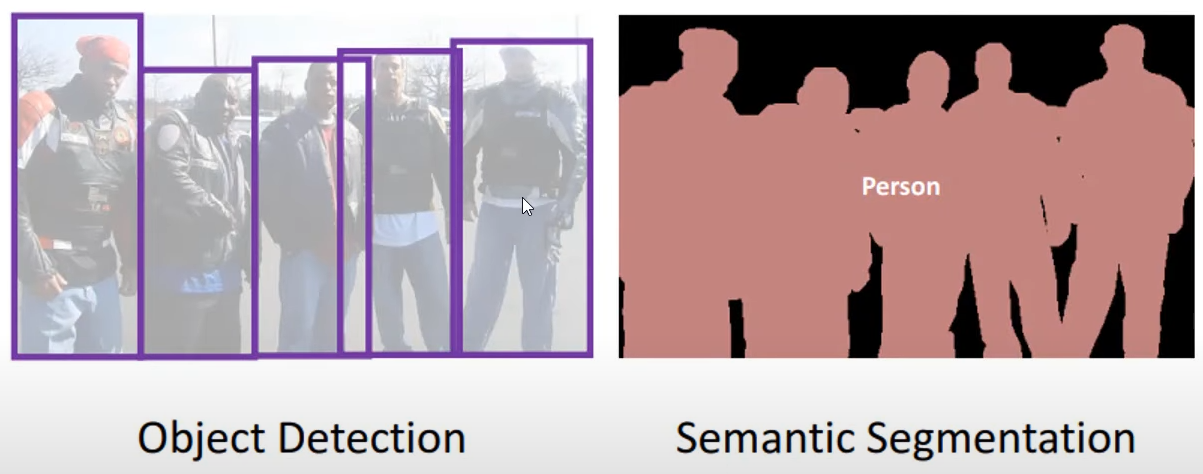
U-Net for Semantic Segmentation

**Sematic Segmentation**

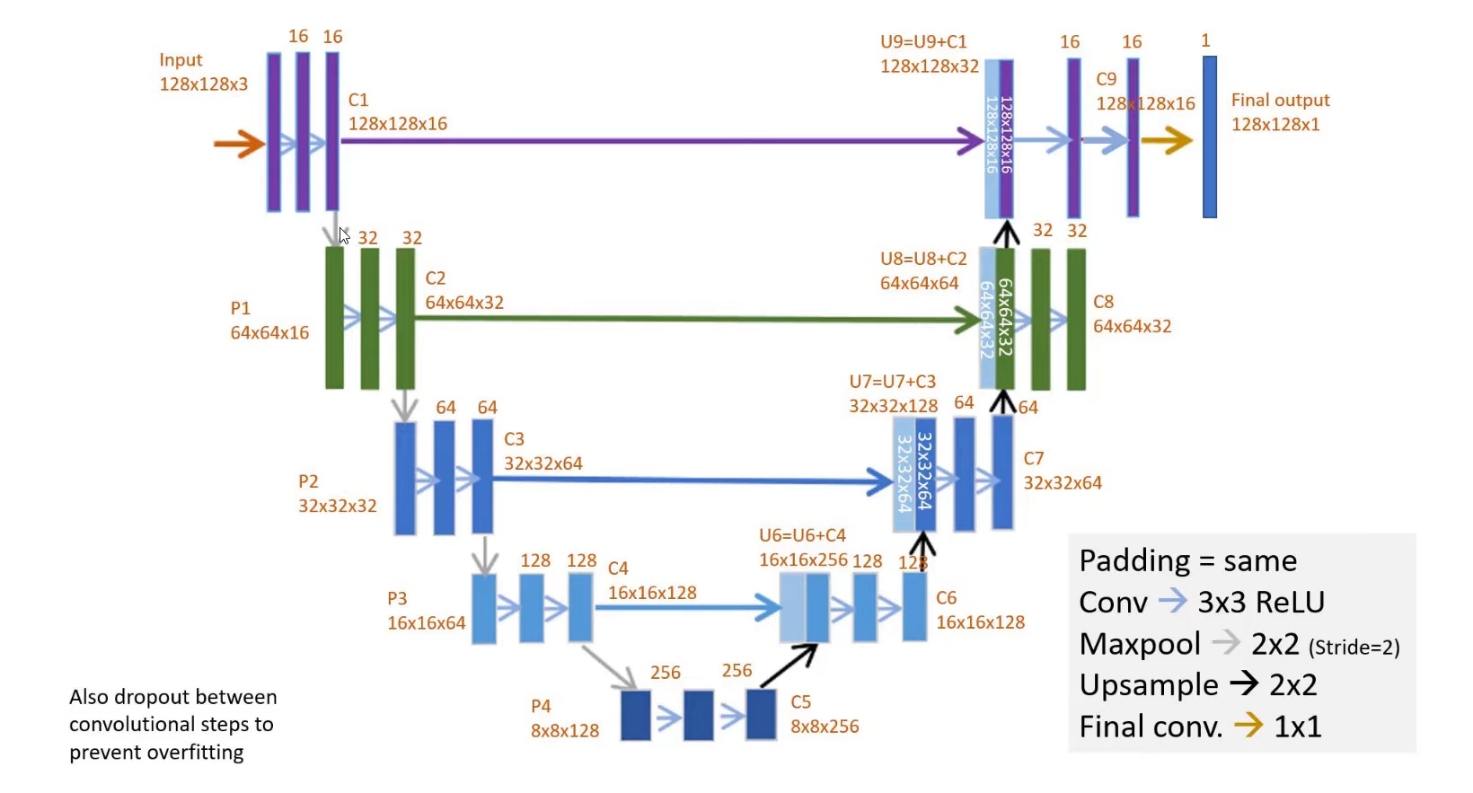
It is a process of partitioning a image into multiple image segments. Let consider an image of group of people in which object detection detects humans in picture while semantic segmentation classifies all humans corresponding to non-human.

