find that buildings and roads changed dramatically in last 20 years using ai technology and satellite imagery which is whole a district transformed from a countryside to an urban area. We can also detect the building change in the city there are two images collected at different times the green parts are the buildiing change. We detected, zoomed, saw which buildings are newly built which buildings are removed and which buildings are rebuilt. By tracking the urbanization track we help local authority to monitor the change of the city and build sustainable communities in the future.

14. So lets built on how segmentation is carried in deeplearning. The satellite image consists of huge amount of pixels which cant be computed directly in the model in one go. Thus the image is first preprocessed by patched smaller dimension grids of images to be fed into the model. The images are taken in accordance with their metadata and respectively refered again while repatching the grids to form the segmentated satellite image later. Taking an example of Unet architecture.The input to the deep learning model is a grid image of specified dimension with numerous number of channels according its metadata. It then runs through the various convolution layers of the model as an encoder and decoder shaped architecture. Encoder is to extract different features of the image and decoder part is to rebuild the segmentated image on the basis of the trainned extracted features, to a shape same as input image.

15. The road segmentation task can be thought of as binary classification, with road pixels being the positive whereas the non-road pixels being negative. To assess the relative efficacy of various segmentation and classification models, quantitative criteria for comparison must be applied. The most common evaluation indexes are namely Accuracy, the Mean Dice Coefficient (mDC), mean Intersection over Union (mIOU), precision, and recall, assisting in the comparison and evaluation of the models.
    a. The dice Coefficient can be used to calculate the similarity between the segmentation images and their ground truth. A larger mDC stands for a better segmentation result.
    b. mIOU refers to the overlap rate of the generated candidate boxes and the original marker boxes, that is, the ratio between the intersection and the union.

16. Here are the results of some of the recent works on how the models have been performing on the publicly available Massachusetts dataset.

17. In summary, a brief overview of what deep learning, semantic segmentation, and remote sensing are covered along with special attention to the application of deep learning in the remote sensing satellite imagery for semantic segmentation of road extraction for urban planning. We also went over state-of-the-art implementations in deep learning techniques along with an introduction to developing a deep learning pipeline for satellite imagery. Build a working deep network pipeline that takes in data to produce semantically segmented maps on the images. The sequence of the preprocessing steps taken to create a usable dataset and reasoned a methodology for building the deep learning models in this work. Compare different neural network structures specified in the existing literature. The models developed were evaluated using standard classifier evaluation techniques. These were compared to provide final assessments of the models.

18. It can be observed that much work can be done in improving these performances. Future works in the field can primarily focus on the topics listed as follows.

a. Foremost is the fact that the dataset used in training is fairly small, and better and more varied augmentation schemes can be used. Publicly available resources can also be leveraged at a larger scale, to obtain more satellite imagery to train the networks. Care must be taken to ensure appropriate corrections for the scaling of images.
b. A manual approach was used for hyperparameter fine-tuning in most cases. An algorithmic approach to exploring the hyperparameter space can future work to ensure that the best hyperparameters are selected.
c. Furthermore, various other variations such as attention mechanisms, hybrid dilations can be used to further improve the performance of automatic road extraction from satellite images.

19. Here is the list of source references that were ever used in the presentation.

20. With this, we come to the end of our presentation. Thank You and have a nice day.