U-Net: A Brief Exploration

Carl Kashuk

Machine Learning Marathon 2025

Informatics Data Scientist, Radiology

What is U-Net?

- U-Net is a convolutional neural network that was developed for image segmentation, with fewer training images needed, and more precise segmentation

What is U-Net?

- U-Net is a convolutional neural network that was developed for image segmentation, with fewer training images needed, and more precise segmentation

Convolutional Neural Network:

input layer -> hidden layers -> output layer

Some of the hidden layers are convolutions:

What is U-Net?

- U-Net is a convolutional neural network that was developed for image segmentation, with fewer training images needed, and more precise segmentation

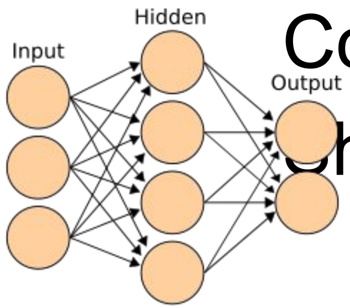
Convolutional Neural Network:

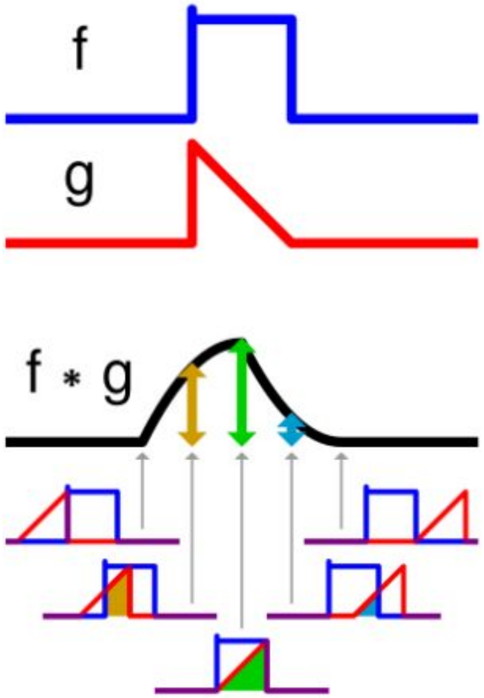
input layer -> hidden layers -> output layer

Some of the hidden layers are convolutions:

Convolution: functions f, g -> f * g

Shows how the shape of one function is modified by the other --- a filter!





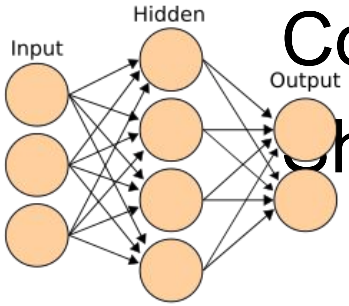
What is U-Net?

- U-Net is a convolutional neural network that was developed for image segmentation, with fewer training images needed, and more precise segmentation

Convolutional Neural Network:

input layer -> hidden layers -> output layer

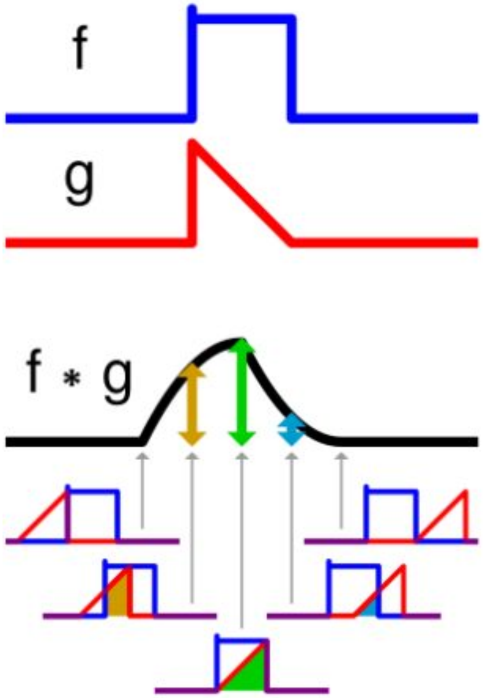
Some of the hidden layers are convolutions:



Convolution: functions f, g -> f * g

Shows how the shape of one function is modified by the other --- a filter!

