RSNA Bone Age Prediction

UC San Diego
Electrical and Computer Engineering
JACOBS SCHOOL OF ENGINEERING

Juan Camilo Castillo, Yitian Tong, Jiyang Zhao, Fengcan Zhu

Abstract

A fast, automated and accurate machine learning model for bone age assessment is proposed in this project. Bone age assessment is a common clinical practice in the diagnosis of child development. The error of BoneXpert, the system in use for now, is about 8.4 months[1]. For our project,we trained various models using over 12k radiological bone pictures with associated labels each corresponding to a patient's sex and age. Both regression based models and CNN with transfer learning, along with multiple image processing and feature extraction methods were used. Finally using the VGG16 pretrained model with attention mapping focused architecture we were able to achieve a mean absolute errors less than a year.

Dataset and Image Preprocessing

The goal our project is to develop an algorithm which can most accurately determine bone age based on X-ray scans of hands.

The original dataset was released by Pediatric Bone Age Challenge organized by RSNA. This training set includes 12611 pediatric hand radiographs. The label files includes corresponding skeletal ages and gender for each radiographs. Some sample hand radiographs are shown in figure 1 below. Watersheding and blob detection method were used to extract hand images shown in figure 2.

Figure 1. Original images

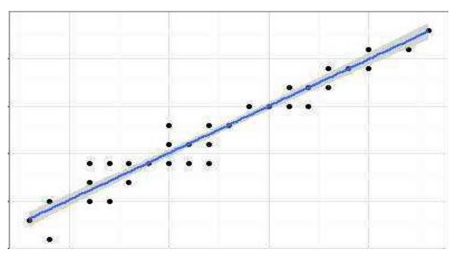
Figure 2. Extracted hand images

Method

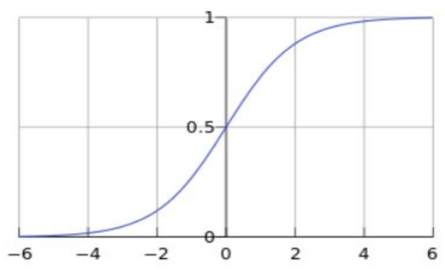
Two major approaches are implemented to predict the bone ages

1. Regression based approach:
   - Linear regression, Logistic regression and preprocessing

   Regression models:



