Anti-Interference Study on Radiographic Bone Age Estimation Based on Artificial Intelligence Model

Songzheng Huang[1], Jianfeng Chen[2*]

[1]Zhejiang Kangyuan Medical Devices Incorporation, Hangzhou, China
[2]Department of Radiology and Medical Imaging, Stritch School of Medicine, Loyola University Medical Center, Chicago, Illinois, USA
Email: *jfchen@live.com

Abstract

In this paper, the interferences of X-ray image noise on a bone age model, Xception model, were studied. We conduct a comparative experiment test according to the output performance of the neural network model using both the original image training and noise-added (Gaussian noise plus salt-pepper noise) training, and analyze the anti-interference ability of the Xception model, hoping to improve it through noise enhancement training and generalize the application ability of the model. The results show that the model trained with noise-added (Gaussian noise plussalt-pepper noise) images can make predictions that are more robust and less affected by the image disturbances, such as image noise.

Keywords

Bone Age, X-Ray Image, Xception Model, Gaussian Noise, Salt-Pepper Noise

1. Introduction

"Bone age" is the abbreviation of skeletal age. It is the developmental age obtained by comparing the skeletal development level of adolescents and children with the bone development standard [1] [2] [3]. It can more accurately reflect the maturity of the body than just based on age, height and weight, and it can more accurately reflect the growth of the individual, including the developmental level and maturity, and can be used to predict future adult height.

In recent years, a deep learning model called Xception model has been developed by Google that was originally designed to solve image classification prob-

lems [3]. Released in 2017, it was one of the first sensational models in computer vision. It is widely used in computer vision, natural language processing, medical image analysis and many other fields. The evolution process of the Xception model can be divided into the following steps:

1) The ImageNet competition promotes the development of deep learning: The ImageNet competition is an image recognition competition initiated by companies such as Google, Microsoft, and Yahoo. This competition produced a large amount of training data and models, some of which provided useful inspiration for the development of Xception.

2) The origin of ResNet: ResNet is a very deep convolutional neural network, which can solve the problem of gradient disappearance when training deep neural networks. Its appearance promoted the development of deep learning and provided the soul for the design of Xception.

3) The introduction of the Bottleneck module: The Bottleneck module is a technology that compresses large-scale neural network models into models with higher performance. By introducing the Bottleneck module, the X receiver model can reduce the number of parameters while maintaining high performance.

4) Adoption of Multi-scale Pyramid Pooling: Multi-scale Pyramid Pooling is a technology for image segmentation, which can adaptively adjust the size and complexity of the network, thus maintaining high performance and supporting more data types.

Herethe Xception model uses a depthwise separable convolution to replace the traditional convolution structure, which greatly improves the performance and efficiency of the model. Its main architecture includes:

- Input layer: receive input images;
- Initial convolutional layer: feature extraction from the input images;
- 13 depthwise separable convolutional blocks: each block includes depthwise separable convolutional layers and a residual connections;
- Global average pooling layer: Average pooling on the output of the last convolutional block;
- Fully connected layer: maps the output of the pooling layer to the classification output.

The bone age model, Xception model (https://www.kaggle.com/datasets/kmader/rsna-bone-age), originally released by RSNA (Radiological Society of North America) is used in this paper. For the same model, we conduct a comparative experiment test according to the output performance of the model trained with the original images and noise-added (Gaussian noise and salt-pepper noise) images, and study the anti-interference ability of the Xception model, improve it through noise enhancement training and generalize the application ability of the model. This study is organized as follows: In Section 2, a method based on the Xception model is presented. In Section 3, the anti-interference properties of the model are investigated by adding Gaussian and salt-pepperimage noise, which is known to be unwanted sig-

nals that corrupt images. The test results are discussed and summarized in Section 4. Section 5 presents the final conclusions.

2. Methods

2.1. Neural Network Modeling

The main structure of the Xception model includes: input layer, initial convolutional layer, 13 depthwise separable convolutional blocks, global average pooling layers and fully connected layers. The outline structure of the model is as follows Figure 1.
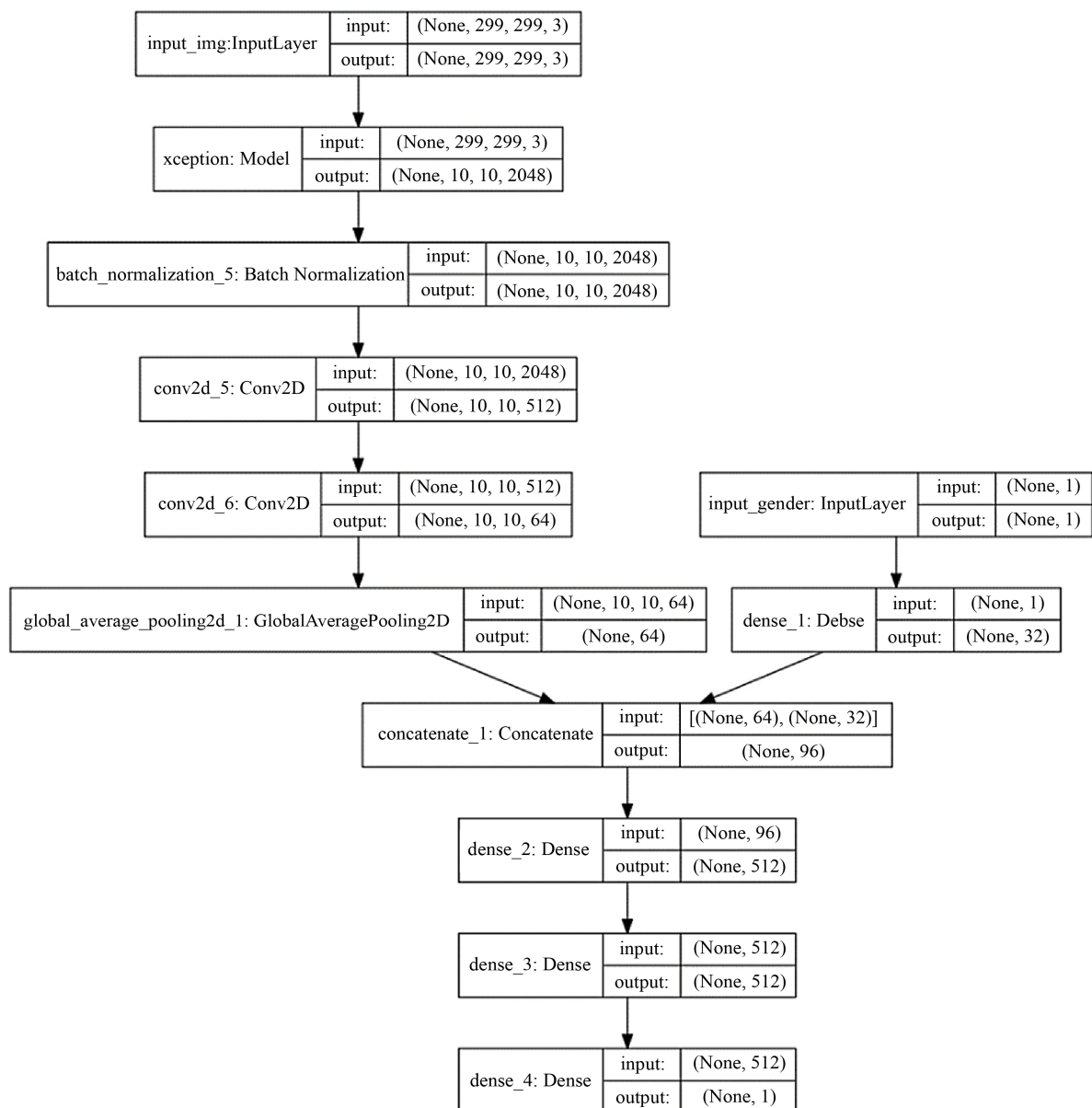