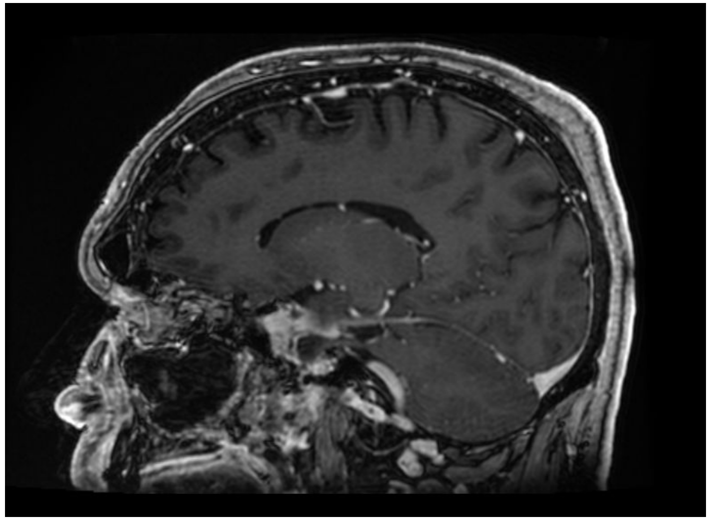
Compares to visual info in the brain!



What is U-Net?

- U-Net is a convolutional neural network that was developed for image segmentation

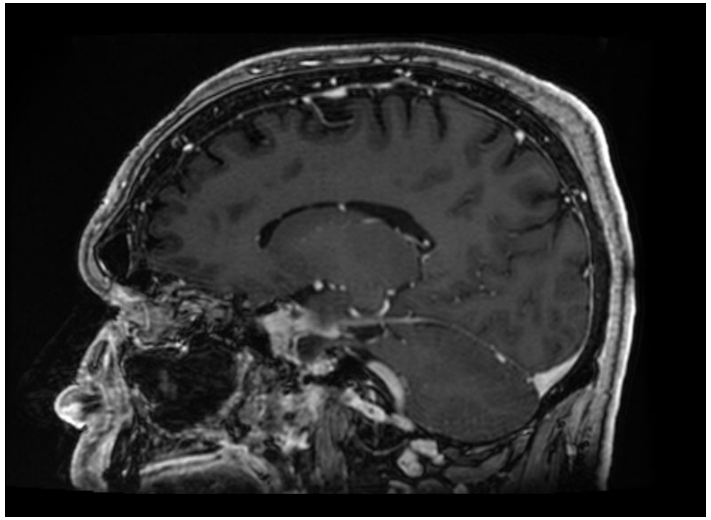
Image Segmentation: Finding regions of interest in an image:

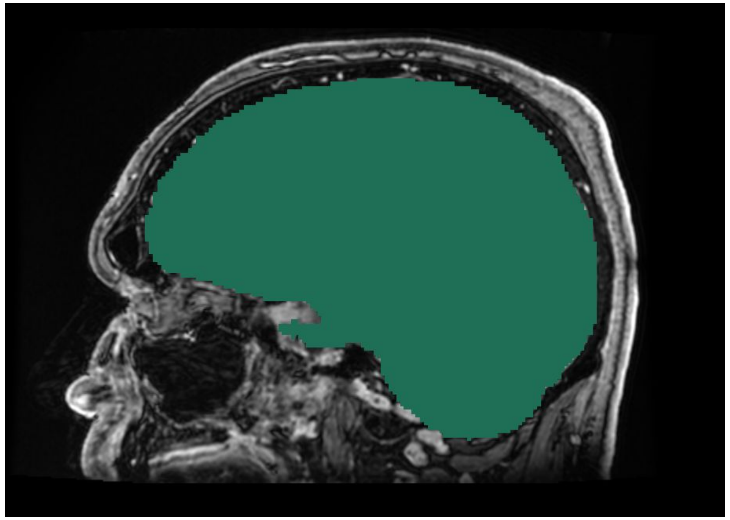


What is U-Net?

