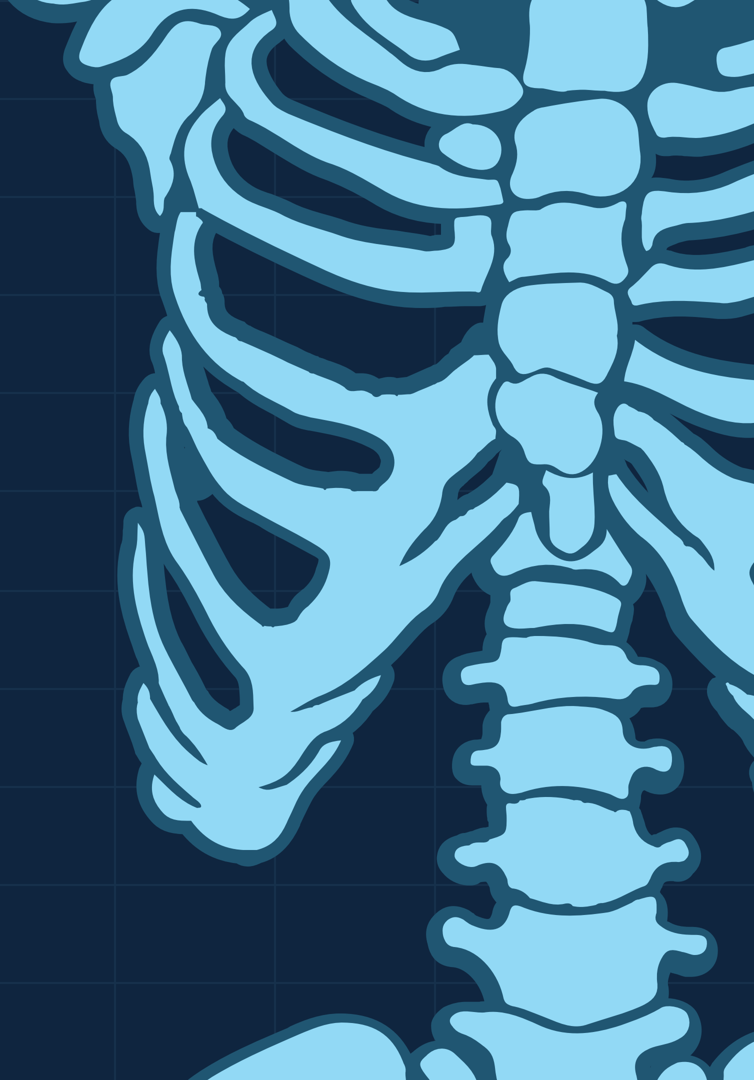
Predictive CT Scan Imaging:
An application of U-Nets

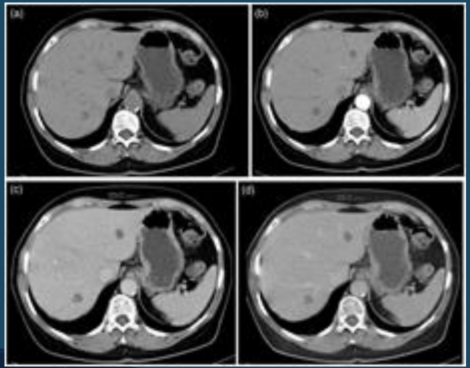
Yilei Li
Mentor: Bin Chen



10cm

Table of contents

10cm



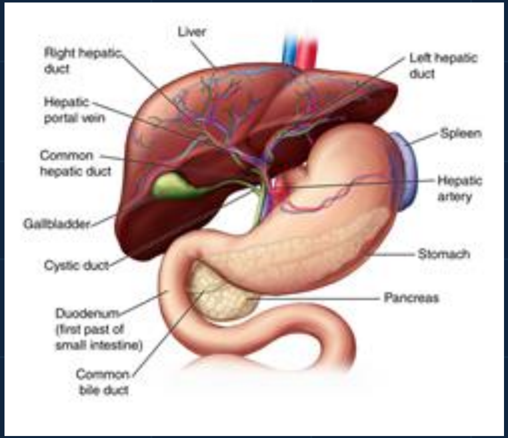
Introduction

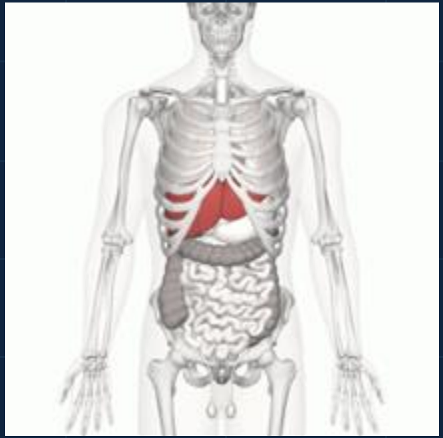
Currently, patients undergo four different scans before a 3D image of the liver is obtained. The objective of the research is to use the data extracted from the available samples to develop a machine learning program which will predict 25%-50% of the CT volumes. If successful, this would allow for patients to be scanned fewer times resulting in reduced exposure to harmful radiations.

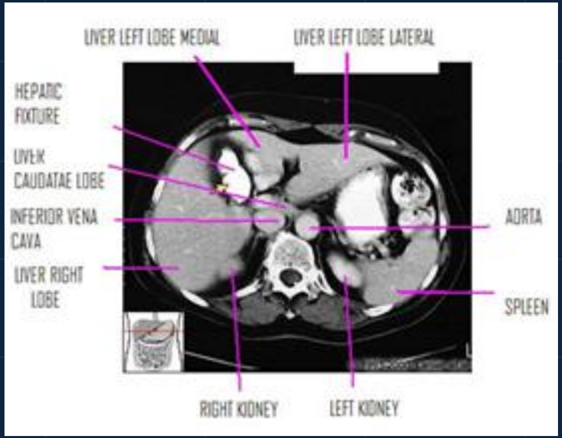
10cm

Liver Anatomy

The liver is located in the upper right-hand portion of the abdominal cavity, beneath the diaphragm, and on top of the stomach, right kidney, and intestines. Shaped like a cone, the liver is a dark reddish-brown organ that weighs about 3 pounds.



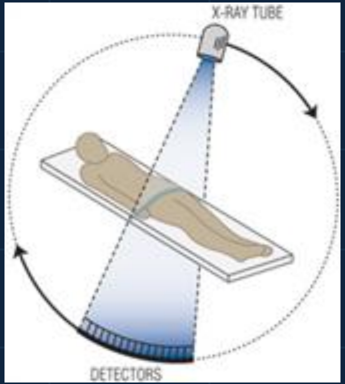




10cm

CT Scanning

During a computerized tomography (CT) scan, a thin X-ray beam rotates around an area of the body, generating a 3D image of the internal structures.



X-ray produces ionizing radiation — high-energy wavelengths or particles that penetrate tissue to reveal the body's internal organs and structures. Ionizing radiation can damage DNA. At the low doses of radiation a CT scan uses, the risk of developing cancer from it is so small that it can't be reliably measured. Because of the possibility of an increased risk, however, the American College of Radiology advises that no imaging exam be done unless there is a clear medical benefit.

