IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 27, 2019, accepted April 16, 2019, date of publication May 22, 2019, date of current version June 6, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2918205

A Lightweight U-Net Architecture Multi-Scale Convolutional Network for Pediatric Hand Bone Segmentation in X-Ray Image

LIAN DING[1], KAI ZHAO[2], XIAODONG ZHANG[2], XIAOYING WANG[2], AND JUE ZHANG[1,3]

[1]Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China
[2]Department of Radiology, Peking University First Hospital, Beijing 100034, China
[3]College of Engineering, Peking University, Beijing 100871, China

Corresponding author: Xiaoying Wang (wangxiaoying@bjmu.edu.cn)

ABSTRACT Bone age assessment (BAA) is a common radiological examination used in pediatrics based on an analysis of ossification centers and epiphyses of hand bones. Segmentation of hand bones could help give specific descriptions of hand bone features in medical records and assess bone age automatically. This study proposes a lightweight U-Net architecture multi-scale convolutional network for pediatric hand bone segmentation in the X-ray image. The compact structure is based on U-Net architecture with two down-sampling and up-sampling operations and multiple filters with different kernel size are adopted for countering hand bone scale variations during growth in children. This is the first-hand bone segmentation study with deep learning and the experiment results indicate promising performance in hand bones segmenting, especially for small bones of the hand.

INDEX TERMS Bone age assessment, U-Net, multi-scale convolutional network, segmentation of hand bones, X-ray.

I. INTRODUCTION

Bone age assessment (BAA) is a radiological examination for pediatrics to determine any discrepancy between a child's skeletal age (the developmental age of their bones) and their chronological age (in years, taken from birth date) [1]. The examination is based on an analysis of ossification centers in the carpal bones and epiphyses of tubular bones including distal, middle, and proximal phalanges as well as ulna and radius (Fig. 1). For example, an illness may cause accelerated or delayed appearance of epiphyses or ossification centers. BAA is commonly used to manage endocrine disorders and pediatric syndromes [2]. Moreover, BAA is used in prediction of the adult height as well as in forensic medicine [3].

There are two methods applied in clinical routine: Greulich and Pyle (GP) [4] and Tanner and Whitehouse (TW) [5]. The G&P method focuses on a set of regions of interest (ROIs) of the hand and wrist joints (Fig. 1). The TW method (also the TW2 and TW3 for the second and third editions, respectively) analyzes 20 ROIs and assigns a staging score to each of them. However, both of these two methods are time-consuming and cumbersome tasks. The average reading time is 84 s and 474 s for GP and TW methods, respectively [6]. Moreover, both methods suffer from high intra- and inter-observer variability. The average spreads of the reading are 0.96 year (11.5 months) for the G&P and 0.74 year (8.9 months) for the TW2 [7]. Hence, automated BAA is desired.

The traditional method for computer-assisted BAA relies on image processing, including hand bones segmenting and relevant ROI feature extracting. Pietka *et al.* [8] proposed a method for epiphyseal/metaphyseal segmentation. The features extracted from the ROIs describe the stage of skeletal development objectively. Giordano *et al.* [9] proposed an Epiphyseal/Metaphyseal ROI segmentation method by using the Difference of Gaussians filter and extracting main features of these bones for the stage TW2 evaluation. They all deal with the problem of segmentation of certain regions within the radiograph. However, a rather low accuracy rate of the bone segmentation may be caused due two reasons. First, a nonuniformity of the image background should be suppressed prior to the image analysis [10]. Second, hand orientation sometimes [11] varies from the standard position without the cooperation of little children, reducing the segmentation robustness.

The associate editor coordinating the review of this manuscript and approving it for publication was Muhammad E. H. Chowdhury.

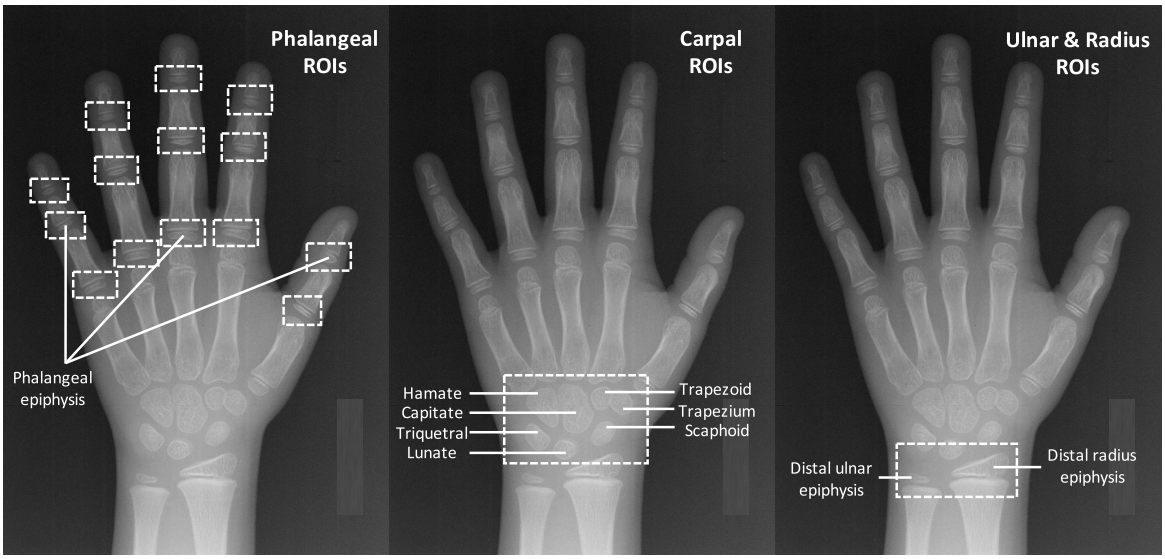


FIGURE 1. An example of hand image radiograph with superimposed regions of interest.