- U-Net is a convolutional neural network that was developed for image segmentation

Image Segmentation: Finding regions of interest in an image:

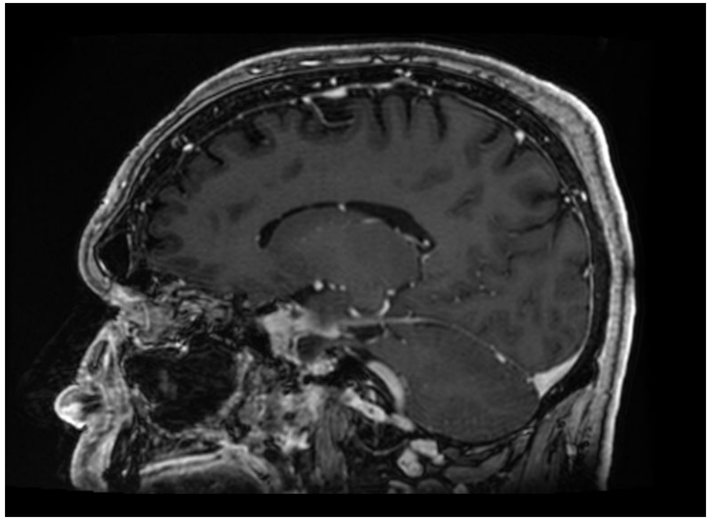


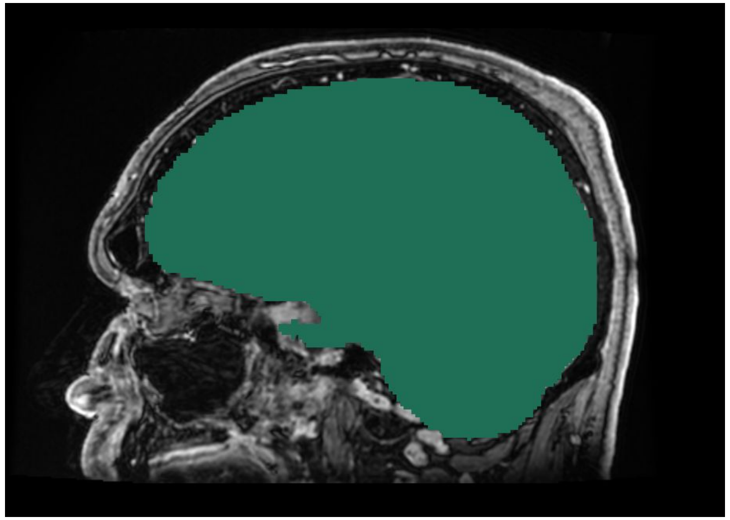


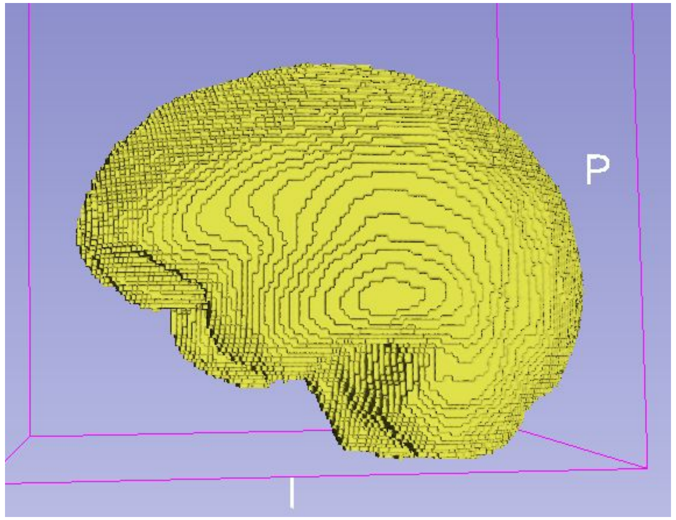
What is U-Net?

- U-Net is a convolutional neural network that was developed for image segmentation

Image Segmentation: Finding regions of interest in an image:







U-Net Reference

U-Net: Convolutional Networks for Biomedical Image Segmentation

Olaf Ronneberger, Philipp Fischer, and Thomas Brox

Computer Science Department and BIOSS Centre for Biological Signalling Studies,
University of Freiburg, Germany
ronneber@informatik.uni-freiburg.de,
WWW home page: http://lmb.informatik.uni-freiburg.de/

