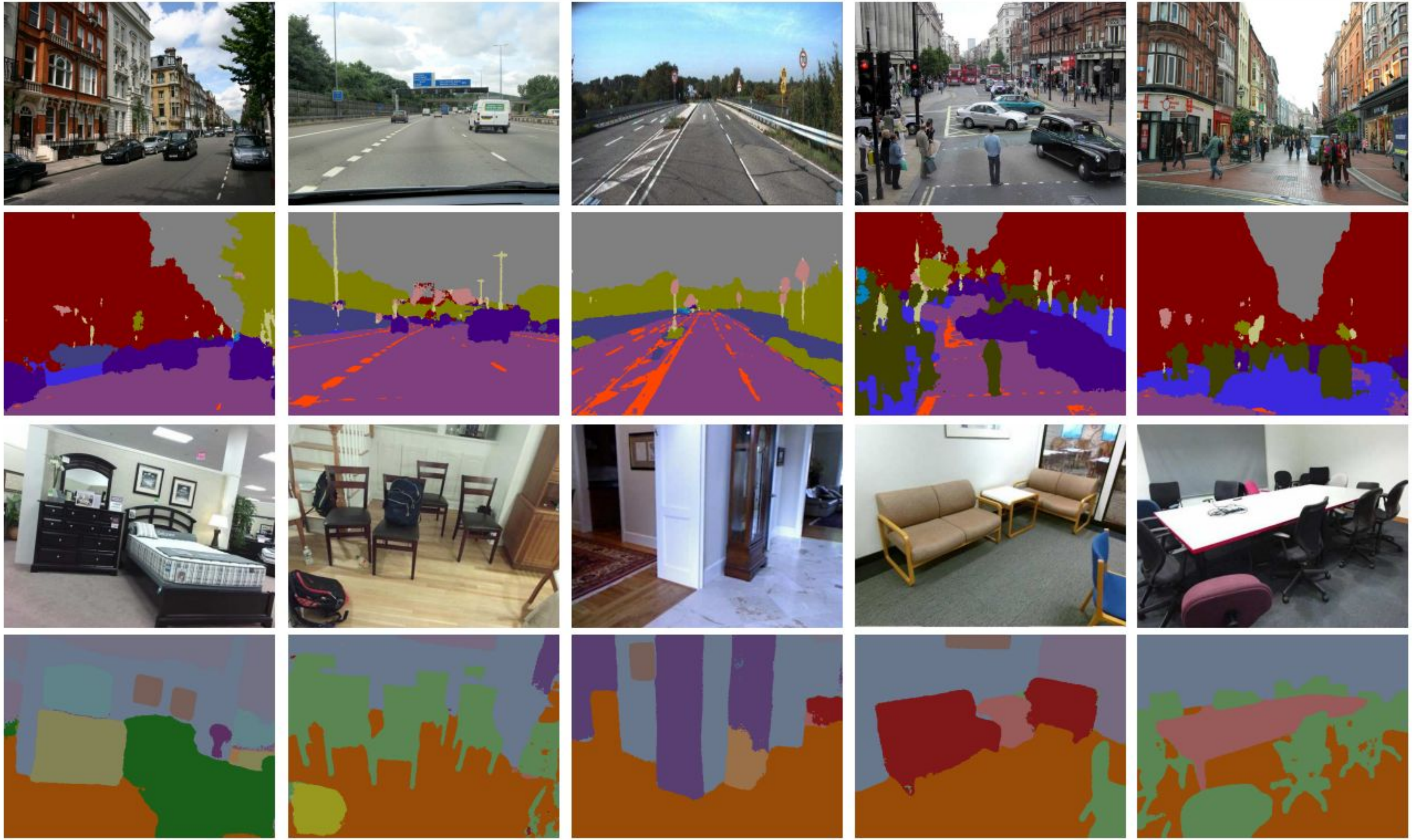
Image Segmentation - UNet

B.Tech. Data Science, NMIMS

By,

Bilal Hungund, Data Scientist, Halliburton



https://arxiv.org/pdf/1511.00561v3.pdf

About UNet

The U-Net is a convolutional neural network architecture designed for image segmentation tasks, specifically for biomedical image segmentation. The architecture of the U-Net consists of an encoder and decoder network with skip connections between corresponding layers.