Figure 1. The bone age assessment model released by RSNA is based on the Xception model. The structure of the model includes: input layer, initial convolutional layer, 13 depthwise separable convolutional blocks, global average pooling layer and fully connected layer, as showns in the following figure.

The principle of the model is to predict the child's age by analyzing the bone development in X-ray images, so as to evaluate the child's growth and development, and help doctors diagnose and treat children's health problems. The model uses a technique called "image alignment" to align different X-ray images for easier comparison and evaluation. The accuracy of the model was tested in RSNA's bone age assessment challenge and achieved very good results. Specifically, the model has a *MAE* (Mean Absolute Error) of 6.09 months and a *MSE* (Mean Standard Error) of 9.27 months on the test dataset. These results demonstrate that the model has high accuracy and precision for bone age assessment and can be widely used in medical practice. In order to reduce the hardware requirements of the model, we simplify the model, which includes reducing the number of feature layers to 512, discarding the gender parameter, and changing the number of input channels to 1, as shown in **Figure 2**.

2.2. Adding Image Noise

Noise is known to be an unwanted signal that can easily destroy the image quality [4]. This unwanted signal in an image can appear in the form of dsitored pixel values, which can show up images as uneven lines and blurry objects and can also be seen in film grain and in the shot noise of a photon detector [5] [6] [7]. Noise can be in the form of additive, multiplicative or impulsivenoise [8] [9]. Here Gaussian noise, and salt-pepper noise are added to original test images to compare the difference of the output results from the model, and to study the anti-interference ability of the model by training the image date with these noise for generalization application ability.

Here Gaussian noise is also referred to amplifier noise. This noise is known as the main part of the "true noise" in the image sensor. The main sources of Gaussian noise are generated during image acquisition, such as sensor noise caused by poor illumination or high temperature. It is known as statistical noise and its *PDF* (Probability Density Function) is comparable to a Gaussian distribution.
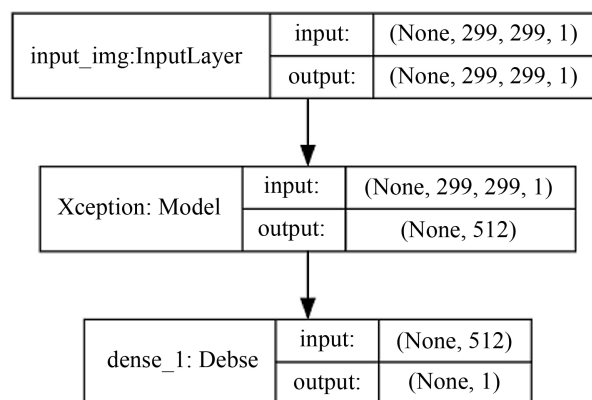
| input_img:InputLayer | input: | (None, 299, 299, 1) |
| | output: | (None, 299, 299, 1) |

| Xception: Model | input: | (None, 299, 299, 1) |
| | output: | (None, 512) |

| dense_1: Debse | input: | (None, 512) |
| | output: | (None, 1) |

Figure 2. We simplified the model, reduced the number of feature layers to 512, discarded the gender parameter, and changed the number of input channels to 1, as shown in the following figure.

