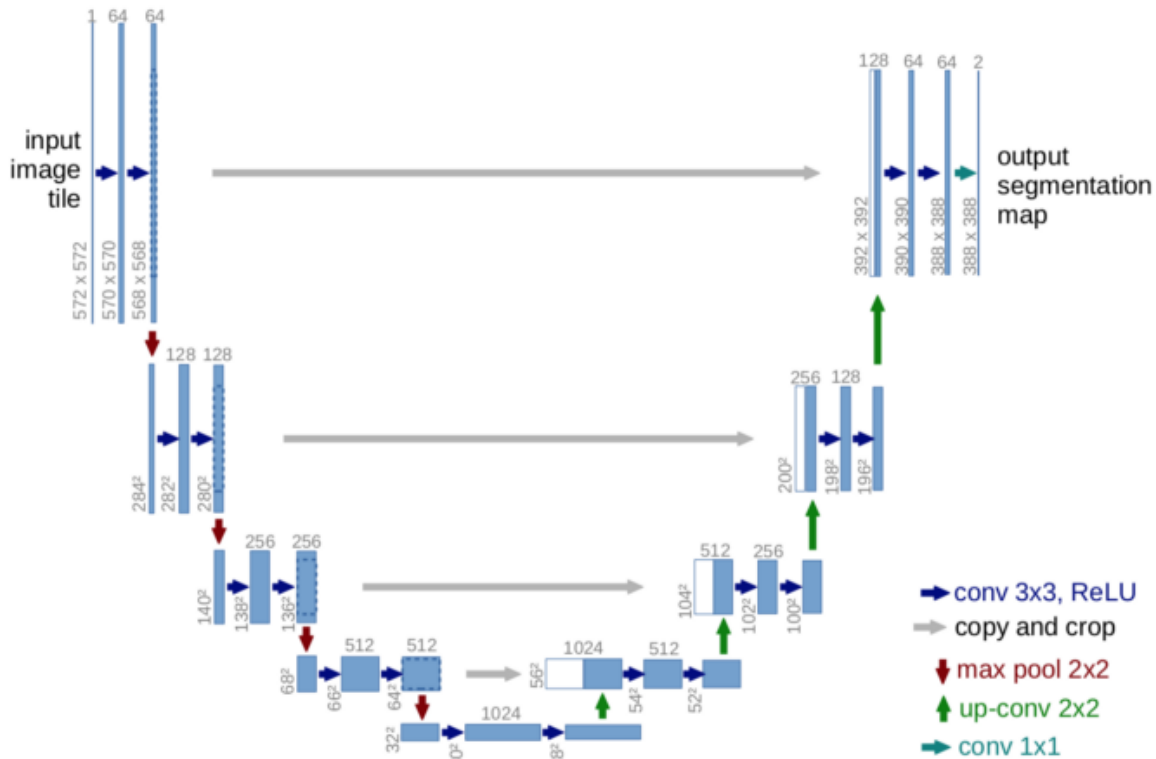


UNet



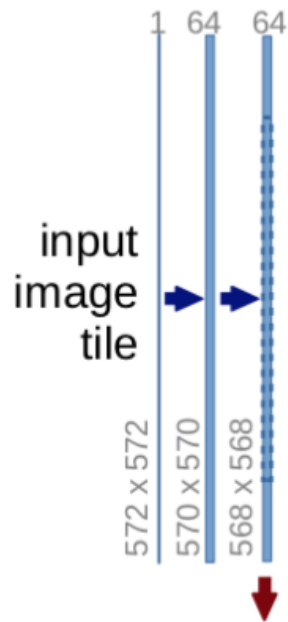
It has a “U” shape. The architecture is symmetric and consists of two major parts — the left part is called contracting path, which is constituted by the general convolutional process; the right part is expansive path, which is constituted by transposed 2d convolutional layers (upsampling).

- **Contracting path**

Conv_layer1 -> conv_layer2 -> max_pooling -> dropout (optional)

Each process constitutes two convolutional layers, and the number of channels changes from 1 → 64, as the convolution process will increase the depth of the image. The red arrow pointing down is the max pooling process which halves the size of the image.

This process is repeated 3 times.



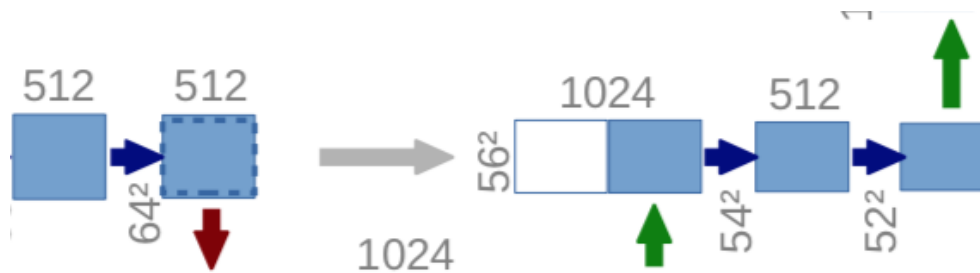
Bottom-most part- 2 convolution layers are built without max-pooling.



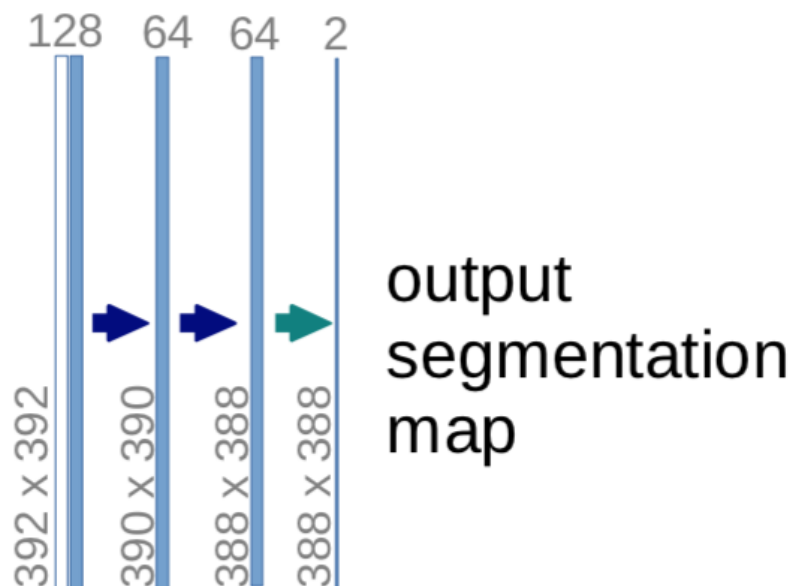
- Expanding part

Transposed convolution is an up-sampling technique that expands the size of images. Basically, it does some padding on the original image followed by a convolution operation.

After the transposed convolution, the image is up-sized from and then, this image is concatenated with the corresponding image from the contracting path. The reason here is to combine the information from the previous layers in order to get a more precise prediction. Same as before, this process is repeated 3 more times.



Uppermost of the architecture- the last step is to reshape the image to satisfy our prediction requirements.



The last layer is a convolution layer with 1 filter of size 1x1.

UNet is able to do image localisation by predicting the image pixel by pixel.