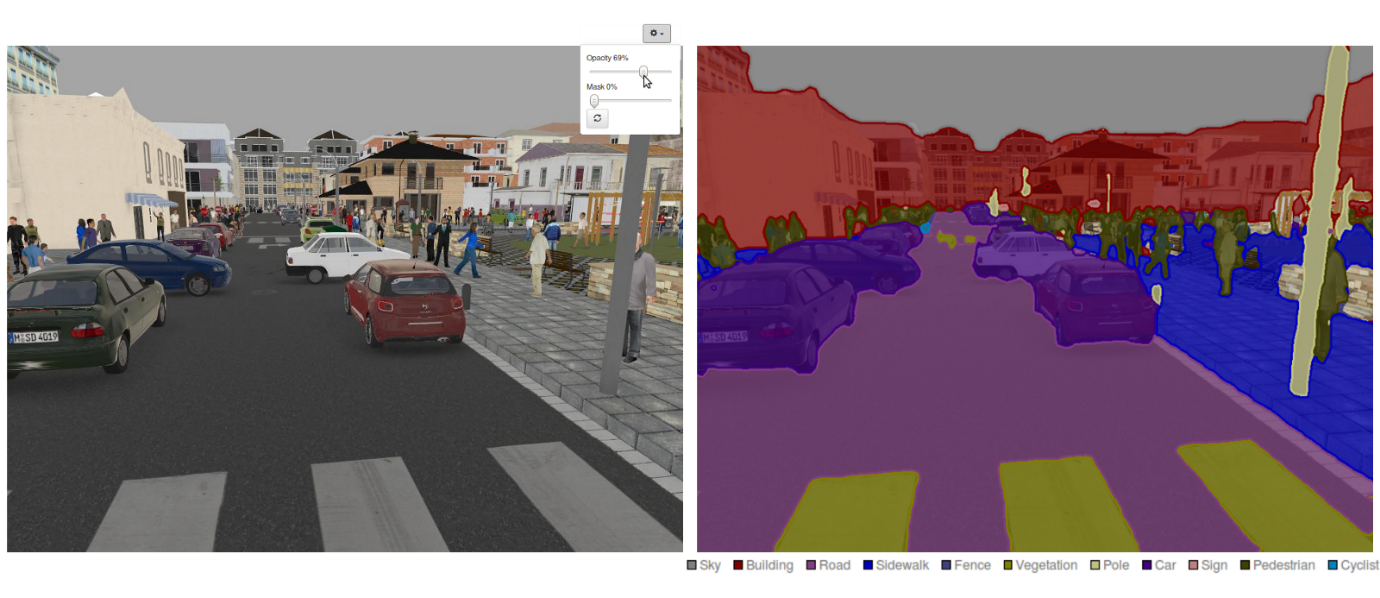
**Semantic Segmentation**

**Introduction**

Deep learning algorithms have solved several computer vision tasks with an increasing level of difficulty. But the object detection task has exceeded the image classification task in term of complexity as simple deep learning algorithms can’t provide a full comprehension of a scene and they only classify a small part of the information.

The semantic segmentation is one of the high-level task that paves the way towards complete scene understanding.



Semantic segmentation is the task of classifying each and very pixel in an image into a class as shown in the image above. Here you can see that all buildings are red, the vehicles are purple, the sidewalk is blue etc.

Some examples where semantic segmentation can be useful are: -

1. Self Driving Cars
2. Medical image diagnostics
3. Damage Detection

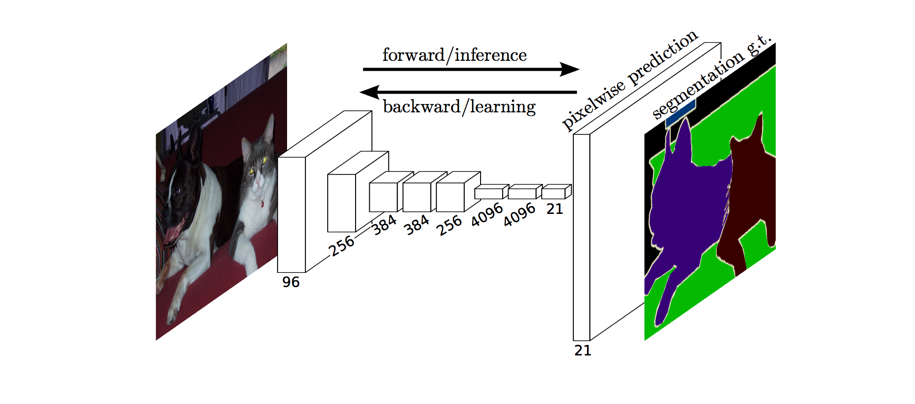
**Deep Learning Model Architectures for Semantic Segmentation**

There are various Deep Learning Model Architectures for Semantic Segmentation which are listed below: -

1. Fully Convolutional Network (FCN)
2. ParseNet
3. U-Net
4. Feature Pyramid Network (FPN)
5. Mask RCNN

Here I will be focusing on Fully Convolutional Network (FCN) and how they are used in autonomous vehicles.

**Fully Convolutional Network (FCN)**

****

The approach of using a "fully convolutional" network trained end-to-end, pixels-to-pixels for the task of image segmentation was introduced by J. Long et al. in late 2014. The FCN takes an image with an arbitrary size and produces a segmented image with the same size. The author proposed adapting existing, well-known architectures (AlexNet, VGG16, GoogLeNet) to act as an encoder while replacing all the fully connected layers by convolutional layers which acts as a decoder. This model uses various blocks of convolution and max pool layers to first decompress an image to 1/32th of its original size. It then makes a class prediction at this level of granularity. Finally, it uses up sampling to resize the image to its original dimensions. There are few different methods to upsample such as unpooling which performs the opposite of pooling. But here another method called transpose convolutions is mostly used. The goal of down sampling steps is to capture semantic/contextual information while the goal of up sampling is to recover spatial information. The final image is the same size as the original image.

However, because the encoder module reduces the resolution of the input by a factor of 32, the decoder module struggles to produce fine-grained segmentations. The author has addressed this problem and due to which skip connections are added which improved the performance of the model.

**Autonomous Vehicles**



In the context of autonomous vehicles, autonomous vehicles also require a deep understanding of the surrounding. As discussed earlier, a simple deep learning approach does not provide spatial information so semantic segmentation plays a very important role in self driving cars. Camera frames are used to recognize the road, pedestrians, cars, sidewalks and many other things at a pixel-level accuracy with the help of semantic segmentation. It is also applied to know the drivable surface and estimate the lane where the car can drive. FCNs are applied which is useful for complete scene understanding (working of FCN explained above) and accordingly the steering angle and speed throttle is controlled. This methodology is very useful and have helped in achieving autonomous vehicles a reality.