Here *add_gaussian_noise* (*image*, *mean*, *std*) is used to add Gaussian noise to the images. It has three parameters, "*image*" is the name of input image, "*mean*" is the mean of the Gaussian distribution, which is defaulted as 0, "*std*" is the standard deviation of the Gaussian distribution. This function first generates a Gaussian random noise with the same size as the input image, then adds the noise to the input image, and finally the image pixel values in the range of 0 to 255 is limited through "np" and "clip". **Figure 3** shows a typical original image and its Gaussian noise with different standard deviation added.
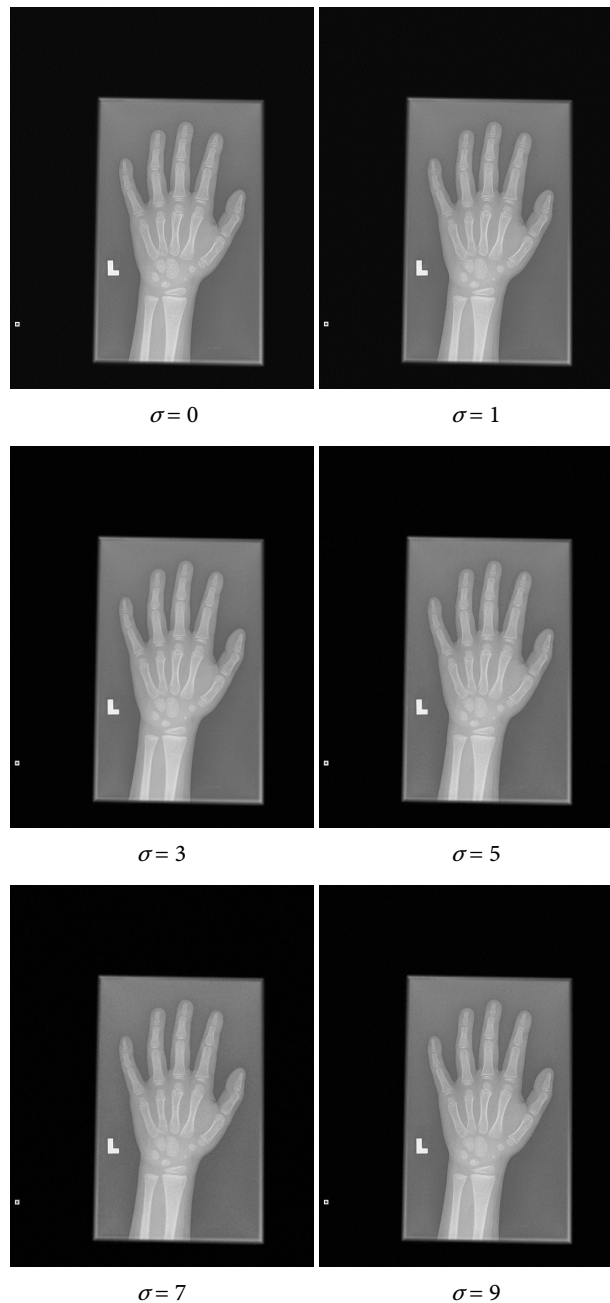
Figure 3. Shows a typical original image and its Gaussian noise with different standard deviation σ added. (σ = 0, 1, 3, 5, 7, 9).

Salt-pepper noise is known as impulse noise, which is a type of noise that tends to modify pixel values randomly. This type of noise is cause by sudden and severe changes in the image signals due to equipment failures. It is characterized by occasional black and white pixels, where there are bright pixles in dark places and dark pixels in bright places [5]. A function, *add_salt_pepper_noise* (*image*, *probability*), is used to add salt-pepper noise to the image. It has two parameters, "*image*" is the name of input image, and "*probability*" is the probability that each pixel becomes salt or pepper noise. First, the function calculates the dimensions of the image, then randomly selects some pixels based on the probability. Then, set these randomly selected pixels to white (255) or black (0) for our 8-bit image, and finally return the image with salt-pepper noise added. **Figure 4** shows a typical original image and the images with its salt-pepper noise added.
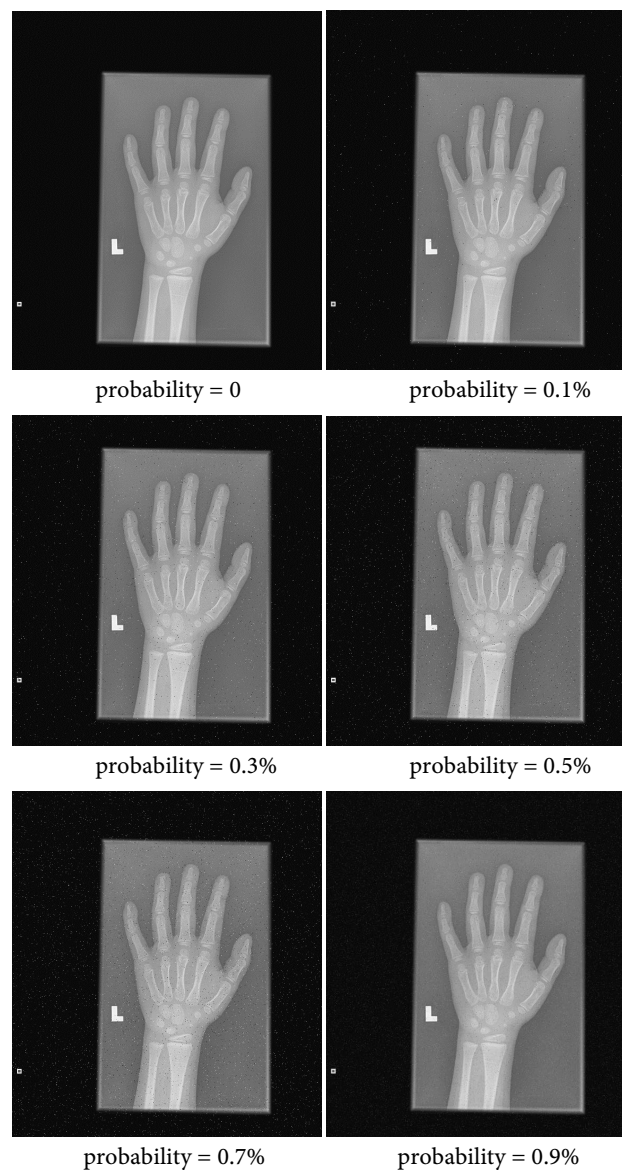

probability = 0   probability = 0.1%

probability = 0.3%   probability = 0.5%

probability = 0.7%   probability = 0.9%

Figure 4. Shows a typical original image and adding with different probability of being salt-pepper noisefor each pixel. (probability = 0, 0.1%,0.3%,0.5%,0.7%, 0.9%).