U-Net Reference

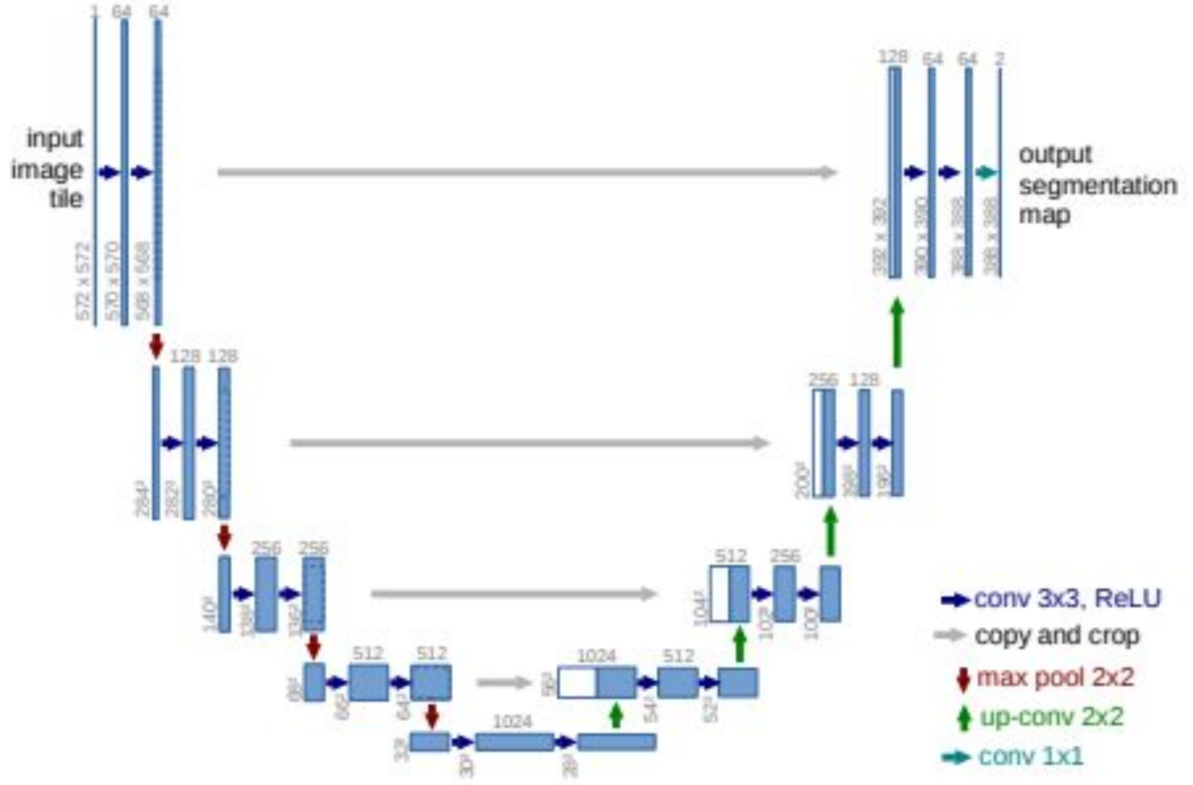
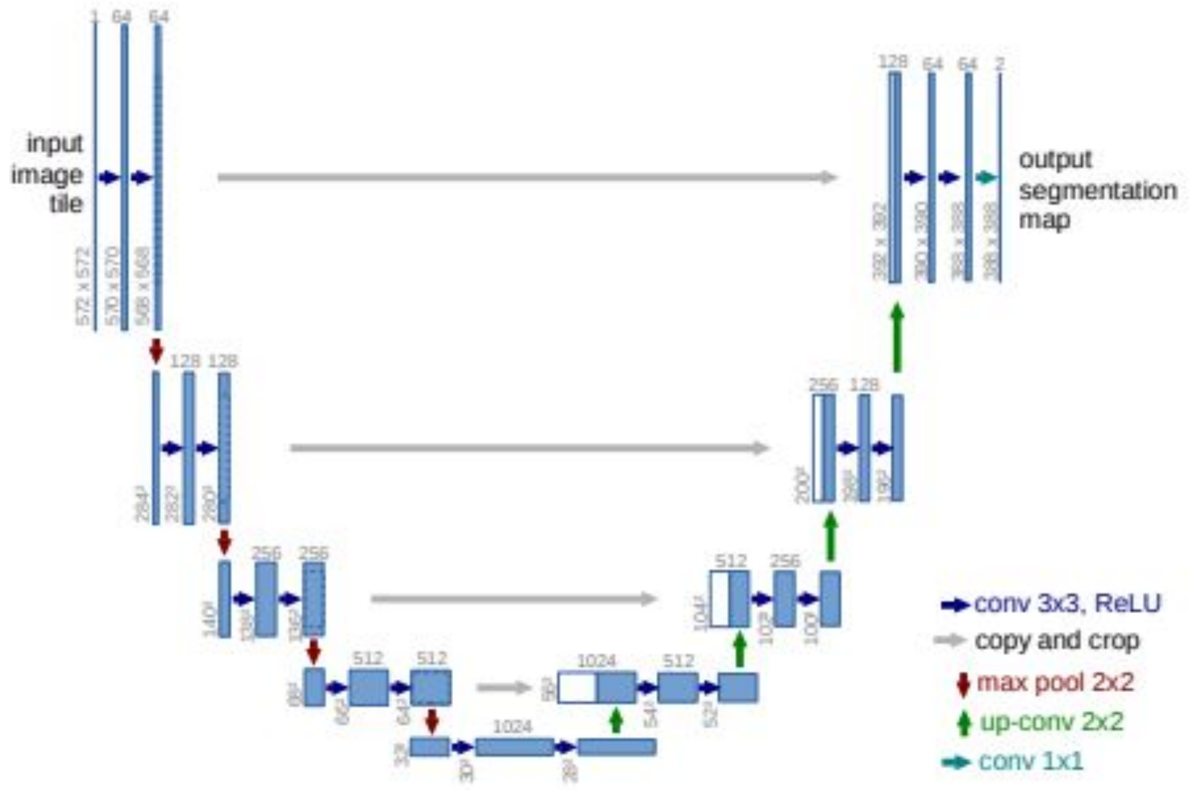


