NYU School of Medicine
NYU LANGONE MEDICAL CENTER

NYU GRADUATE SCHOOL OF ARTS & SCIENCE

DEEP CONVOLUTIONAL NEURAL NETWORKS FOR PREDICTION OF PEDIATRIC BONE AGE

Benedict Vincent C. Guzman*[1], Yuhan Wang*[2], Yun Qin *2

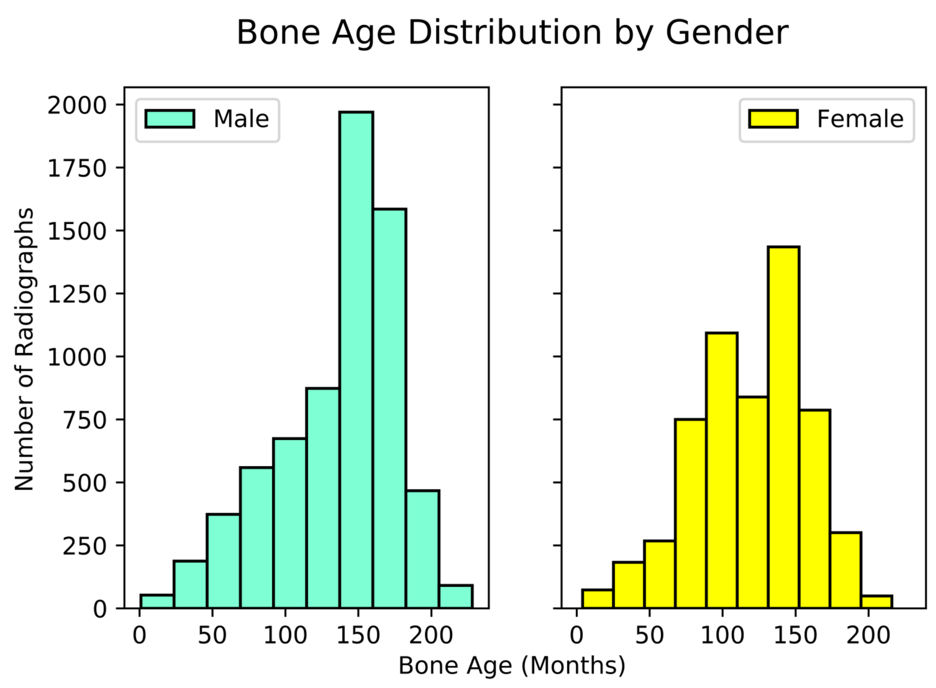
NYU Langone Health[1], New York University, Department of Biology[2],

BACKGROUND & PROBLEM DEFINITION:

- Pediatric bone age assessment via hand radiographs has been used in the clinical practice for > 75 years to determine chronological age, nutrition, genetics, and disease states.
- 2 conventional manual methods to perform bone age assessment .
  A. Tanner Whitehouse (TWS2): designates scores to different bones of the hand and wrist (takes average of 7.9 minutes to complete).
  B. Greulich-Pyle (GP): compare radiograph with atlas (takes average of 1.4 minutes but tedious)
- Novel and automated tools exist such as BoneXpert, but problematic because of its sensitivity to image quality and does not utilize carpal bones for bone age determination.
- Deep learning: feasible solution for prediction with high efficiency and accuracy, thus the RNSA launched the Pediatric Bone Age Challenge in 2017 to challenge the community to create models to perform the task automatically.

DATASET

- 12,611 radiographs (PNG file) for training set and 200 radiographs (PNG file) for test set.
  *6833 males and 5778 females in training set.
  *127.32 months overall average age.
  *135.03 months for males, 117.88 for females.



Bone age distribution by gender

IMAGE AUGMENTATION:

- Training dataset was split as 90% (11439 radiographs) for training and 10% (1262 radiographs) as validation.
- Resized images to 3x512x512 (from 1x2000x1500).
- Performed random horizontal flips, random cropping, and normalization using values of mean = [0.485, 0.456, 0.406] and std = [0.229, 0.224, 0.225].

DEEP CONVOLUTIONAL NETWORK MODELS:

- Explored three different models: Inception V3, ResNet-34, and VGG16.
- Performed bone age prediction on these baseline models.
- Produced subpar results of an average of 31.5 months as calculated by mean absolute error (MAE), which is the challenge's evaluation metric.