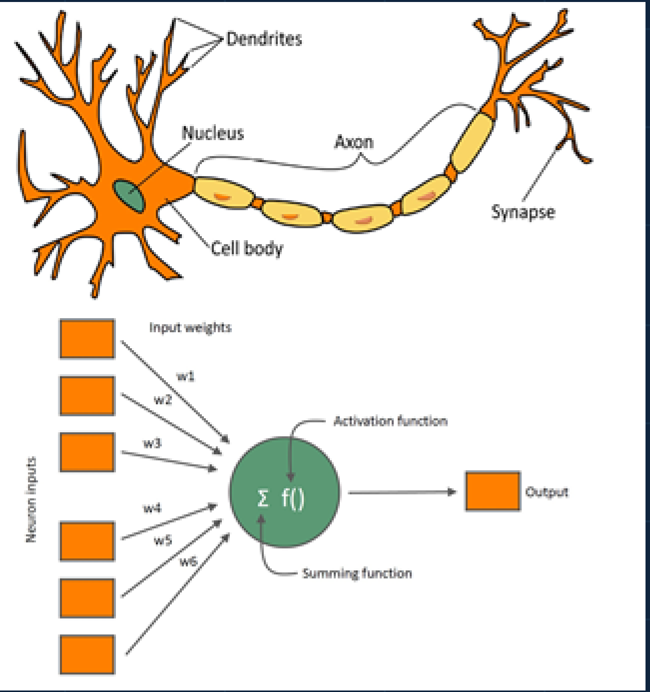
10cm

Reference images from: [CAT Scanning by on emaze]

What is Neural Network?

Neural networks (or Artificial neural networks) are computing systems inspired by the biological neural networks that constitute animal brains.

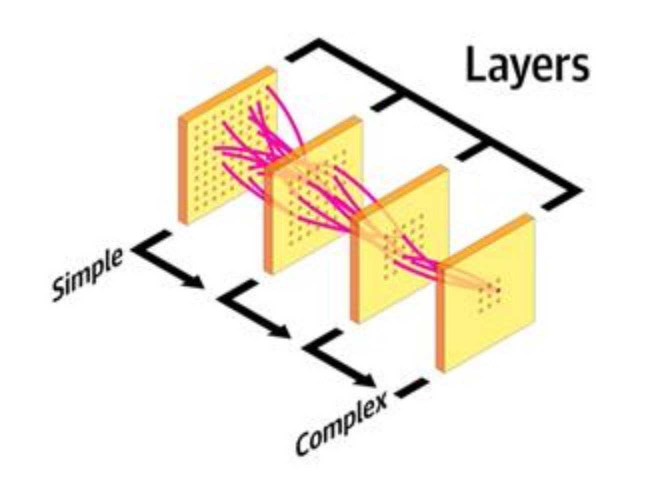
Neural networks reflect the behavior of the human brain, allowing computer programs to recognize patterns and solve common problems in the fields of AI, machine learning, and deep learning.



10cm

Convolutional Neural Network

CNN is a complex network of interconnected processes, organized in layers. With each layer, the CNN can detect higher-level, more abstract features. When the CNN is identifying these features, it uses something called a filter.



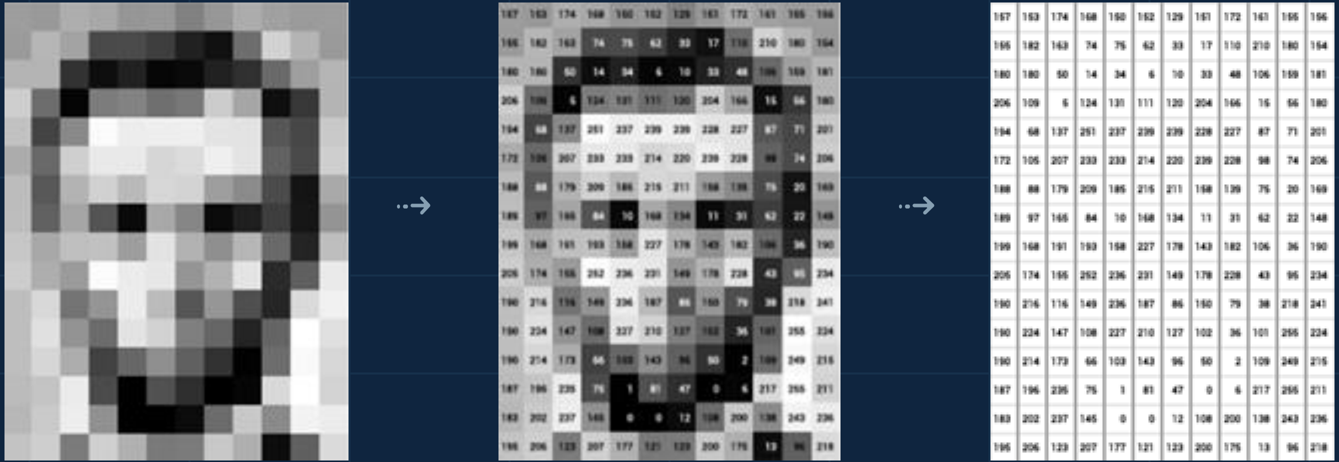
Convolution

$$(f * g)(t) \stackrel{\mathrm{def}}{=} \int_{-\infty}^{\infty} f(\tau) g(t-\tau)\, d\tau$$
$$= \int_{-\infty}^{\infty} f(t-\tau) g(\tau)\, d\tau.$$

10cm

Convolutional Neural Network

In a digital context, a pixel is the smallest, controllable unit of an image represented on a display. Each pixel is a representation of a tiny portion of the original image.



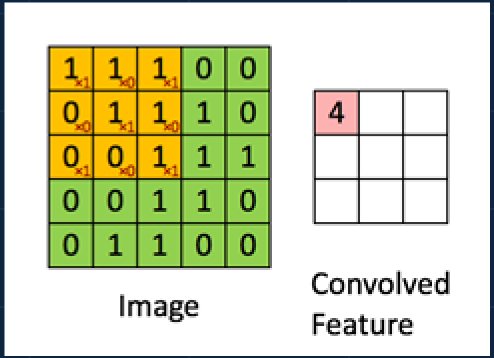
10cm

Reference images from https://ai.stanford.edu/~syyeung/cvweb/tutorial1.html

Convolutional Neural Network

In a digital context, a pixel is the smallest, controllable unit of an image represented on a display. Each pixel is a representation of a tiny portion of the original image.

A medical image is characterized by local structures such as textures, edges, curves, and corners. And what these filters are doing are constituting little miniature versions of each of these little building blocks. The way that the CNN looks for these building blocks is through convolution.