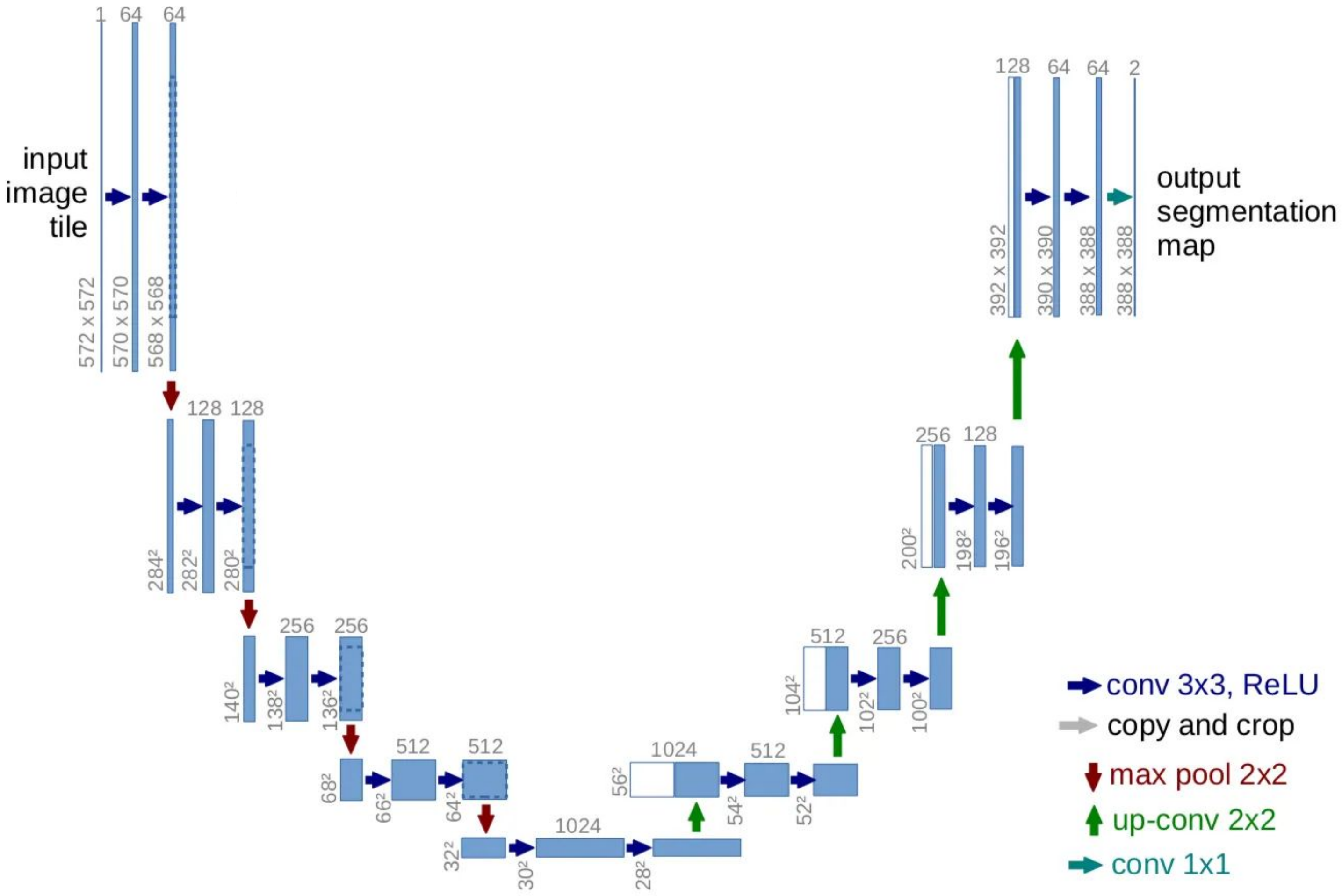
The **encoder** network of the U-Net performs feature extraction from the input image, similar to other convolutional neural networks. The difference in the U-Net is the use of max pooling layers to downsample the feature maps, which reduces the spatial dimensions and increases the receptive field of the network.