Fig. 1. U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations.

U-Net Reference

Downsampling



Upsampling

Fig. 1. U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations.

U-Net Reference

Skip Connections

Downsampling

Upsampling

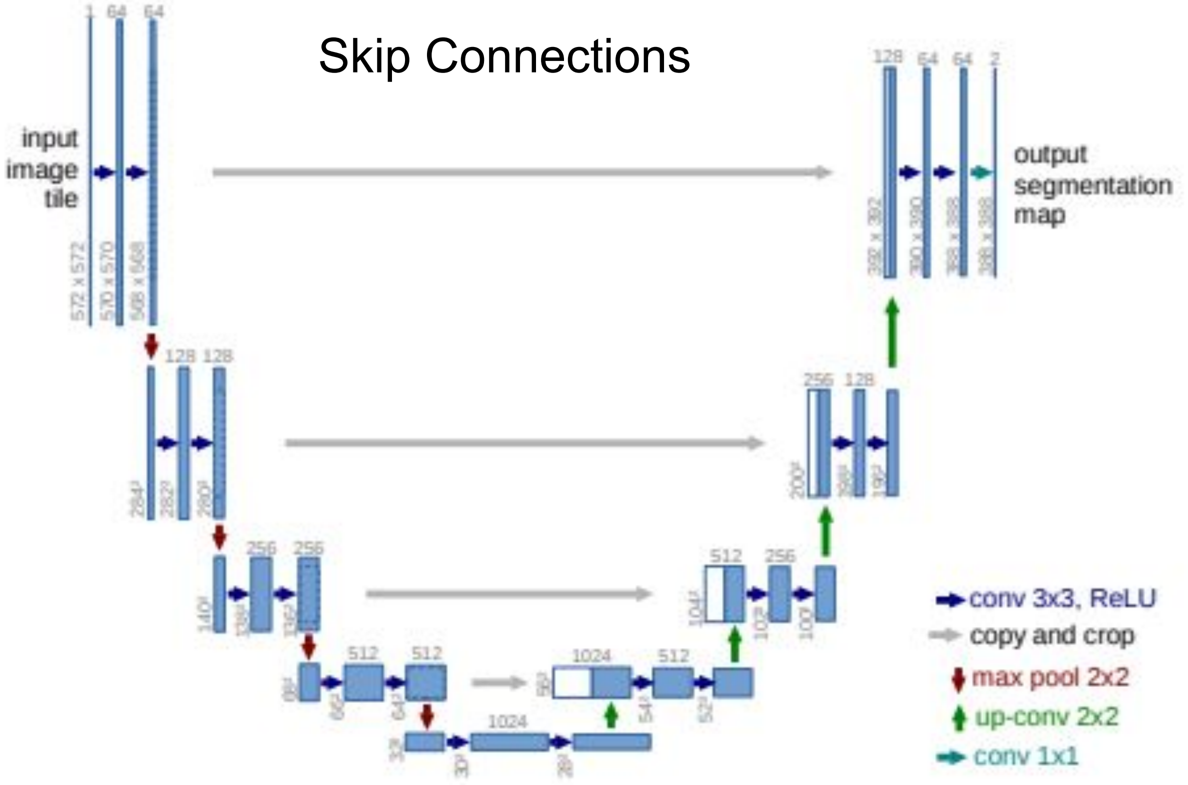
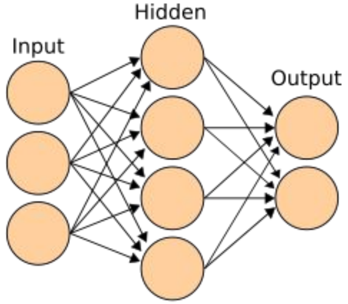
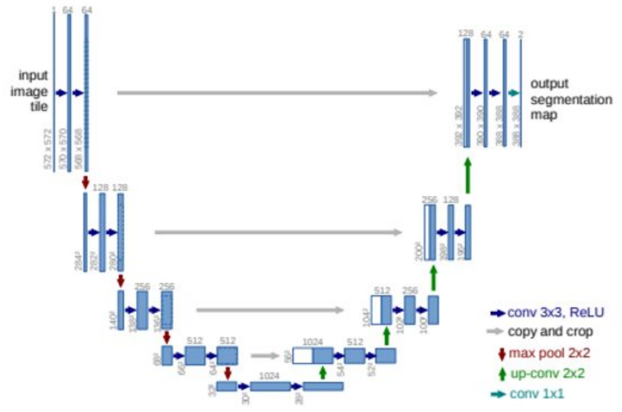