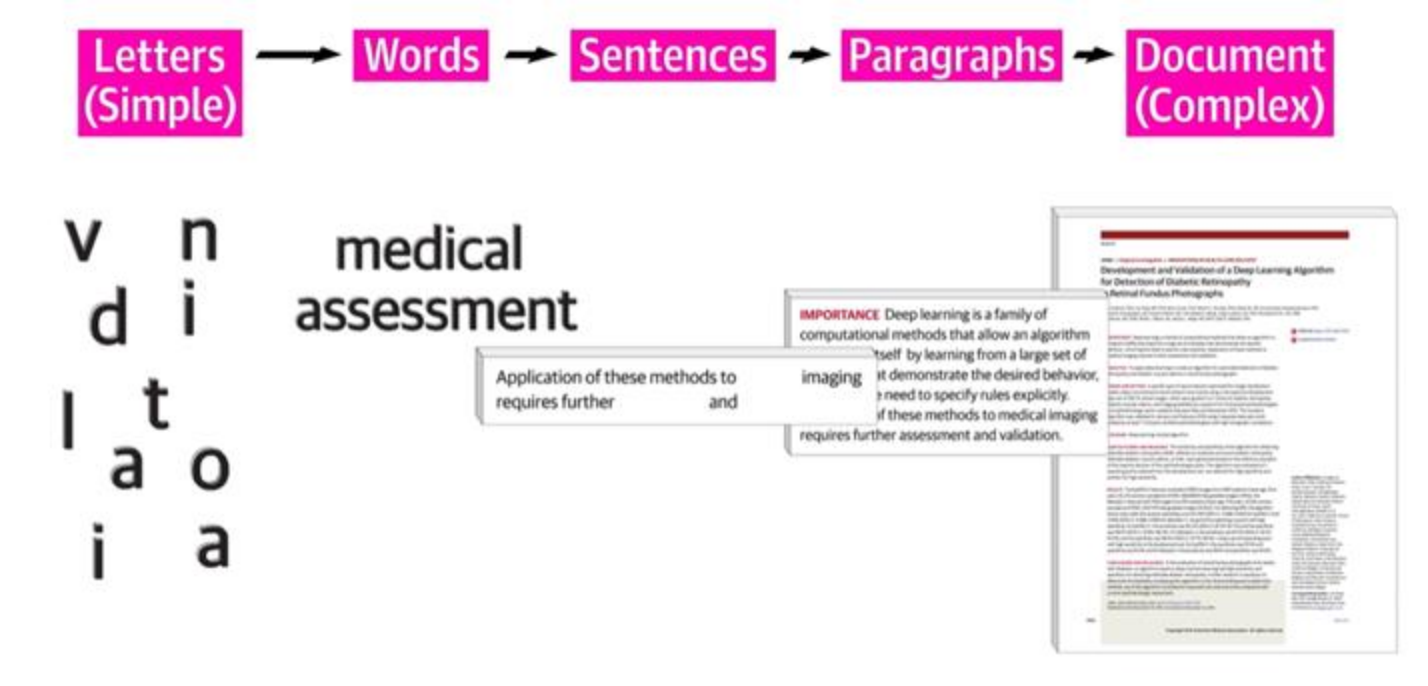


You have a filter, and you're walking to every part of the image, and you're just comparing how much does this image look like that filter.

10cm

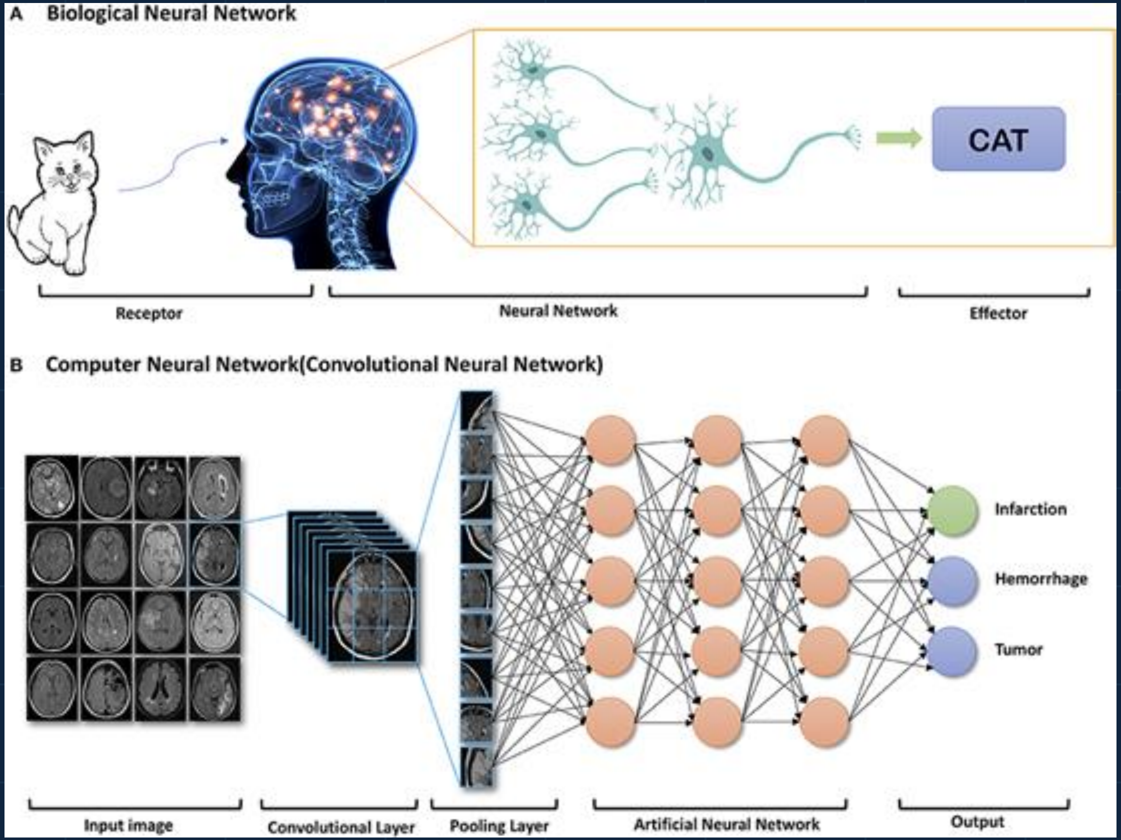
Reference images from https://ai.stanford.edu/~syyeung/cvweb/tutorial1.html

Convolutional Neural Network



10cm

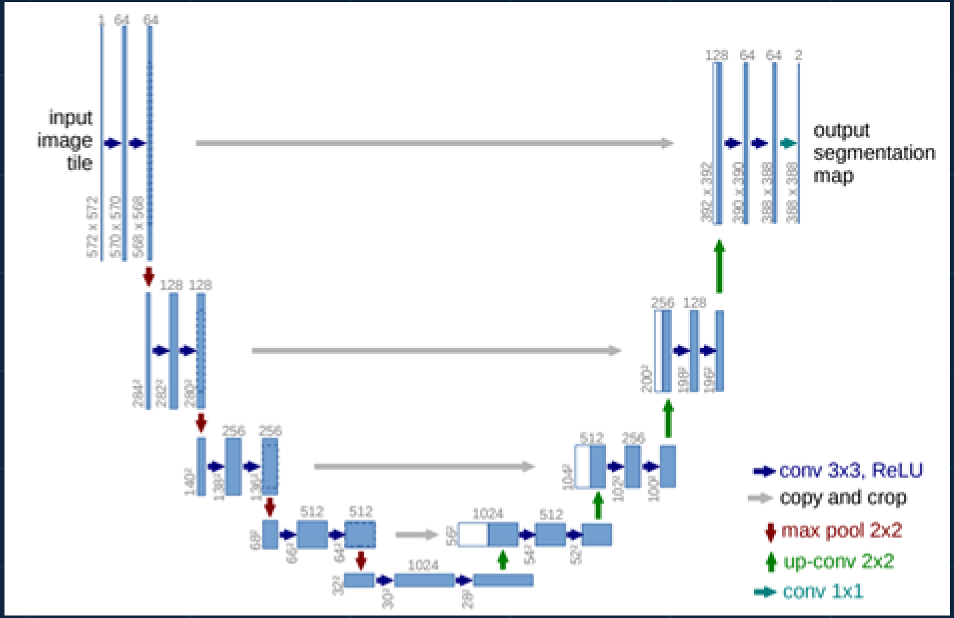
How does machine learn?