Linear regression

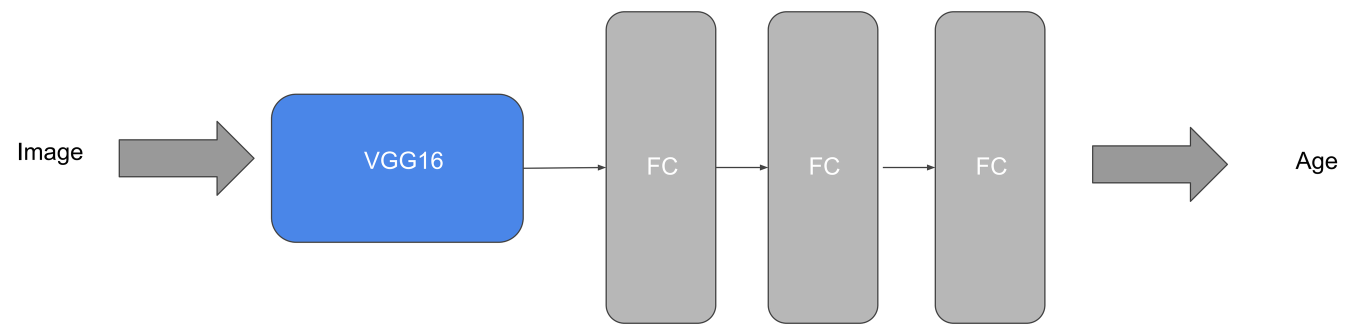


Logistic regression

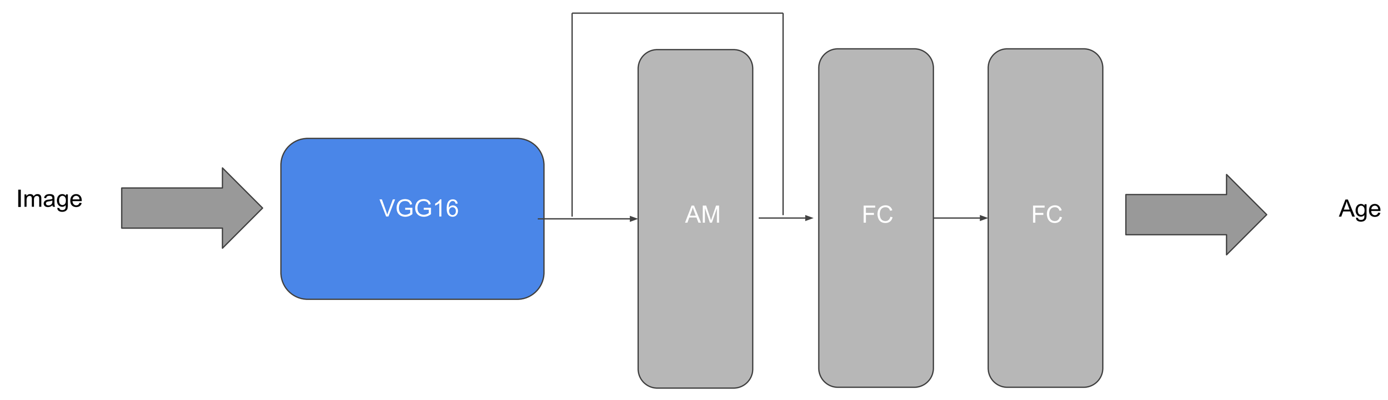
Preprocessing for the model:
- Group data by male and female
- Resize and random crop to 128*128
- PCA with 60% features left

2. Deep learning:

   For feature extraction, the pretrained VGG model is used
   - Model 1: VGG + 3 Fully Connected layers
     - 1,576,961 trainable Parameters
     - (256 x 256) input image size
     - GAP used to compress output of VGG



   - Model 2: VGG + Attention Mapping + 2 Fully Connected layers
     - 565,058 trainable Parameters
     - (500 x 500) input image size ~ higher resolution
     - Added attention mapping based on locally connected layer and GAP



Results

Mean absolute error (MAE) is used to measure the difference between predictions and the labeled ages. Performances of two regression models and two deep learning models are shown below in table 1. The best performance was achieved using deep learning model 2. Training history of model 2 is shown in figure 3 and a sample of its prediction results are shown in figure 4.

	Male	Female	Both Gender
Validation MAE Linear Regression (Months)	30	27	32
Validation MAE Logistic Regression	34	31	36
Train MAE Model1	17.34	17.58	18.07
Validation MAE Model1	15.21	16.15	16.88
Train MAE Model2	8.14	8.38	11.19
Validation MAE Model2	9.82	10.78	11.45

Table 1. Training results.

