Classifying Knee Bony Anatomy Using U-Net (May 2020)

Matthew Vue and Maryam Vazirabad

***Abstract*— This paper explores biomedical data science with an image segmentation problem: identifying knee bony anatomy in order to later develop an abnormality diagnosis model. Almost 500 samples of knee x-ray images were taken from the dataset, Osteoarthritis Initiative (OAI). The OAI dataset features over 4,000 publicly available images of knee radiographs. The extracted images were manually annotated using OpenCV's Computer Vision Annotation Tool (CVAT), and the annotations converted into PNG masks. Using a U-Net convolutional neural network with Keras TensorFlow in the backend, this paper attempts to interpret the results of the semantic segmentation model in order to determine its worth in a practical scenario. The U-Net model's validation set obtained almost 97% accuracy, suggesting that it was a success.**

I. INTRODUCTION

Today, more than ever before, people are realizing the importance of managing and analyzing healthcare data. Harnessing this data is incredibly important in applications like medical imaging and clinical decision-making. Image processing is a field of study that requires attention in order to improve patient outcomes and achieve higher quality healthcare. The Osteoarthritis Initiative (OAI) is a 10-year observational study [1] of older men and women aimed at improving diagnosis and clinical care, specifically for treatment of osteoarthritis. This collection of radiographs from almost 5,000 participants can also be helpful for the classification of the bones of the knee: the femur, patella, tibia, and fibula. Datasets like this are useful for developing machine learning algorithms that can be beneficial in the medical field and have a significant impact on our world. With this dataset and the annotations created from the images, a convolution neural network, U-Net, was built in order to classify the bones into their respective categories.

II. RESEARCH/RELATED WORK

Before being introduced to this project, we were unaware of the convolutional neural network, U-Net. We were informed that this relatively new CNN is known to be one of the most powerful types of architecture today for biomedical image segmentation, and has been shown to reach human level performance in prediction. It is based off fully convolutional neural networks but enhanced to work with limited training images and yield accurate segmentation.

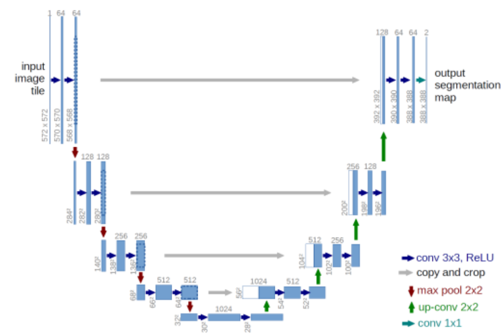

Figure 1: U-Net architecture

The architecture looks like a 'U' shape which justifies its name. The architecture consists of a contracting path to capture context while being computationally efficient, and a symmetric expanding path in which the pooling layers are replaced with upsampling operations. This retains image resolution and therefore allows for a more precise output [2].

While researching applications of U-Net models in order to better understand them, we came across an article that discussed implementing a convolution neural network to the same OAI dataset [3]. The researchers were attempting to automate diagnosis of osteoarthritis via machine learning algorithms, specifically U-Net. Unlike in our research, they did not take data solely from the OAI dataset. They also had access to the Multicenter Osteoarthritis Study, MOST dataset, which contains a similar dataset of knee radiographs. When developing image classification algorithms, it is important to not only use a relatively large number of images, but also limit bias by using data from different sources. Unfortunately, the organizers from MOST were not able to provide us access to their dataset at the time of our research, so OAI remained our only source of data.

The researchers had randomly selected a certain number of participants to use in their model. While we have used only around 500 images for our annotations, they were able to annotate 3,000. They achieved an accuracy score of 63% and reported a radiological OA diagnosis area under their ROC curve of around 0.93, suggesting the relatively high performance but recognizing that improvements could be made to enhance the model.

III. Dataset and Features

Before delving into manual annotation, preprocessing of the dataset, and developing the U-Net model, we must inspect the data we are given. There were exactly 4,926 participants in the Osteoarthritis Initiative study, and each participant had a set of knee radiographs. Both right and left knees were x-rayed, and images were taken throughout the 10-year study. We downloaded the first dataset, which were screening images taken soon after the participants had agreed to partake in the study.

