Improved Deep Learning Model for Bone Age Assessment using Triplet Ranking Loss

Byeong-Uk Bae
Vuno Inc.
Seoul, South Korea
byuk.bae@vuno.co

Woong Bae
Vuno Inc.
Seoul, South Korea
iorism@vuno.co

Kyu-Hwan Jung
Vuno Inc.
Seoul, South Korea
khwan.jung@vuno.co

Abstract

We propose a deep learning model using triplet ranking loss for bone age assessment. We build a hand segmentation network and transformation network for preprocessing to normalize x-ray images. We also added a triplet ranking loss to regression loss so that the embedded feature can be ordered. As the results, the model learns with ordered features show better performance. We evaluated our model with RSNA bone age assessment competition dataset.

1 Introduction

The established approach to bone age assessment (BAA) using x-ray images of the left hand are the Greulich-Pyle (GP) and Tanner-Whitehouse (TW2) methods. However, both methods take considerable time for assessment and have different accuracies depending on the experience of the reader. In order to solve the above problems, an automatic bone age assessment system based on computer vision and machine learning was developed (BoneXpert). Another approach has been proposed using convolution neural networks (CNNs) with leading performances in the RSNA BAA competition.

In this paper, we propose an improved deep learning model using triplet ranking loss along with the regression loss to perform bone age assessment. The proposed triplet ranking loss regularizes the learning process by making the distance of the feature vector of similar labeled data close to each other and further away from opportunity cases. We trained our model and evaluated with the RSNA BAA competition dataset. The trained model using triplet ranking loss reduces the mean absolute distance (MAD) from 5.82 to 4.77 months.

2 Methods

We describe our proposed deep network architecture for bone age prediction using triplet loss. As a first step, we constructed a hand segmentation network for cropping the hand region and a transformation network to center the hand position. In the prediction phase, we extract the high layer embedded features of arbitrarily sampled triplets and calculate the triplet ranking loss between them. The final loss is calculated as the weighted sum of regression loss and ranking loss (See Figure 1).

2.1 Hand x-ray image preprocessing

Most x-ray images have different appearance types because they are taken with different equipments. Therefore, in order to improve the performance of the model, it is necessary to normalize the images.

1st Conference on Medical Imaging with Deep learning (MIDL 2018), Amsterdam, The Netherlands.

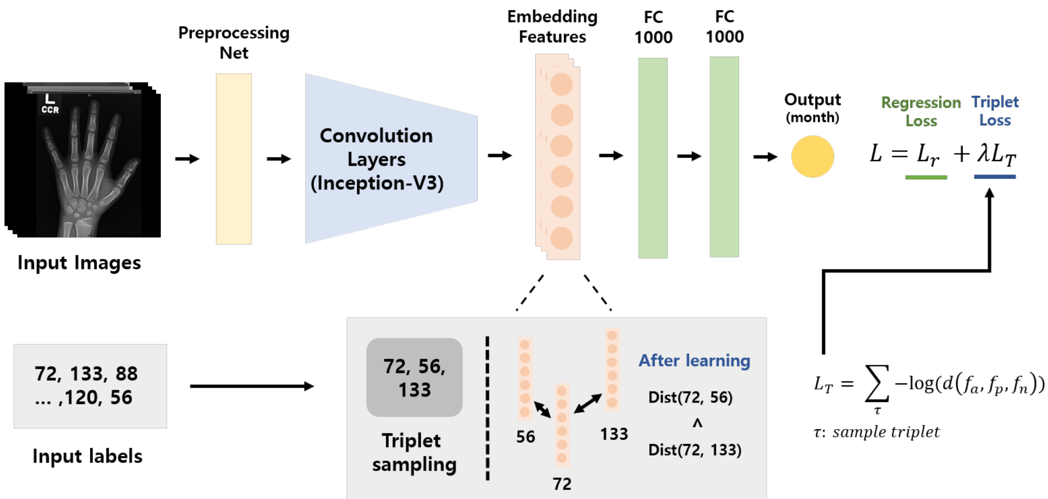


Figure 1: Illustrate proposed deep learning architecture.

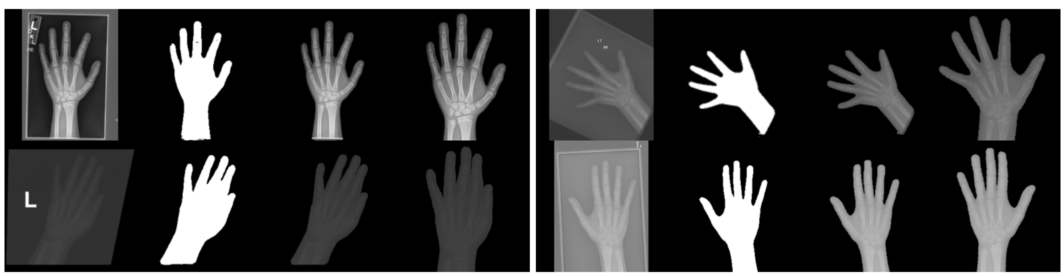


Figure 2: Preprocessing network results. left to right: input images, generated hand mask, applied mask to remove background, transformed hand to locate center of image

We applied the idea suggested in [2] that proposed a hand segmentation network and transformation network to relocate hand position by computing a similarity transform matrix. Figure 2 shows the input image and preprocessing pipeline results.

