We went through the paper Pyramid Scene Parsing Network and learned about the PSP network and how it works.

The main goal of this network is to assign each pixel of an image in a category label. This network provides a complete understanding of the scene. It predicts the label, location as well as shape of each element in the image.

Difficulty of this network's computational work depends on the scene and label variety.

This kind of semantic segmentation network works by having two parts: an encoder and a decoder. The encoder is responsible for extracting features from the scene and the decoder predicts the labels of the different segments of the scene.

Some key characteristic of the PSPNet, which we discovered so far is:

# CNN backbone

# Dilated convolutions

# pyramid pooling.

For implementing this kind of network we learned about an API called arcgis.learn.

* https://developers.arcgis.com/python/api-reference/arcgis.learn.toc.html

A another API is Deeplab

* https://chainercv.readthedocs.io/en/stable/reference/links/deeplab.html