Fig. 1. U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations.



U-Net Architecture



The architecture is symmetric and has three key parts:

Contracting Path (Encoder):

- Uses small filters (3×3 pixels) to scan the image and find features.
- Apply an activation function called ReLU to add non-linearity help the model to learn better.
- Uses max pooling (2×2 filters) to shrink the image size while keeping important information. This helps the network focus on bigger features.

Bottleneck:

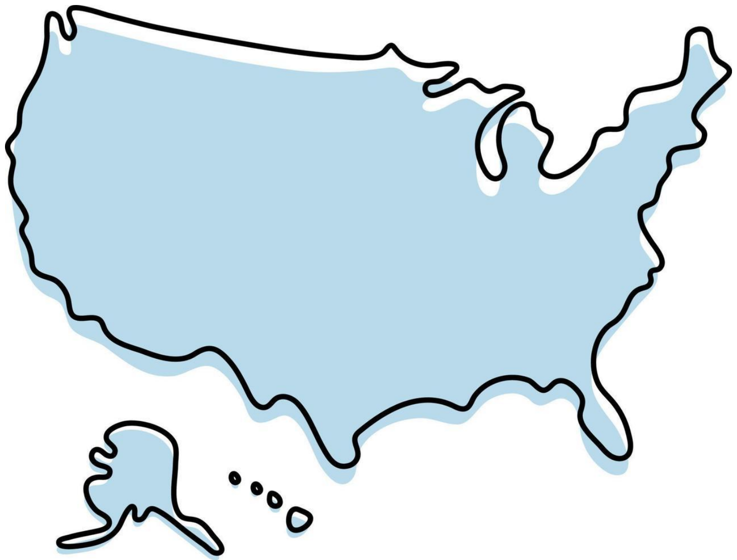
- The middle of the “U” where the most compressed and abstract information is stored. It links the encoder and decoder.

Expansive Path (Decoder):

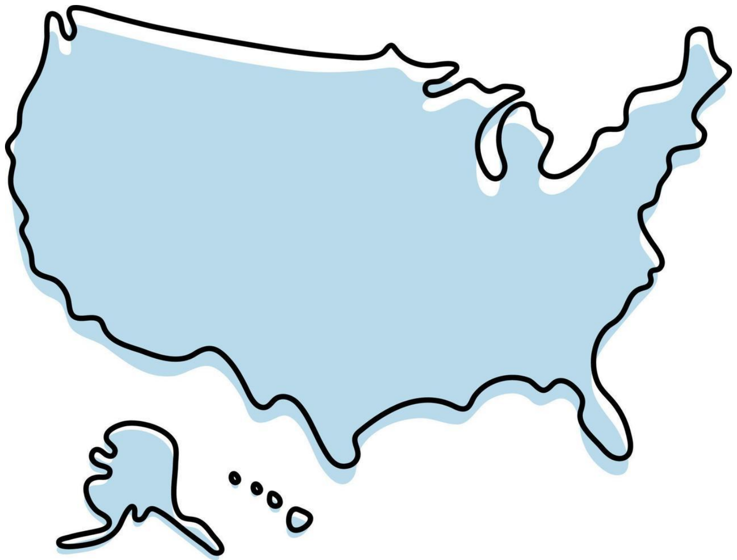
- Uses upsampling i.e. increasing image size to get back the original image size.
- Combines information from the encoder using “skip connections.” These connections help the decoder get spatial details that might have been lost when shrinking the image.
- Uses convolution layers again to clean up and refine the output.