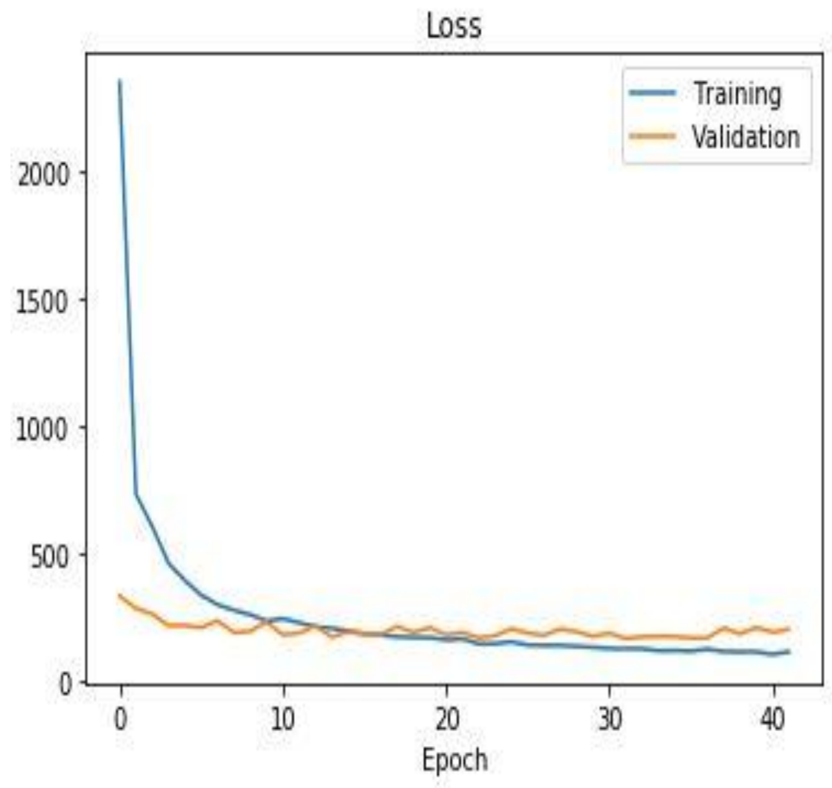


Figure 3. Training history for model 2 .

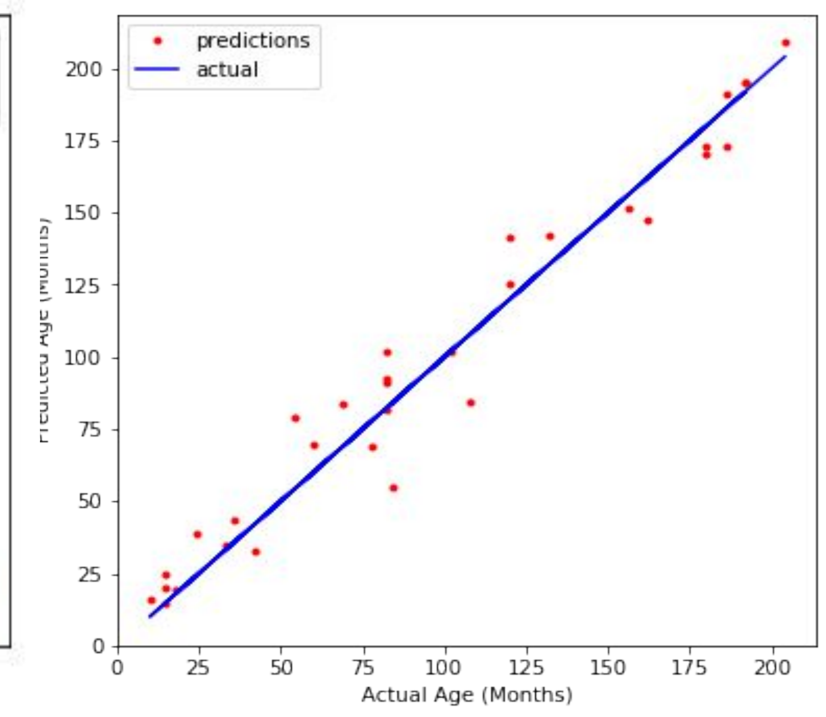


Figure 4. Predictions for model2 with both gender

Discussion

Referring to the training history in figure 3, the validation loss reached its lowest point in a few epochs and oscillated while the training loss still decreased. This indicated that the trained model was overfitting to the training set.

In figure 4, predictions were more accurate at two ends than in the middle. This may caused by the fact that girls grow up much faster than boys around the age 10, which leads to high variances when train two genders together. This phenomenon cannot be observed when two genders were trained separately.

Four samples of attention maps output from Model 2 and their original images are shown in figure 5. It shows that the carpal and metacarpal bones have more information on the bone age prediction than other areas of the hand.