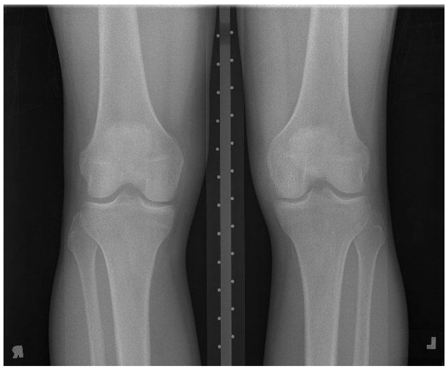

Figure 2: sample knee radiograph

The dataset includes data from men and women aged between 50–79 years old and 45–79, respectively. The mean age is around 61 years of age. There were 2804 females and 1992 males, slowing slight gender imbalance.

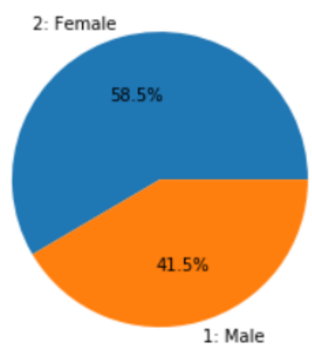

Figure 3: Slight gender imbalance in OAI dataset

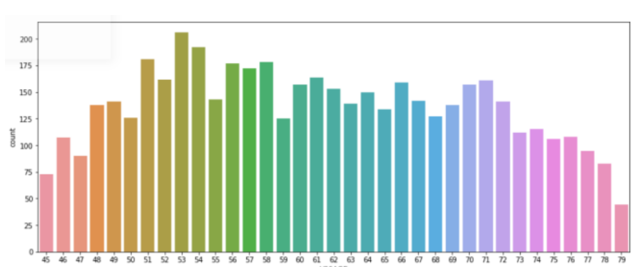

Figure 4: Age distribution in OAI dataset

IV. Annotation

In order to apply semantic segmentation to our selected images in the dataset and build our U-Net, we needed to manually annotate them. Every pixel must be attributed to a class label. To create our masks, we used OpenCV's Computer Vision Annotation Tool (CVAT) [4]. CVAT allows the user to upload images and create annotations that can be downloaded as various file types. Within CVAT, we used a polygon tool which allowed us to trace around the different segments of the radiograph. After we completed the annotations, we were able to download them in both an XML and PNG format, both of which were helpful to us. Each class (type of bone) is represented by a different color. We were able to annotate almost 500 images.

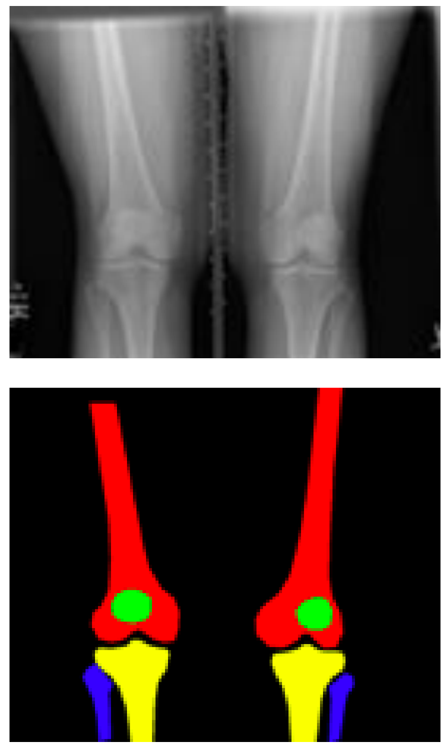

Figure 5: Radiograph and its corresponding mask

V. Data Preprocessing

At first, the radiographs and corresponding masks were not the same size, which can cause issues down the road of implementing our model. We resized the radiographs and masks to the following dimensions: (256 x 256 x 3). The three represents the three-color channels, RGB. Therefore, each pixel in the radiographs and masks is a (1 x 3) array. For example, [255, 0, 0]. Since the radiographs and masks are in RGB, the values within the (1 x 3) array range from 0 to 255. Based on our annotations, the table below shows pixels values and their corresponding class.

Pixel	Class
[0, 0, 255]	Femur
[255, 0 ,0]	Fibula
[0, 255, 0]	Patella
[255, 255, 0]	Tibia
[0, 0, 0]	Background

Figure 6: Pixel value and corresponding class

Next, we divided each mask by 255 in order to get only zeros and ones as values within the (1 x 3) arrays. For example, [255, 0, 0] / 255 = [1, 0, 0]. We then proceeded to create a new mask set with each mask having the following dimensions, (256 x 256 x 5). The five is necessary in order to segment out 5 classes within each radiograph. Therefore, each pixel is now a (1 x 5) array. For example, a pixel could be converted into [1, 0, 0, 0, 0]. The index at which the one is located indicates what class the pixel belongs to. The table below shows the five (1 x 3) pixels and their corresponding (1 x 5) pixel.