10cm

UNet

The image below represents the U-Net. As the network is composed of layer groups that are shaped like an U, it's not surprising where the name comes from. The left part and foot of the U is called the contracting path, whereas the right part is called the expansive path. U-nets downsample an input image to learn about its salient features, to reconstruct the input via upsampling.



10cm

UNet

The image below represents the U-Net. As the network is composed of layer groups that are shaped like an U, it's not surprising where the name comes from. The left part and foot of the U is called the contracting path, whereas the right part is called the expansive path. U-nets downsample an input image to learn about its salient features, to reconstruct the input via upsampling.

An adaptation of the U-Net to output pixel-wise regression values, instead of class labels, based on multichannel input data, has been developed in the remote sensing satellite imaging research domain.

The pixel-wise regression U-Net has only received limited consideration as a deep learning architecture in medical imaging for the image estimation/synthesis problem, and the limited work so far did not consider the application of 3D multichannel inputs.

10cm

Data

For each patient, we have four volumes of two-dimensional CT images of liver slices. Each volume of such forms a 3D image of the liver. These two-dimensional images need to be manually annotated to identify the regions of interest, the liver.

2020

3020

4020

5020

10cm

Methodology: Preparing the data

Segmentation

Identify region of interest

Registration

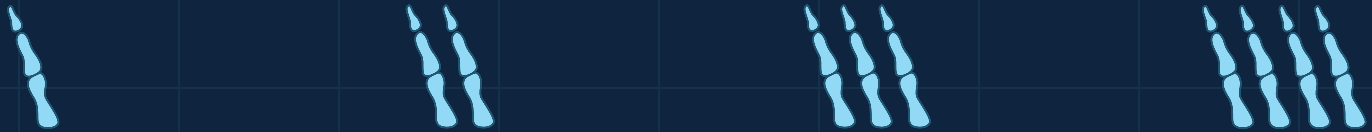
All images in the same coordinate system

Normalization

Normalize pixel value to be between 0 and 1

Create Dataset

Set up training data and validation data

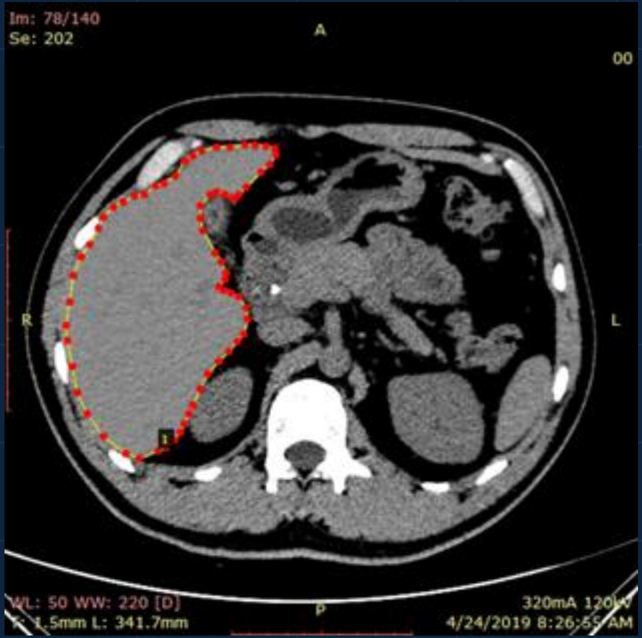


10cm

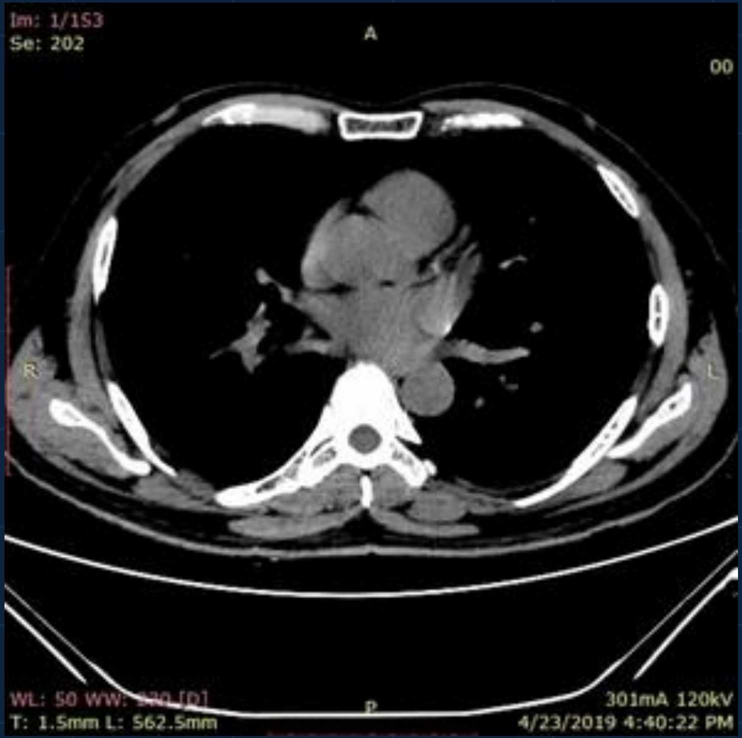
Segmentation

For each patient, we have four volumes of two-dimensional CT images of liver slices. Each volume of such forms a 3D image of the liver.

These two-dimensional images need to be manually annotated to identify the regions of interest, the liver.