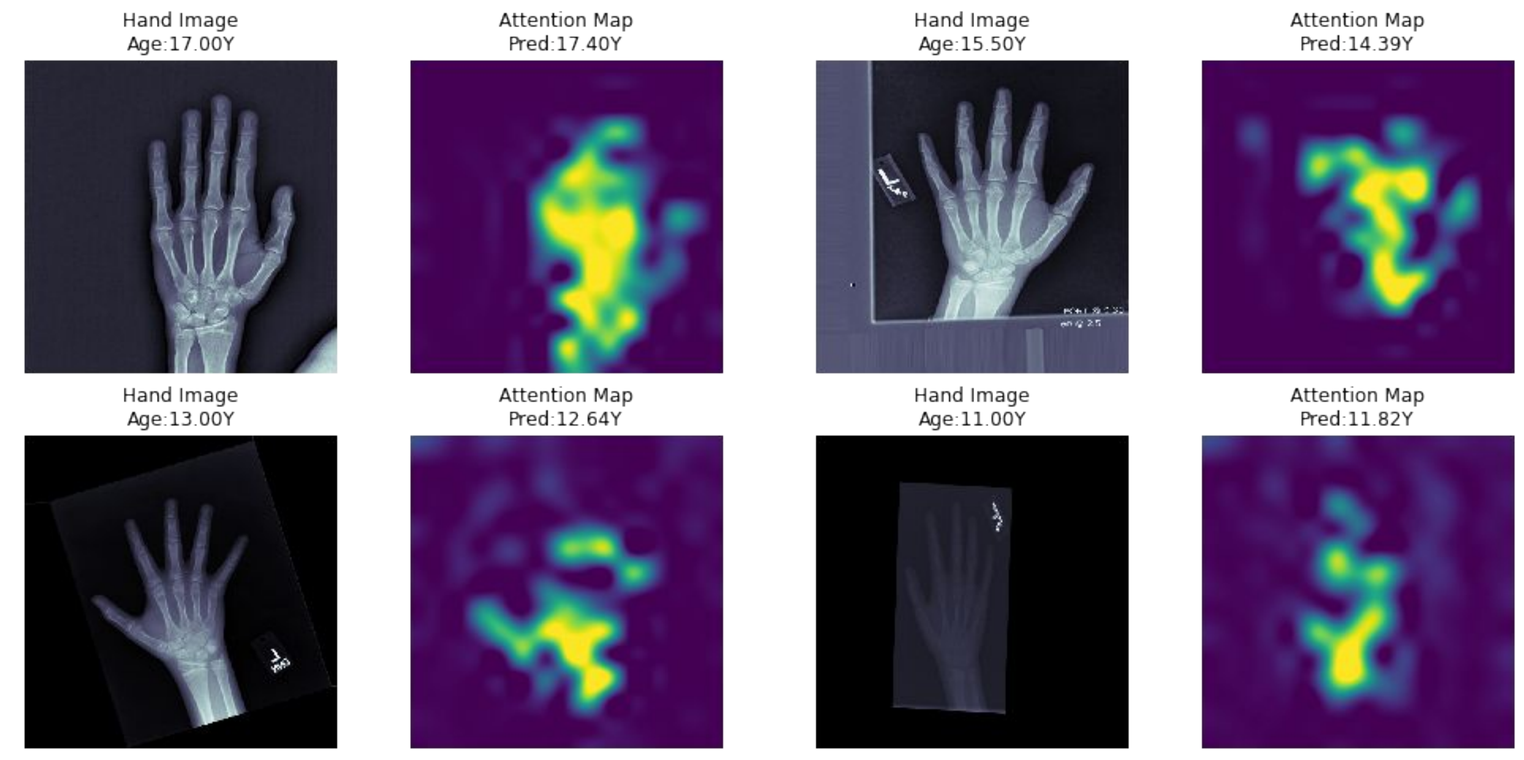


Figure 5. Attention map samples

Conclusions

We achieved MAE of 9.82/10.75 months for male and female using VGG16 pretrained model and attention mapping. The result is similar to the 9.84/11.16 achieved by Fully Automated BAA[2] using the same dataset. The most salient features for predicting the age of an individual clearly seems to be the bones found in the wrist and middle of the hand. Future work can include trying different architectures and analyzing the associated efficacy of the implemented designs.

Contact

- J. Castillo : jcc040@eng.ucsd.edu
- Jiyang Zhao: jiz381@ucsd.edu
- Yitian Tong: yit050@ucsd.edu
- Fengcan Zhu: f9zhu@eng.ucsd.edu

Reference

[1] Vincenzo De Sanctis, Salvatore Di Maio, Ashraf T. Soliman, Giuseppe Raiola, Rania Elalaily, and Giuseppe Millimaggi. *Hand X-ray in pediatric endocrinology: Skeletal age assessment and beyond.* 2014 Nov; 18(Suppl 1): S63–S71.

[2] Hyunkwang Lee, Shahein Tajmir, Jenny Lee, Maurice Zissen, Bethel Yeshiwas, Garry Choy, Tarik Alkasab, and Synho Do. *Fully-automated Deep Learning System for Bone Age Assessment* .Journal of Digital Imaging, Pp. 1-15. 3/8/2017.