Original (1 x 3) Pixel	(1 x 3) Pixel after dividing by 255	Corresponding (1 x 5) pixel
[0, 0, 255]	[0, 0, 1]	[1, 0, 0, 0, 0]
[255, 0 ,0]	[1, 0, 0]	[0, 1, 0, 0, 0]
[0, 255, 0]	[0, 1, 0]	[0, 0, 1, 0, 0]
[255, 255, 0]	[1, 1, 0]	[0, 0, 0, 1, 0]
[0, 0, 0]	[0, 0, 0]	[0, 0, 0, 0, 1]

Figure 7: (1 x 3) Pixels and their corresponding (1 x 5) pixels

Once we have a new mask set with each mask having the following dimensions, (256 x 256 x 5), we can feed the data into our U-Net model. We divided the data into a test and train set with 10 percent of the data going into the train set and 90 percent going into the training set.

VI. U-Net Model

We utilized the neural network library, Keras [5], to implement our U-Net model. The model used for this project was based on a GitHub repository by Harshall Lamba [6]. The U-Net model had 1,179,477 parameters. The model starts out with 16 filters and that number increases as the model progresses. The dimensions of the filters in the convolutional layers are (3 x 3). To add, the filters in the max pooling layers are (2 x 2). The dimensions of the input are (256 x 256 x 3) and the dimensions of the output is (256 x 256 x 5). The activation function used was the Relu activation function. To prevent overfitting a dropout rate of .05 was applied to the U-Net model. The max pooling layers are used to obtain the most important information from the feature maps. In order to do so, the max pooling layers reduce the dimensions of the feature maps. Total, there are 19 convolutional layers in our Unet model. In the output layer the softmax function is used as the activation function. The softmax function maps each pixel to one of the five classes. The probabilities should therefore add up to one. To evaluate how well the model is doing, the binary cross entropy function was chosen as our loss function. In addition, the Adam optimizer was chosen as our model optimizer. The accuracy of the model was evaluated during training and testing.

```
Model: "model_1"
__________________________________________________________________________________________________
Layer (type)                    Output Shape         Param #     Connected to
==================================================================================================
img (InputLayer)                (None, 256, 256, 3)  0
__________________________________________________________________________________________________
conv2d_2 (Conv2D)               (None, 256, 256, 16) 448         img[0][0]
__________________________________________________________________________________________________
batch_normalization_2 (BatchNor (None, 256, 256, 16) 64          conv2d_2[0][0]
__________________________________________________________________________________________________
activation_2 (Activation)       (None, 256, 256, 16) 0           batch_normalization_2[0][0]
__________________________________________________________________________________________________
max_pooling2d_1 (MaxPooling2D)  (None, 128, 128, 16) 0           activation_2[0][0]
__________________________________________________________________________________________________
dropout_1 (Dropout)             (None, 128, 128, 16) 0           max_pooling2d_1[0][0]
__________________________________________________________________________________________________
conv2d_4 (Conv2D)               (None, 128, 128, 32) 4640        dropout_1[0][0]
__________________________________________________________________________________________________
```

Figure 8: Segment of the U-Net model

VII. Experiments/Results/Discussion

We annotated around 500 radiographs. At first, we tried running different numbers of epochs. We decided to run 43 epochs which gave us sufficient results. The model for each pixel gave us a (1 x 5) array with probabilities. The index with the max probability indicates what class the pixel was predicted to belong to. For example, the pixel below indicates that the pixel belongs to the background. Notice that the probabilities add up to one.

```
[3.2077790e-03, 2.2153831e-03, 3.7505671e-03, 5.9767539e-04, 9.9022865e-01]
```

By running 43 epochs, we achieved around a 97 percent validation accuracy.

