A Beginner's Guide for

Machine Learning Engineers on Medical Imaging

Abstract

Artificial intelligence techniques, especially machine learning plays an important role in healthcare analytics. This article introduces a guide for engineers to understand the background necessary for applying machine learning models in medical imaging.

Machine Learning in Healthcare

Machine learning in healthcare has recently made headlines. For example, Google has developed a machine learning algorithm to help identify cancerous tumors on mammograms. Stanford is using deep learning algorithms to identify skin cancer [5]. These machine learning models are trained to computationally analyze the images to identify abnormalities (classification) and point out the areas (segmentation) that need attention.

The ethics of using machine learning in healthcare has been discussed before. The best machine learning tool so far is seen as the doctor's experience in making determinations about the patient than algorithms. This is similar to self-driving vehicles, which are very popular, but not fully trustworthy as people never want to risk their lives. Also, patients always need a human touch and the compassionate relationship of people who deliver care. Therefore, machine learning models can be seen as tools which support doctors and physicians (often overworked) in making decisions.

What is Medical Imaging?

Medical imaging is the process of visualizing the human body parts to help doctors or physicians to diagnose, monitor, or treat a disease or an injury [2]. Every year, we see the introduction of new scanners with shorter scanning times and higher spatial resolution, producing more information in detail [4]. Medical imaging has many techniques that are developed for scientific and industrial applications such as Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI), Computed Tomography (CT), or Computed Axial Tomography (CAT).

CT produces a 3D image that is sliced from 1 to 10mm, which helps the doctors see every detail of the object from multiple angles. CT is very beneficial to detect smaller diseases such as tumors or cancer cells. MRI is another form of imaging which does not use radiation, and relies on radio waves as well as a magnetic force.

PET scans are similar to CT because they use the radiation, however, the effects are smaller. PET can be combined with CT and MRI to produce more accurate 3D images.

Voxel

A voxel is used in the visualization and analysis of medical and scientific images. It is the value represented in the volumetric display. Voxels can contain multiple scalar values, essentially vector (tensor) data. The word voxel originated by analogy with the word "pixel", with vo representing "volume" and el representing "element."

- Pixel - picture element
- Resel - resolution element
- Texel - texture element
- Maxel - material element
- Tixel - tactile element
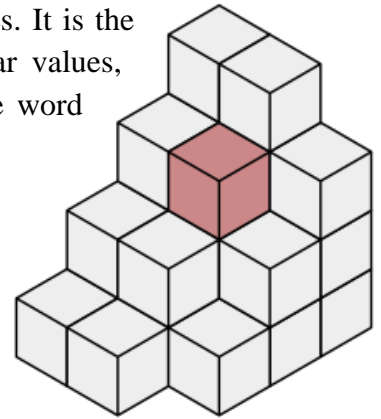
The 3D DICOM and NIFTI Formats

Medical Images are often stored in DICOM or NIFTI formats. Below are details about each of the formats.

1. DICOM Image

The DICOM standard is useful for integrating all modern imaging equipment, accessories, networking servers, workstations, printers, and picture archiving and communication systems (PACS) that may have been installed by multiple manufacturers. The integration and continuous evolution of this communication standard have over the years achieved a nearly universal level of acceptance among vendors of equipment to view these images on computers when a proprietary viewer is not supplied with the system. DICOM differs from other image formats in aggregating information into data sets [6].
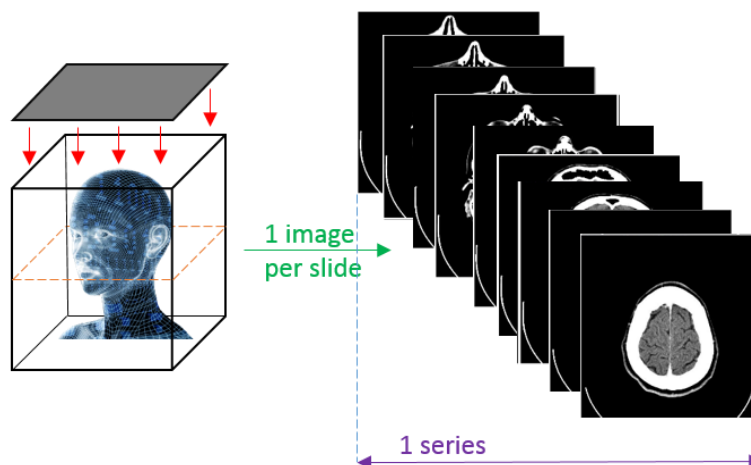
Figure 1: 3D Medical Image Slicing

