SURGICAL VIDEO ANALYSIS

SEMANTIC SEGMENTATION

Jahnavi Challa(IMT2020103)

Dataset: cholecseg8k

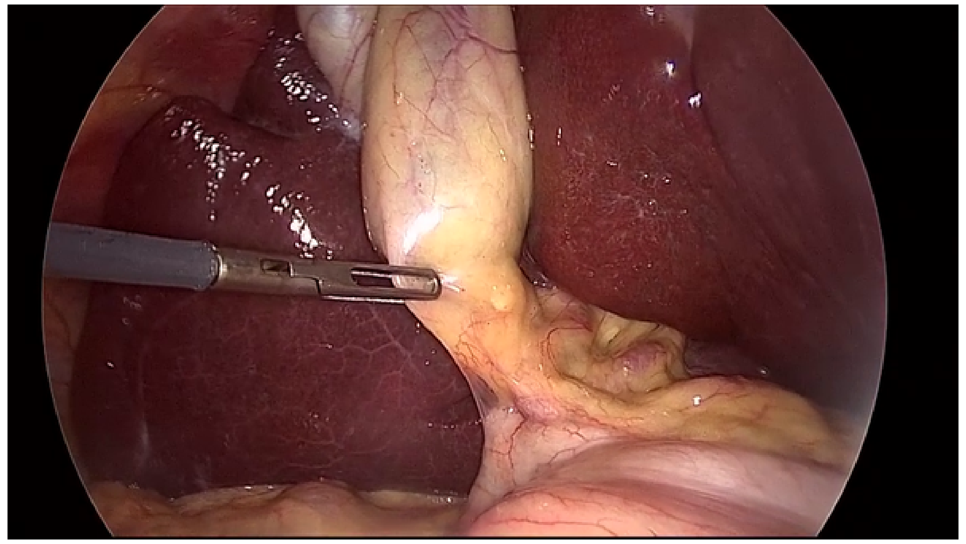
contains:

Original image
Semantic segmentation of that image
Water shed Masking
Masking

SEMANTIC SEGMENTATION

→ Semantic segmentation is the process of classifying each pixel of an image into a category or class.

OUTPUT



Class Number	Class Name	RGB hexcode
Class 0	Black Background	#505050
Class 1	Abdominal Wall	#111111
Class 2	Liver	#212121
Class 3	Gastrointestinal Tract	#131313
Class 4	Fat	#121212
Class 5	Grasper	#313131
Class 6	Connective Tissue	#232323
Class 7	Blood	#242424
Class 8	Cystic Duct	#252525
Class 9	L-hook Electrocautery	#323232

Steps We Follow During SEMANTIC SEGMENTATION:

01 Data collection

Gather a dataset of images with corresponding pixel-level annotations that indicate the class of each pixel.

02 Preprocessing

Preprocess the images by resizing them to a fixed size, normalizing pixel values and Augmenting the dataset

03 Model selection

we choose suitable architecture(U-Net) to learn a mapping between image pixels and class labels.

04 Training:

Train the model on the dataset using an appropriate loss function. Monitor the training process using metrics such as accuracy and loss.

05 Validation

Evaluate the model on a validation dataset to check its generalization performance.

06 Testing

Apply the trained model to new images for semantic segmentation.

Semantic Segmentation

- In semantic segmentation tasks, it is important to capture both local and global spatial information in order to accurately identify and label different objects within the image.

What is Spatial information in images?

- Local Spatial information.
- Global Spatial information.