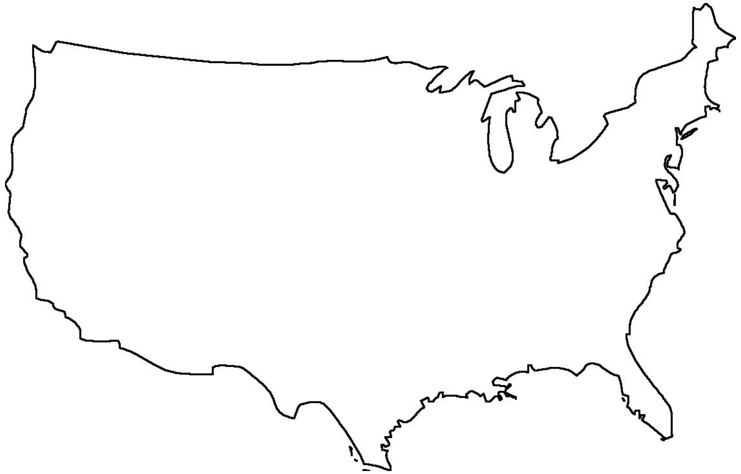
https://www.geeksforgeeks.org/machine-learning/u-net-architecture-explained/

Best of Both Worlds

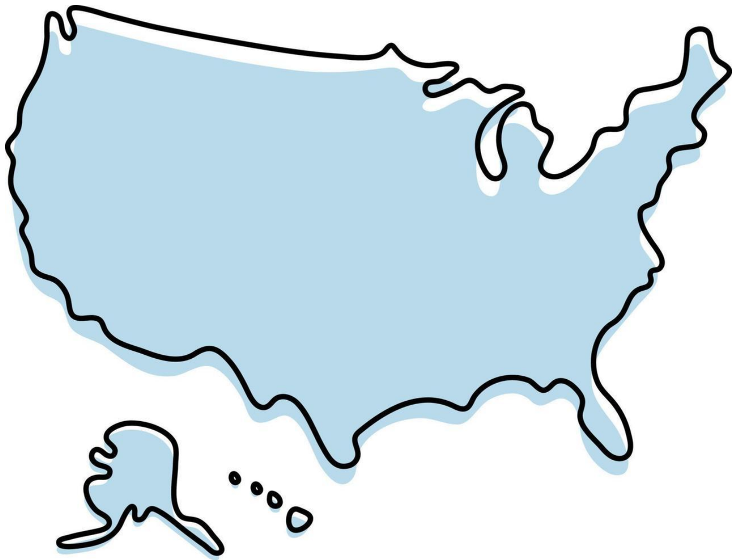
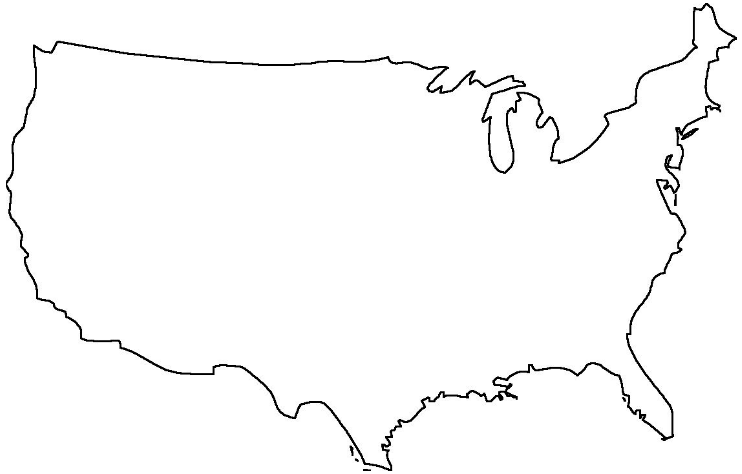