Segmentation Result I

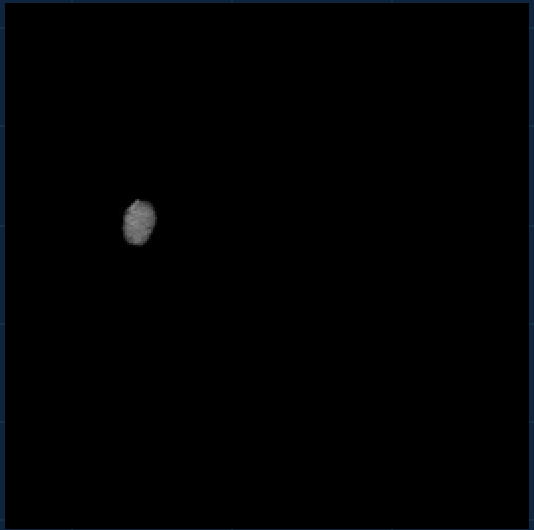


10cm

Segmentation Result II

S13430-S2020-0029

Original image

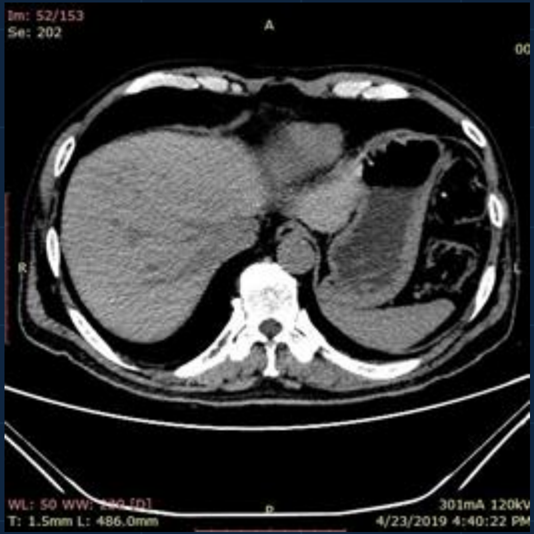


Mask

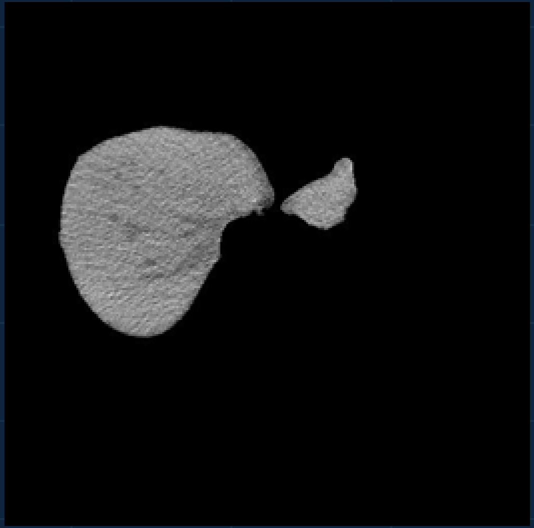
10cm

Segmentation Result III

S13430-S2020-0052



Original image

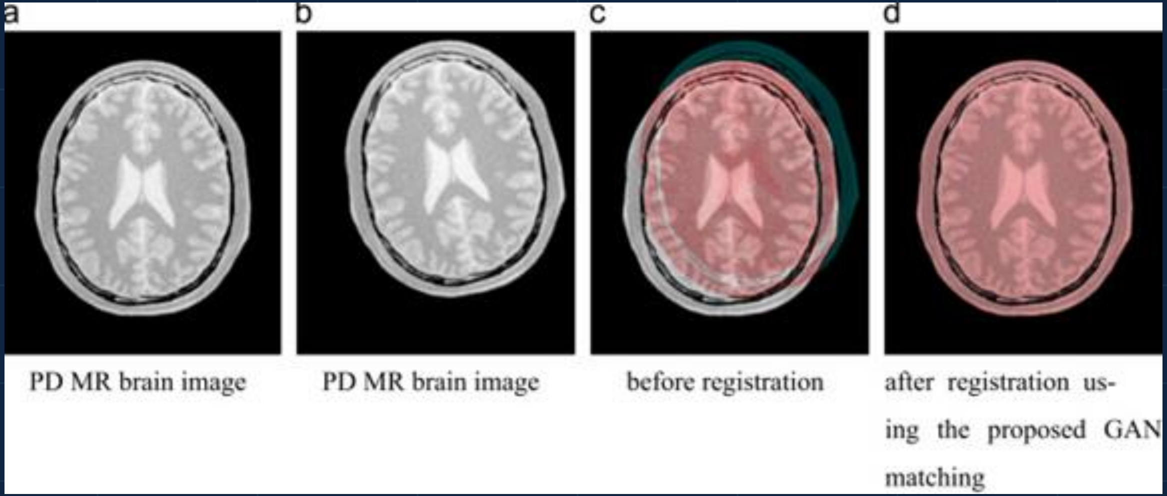


Mask

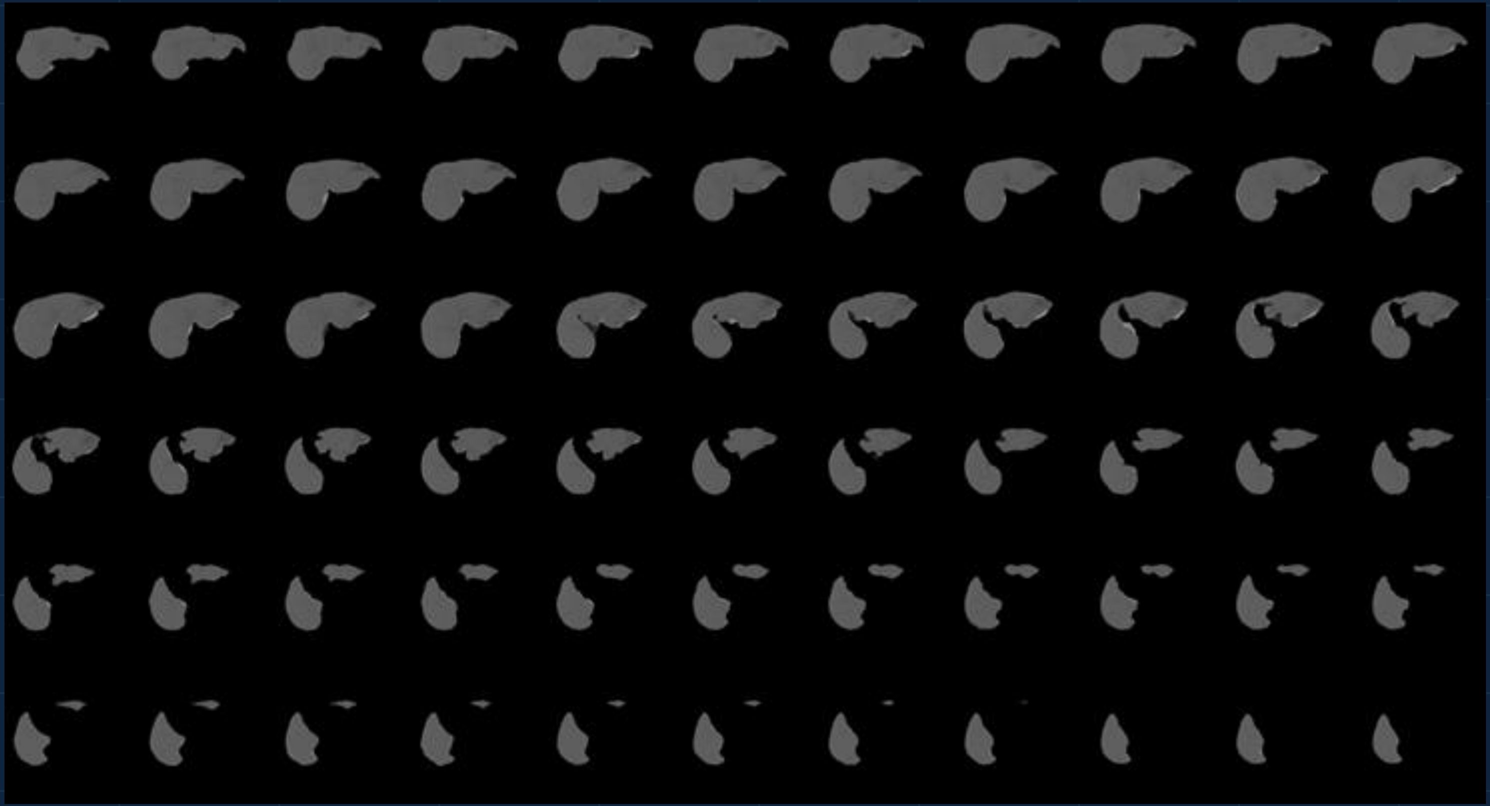
10cm

Registration

Image registration is the process of transforming different images of one scene into the same coordinate system. These images can be taken at different times (multi-temporal registration), by different sensors (multi-modal registration), and/or from different viewpoints. This step is necessary in order to compare or integrate the data obtained from these different CT volumes.