The DICOM Image structure (figure 1) has four components: patient, studies, series, and images. Every patient has many studies which refer to examination and procedures of brain, lung, etc. The study entity is a part of a scanning session that consists of many series. Each series is a sequence of images [7] that represents how the 3D image sliced into 2D images. The series has an attribute called modality; which is the type of equipment (MRI or CT), used to create the images in sequence. Every image from the sequence has the same number of columns and rows and is made up of many pixels in an image, each of which is represented using three color components. In general, the DICOM image contains all the patient information in detail.
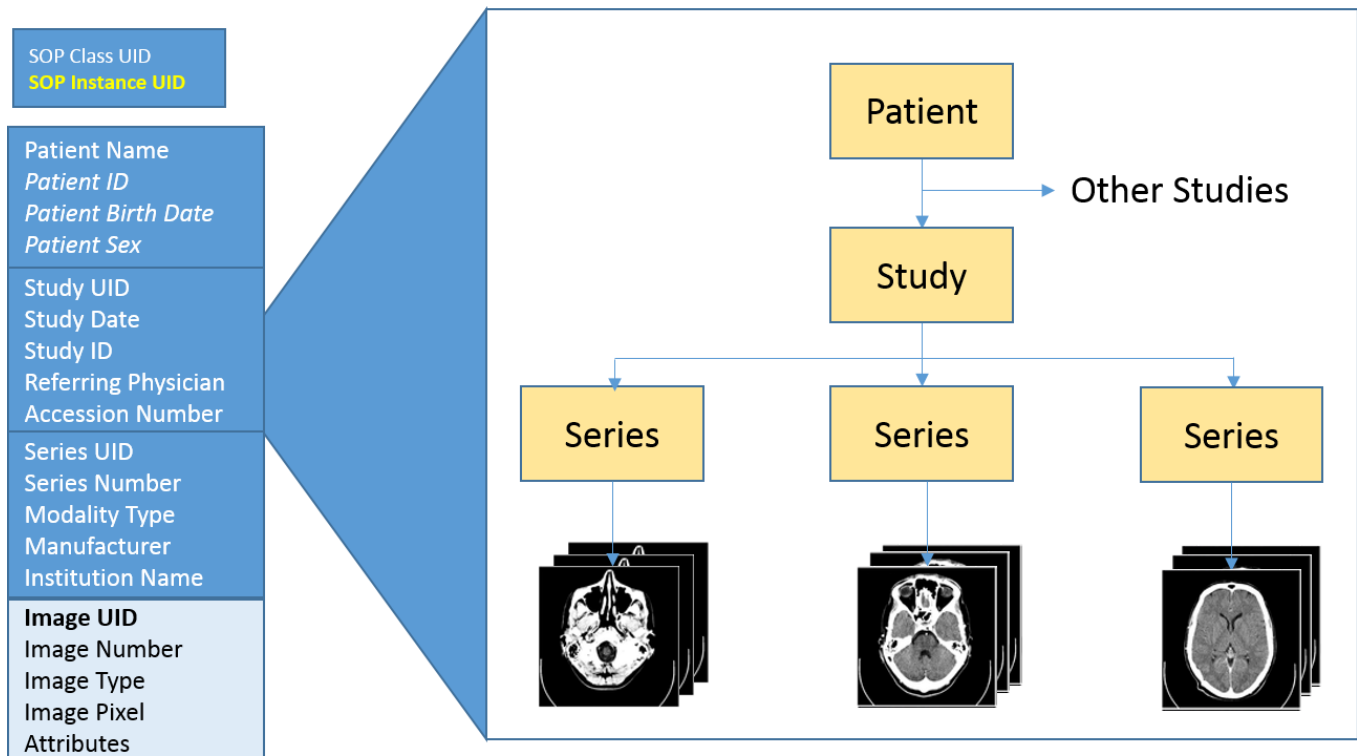

Figure 2: The DICOM Image Structure

2. NIFTI Image

NIFTI (The Neuroimaging Informatics Technology Initiative) is an analyze-styled data format, proposed by the NIFTI Data Format Working Group as a "short-term measure to facilitate inter-operation of functional MRI data analysis software packages". There are many tools available online to convert from DICOM to NIFTI and to view NIFTI formats such as dcm2nii and MRIcron.

