CS 766 - Project Midterm Report

Lanston Chu (hchu34@wisc.edu)
Ying Ji Chuang (ychuang6@wisc.edu)
Mehmet Furkan Demirel (demirel@wisc.edu)
Kendall Park (kpark229@wisc.edu)

March 27, 2019

1 Summary

The goal of our project is to reconstruct 3D volume of radiodensities from 2D chest X-ray using a CNN approach. Our encoder-decoder CNN model will be trained using synthetic X-rays (also known as digitally reconstructed radiographs) generated from a database of chest computed tomography (CT) scans.

2 Synthetic X-ray Generation

The dataset used in this project consisted of over 1000 series of CT studies. To train a neural network model that could reconstruct 3D volumes of these CT studies using 2D X-ray images as input, we need to create synthetic X-rays of these volumes. In our initial proposal, we referred to Moturu and Chang's [2] method of digital radiography reconstruction using the Beer Lambert Law to perform ray tracing interpolation. We have since gained access to the authors' code and managed to obtain good synthetic X-ray images using a small portion of the CT studies. However, we realized that the ray tracing function in the script was performed using parallel-ray approximation, which used a point source at an infinite distance from the patient. Ideally, we wanted our 2D X-ray images to be generated from a point-source geometry approach, allowing slight variations in the data as real-life X-ray images are taken from various distance away from the patient.

While researching digitally reconstructed radiography techniques that used a non-infinite X-ray source, we discovered the Siddon-Jacobs fast ray-tracing algorithm in the well-known Insight Segmentation and Registration toolkit (ITK) imaging package [4]. This algorithm allows quick computation of the resulting projection image in addition to more flexibility in manipulating the projection angle. The resulting synthetic X-ray images appeared faint compared to the result in [2]. This is expected since [2] had applied several post-processing steps such as gamma correction for image contrast enhancement.

Initially we had hoped that the algorithm used in [2] would be sufficient for the setup of our project despite the use of parallel ray, but with consideration of the scale of our project, we prefer the ITK Siddon-Jacobs implementation [4] due to better compatibility with Apache Beam. The ITK implementation would require conversion of CT DICOM series into NIfTI file format. This can be completed using the Python packages *dicom2nifti*, *dipy* and *nibabel*. We have also resampled the DICOM series into 3D volume with isotropic voxel size using trilinear interpolation in order to generate undistorted 2D images. This step is especially critical as the slice thickness between CT slices are usually larger than their respective pixel spacing.

We were able to get the maintainers of the Siddon-Jacobs package [4] in ITK to wrap the module in a PyPI package. This Python wrapper has allowed us to work in a dynamic scripting language instead of the C++ code that the original algorithm was written in. We are currently working on the parallelization of the synthetic X-ray generation using Apache Beam and Google Cloud Dataflow. This task is proving more complex that it initially seemed as most toolkits for DICOM to NIfTI conversion assume the existence of a local filesystem. Some additional work will be required to get the algorithms running en masse in the cloud.

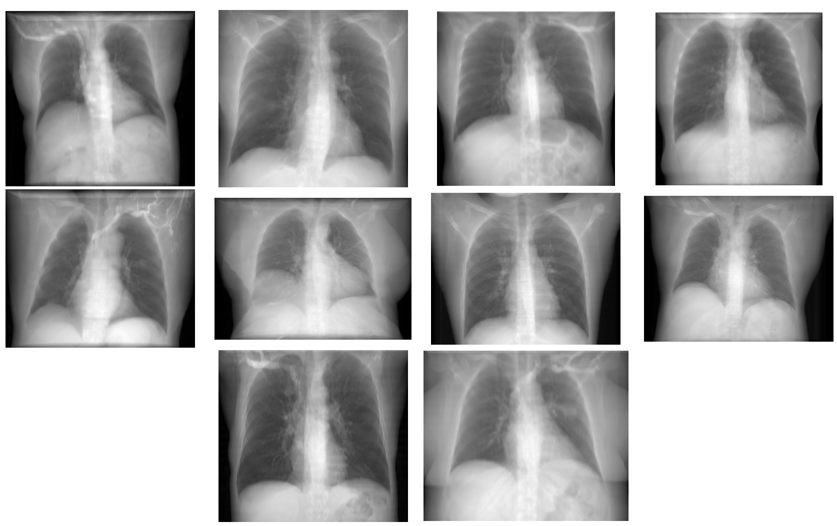

Figure 1: Synthetic chest X-rays from ten different patients generated from CT volumes using the ITK Siddon Jacobs ray-tracing algorithm.[4]

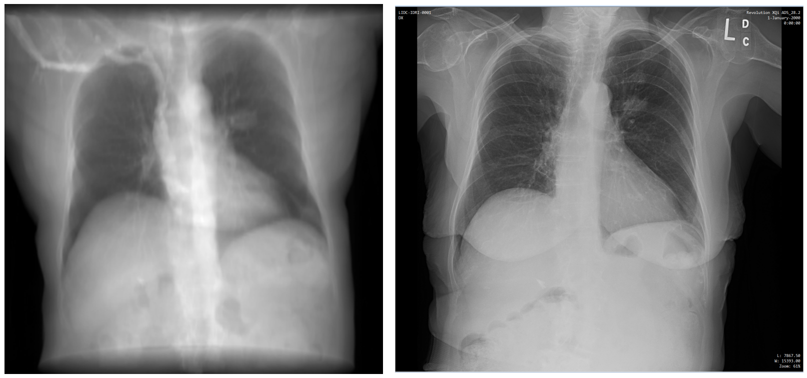

Figure 2: Digitally reconstructed chest X-ray from volumetric data (left) versus actual X-ray (right) from the same patient. Nodule in the upper left lung is clearly visible in both images.