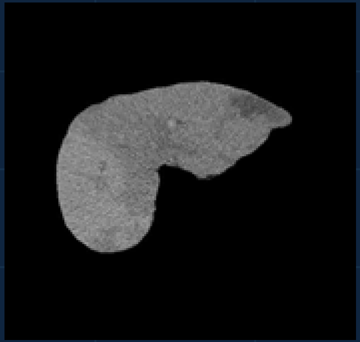
S14890: register 3020 onto 2020's coordinate system



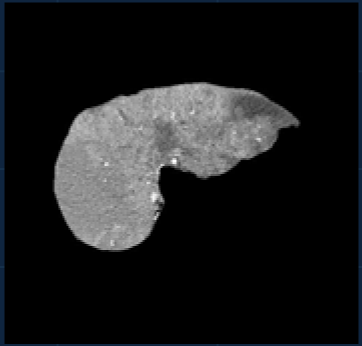
10cm

S14890: register 3020 onto 2020's coordinate system

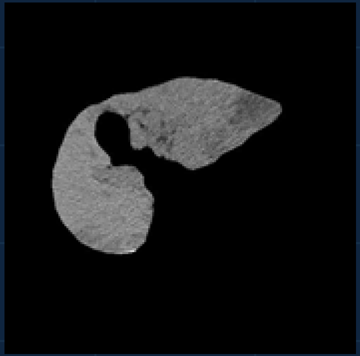
After registration:



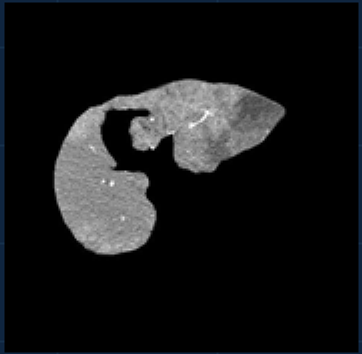
2020 slide 56



3020 slide 56



2020 slide 73



3020 slide 73

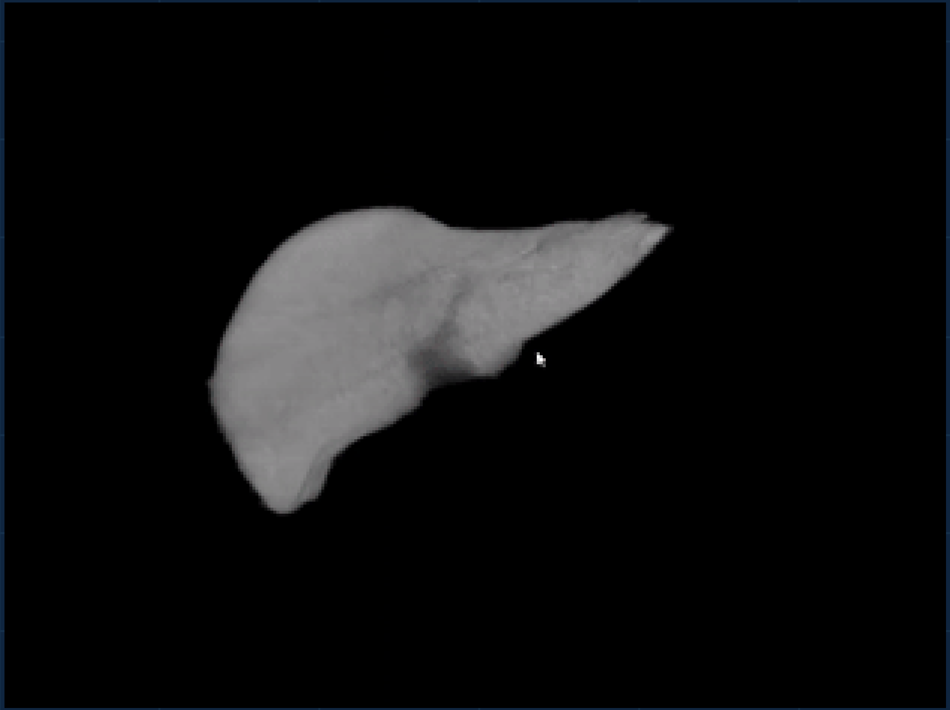
10cm

3D model of a liver

10cm

Liver HD resolution no animation by flarar-01 on Sketchfab: Liver 3D models - Sketchfab

3D model we acquired



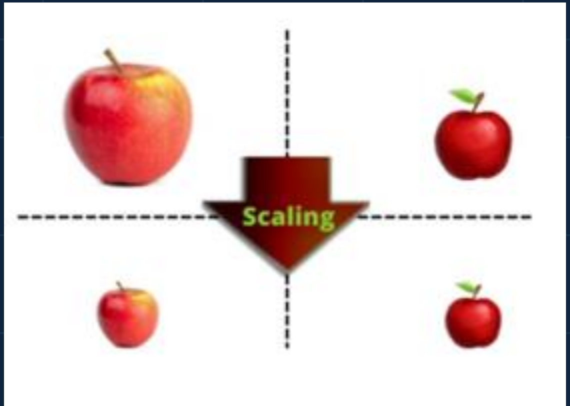
10cm

Brain Viewer Webapp (umich.edu)

Data Normalization

Machine learning algorithms like neural network, that use gradient descent as an optimization technique require data to be scaled. Neural networks process inputs using small weight values, and inputs with large integer values can disrupt or slow down the learning process.

For most image data, the pixel values are integers with values between 0 and 255. As such it is good practice to normalize the pixel values so that each pixel value has a value between 0 and 1.

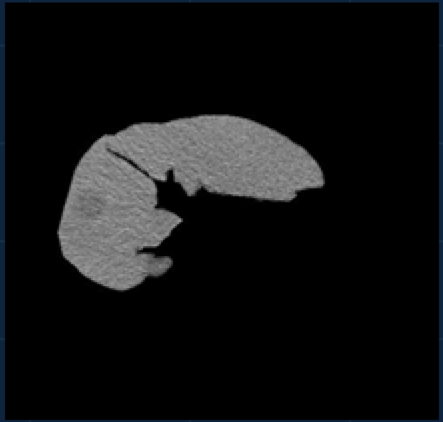


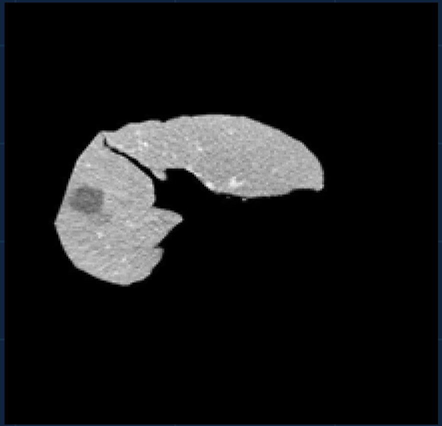
10cm

Data Normalization

Machine learning algorithms like neural network, that use gradient descent as an optimization technique require data to be scaled. Neural networks process inputs using small weight values, and inputs with large integer values can disrupt or slow down the learning process.

For most image data, the pixel values are integers with values between 0 and 255. As such it is good practice to normalize the pixel values so that each pixel value has a value between 0 and 1.