An affine coordinate system on a plane is defined by an ordered pair of non-collinear vectors. The NIFTI image affine coordinate definitions relating to voxel index to spatial location. It also standardizes the way to store vector-based values in a dataset. The NIFTI image usually contains in dual files .hdr and .img or a single file .nii. The NIFTI image's coordinate system consists of three planes to describe the standard anatomical position of a human:

- A sagittal plane, (also known as a median plane) is a y-z plane, perpendicular to the ground, which separates left from right. The mid-sagittal plane is the specific sagittal plane that is exactly in the middle of the body.
- A coronal plane, (also known as a frontal plane) is an x-z plane, perpendicular to the ground, which (in humans) separates the anterior from the posterior, the front from the back, the ventral from the dorsal.
- A transverse plane, (also known as an axial or horizontal plane) is an x-y-z plane, parallel to the ground, which (in humans) separates the superior from the inferior, or put another way, the head from the feet.
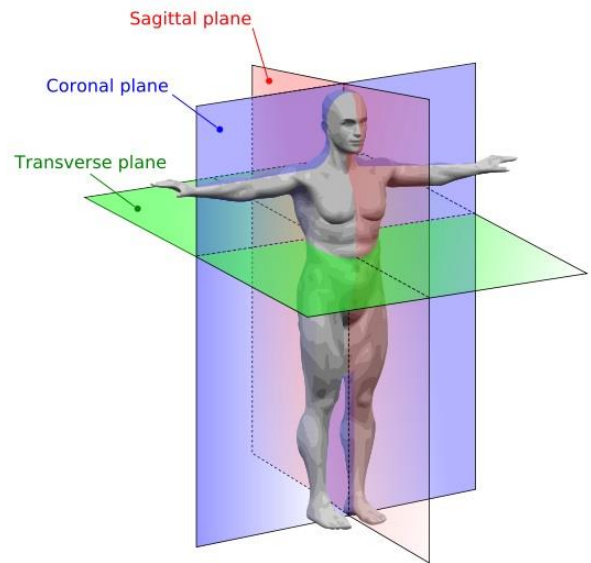
Modalities

Medical images can be captured using different modalities, i.e., different devices.

1. Computed Tomography (CT)

Computed Tomography (CT), also commonly known as a CAT scan, is a medical imaging method that combines multiple X-ray projections taken from different angles to produce images of the areas inside the human body in detail. CT images allow doctors to get very precise, 3-D views of certain parts of the body, such as soft tissues, the pelvis, blood vessels, the lungs, the brain, the heart, abdomen, and bones. CT is also often the preferred method of diagnosing many cancers, such as liver, lung and pancreatic cancers.

CT is often used to evaluate: presence, size and location of tumors, organs in the pelvis, chest and abdomen, colon health (CT colonography), vascular condition/blood flow, pulmonary embolism (CT angiography), abdominal aortic aneurysms (CT angiography), bone injuries, cardiac tissue, traumatic injuries, cardiovascular disease.

CT images are acquired only in the axial plane. The axial data set can then be used to reconstruct images in other planes; Sagittal and coronal are the most commonly reconstructed planes.

2. Magnetic Resonance Imaging (MRI)

Magnetic Resonance Imaging (MRI) is a medical imaging technology that uses radio waves and a magnetic field to create detailed images of organs and tissues. MRI has proven to be highly effective in diagnosing several conditions by showing the difference between normal and diseased soft tissues of the body.

MRI is often used to evaluate: blood vessels, abnormal tissue, breasts, bones and joints, organs in the pelvis, chest, and abdomen (heart, liver, kidney, and spleen), spinal injuries, tendon, and ligament tears.

MRI can be acquired in any plane, not just axial. The primary MR sequences include T2, T1, and T1 with contrast, Diffusion, and FLAIR. Each sequence has to be acquired separately, which means that an MRI will take a lot longer to perform than a CT.

3. Position Emission Tomography (PET)

Positron Emission Tomography (PET) is a nuclear imaging technique that provides physicians with information about how tissues and organs are functioning. PET, often used in combination with CT imaging, uses a scanner and a small number of radiopharmaceuticals which is injected into a patient's vein to assist in making detailed, computerized pictures of areas inside the body.