Best of Both Worlds

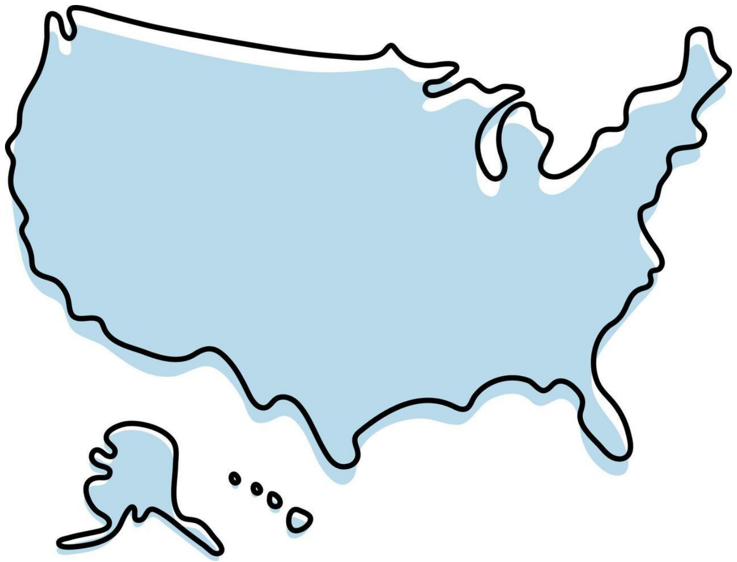


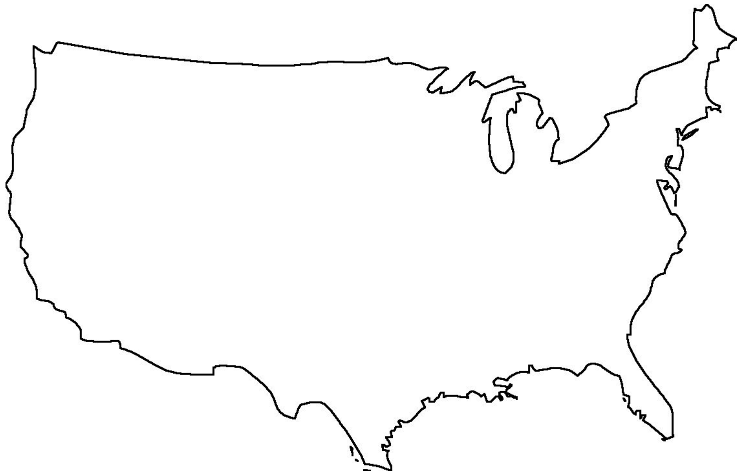


Best of Both Worlds

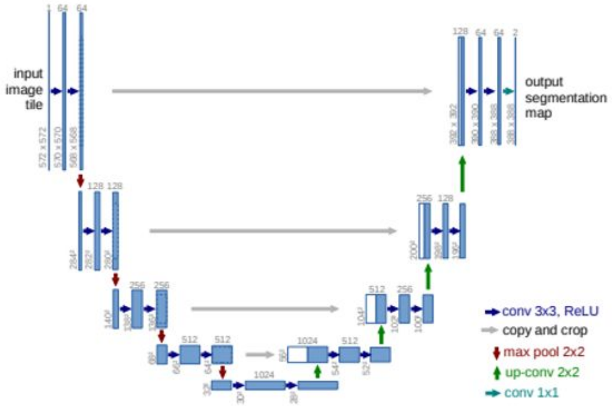




Best of Both Worlds





Modifications to U-Net



- nnU-Net – easily customizable depending on data
  - TotalSegmentator: finds many anatomical structures in CT and MR images

- swinUNet – convolutions replaced by transformers:
  - Transformer is an alternative scheme based on multi-head attention mechanism
  - OSCAR toolkit – segments fat and other soft tissue
  - “radsurv” model – segments areas of brain tumor from head MR series
- W- and V- nets should be self explanatory by now?

- Good recent review of modifications: https://arxiv.org/pdf/2502.06895

Coding U-Net(s)

- Many examples all over the net
- A very good code tutorial exists at:
  https://github.com/JianZhongDev/UNetPyTorchTutorial/tree/main