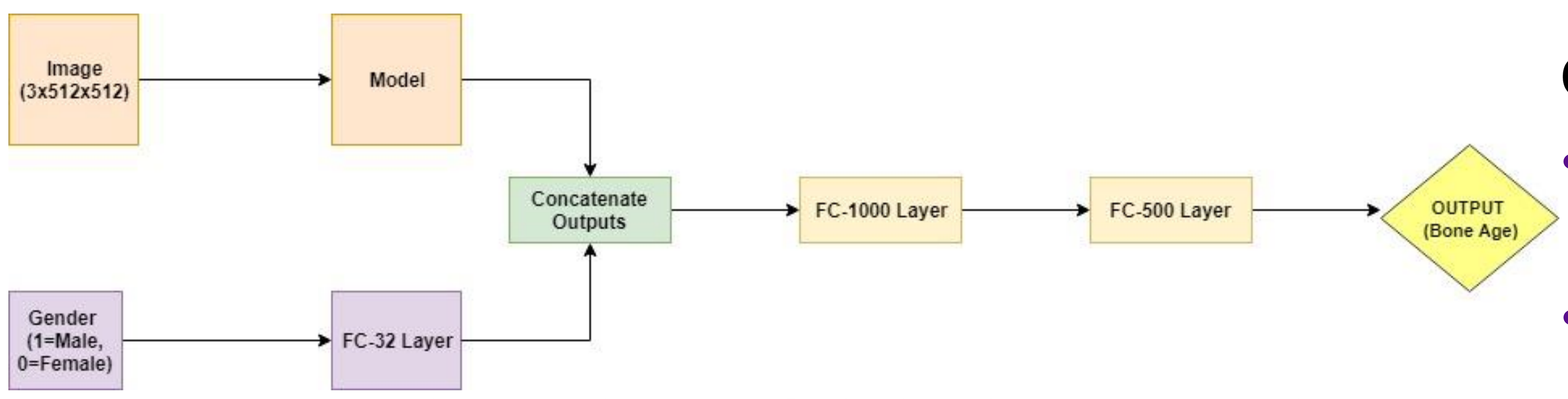
$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|x_i - x|$$

- Improved the baseline models by incorporating the gender variable.

MODEL HYPERPARAMETERS:

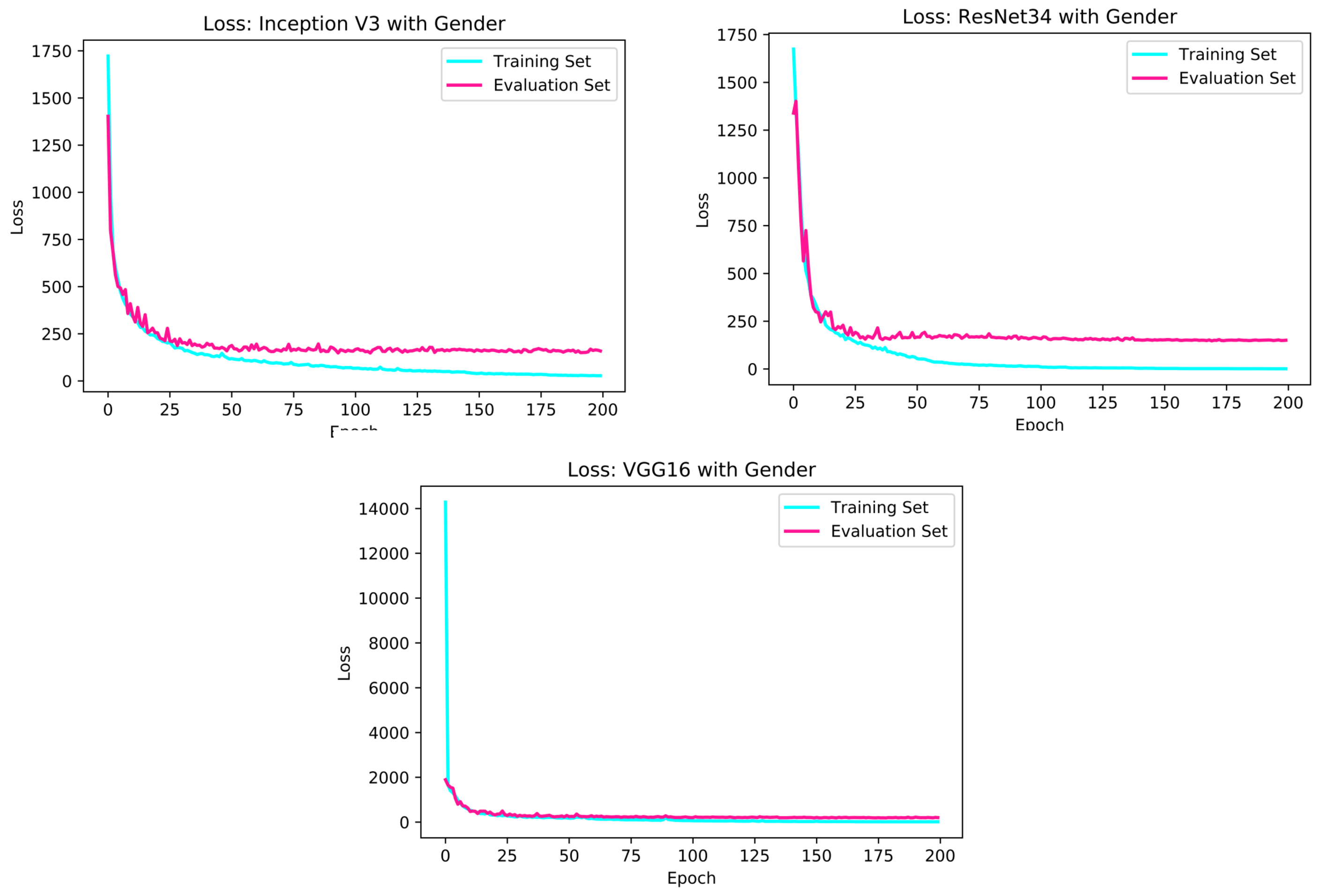
- Mean Squared Error (MSE) as loss function.
- Adam as optimizer with learning rate=0.001, β_1=0.1, β_2=0.1, and ϵ=1e-08.
- Batch size=16, Epoch=200.
- ReduceLROnPlateau() with factor=0.1, min_lr=0.0001, patience=9, and cooldown=5.