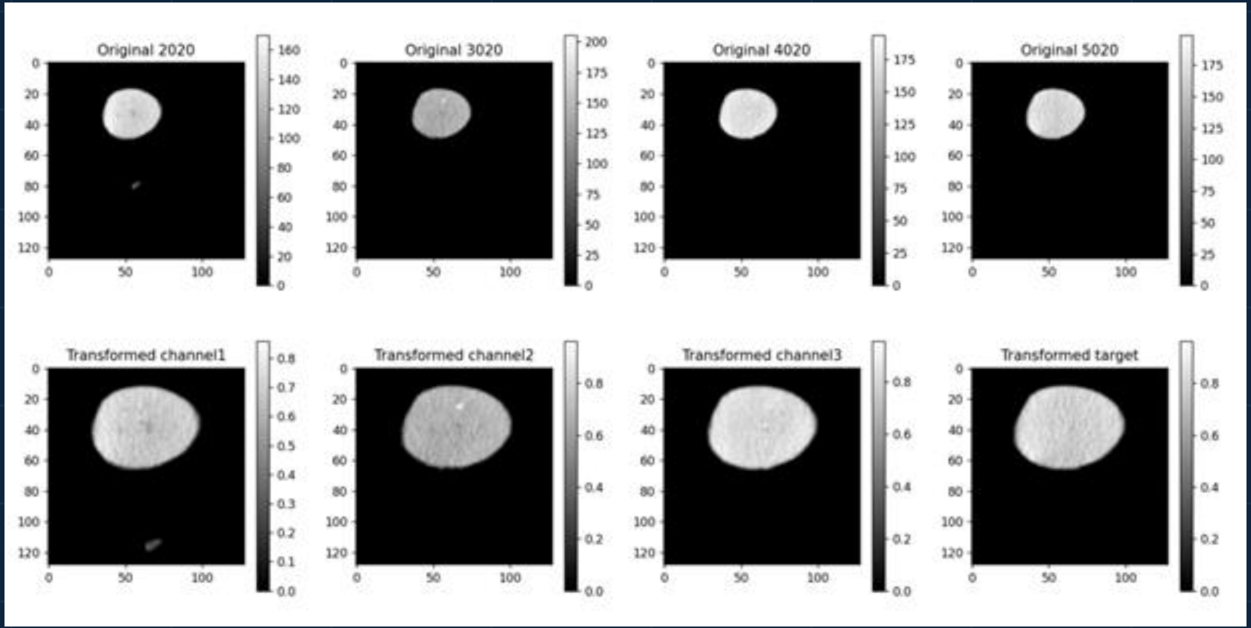




10cm

Data Preprocessing

This combination of transformation functions allowed us to normalize the intensity of each 3D volume to between 0 and 1, crop off the empty regions of the image in the background, and resize the images to 128*128*128.



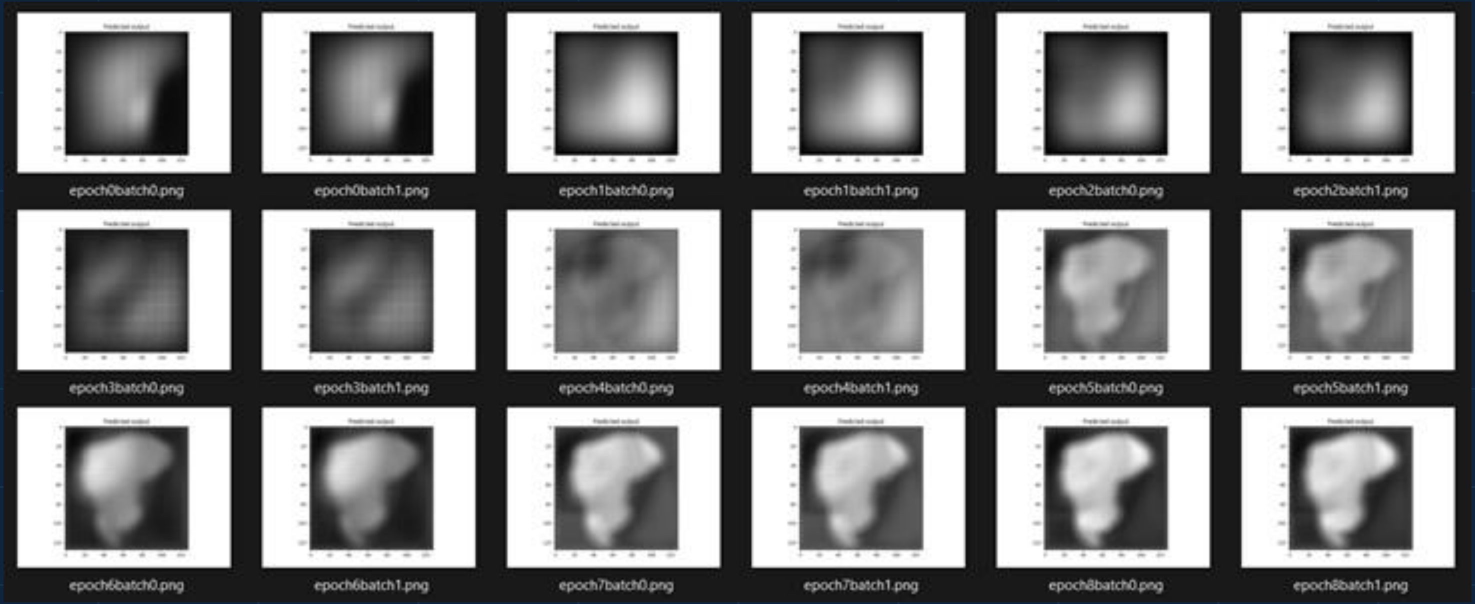
Create Dataset

For each subject, the 2020, 3020, and 4020 volumes are used for training and the 5020 volume is the corresponded target images. It is necessary to combine the three training images into one during the preprocessing step. To achieve that, we resliced all the input volumes and combined each set of 2020, 3020, and 4020 slices into a three-channeled 2D image. Pairing that with the corresponding 5020 slice as the target gave us the desired setup of training data.

```python
for data in self.train_dataset_3d:
    for i in range(128):
        ch1, ch2, ch3 = data['channel1'][0, : ,: ,i], data['channel2'][0, : ,: ,i], data['channel3'][0, : ,: ,i]
        channel1_slices.append(ch1)
        channel2_slices.append(ch2)
        channel3_slices.append(ch3)
        trgt = data['targets3d'][0, : ,: ,i]
        train_target.append(trgt)
        # TEST
        print("Number of slices: ", len(channel1_slices))
```