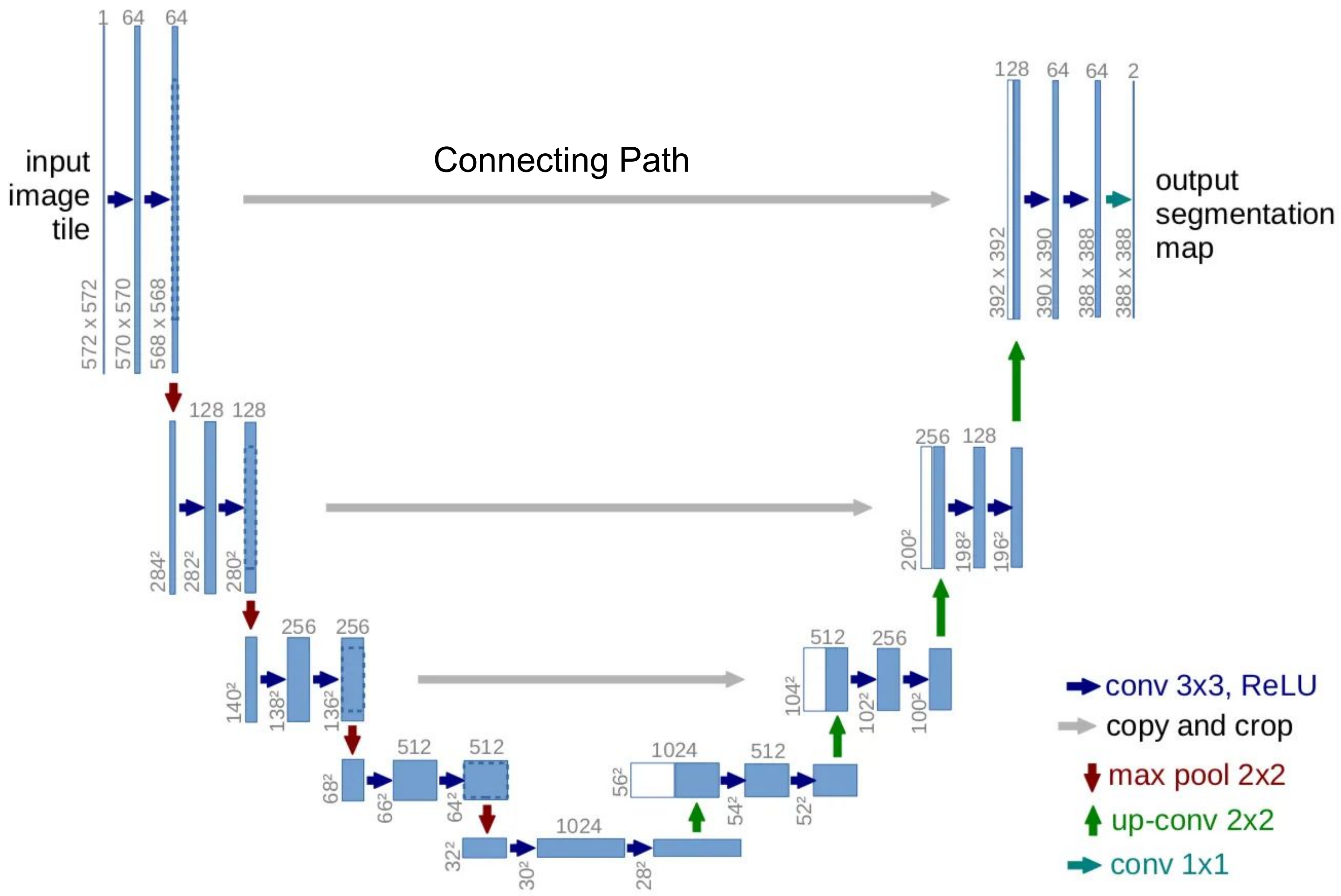
The **decoder** network of the U-Net performs upsampling of the feature maps to generate the final segmentation map. This is done using transposed convolution layers, which perform the reverse operation of a regular convolution by increasing the spatial dimensions of the feature maps. Skip connections are then used to concatenate the feature maps from the corresponding layer in the encoder network to the upsampled feature maps in the decoder network. This helps to preserve the spatial information lost during the downsampling process.

The U-Net also uses custom convolutional blocks in both the encoder and decoder networks, which consist of two convolutional layers followed by batch normalization and ReLU activation. This helps to improve the performance of the network by reducing the vanishing gradient problem and accelerating the training process.

