Anomaly Detection of Arm X-Ray Based on Deep Learning

Jingchen Wang
Phillips Exeter Academy, 20 Main St., Exeter, NH, 03833
hzhc_wjdi@163.com

Abstract. The goal of this paper is to determine whether the arm has a fracture by detecting the X-ray of the human arm. This paper used the Keras deep learning framework and use the NASNetMobile model for training. The data set is MURA-v1.1, and the test accuracy on the verification set is about 70%. After downloading X-ray photographs of fractured arms, this paper performed an anomaly detection of the single image to test the accuracy of the model.

1. Introduction
In 2015, the misdiagnosis rate of common diseases in China was as high as 27%, and that of major diseases was as high as 40%. The misdiagnosis of fractures is the most prominent issue doctor’s encounter at the early periods of a patient's treatment. According to global news and reports, more than 1.7 billion people worldwide have musculoskeletal problems, and there are about 30 million emergency cases every year, in which its number increases with an exponential rate. Skeletal problems have become the most common and severe chronic disease. Children and the elderly are the two groups of people most likely to be exposed to upper limb injuries. Since the difficulty of diagnostics hinders the doctors' correct judgment by only examining the outside of one's arm, an X-ray machine become necessary during this process.

The attempt of using machines to determine the condition of patients' bones by observing the patient's X-ray film was introduced in many hospitals across the globe. With multiple tryouts and failures, there still hasn't been a significant decrease in the misdiagnosis rate.

This project is aimed to extract details from the upper limbs of patients, including the hands, wrists, arms, elbows, and arms, and produce a result as to the bone's status. Its significance lies in which the system was built exclusively for the doctors to diagnose only the few body parts with high accuracy simultaneously. Deep learning was chosen to be the key framework for the machine to judge the fracture of the upper limbs of the patient. This project not only achieves high accuracy of judgment, but also realizes the convenience of operation. The user can automatically recognize the X-ray film by placing it in a folder on the desktop.

Determining whether a radiographic study is normal or abnormal is a critical radiological task: a study interpreted as normal rules out disease and can eliminate the need for patients to undergo further diagnostic procedures or interventions. The musculoskeletal abnormality detection task is particularly critical as more than 1.7 billion people are affected by musculoskeletal conditions worldwide (BMU, 2017). These conditions are the most common cause of severe, long-term pain and disability (Woolf & Pfleger, 2003), with 30 million emergency department visits annually and increasing. Our dataset, MURA, contains 9,045 normal and 5,818 abnormal musculoskeletal radiographic studies of the upper
extremity including the shoulder, humerus, elbow, forearm, wrist, hand, and finger. MURA is one of the largest public radiographic image datasets.

This paper found that model performance is comparable to the best radiologist’s performance in detecting abnormalities on finger and wrist studies. However, model performance is lower than best radiologist’s performance in detecting abnormalities on elbow, forearm, hand, humerus, and shoulder studies. This paper made the dataset freely available to encourage advances in medical imaging models.

2. Data Set
MURA (musculoskeletal radiographs) is a large dataset of bone X-rays. Algorithms are tasked with determining whether an X-ray study is normal or abnormal.

Musculoskeletal conditions affect more than 1.7 billion people worldwide, and are the most common cause of severe, long-term pain and disability, with 30 million emergency department visits annually and increasing. This paper hopes that the dataset can lead to significant advances in medical imaging technologies which can diagnose at the level of experts, towards improving healthcare access in parts of the world where access to skilled radiologists is limited.

MURA is one of the largest public radiographic image datasets. This dataset is made available to the community and hosting a competition to see if your models can perform as well as radiologists on the task.

This paper chose to use the MURA-v1.1 data set, which consisted of 36,808 training samples and 3,197 validation samples. It contained 14982 cases of upper extremity musculoskeletal X-rays. Each case contains one or more images that are manually labeled by a radiologist.

3. The Architecture
The MURA dataset contains a training file and a validation file. In each files lies the different cases of each patient, and their X-ray photographs. The MURA dataset was then processed through this file called download and convert MURA, which converts it into this data file that also has a train and a validation file. However, the images are now classified not by the number of cases, but whether the patients’ situations are normal. By doing this, the neural network can classify this information more clearly and solves the problem that certain directories cannot be found.

Convolutional Neural Network trained the dataset, and the system uses NASNetMobile because it gave the best output of accuracy for the system. The batch size, which is how many images are being dealt every cycle is 32, and the system reshapes all the images into 224 times 224 with all the red, green, and blue channels. The model is then being transferred and used through the functions defined below.

Then, the loss of the entire system can be calculated by using the cross entropy loss function, which later enables us to calculate the accuracy of the system’s prediction. And by obtaining the images placed on the desktop, the system will do its prediction.