PET is often used to evaluate: neurological diseases such as Alzheimer's and Multiple Sclerosis, cancer, the effectiveness of treatments, heart conditions.

# *Deep Learning Toolkit for Medical Imaging (DLTK)*

Deep Learning Toolkit for Medical Imaging (DLTK) is a neural networks toolkit written in python, on top of TensorFlow. It is developed to enable fast prototyping with a low entry threshold and ensure reproducibility in image analysis applications, with a particular focus on medical imaging. Its goal is to provide the medical community with operations and functions to build machine learning models to accelerate research in this exciting field. DLTK uses the NifTI (or .nii format), originally developed for brain imaging, but widely used for most other volumetric images.

1. How does DLTK work?

DLTK requires specialty header information and some of which are mentioned below:

- Dimensions (3D) and size vector
- Analyzed data type
- Voxel spacing (typically in mm)
- Physical coordinate system origin
- Direction (sagittal, coronal, transverse)

Apart from exposing low-level operations (e.g. tensor multiplications, etc.) DLTK also allows developers to use the higher-level operations for volumetric images (e.g. differentiable 3D upsampling layers, etc.)
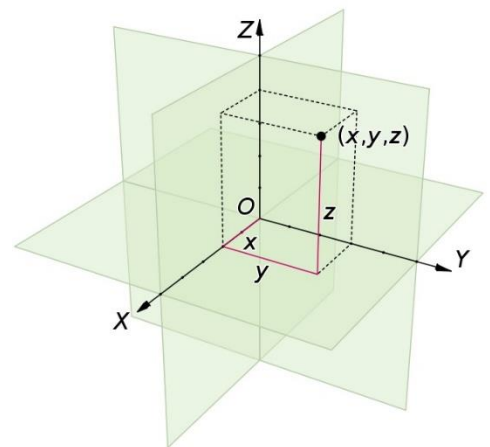
2. Training Using DLTK

The network will train in the space of voxels, meaning we will create tensors of shape and dimensions [batch size, dx, dy, dz, channels/features] and feed it to the network. The network will train in that voxel space and assume that all images (also unseen test images) are normalized in that space or might have issues to generalize.

In that voxel space, the feature extractors (e.g. convolutional layers) will assume that voxel dimensions are isotropic (i.e. are the same in each dimension) and all images are oriented the same way. All information to transform are vectors stored in the .nii header.

We chose SimpleITK, a python wrapper around the ITK library, which allows us to import additional image filters for pre-processing and other tasks. Depending on the size of the training database, there are several options to feed .nii image data into the network graph.

3. Data Input and Output

In memory & feeding dictionaries: This is a traditional way: read all dataset into memory before training. This direct approach is typically the fastest and easiest to implement, as it avoids continuously reading the data from disk, however, requires to keep the entire database of training examples (and validation examples) in memory, which is not feasible for larger databases or larger image files.

Using a TFRecords database: TFRecords are fast means of accessing files from disk but require to store yet another copy of the entire training database. If we are aiming to work with a database of several TB size, this could be prohibitive.

Using native python generators: Create a method to directly load the image data and use the built-in method Dataset.from_generator() to queue the examples. It avoids creating additional copies of the image database, however, is considerably slower than TFRecords, since the generator cannot parallel read and map functions.

Which method should we choose for the program? It depends on how much multi-GPU processing, the replicated data centers you have. You also can use the native python generator method if you want to obtain the result without causing an issue (poor computer) and care nothing about time complexity.

# Algorithms for Data Preprocessing

1. What is Normalization (ANTs)

ANTs [9] is software for biomedical image analysis with a focus on registration, segmentation, geometric quantification, and statistics. ANTs became well known because they performed well in a variety of open competitions related to image registration. Most of these successes occurred in the days before the community had many such competitions:

- Finishing in the top rank in the Klein 2009 evaluation on brain MRI
- Finishing in the top rank in the Murphy 2011 lung CT evaluation
- Top SATA 2012 and 2013 finishers used ANTs

- Performing well in a cardiac motion estimation competition
- Well-known robust performance on large datasets