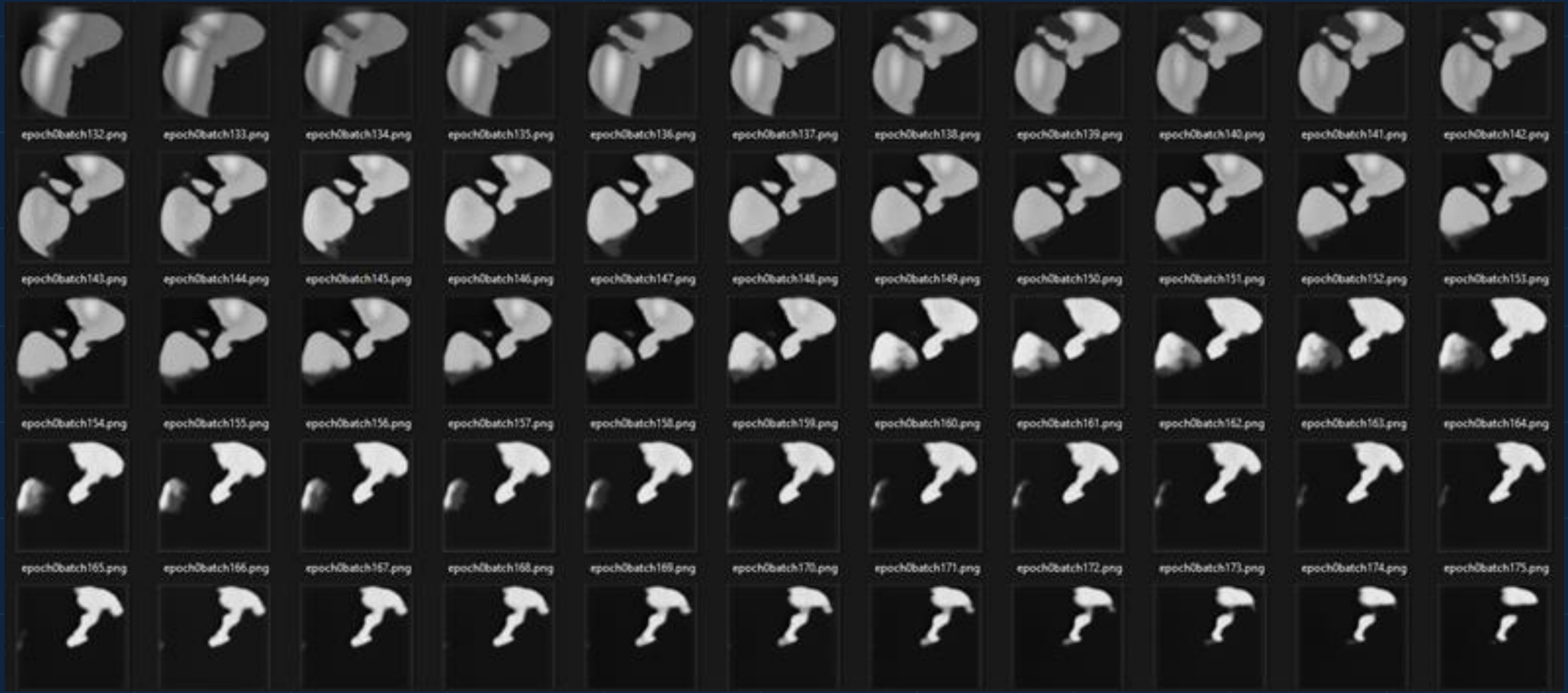
```python
for i in range(len(channel1_slices)):
    # Create a blank image that has three channels and the same number of pixels as your original input
    multi_channel_img = torch.zeros((3, 128, 128))
    # Add the channels to the needed image one by one
    multi_channel_img[0, :, :] = channel1_slices[i]
    multi_channel_img[1, :, :] = channel2_slices[i]
    multi_channel_img[2, :, :] = channel3_slices[i]
    train_image.append(multi_channel_img)
```

Initial Output from the UNet



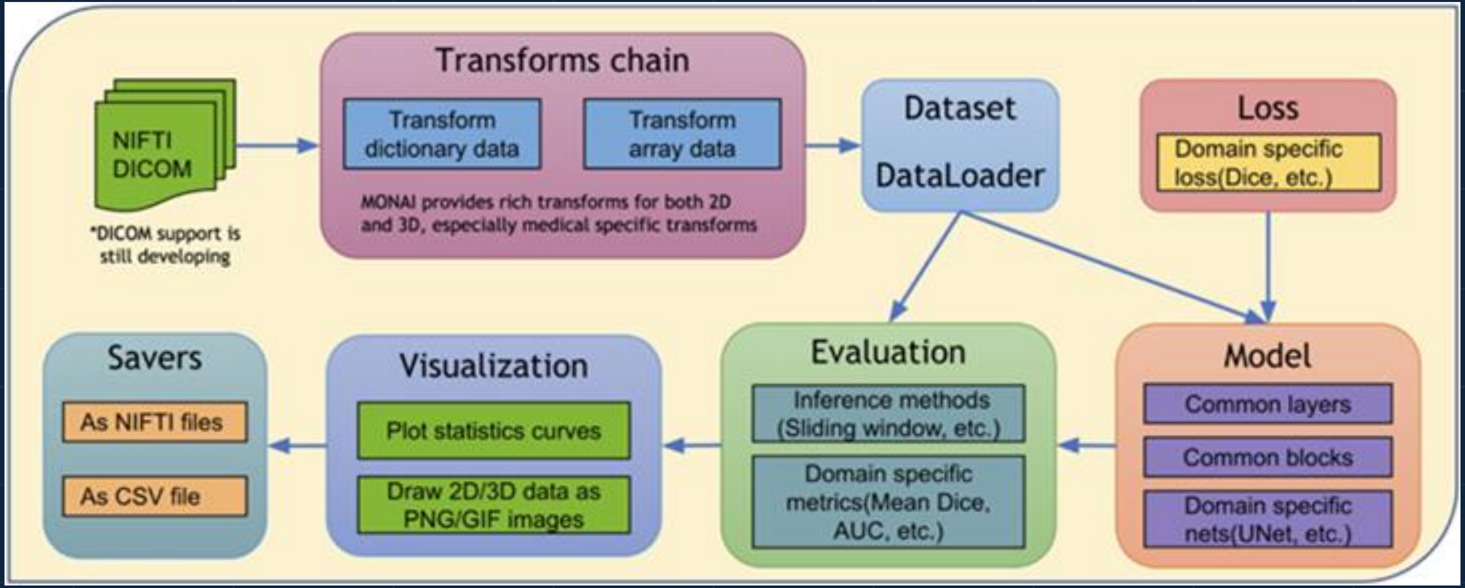
10cm

Current Output From the UNet



10cm

The Next Step



10cm

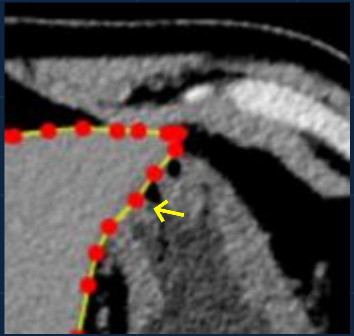
Credits

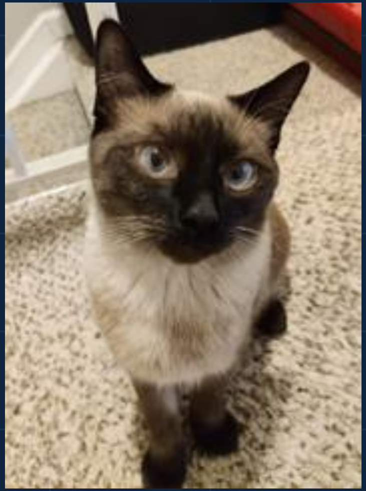
Bin Chen

Farah Combs

Rishi Mitra

Kashyab Ambarani





Turing

References

- “VGG Image Annotator (via).” Visual Geometry Group - University of Oxford. Accessed February 3, 2022. https://www.robots.ox.ac.uk/~vgg/software/via/.
- “Registration Overview.” Registration Overview - SimpleITK 2.0rc2 documentation. Accessed February 3, 2022. https://simpleitk.readthedocs.io/en/master/registrationOverview.html.
- Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation.” arXiv.org, May 18, 2015. https://arxiv.org/abs/1505.04597.
- Siddique, Nahian, Paheding Sidike, Colin P Elkin, and Vijay Devabhaktuni. “Michigan Technological University Digital Commons ...” Accessed February 3, 2022. https://digitalcommons.mtu.edu/cgi/viewcontent.cgi?article=34342&context=michigantech-p.
- Wang, Jueqi, David Mattie, Derek Berger, and Jacob Levman. “Multichannel Input Pixelwise Regression ... - Openreview.net.” Accessed February 3, 2022. https://openreview.net/pdf?id=JG895xlWsfA.
- https://jamanetwork.com/journals/jama/article-abstract/2702856?utm_campaign=video&utm_medium=social&utm_source=youtube&utm_content=public
- https://www.3blue1brown.com/topics/neural-networks

10cm