3 Neural Network Model for 3D Reconstruction

The second part of our project entails the 3D reconstruction of human chest X-rays. For this task, we are designing a neural network model that can take a 2D X-ray image as input and output the respective 3D volume representation. There are two approaches for building such a neural network, namely: (1) building the model from scratch, (2) using transfer learning from an existing model for a similar task. In view of the industrial practices, the second option is more promising—especially when the resources are limited. We adopted an existing model with a slight network surgery and will perform our training with the synthetic X-ray images that we have created.

Our model is based on Henzler et al. [3], whose model is designed to generate 3D volumes from 2D scans of mammal skulls and has an encoder-embedding-decoder structure. In fact, several recent models for 3D reconstruction of various areas are also structured in this encoder/decoder manner. While the model in [3] has been developed using the Caffe framework, we wanted to work under Tensorflow/Keras framework so that the data pipeline will be more in line with our project setting (e.g. using TFRecord protobufs). We used two publicly available projects (i.e. caffe-tensorflow and MMdnn) to extract the pre-trained weights and to analyze the model structure under Caffe.

However, since some layers (e.g. deconvolutions) were not supported by the original projects, we needed to customize the source code for the model extraction/conversion. Furthermore, due to the incompatibility between Tensorflow and Caffe, there were some other technical difficulties in directly converting such a complicated and colossal network. Because of this, we constructed the model from scratch in Keras/Tensorflow instead of using the conversion tool, using the structure and weights we extracted from Caffe as a guide. In parallel with the model in [3], we manually performed network surgery on some of the layers.

The model consists of 139 layers, separated into the encoder (114 layers) and decoder (25 layers) sections. The encoder mainly consists of the basic block combinations, i.e. convolution-batch normalization-ReLU, shrinking the input from size 256×256 to a $8 \times 8 \times 256$ embedding tensor. On the other side of the model, the decoder consists of several deconvolution layers that expand the $8 \times 8 \times 256$ embedding tensor to a $128 \times 128 \times 128$ 3D volume as output. The details of the model structure can be found here.

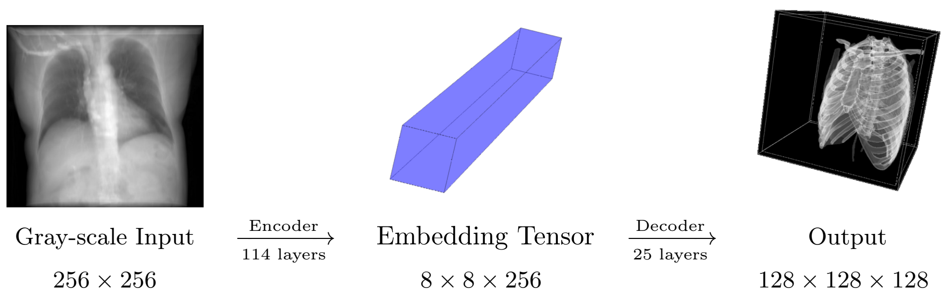

Figure 3: Overall structure of 3D reconstruction model

Upon implementing the model from scratch, we first tested our model based on the 2D X-ray of a mammalian skull, which is the original task of [3]. Our network was able to create a 3D volume based on the 2D X-ray image given. At the moment, the results have not reached the desired quality level, but we believe that the training with our synthetic human X-ray dataset will bring us to achieve a model that is able to reconstruct 3D volumes from human chest X-rays.

The current complexity of our model is slightly higher than that of [3]; thus, the Vapnik–Chervonenkis (VC) dimension of our model should be sufficient for the human chest 3D reconstruction task, which is similar to [3]. In view of the recent study in learning theory on neural networks [1], we can expect that there is a local optimum nearby the initial point under a large neural network, which is also in line with the effective training process that [3] had experienced for their tasks. Hence, we are optimistic for better results after some training.

In the next step, we will use our synthetic X-ray data to do batch training in the cloud and fine-tune our neural network for the specific task that we want to achieve. One of the ongoing tasks regarding the connection between the neural network model and the data are the appropriate scaled/cropped data of the 3D scan for the training process. We are also working on the appropriate visualization approach for the output.

References

[1] Zeyuan Allen-Zhu, Yuanzhi Li, and Yingyu Liang. Learning and Generalization in Overparameterized Neural Networks, Going Beyond Two Layers. *arXiv preprint arXiv:1811.04918*, 2018.

[2] Alex Chang and Abhishek Moturu. Detecting Early Stage Lung Cancer Using a Neural Network Trained with Patches from Synthetically Generated X-Rays. 2019.

[3] Phlipp Henzler, Volker Rasche, Timo Ropinski, and Tobias Ritschel. Single-Image Tomography: 3D Volumes from 2D Cranial X-Rays. 37(2):377–388, 2018.

[4] Jian Wu. Itk-based Implementation of Two-Projection 2D/3D Registration Method with An Application in Patient Setup for External Beam Radiotherapy. *Insight Journal*, page 784, 2010.