Although ANTs have often performed well without using domain knowledge, it is still valuable to use problem-specific solutions when feasible. ANTs gives a quick deformable registration option that initializes the position, aligns the image, and fit the area.
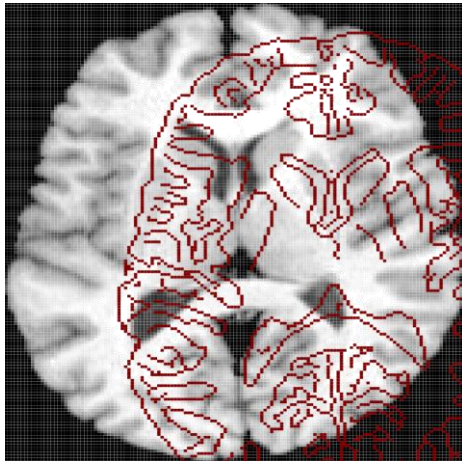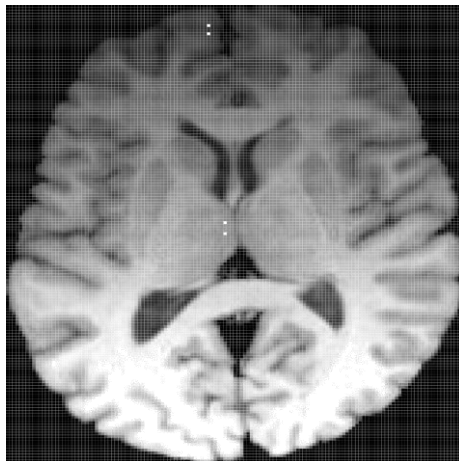

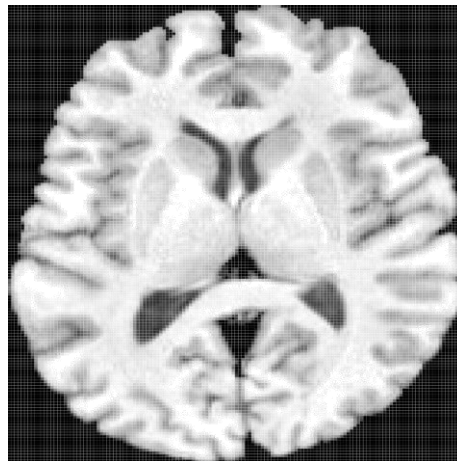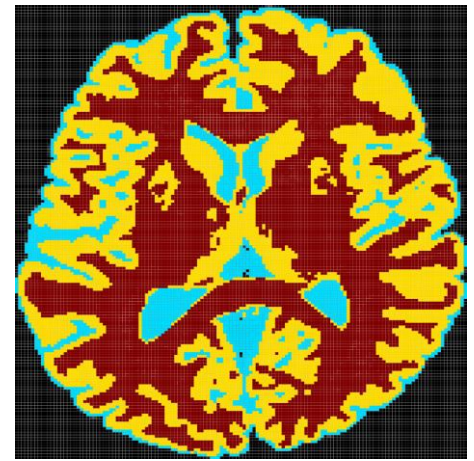| Initial Relative Image Positions | Rigidly Aligned Image Positions | Quick Deformable Registration |

ANTs also provides a method to correct or improve original images that called N4 Bias Field Correction.

| Original Corrupted Image | N4 Correction | K-means Result |

Expectation maximization segmentation with a variety of likelihood models and initialization strategies. Incorporates multiple modalities and allows control of prior strength. The finite mixture modeling (FMM) segmentation approach is the most popular, etc.

2. What is Augmentation (DLTK)

Augmentation is the process of making or becoming greater in size or amount. Augment training images by simulating a variation in the data to be robust against. Similar to normalization methods, we distinguish between intensity and spatial augmentations.

Examples of intensity augmentations:

- Adding noise to training images generalize to noisy images

- Adding a random offset or contrast to handle differences between images

Examples of spatial augmentations:

- Flipping the image tensor in directions on where to expect symmetry (e.g. a left/right flip on brain scans)

- Random deformations, (e.g. for mimicking differences in organ shape)

- Rotations along axes (e.g. for simulating difference ultrasound view angles)