3.1. Framework
Keras is a high-level neural networks API, written in Python and capable of running on top of TensorFlow, CNTK, or Theano. It was developed with a focus on enabling fast experimentation. Being able to go from idea to result with the least possible delay is key to doing good research. Keras was chosen as our deep learning framework because of the conveniences of the system.

3.2. Model
The paper adopted NASNetMobile model for pre-training. The paper set batch-size as 32 and epoch equal to 5, loaded data with the pre-trained model NASNetMobile, and saved the new model and weights. Google's recently launched NASNet is the best model in the field of image recognition. Developing neural network image classification models often requires significant architecture engineering. NASNet is built using the Neural Structure Search (NAS) framework. It is aimed to build a network architecture using data-driven and intelligent methods rather than intuition and experimentation. The composition of NasNet is composed of two network elements.
Figure 1. Architecture of the best convolutional cells (NASNet-A) with $B = 5$ blocks identified with CIFAR-10. The input (white) is the hidden state from previous activations (or input image). The output (pink) is the result of a concatenation operation across all resulting branches. Each convolutional cell is the result of $B$ blocks. A single block is corresponds to two primitive operations (yellow) and a combination operation (green).

The stacking scheme of these two cells are as follows:

Figure 2. Scalable architectures for image classification consist of two repeated motifs termed Normal Cell and Reduction Cell. This diagram highlights the model architecture for CIFAR-10 and ImageNet. The choice for the number of times the Normal Cells that gets stacked between reduction cells, $N$, can vary in our experiments.
The reason of choosing NASNet is that it does well in accuracy. Here is the target detection using NASNet in figure 3.

![Target detection using NASNet](image)

**Figure 3.** Target detection using NASNet

### 3.3. ReLU Nonlinearity

The standard way to model a neuron’s output $f$ as a function of its input $x$ is with $f(x) = \tanh(x)$ or $f(x) = (1 + e^{-x})^{-1}$. In terms of training time with gradient descent, these saturating nonlinearities are much slower than the non-saturating nonlinearity $f(x) = \max(0, x)$. Thus the relu activation function was selected. Activation functions are demonstrated below.

![Sigmoid](image)

**Figure 4.** Sigmoid

![Tanh](image)

**Figure 5.** Tanh

![Relu](image)

**Figure 6.** Relu

![Leaky-Relu](image)

**Figure 7.** Leaky-Relu
3.4. Training on Mac CPUs
The entire training process lasted for three days on the Mac OS system, where each of the 36808 images were trained five times, one in each epoch. The screenshot of the training process is shown in the figure 8.

![Figure 8. Part of the training process](image)

4. Results on single X-pictures
After downloading X-ray photographs of fractured arms, this paper performed an anomaly detection of the single image to test the accuracy of the model. The testing includes 50 subjects, which are manually labeled from 1 to 50 and path set to the desktop. The images were X-ray films of specific parts of the upper limb which were already diagnosed, and the system provides if the bone is fractured. It then the system detects if the prediction matches the labels, and yields a accuracy percentage in the end. The accuracy of this system is about 73%. Here is the part of testing process in Figure 9.

![Figure 9. Part of the testing result](image)

5. Discussion
The objective of this project is to determine whether the arm has a fracture by detecting the X-ray of the human arm. Applying the Keras deep learning framework and the NASNetMobile model for training provided high accuracy and thus yielded better results. The data set is MURA-v1.1, and the test accuracy on the verification set is about 70%. After downloading X-ray photographs of fractured arms, this paper performed an anomaly detection of the single image to test the accuracy of the model. Due to the uneven number of positive and negative samples, the training accuracy is not perfected. Therefore obtaining more X-ray photographs to balance the number of positive and negative samples and continue training based on the existing model is key to the prospect of this project.

References
[1] Barret Zoph, Vijay Vasudevan, Jonathon Shlens and Quoc V. Le. Learning Transferable Architectures for Scalable Image Recognition.
[2] AIMI. Artificial intelligence in medicine & imaging: Available labeled medical datasets. https://aimi.stanford.edu/available-labeled-medical-datasets. [Online; accessed 2- December-2017].

[3] Bejnordi, Babak Ehteshami, Veta, Mitko, van Diest, Paul Johannes, van Ginneken, Bram, Karssemeijer, Nico, Litjens, Geert, van der Laak, Jeroen AWM, Hermsen, Meyke, Manson, Quirine F, Balkenhol, Maschenka, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. Jama, 318 (22): 2199–2210, 2017.

[4] Deng, Jia, Dong, Wei, Socher, Richard, Li, Li-Jia, Li, Kai, and Fei-Fei, Li. Imagenet: A large-scale hierarchical image database. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pp. 248–255. IEEE, 2009.

[5] Woolf, Anthony D and Pfleger, Bruce. Burden of major musculoskeletal conditions. Bulletin of the World Health Organization, 81 (9): 646–656, 2003.