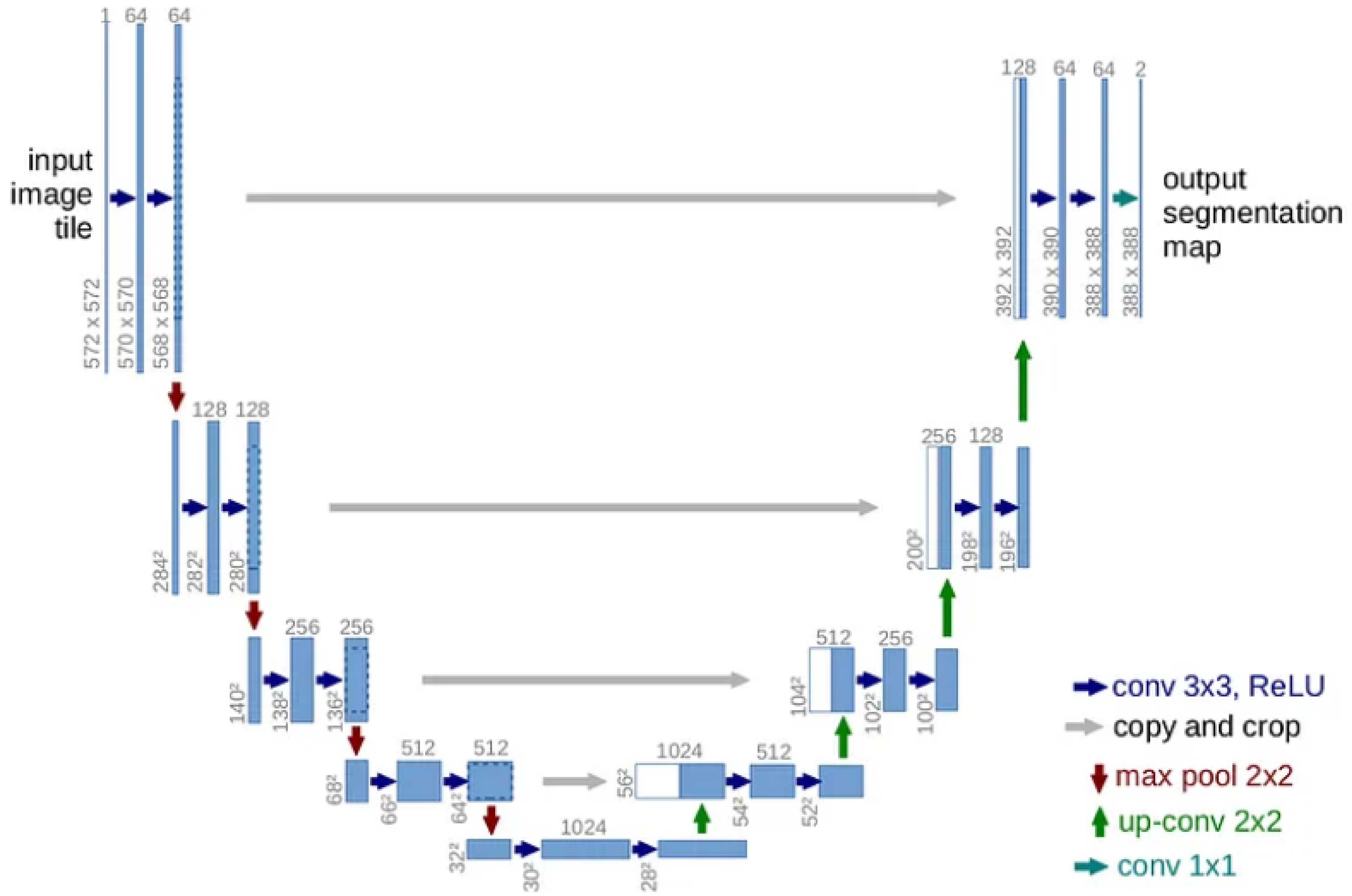
Why U-NET???

Brief Overview of U-NET

- → It is a modified version of the standard CNN architecture

- → U-Net has a symmetric encoder-decoder architecture that enables it to capture both low-level and high-level features of an image. The encoder part of the network captures the context of the input image, while the decoder part of the network reconstructs the segmented image.

- → U-Net uses skip connections between the encoder and decoder layers. These connections help to preserve the spatial information lost during the down-sampling operation of the encoder.