3. Results

In our study, we split the training set and the verification set at a ratio of 3:1 from the data set of 12668 children's palm frontal X-ray images. That means that 9501 images were used to train the neural network model, and 3167 images were used to validate the model. Then an independent data set consisting of 200 children's palm frontal X-ray images (pictures in png format, folder name *bone-age-test-dataset*) was used as the test data set. The neural network model is trained using the above training set to obtain the function of *MAE*, which can be quantitatively used to study the model's performance and is given by the difference between the model prediction results and the actual observations. *MAE* is defined by

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|Predict\ Value - Actual\ Value\right|$$

where N is the total number of samples, Σ represents the sum operation, and | | represents the absolute value operation. It measures the average error between the predicted value and the real value, and the smaller the value, the closer the predicted result is to the real value. At first, 9501 original children's palm frontal X-ray images were used to train the neural network model, then 3167 original images were used to validate the model. The validating results for *MAE* changed with iterating number are shown in **Figure 5(a)**. In order to train with noise-added images more generally, first generate a uniformly distributed randonnumner a with values between 0 and 10, which is used as the standard deviation of Gaissian noise. Then, *noisy_image = add_gaussian_noise* (*image*, *mean*, *a*) is used to generate images with Gaussian noise-added. Here *image* is the original image. Finally, use *image_mix = add_salt_pepper_noise* (*noisy_image*, *probability*) to create images with both Gaussian noise and salt-pepper noise. Similarly, 9501 noise-added images are used to train the neural networkmodel. Noise-added 3167 images were used to validate the model respectively. The results of *MAE* as a function of the iterating number are shown in **Figure 5(b)**. Here *MAE* is based on 3167 validating image data. Both results show that the *MAE* approach stable values after 20 iterations for the original validating images and 30 iterations for the noise-added validating images.

The standard error (*Std Error*) is used here to evaluate the performance difference between the model trained with the original images and the noise-added images. The standard error is defined as

$$Std\ Error = \sqrt{\frac{\sum_{i=1}^{N}\left(Predict\ Value_1 - Predict\ Value_2\right)^2}{(N-1)}}$$

Where N is the total number of samples, Σ is the summation of all test images, *Predict Value*$_1$ is the prediction of the model trained using the original images, and *Predict Value*$_2$ is the prediction of the model trained using noise-added images. The test results of the model trained using original images and Gaussian noise-added images are shown **Figure 6**. The results of the model trained with
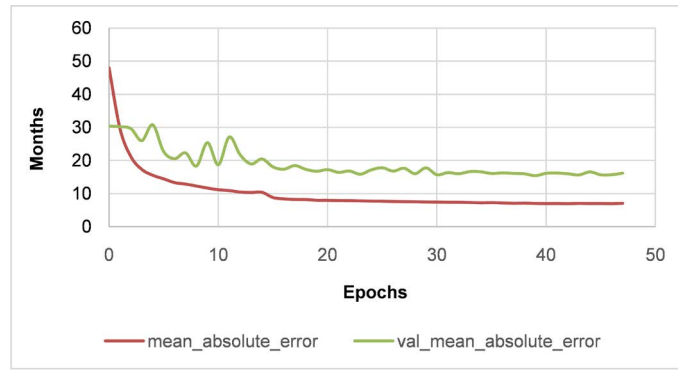
(a)
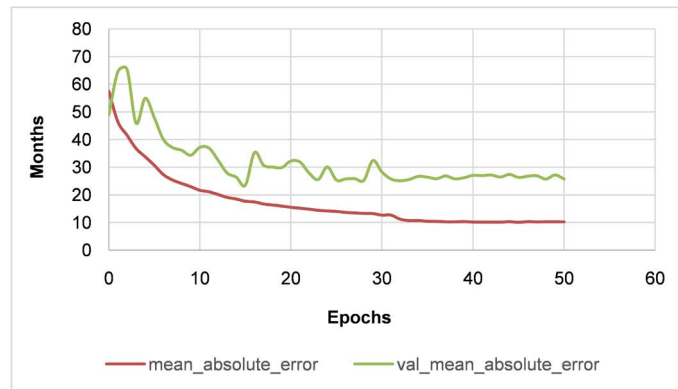
(b)

Figure 5. At first, 9501 original children's palm frontal X-ray images were used to train the model, and the MAE changed with the iterating number for 3167 validating and 200 testing images are shown in **Figure 5(a)**. Then, noise was added to all these original images, and the noise-added 9501 images were used to train the model, and the results of *MAE* as a function ofiterating number are shown in **Figure 5(b)**. Also wo got two trained model weights, We named them W_a trained with original images and W_b trained with noise-added images.