2.2 Using Triplet ranking loss with regression loss

We design triplet sampling methods to compute the triplet ranking loss which is useful for regularize the learning process. The classical triplet samples consists of a target, a positive sample that is similar to the target and a negative sample that is different to the target. However we use relative triplet sampling following the approach of [3]. That sampling satisfies $|y_a - y_p| < |y_a - y_n|$ for every possible triplet $((x_a, y_a), (x_p, y_p), (x_n, y_n))$ where x is the input and y is the label. Relative triplet sampling can take into account all possible triplets in a batch, making model deal with more triplets.

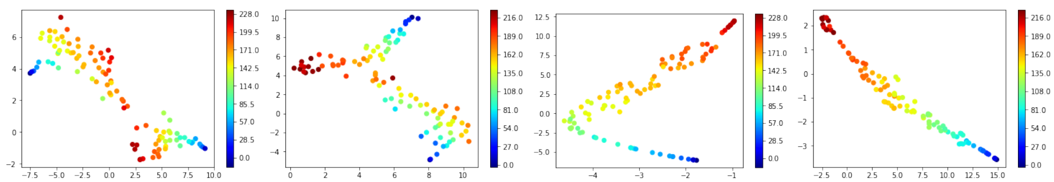
To compute the loss given triplet, we take embedded features of triplets obtained through convolution layers and global average pooling. We use L2 distance between features to obtain positive distance (target and positive) and negative distance (target and negative) and calculate cross-entropy loss with their softmax results. The total triplet loss of a batch is the sum of all triplets's loss that can be generated in that batch. The final network has two weighted losses: regression and triplet loss. In the backpropagation phase we apply the triplet loss only to convolution layers. Because triplet loss is obtained from embedding features, it is not related to fully connected layers.

3 Experiments and results

We run experiments on the RSNA bone age competition dataset that was collected from multiple centers and labeled by radiologists [1]. The dataset consists of 12611 training images and 200 test images with gender information. The label is recorded in months.

Table 1: Mean absolute distance result in RNSA competition test dataset

Mean Absolute Distance Month (MAD)			
Methods	Male	Female	All
Human reviewer[4]	-	-	7.32
David. B[4]	-	-	6.24
Ours without triplet loss	5.74	5.89	5.82
Ours with triplet loss	4.69	4.85	4.77



(a) Male: Only regression (b) Female: Only regression (c) Male: Proposed loss (d) Female: Proposed loss

Figure 3: Embedding feature visualization through t-SNE and projected onto a 2D space. Each data points computed from RSNA test dataset (200 images). From blue to red, it indicates older.

For the network architecture, we use U-Nets for hand segmentation and Inception-V3 for the transformation network and embedded feature generation. We set the weight for triplet loss to 0.01.

To evaluate our model, we compute the mean absolute distance (MAD), which is the L1 distance between model estimated age and radiologist annotated age. We trained the model separately for each gender and calculated the average of the test results. Table 1 shows MAD results comparison between the proposed model and previous studies. When using regression loss alone, the results are similar to the previous studies. But when triplet loss is added, the MAD is reduced from 5.82 to 4.77.

We visualized the t-SNE results of the embedded feature to confirm the effect of our triplet loss (Figure 3). When using only regression loss, features are separate but not ordered. After triplet loss is added, features of similar labels are gathered and ordered.

4 Conclusion

We proposed a deep learning model using triplet ranking loss for bone age assessment. The embedded features were better structured when triplet loss was used. Also it shows a significant improvement in performance in the RSNA competition test set. Though not as good as 4.265 MAD, which is the result of challenge winner, the ensemble will increase our model performance on the test set.

References

[1] Rsna pediatric bone age challenge. http://rsnachallenges.cloudapp.net/ competitions/4 (2017), online; accessed december 12, 2017.

[2] Vladimir Iglovikov et al. Pediatric bone age assessment using deep convolutional neural networks. *arXiv preprint arXiv:1712.05053*, 2017.

[3] Woobin Im et al. Using triplet ranking jointly with classification for facial age estimation. *Submitted to Computer Vision and Pattern Recognition (CVPR)*, 2018.

[4] David B Larson et al. Performance of a deep-learning neural network model in assessing skeletal maturity on pediatric hand radiographs. *Radiology*, page 170236, 2017.