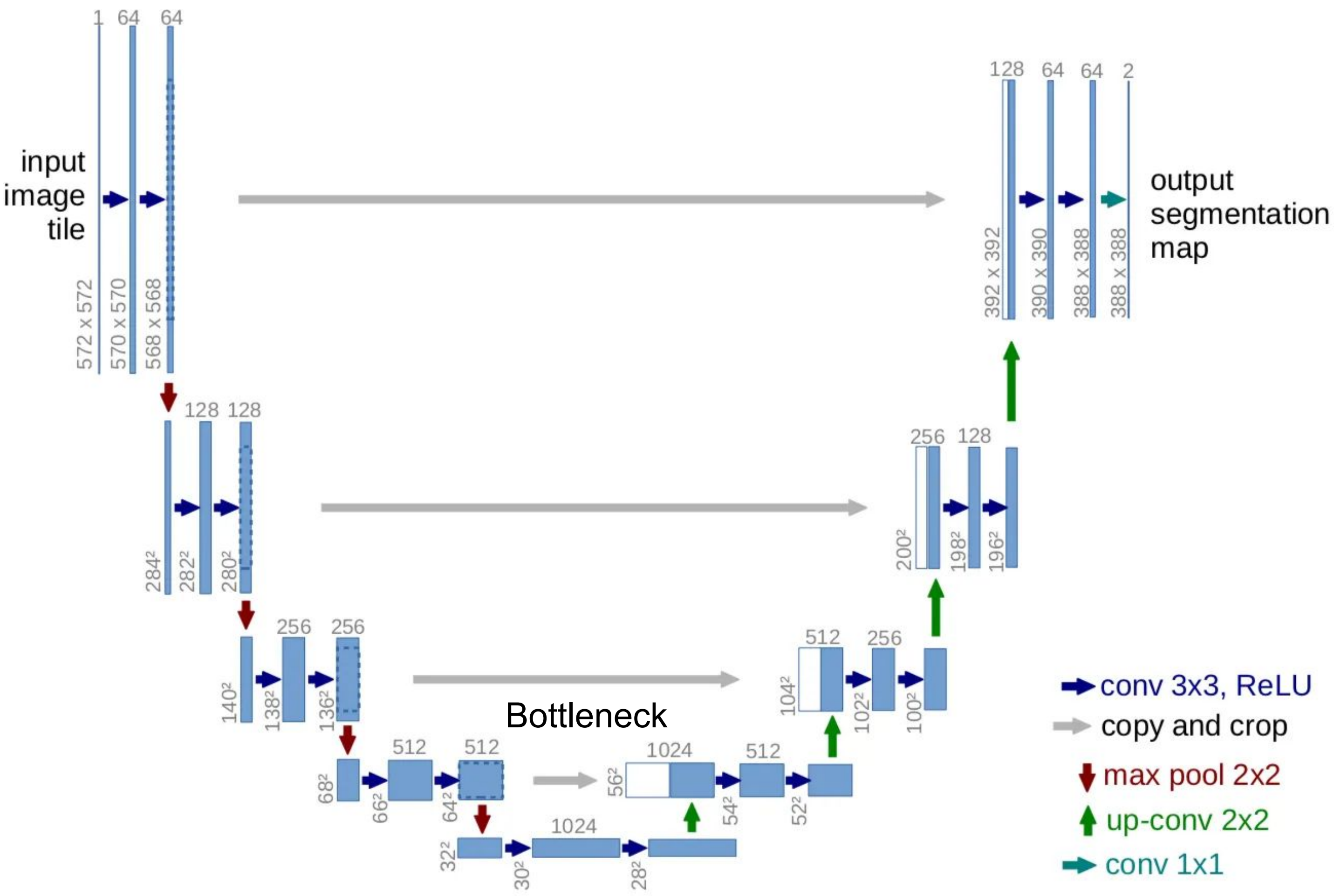
Finally, the output of the U-Net is a single-channel segmentation map, where each pixel value represents the probability of that pixel belonging to the foreground or background class. The sigmoid activation function is used in the final layer to ensure that the output values are between 0 and 1.

https://doi.org/10.48550/arXiv.1505.04597 https://doi.org/10.48550/arXiv.1505.04597