UNet architecture

Detailed Overview

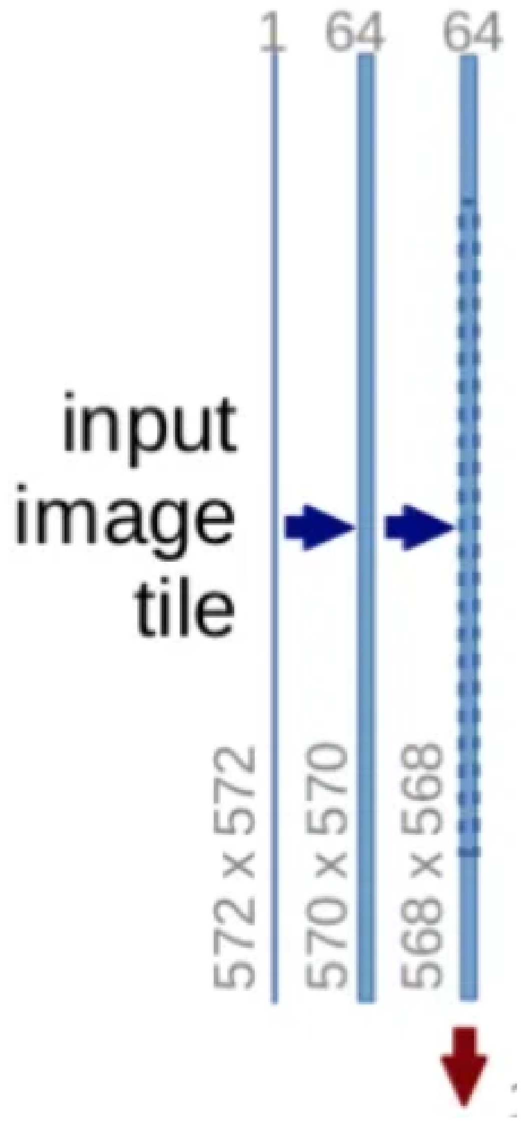
The architecture contains two paths.

1) First path is the contraction path (also called as the encoder) which is used to capture the context in the image. The encoder is just a traditional stack of convolutional and max pooling layers.

2) The second path is the symmetric expanding path (also called as the decoder) which is used to enable precise localization using transposed convolutions.