- Random cropping and training on patches
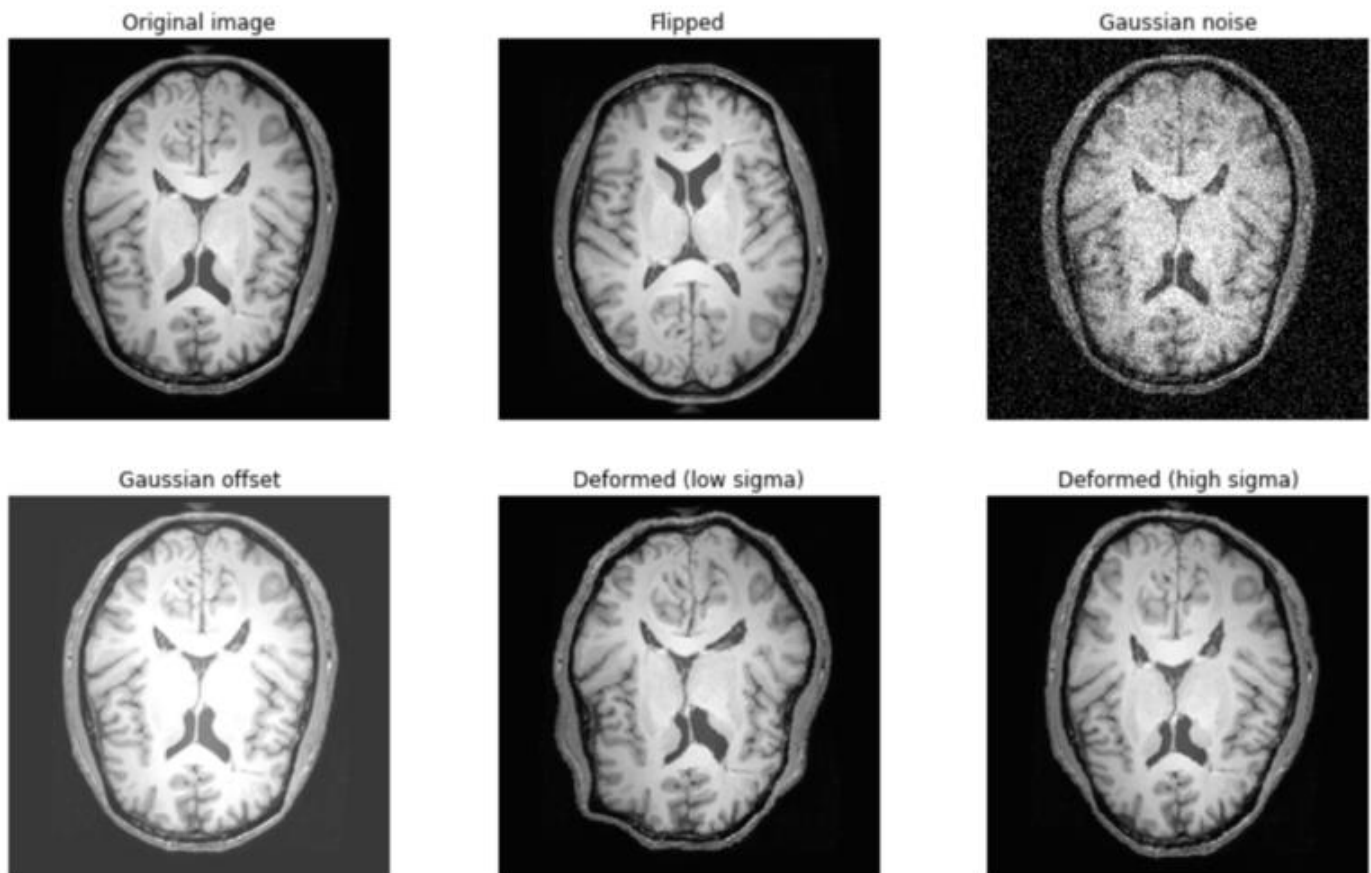
3. Segmentation (U-NET)