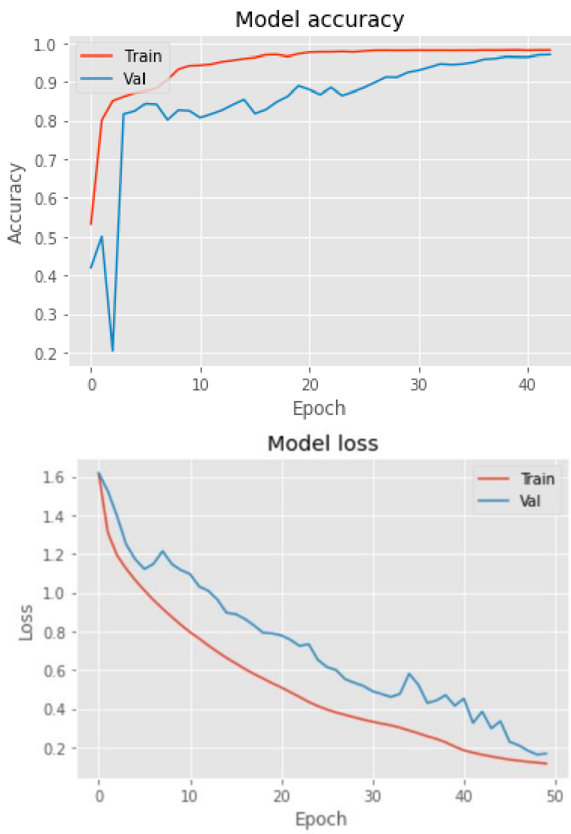

Figure 9: Model accuracy/loss versus number of epochs

One can see in Figure 7, as the number of epochs increase, the validation accuracy becomes close to the training accuracy. This is desirable. We noticed the model sometimes had issues classifying the area around the femur and patella. The model did have a high number of parameters. To combat this issue, future work could be to continue to run the model with different specifications and try using a lower number of parameters.

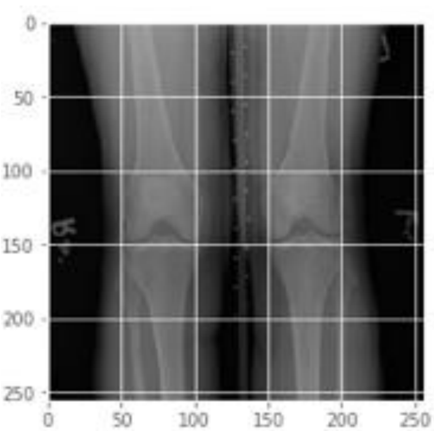

Figure 10: Radiograph in the validation set

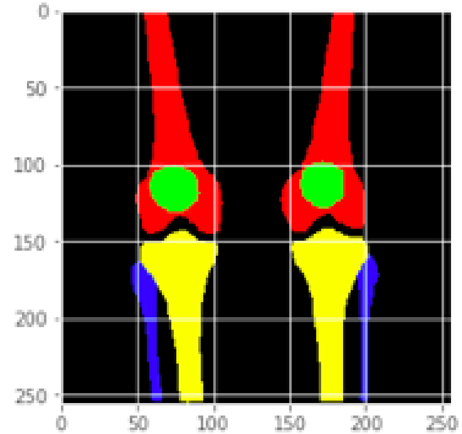

Figure 11: Predicted mask of the radiograph in Figure 8

VIII. Conclusion

In all, our task was to use a U-Net convolutional neural network to segment out the different parts of a knee radiograph: the femur, patella, tibia, and fibula. Our dataset consisted of around 500 radiographs and masks. We obtained the radiographs from the Osteoarthritis Initiative (OAI)

The implementation mentioned in this report performed well with the dataset and annotated images. The U-Net model was able to segment the various parts of the knee, as well as the background. We achieved around 97 percent validation accuracy. Future work includes lowering the number of parameters and building a more robust model to create crisp segmentation. Since there can be much variation in these radiographs, in the future more radiographs and masks could be added to our dataset to make the model more versatile.

IX. References

[1] OAI. (n.d.). [Online].

[2] Ronneberger, O., Fischer, P. and Brox, T., 2015. *U-Net: Convolutional Networks For Biomedical Image Segmentation*. [online] arXiv.org.

[3] F. Ambellana, A. Tacka, M. Ehlkeb, S. Zachowa, "*Automated Segmentation of Knee Bone and Cartilage combining Statistical Shape Knowledge and Convolutional Neural Networks*," MIDL 2018 Conference, 11-April-2018.

[4] Intel. 2019. *Computer Vision Annotation Tool: A Universal Approach To Data...*. [Online].

[5] "Keras: The Python Deep Learning library," Home - Keras Documentation. [Online].

[6] Harshall, L, *Applying UNET Model on TGS Salt Identification*. Challenge hosted on Kaggle, https://github.com/hlamba28/UNET-TGS