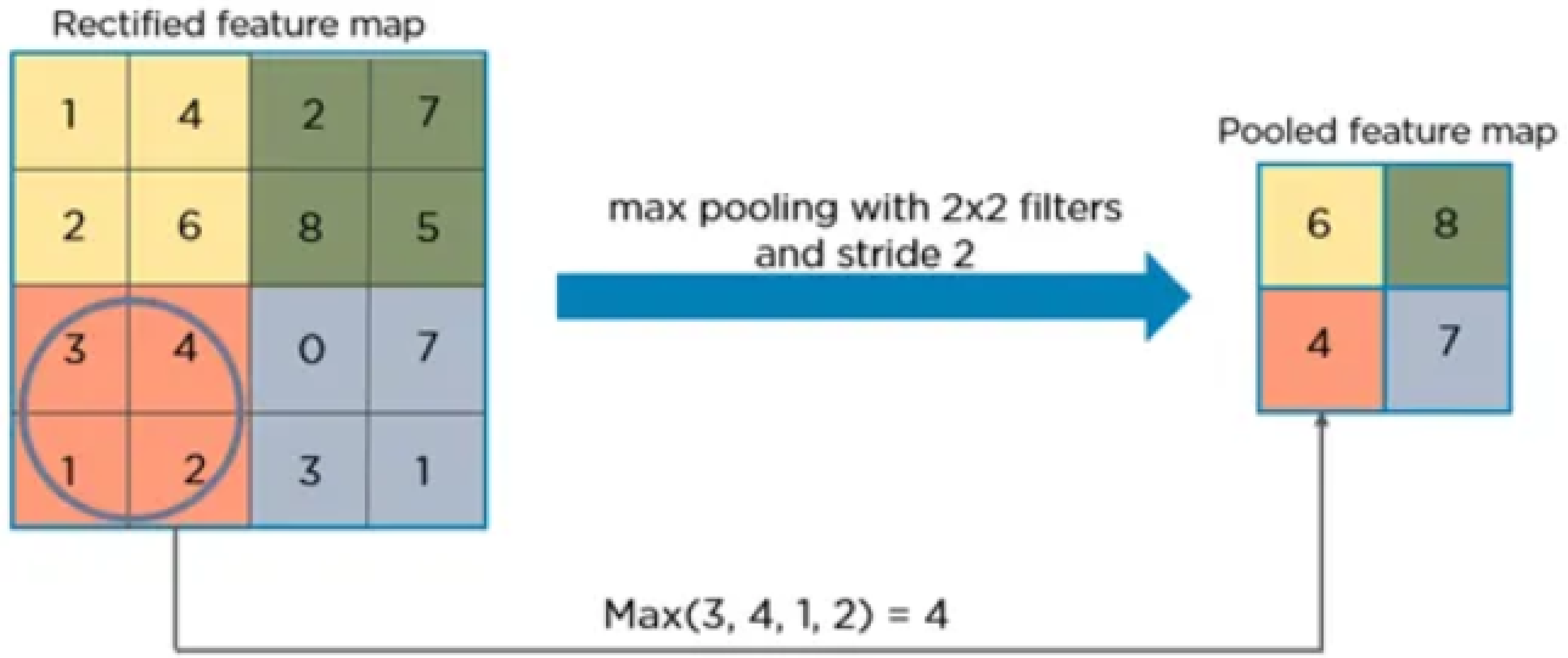
Contracting Path

The contracting path follows the formula:
conv_layer1 -> conv_layer2 -> max_pooling -> dropout(optional)