$\sigma = 1$, std error = 2.84 month

$\sigma = 1$, std error = 1.33 month
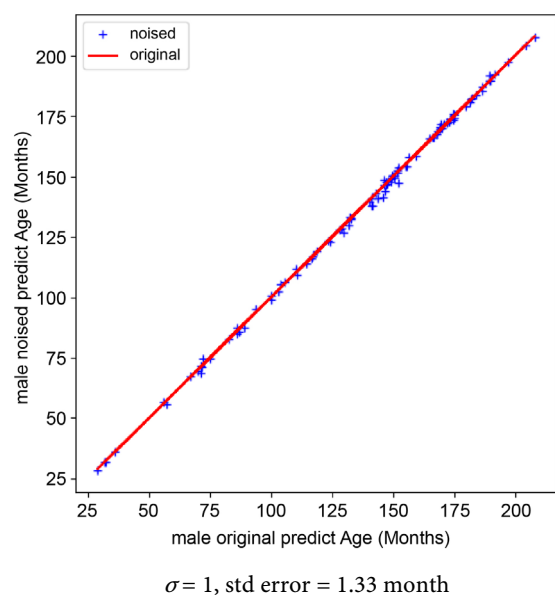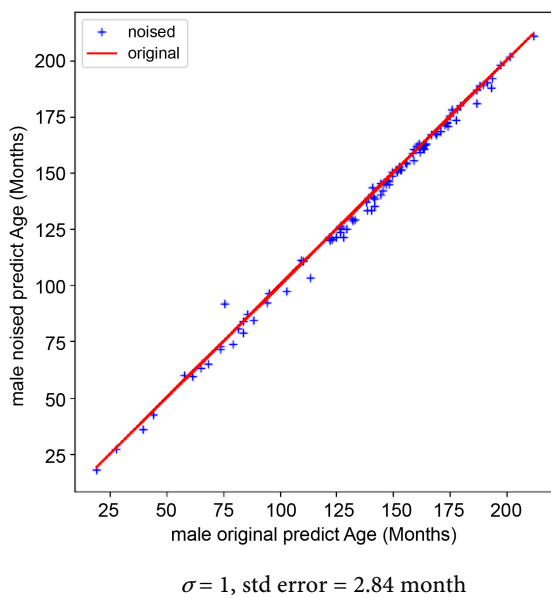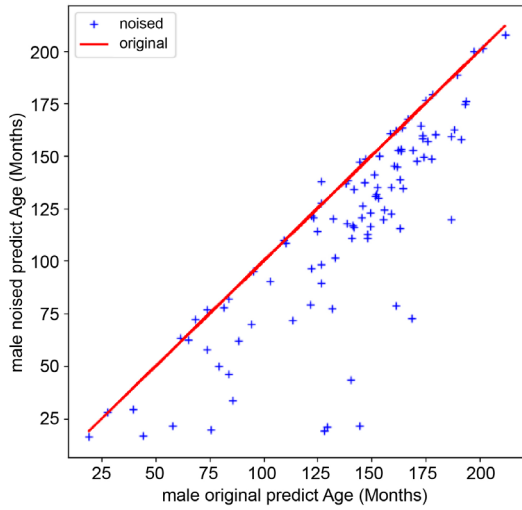
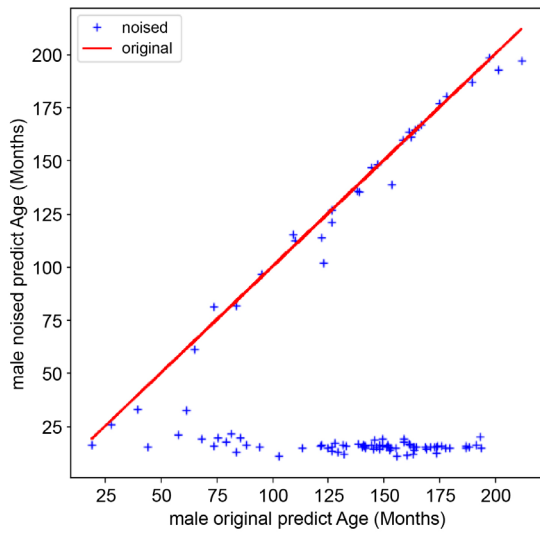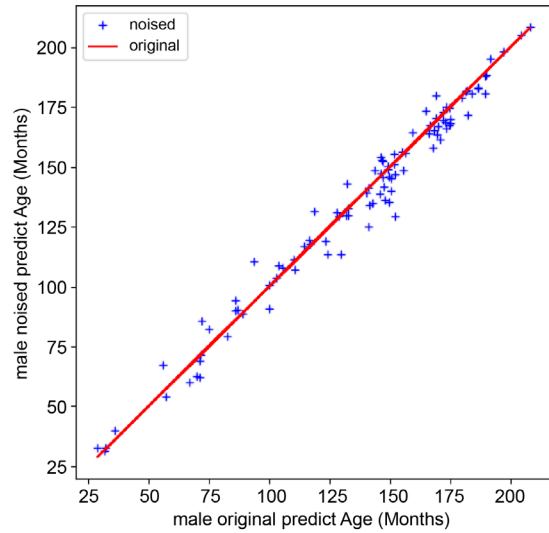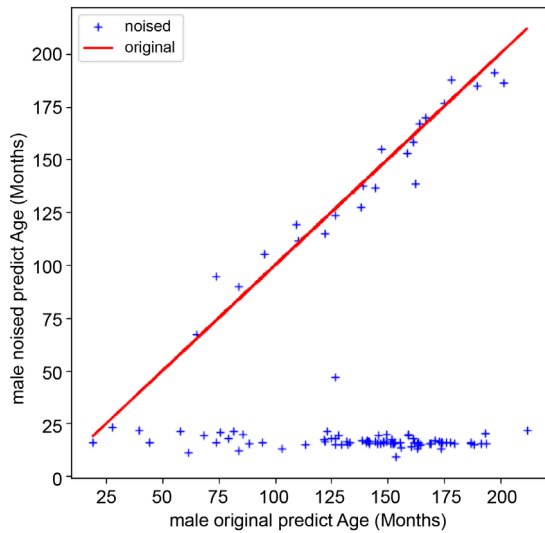$\sigma = 3$, std error = 25.65 month
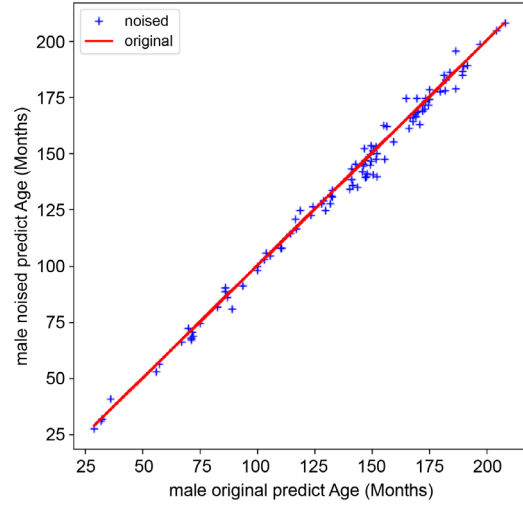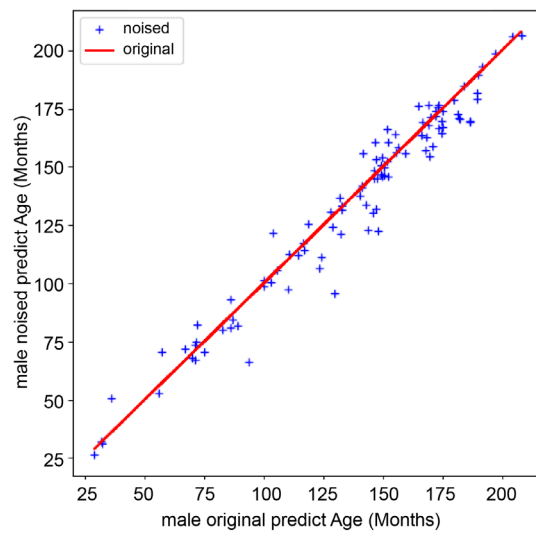