Well in classic FCN we required huge amount of dataset to perfectly train our model which also takes long period of time for training. So, researches modify and extends this architecture such that it works with very few trainings dataset and yields more precise segmentations introduces U-Net architecture.

The main idea behind this theory is to supplement a usual contracting network by successive layers, where pooling operators are replaced by upsampling operator. Hence, these layers increase the resolution of the output.

**U-Net Architecture is Design for Semantic Segmentation**

The is derived from its shape like **U.** The left side of the structure is called as contraction/encoder path to capture context and the right side is called expansion/decoder path that enable precise localization.

It starting with the input of coloured image on which we perform 3x3 convolutional operations with features, each followed by a rectified linear unit (ReLU). New step is to 2x2 MaxPool (creating less size metric, with max values of part). It is done multiple times. At each downsampling step we double the number of feature channel.

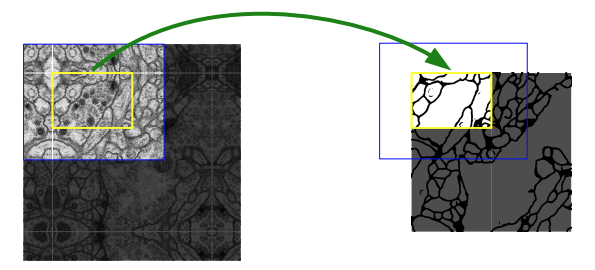
***Example for 8x8 pixels in the lowest resolution***

So, now we have to do upsample with similar parameters we used before. The only difference will now that a concatenation with the correspondingly cropped feature map from the contracting path, and two 3x3 convolutions, each followed by a ReLU. The cropping is necessary due to the loss of border pixels in every convolution, this will continue number of times we perform encoding.

Finally, the output will be the convolutional layer with one feature. In between of the convolutional layer we must drop out some pixels to avoid overfitting.

One important modification in our architecture is that in the upsampling part this have also a large number of feature channel, which allow the network to propagate context information to higher resolution layers. As a consequence, the expansion path is symmetric to the contracting path and yields U-shaped architecture.

The segmentation map only contains the pixels, for which the full context is available in the input image. This strategy allows the seamless segmentation of arbitrarily large image by an overlap-tile strategy.



***Overlap-tile strategy for seamless segmentation of arbitrary large images.***

Prediction of the segmentation in the yellow area, requires image data within the blue area as input. Missing input data is extrapolated by mirroring