Sample radiographs. L-R: Ground truth: 138 months, 126 months, and 120 months



Our general approach to modify the baseline DCNN models

RESULTS:

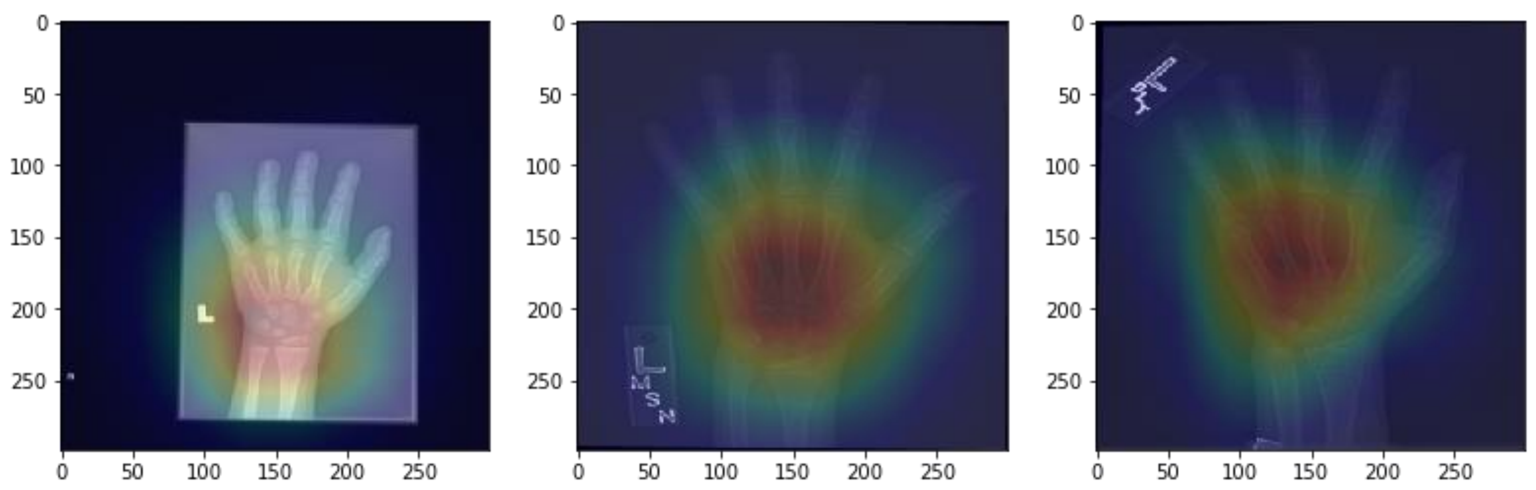


Model	MAE (in Months)	MAE (in Years)
Inception V3	12.09	1.01
ResNet-34	13.36	1.11
VGG16	9.63	0.80

Performance of the modified DCNN for test set

- Best Model for Test Set: VGG16

GRADCAM ANALYSIS



Sample Grad-CAM analysis. L-R: Ground truth: 73.26 months, 129.5 months, 167.69 months

- Grad-CAM analysis of the feature map of last convolutional layer before the fully connected layers of the best model.
- VGG16 does not look at one fixed spot in the radiograph to determine bone age.
- Model focuses on more distal parts as the age increases.

CONCLUSION AND FUTURE WORK

- Best Model of VGG16 with 9.63 MAE, which is 30th place out of 260 participants in challenge.
- Improve model by increasing dataset, extend training time, and create an ensemble of transfer learning models.