$\sigma = 3$, std error = 3.91 month

$\sigma = 5$, std error = 63.74 month

$\sigma = 5$, std error = 6.59 month

$\sigma = 7$, std error = 62.11 month

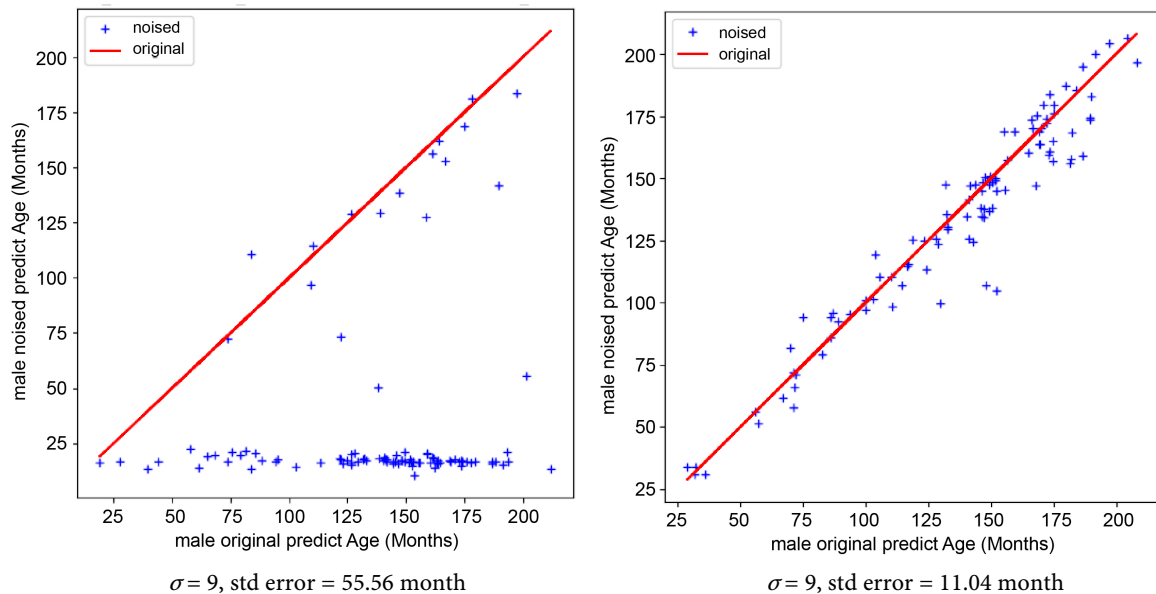
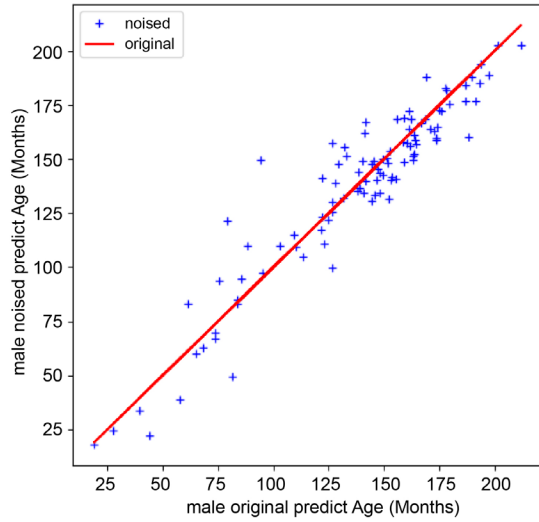
$\sigma = 7$, std error = 8.95 month

$\sigma = 9$, std error = 55.56 month $\sigma = 9$, std error = 11.04 month

Figure 6. The results of the neural network model trained with the original images are shown on the left side, with Gaussian noise added images are shown on the right side.
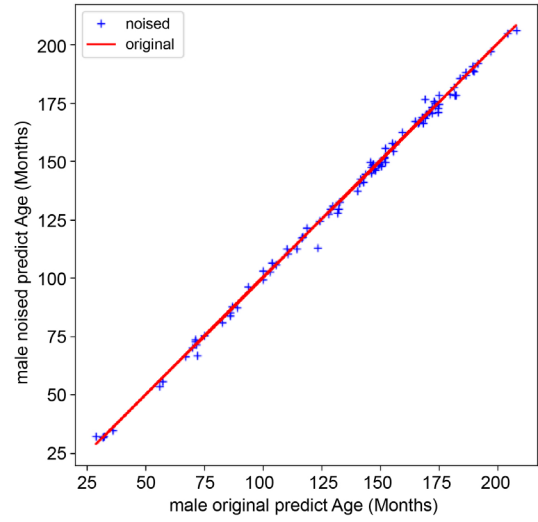
the original images are shown on the left side, with Gaussian noise-added images are shown on the right side. The test results of the model trained using original images and salt-pepper noise-added images are shown **Figure 7**. The results of the model trained with the original images are shown on the left side, with salt-pepper noise-added images are shown on the right side. Both **Table 1** and **Table 2** show that the age prediction error, which is defined as *MAE*, as the strength of image noise. The results clearly show that the model trained with noise-added image, compared with the model trained with the original images, can make predictions more robustly and less affected by the image noise.