Segmentation is the process of dividing the subject into separate parts or sections. The U-NET is convolutional network architecture for fast and precise segmentation of images. It currently has outperformed the prior best method (a sliding-window convolutional network) and has won many prizes.
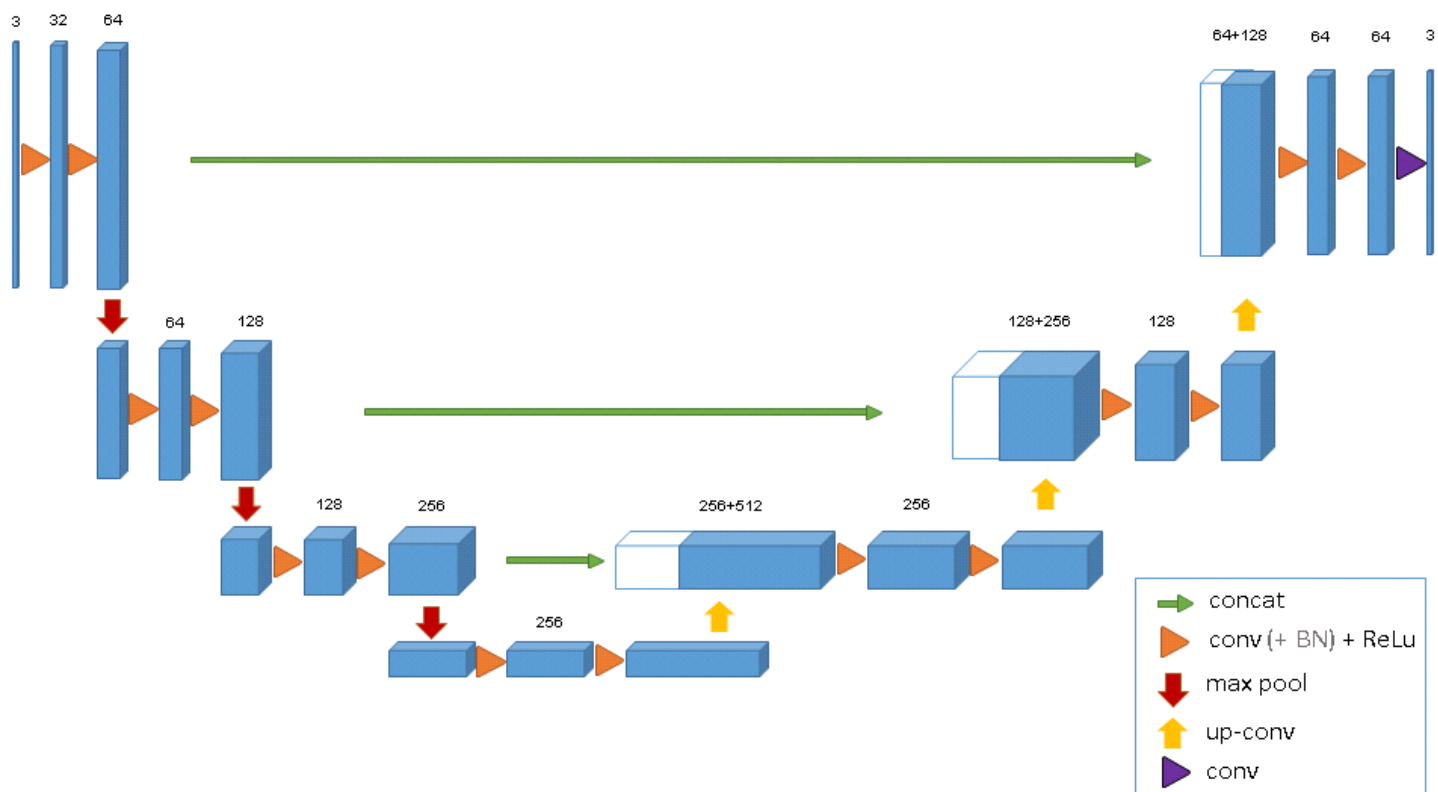
The U-Net has two parts: contracting (downsampling path) and expanding (upsampling path)

The blue boxes represent feature maps. The number of channels is denoted above each feature map. In the last layer, a 1×1×1 convolution reduces the number of output channels (the output of the segmentation map). We also introduce batch normalization (BN) before each ReLU during training with its mean and standard deviation and global statistics are updated using these values

Normalization is done via these computed global statistics and the learned scale and bias (ANTs). We can see that the network is composed of Convolution Operation, Max Pooling, ReLU Activation, Concatenation and Up Sampling Layers.

References

[1] https://reverehealth.com/live-better/mri-ct-pet/

[2] https://www.fda.gov/radiation-emitting-products/radiation-emitting-products-and-procedures/medical-imaging

[3] https://www.medicalimaging.org/about-mita/medical-imaging-primer/

[4] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2398810/pdf/postmedj00076-0015.pdf

[5] https://www.healthcatalyst.com/clinical-applications-of-machine-learning-in-healthcare

[6] https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net/

[7] NEMA: A DICOM publication. (2019) Introduction and Overview. NEMA. 2019

[8] Damien D. A Very Basic DICOM Introduction. DMC4CHE.

[9] http://stnava.github.io/ANTs/