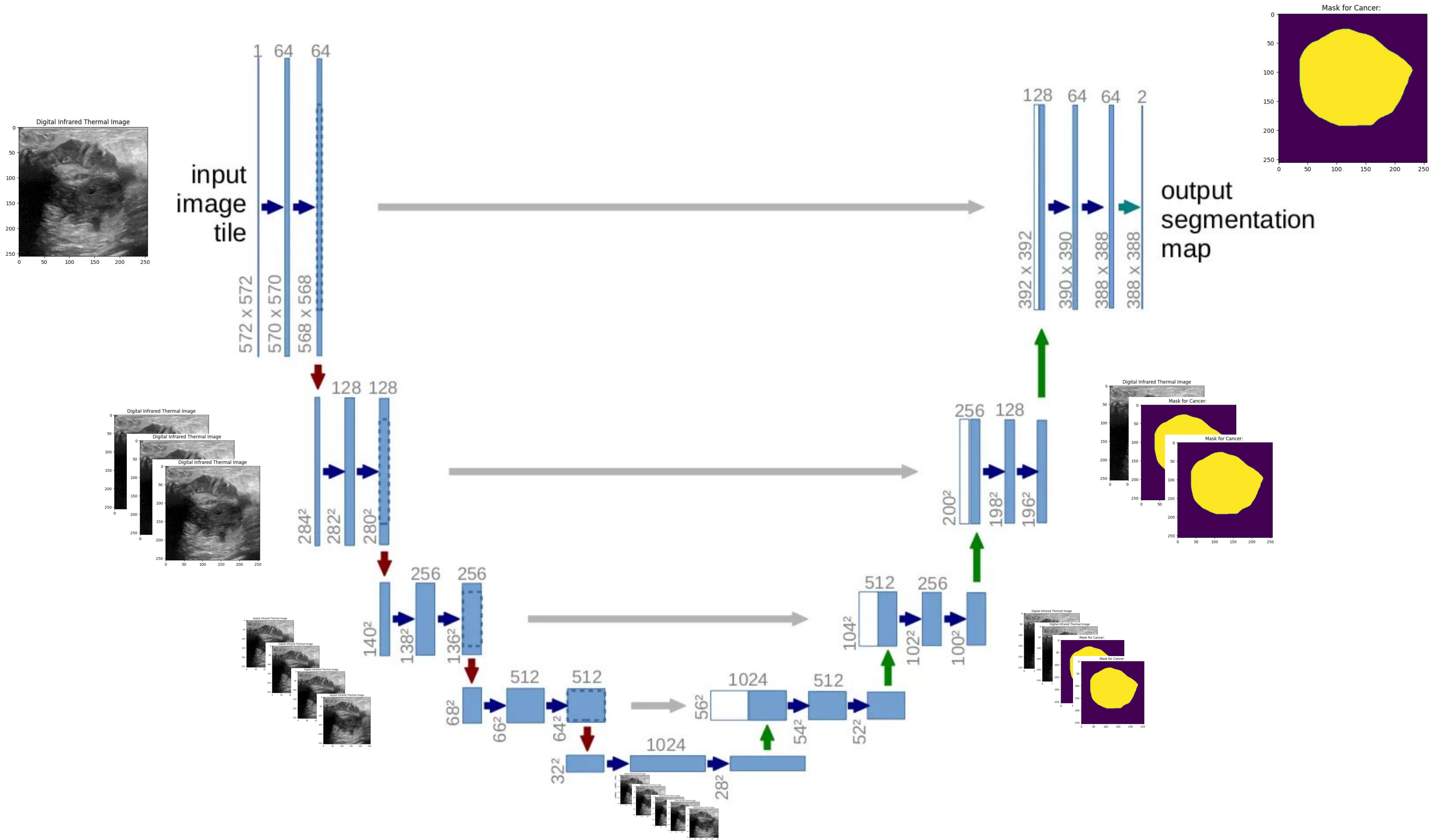




Connecting Path
input
image
tile
output
segmentation
map
1
64
64
572 x 572
570 x 570
568 x 568
128
128
284²
282²
280²
256
256
140²
138²
136²
512
512
68²
66²
64²
1024
32²
30²
28²
1024
56²
512
54²
52²
512
256
104²
102²
100²
256
128
200²
198²
196²
128
64
64
2
392 x 392
390 x 390
388 x 388
388 x 388
conv 3x3, ReLU
copy and crop
max pool 2x2
up-conv 2x2
conv 1x1



input image tile
output segmentation map
Bottleneck
1
64
64
572 x 572
570 x 570
568 x 568
128
128
284²
282²
280²
256
256
140²
138²
136²
512
512
68²
66²
64²
1024
32²
30²
28²
1024
56²
512
54²
52²
512
256
104²
102²
100²
256
128
200²
198²
196²
128
64
64
2
392 x 392
390 x 390
388 x 388
388 x 388
conv 3x3, ReLU
copy and crop
max pool 2x2
up-conv 2x2
conv 1x1



Digital Infrared Thermal Image
input
image
tile
1
64
64
572 x 572
570 x 570
568 x 568
128
128
284²
282²
280²
256
256
140²
138²
136²
512
512
68²
66²
64²
1024
32²
28²
1024
512
56²
54²
52²
512
256
104²
102²
100²
256
128
200²
198²
196²
128
64
64
2
392 x 392
390 x 390
388 x 388
388 x 388
output
segmentation
map
Mask for Cancer:

IoU Intersection of Union

Overlap between prediction and ground truth (union)