4. Discussions

The principle of the current neural network model is to predict the child's age by analyzing the bone development in X-ray images, thereby evaluating the child's growth and development, and helping doctors diagnose and treat children's health problems. The accuracy of predicting age depends on the quality of X-ray images, especially image noise. In this paper, we train the model by using the original images as well as noise-added images. Here image noise includes Gaussian noise as well as salt-pepper noise. Then these trained models were used to study the anti-interference ability of the model. The experimental results show that the model trained with noise-added image, compared with the model trained with the original images, can make predictions more robustly and less affected by the image noise. While current neural network model can accurately predict bone age, training with noise-added images is more time-consuming than the training with the original images. Practical approaches, such as simplifying the model to focus on some sensitive bones instead of treating all hand bones equally still need to be investigated.
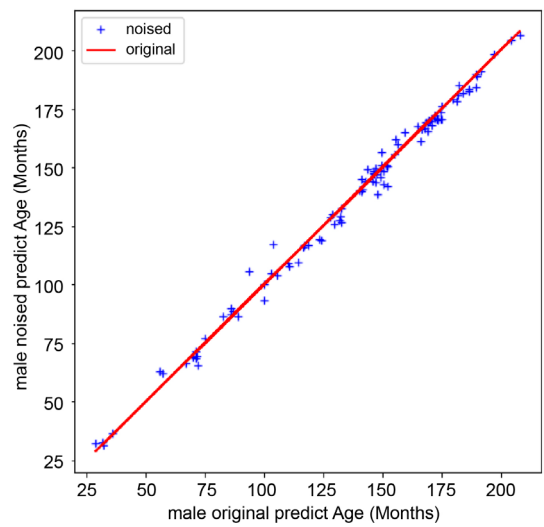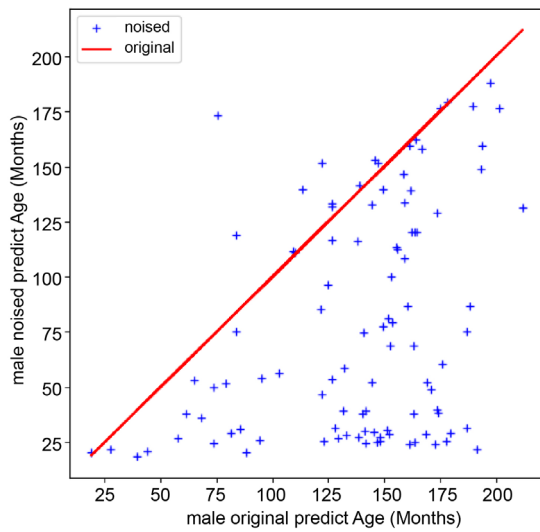
PDF = 0.1%, std error = 13.55 month
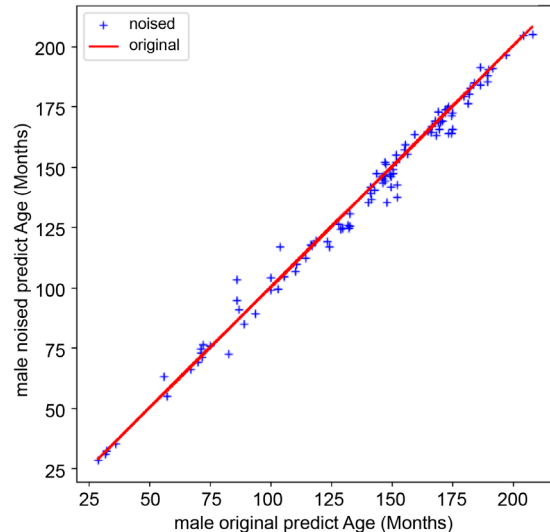
PDF = 0.1%, std error = 2.19 month
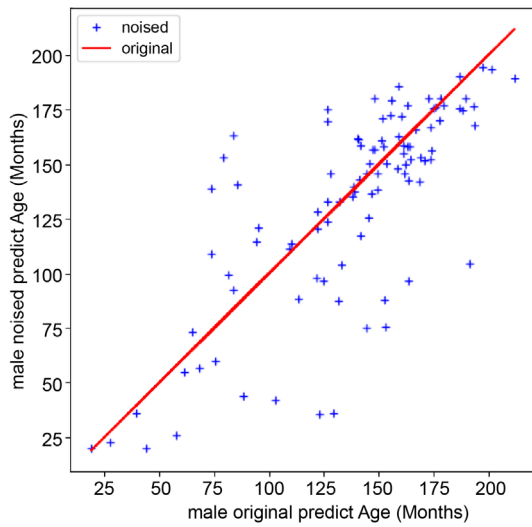
PDF = 0.3%, std error = 30.21 month