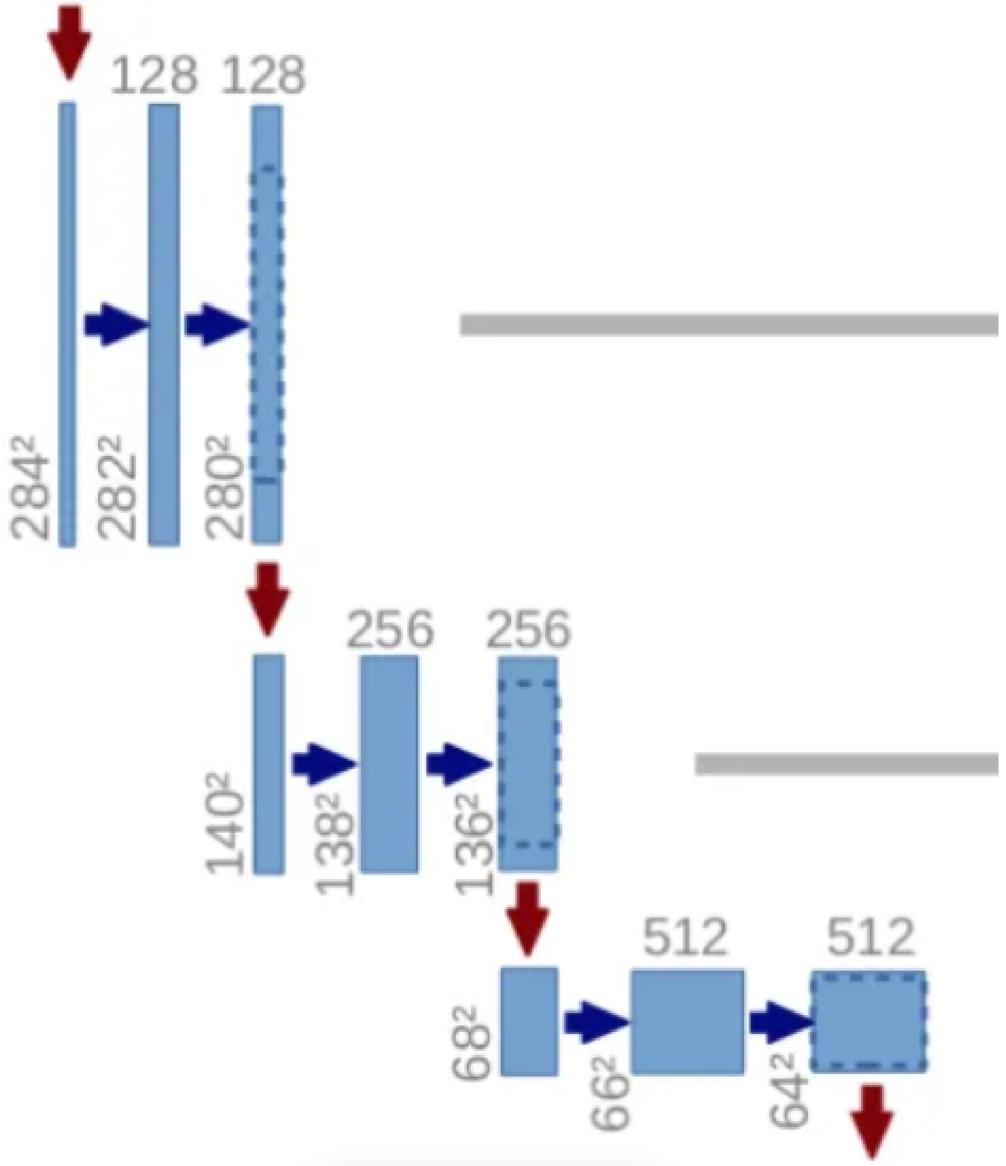


MAX POOLING

Max pooling is typically applied after a convolutional layer to reduce the spatial dimensions of the feature map while preserving the most important information.



Process is continued 3 times



Last layer

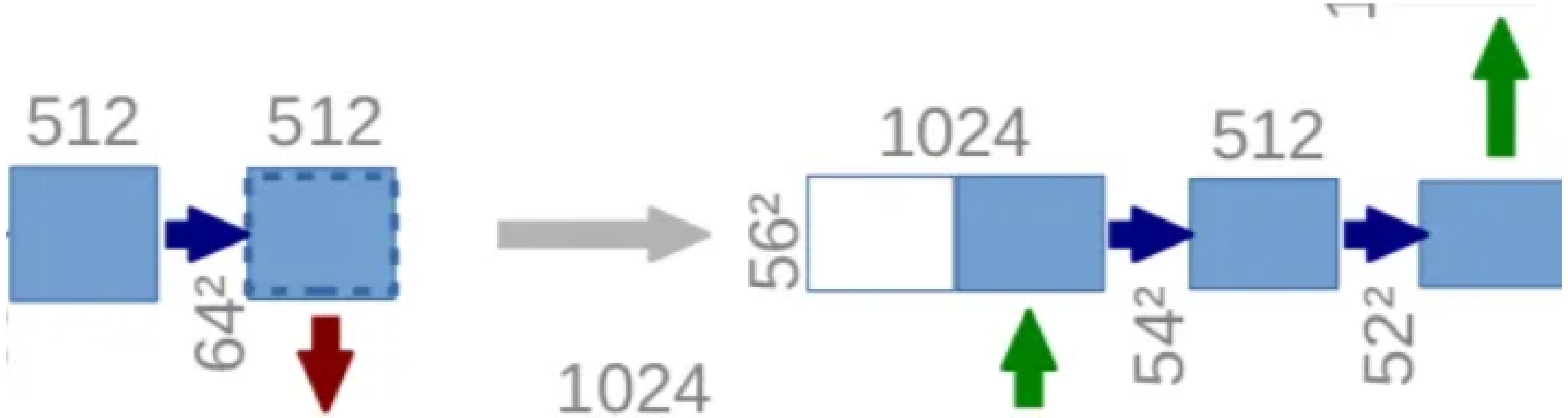
- By down sampling, the model better understands "WHAT" is present in the image, but it loses the information of "WHERE" it is present.

→ In case of segmentation we need both "WHAT" as well as "WHERE" information.

- Hence there is a need to up sample the image, i.e. convert a low resolution image to a high resolution image to recover the "WHERE" information.

Expansive Path

```
conv_2d_transpose -> concatenate -> conv_layer1 -> conv_layer2
```



Decoder recovers the “WHERE” information (precise localization) by gradually applying up-sampling

- To get better precise locations, at every step of the decoder we use skip connections by concatenating the output of the transposed convolution layers with the feature maps from the Encoder at the same level.

- After every concatenation we again apply two consecutive regular convolutions so that the model can learn to assemble a more precise output.

THANK YOU!!!