PDF = 0.3%, std error = 3.75 month

PDF = 0.5%, std error = 52.95 month

PDF = 0.5%, std error = 4.69 month

PDF = 0.7%, std error = 57.78 month

PDF = 0.7%, std error = 4.02 month

PDF = 0.9%, std error = 55.39 month

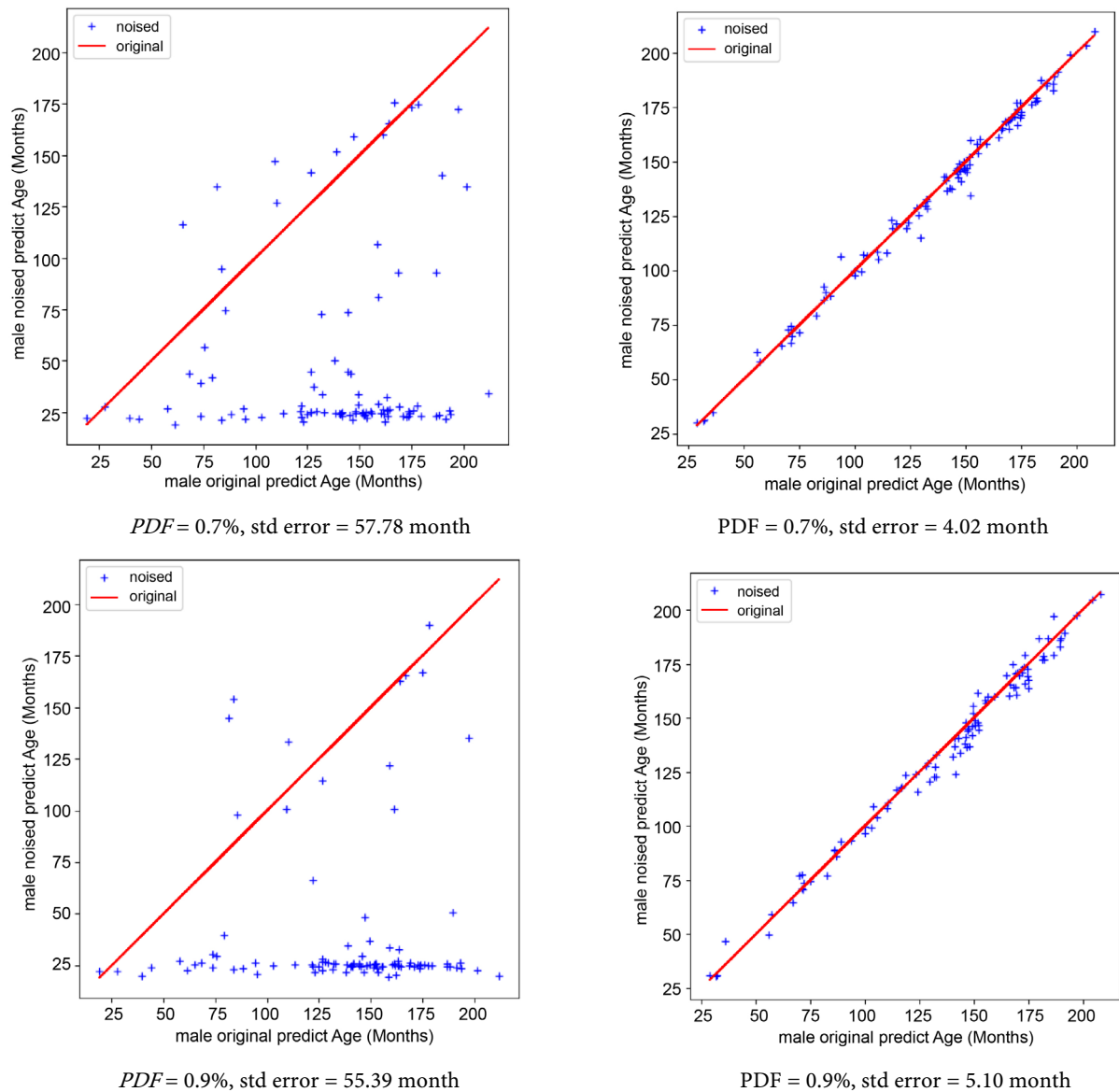PDF = 0.9%, std error = 5.10 month

Figure 7. For 200 test images, the results of the neural network model trained with the original images are shown on the left side, with salt-pepper noise added images are shown on the right side.

Table 1. The neural network model trained with the original images and Gaussian noise added images. The *Std error* of age prediction changed with 200 test original images and the 200 images contained different Gaussian noise strength.

The standard Deviation of Gaussian noise	*Std error* using the Model trained with original images (month)	*Std error* using the Model trained with noise-added images (month)
1	2.84	1.33
3	25.65	3.91
5	63.74	6.59
7	62.11	8.95
9	55.56	11.04

Table 2. The neural network model trained with the original images and salt-pepper noise added images. The *std error* of age prediction changed with the PDF of salt-pepper noise. The *Std error* of age prediction changed with 200 test original images and the 200 images contained different *PDF* of salt-pepper noise.

The *PDF* of salt and pepper noise	*Std error* using the Model trained with original images (month)	*Std error* using the Model trained with noise-added images (month)
0.1%	13.55	2.19
0.3%	30.21	3.75
0.5%	52.95	4.69
0.7%	57.78	4.02
0.9%	55.39	5.10

5. Conclusion

The results show that the model trained with noise-added images, compared with the model trained with the original images, can make predictions more robustly and less affected by the image noise.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Zhang, S.-Y., Liu, L.-J., Wu, Z.-L., Liu, G., Ma, Z.-G., Shen, X.-Z. and Xu, R.-L. (2008) Standards of TW3 Skeletal Maturity for Chinese Children. *Annals of Human Biology*, **35**, 349-354. https://doi.org/10.1080/03014460801953781

[2] Liu, J., Qi, J., Liu, Z., Ning, Q. and Luo, X.P. (2008) Automatic Bone Age Assessment Based on Intelligent Algorithms and Comparison with TW3 Method. *Computerized Medical Imaging and Graphics*, **32**, 678-684. https://doi.org/10.1016/j.compmedimag.2008.08.005

[3] http://www.kaggle.com/datasets/kmader/rsna-bone-age

[4] Owotogbe, J.S., Ibiyemi, T.S. and Adu, B.A. (2019) A Comprehensive Review on Various Types of Noise in Image Processing. *International Journal of Scientific & Engineering Research*, **10**, 388-393.

[5] Kaur, J. (2012) Salt & Pepper Noise Removal Using Fuzzy Based Adaptive Filter. *International Journal of Science, Engineering and Technology Research*, **1**, 24-26.

[6] Kayhan, S.K. (2014) An Effective 2-Stage Method for Removing Impulse Noise in Images. *Journal of Visual Communication and Image Representation*, **25**, 478-486. https://doi.org/10.1016/j.jvcir.2013.12.016

[7] Koli, M. and Balaji, S. (2013) Literature Survey on Impulse Noise Reduction. *Signal & Image Processing*, **4**, 75-95. https://doi.org/10.5121/sipij.2013.4506

[8] Kumar, J. and Abhilasha (2014) An Iterative Unsymmetrical Trimmed Midpoint-Median Filter for Removal of High Density Salt and Pepper Noise. *International Journal of Research in Engineering and Technology*, **3**, 44-50. https://doi.org/10.15623/ijret.2014.0304008

[9] Li, Y., Sun, J. and Luo, H. (2014) A Neuro-Fuzzy Network Based Impulse Noise Filtering for Gray Scale Images. *Neurocomputing*, **127**, 190-199.
https://doi.org/10.1016/j.neucom.2013.08.015