Another type of BAA method relies on the deep learning technique [12]. This method aims at encoding visual features directly. Spampinato *et al.* [11] proposed and tested several deep learning approaches to assess skeletal bone age automatically. Ren *et al.* [10] propose a regression convolutional neural network (CNN) to automatically assess the pediatric bone age from hand radiograph. They first adopt attention maps as inputs for the regression network. Although these studies open a new paradigm for automated BBA, they are not sufficient since it is hard to give specific descriptions of hand bone features in medical records. Additional, only analysis with each single hand bone is in accordance with the TW method. Therefore, not only the estimated bone age but also the hand bones segmentation is needed to clinical application.

Recently, with the great development of deep learning [13], [14], many conventional methods, such as the graph-based segmentation approaches [15] or those based on handcrafted local features [16], have been replaced by deep segmentation networks, which typically produce higher segmentation accuracy [17]. For example, Ronneberger *et al.* [18] introduced the U-Net, which used the skip-architecture that combined the high-level representation from deep decoding layers with the appearance representation from shallow encoding layers to produce detailed segmentation. Moreover, it adopted architectures with pyramidal shapes, including 4 down-sampling and up-sampling operations. It employed low-resolution feature to help locating the object and normal-resolution feature to improve segmentation accuracy for details [19]. Although the U-Net architecture perform well in cell [18] and brain tumor [20] segmenting tasks, it still underperforms in details [21]. Arguably, the number of down-sampling operations adopted in CNN for higher segmentation accuracy depends on the specific problem. For example, Jégou et al. [22] extended DenseNets to deal with the problem of semantic segmentation with 3 down-sampling operations and achieved state-of-the-art results on urban scene benchmark datasets such as CamVid and Gatech. Zhou *et al.* [23] presented U-Net^{++} architecture where the encoder and decoder sub-networks are connected through a series of nested. In theory, this architecture included 4 sub-networks with 1-4 numbers of down-sampling operations separately and achieved intersection over union (IoU) gain in polyp, liver, and cell nuclei segmentation tasks. In BAA, the appearance of small bones of the hand, especially small ossification centers in the carpal bones and epiphyses of tubular bones, are vital for bone age assessment, especially for 0-7 years old children [4].

To obtain accurate segmentation results of hand bones in X-ray, we proposed a lightweight U-Net architecture multi-scale convolutional network. Different number of down-sampling operations adopted in U-Net architecture were compared for hand bones segmentation task and we choose 2 in this work. Moreover, as hand bones becoming larger within growth in children (Fig. 2), same sized kernels combination may not counter hang bones scale variations [24]. Szegedy *et al.* [25]. proposed an Inception module to process visual information at various scales. Motivated by it, we introduce a multi-scale block with different kernel sizes to extract scale-relevant features. We evaluate the ability of different networks for small bones segmentation with the accuracy of small bones detection. This is the first hand bone segmentation study with deep learning and the results indicate promising performance in hand bones segmenting, especially for small bones of the hand.

II. MATERIALS AND METHODS

We propose a lightweight U-Net architecture multi-scale convolutional network to complete the hand bones

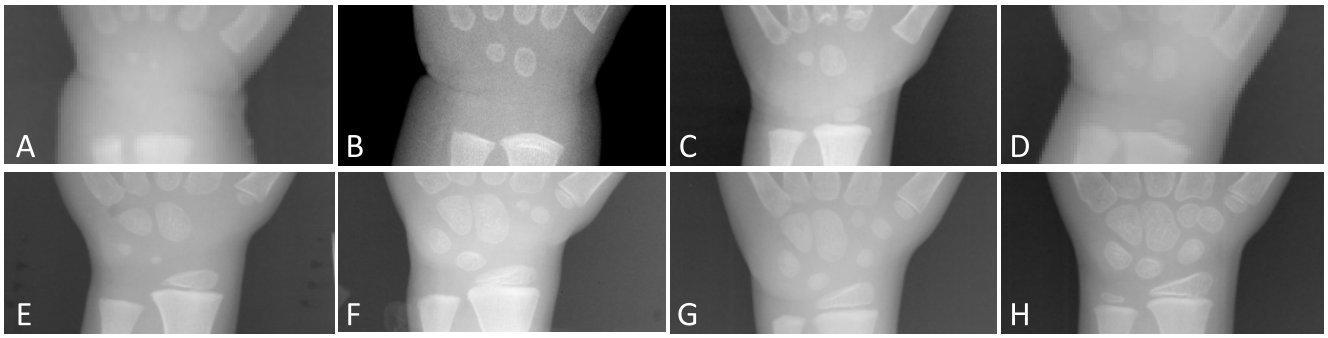


FIGURE 2. Growth pattern of carpal bones from newborn to 7-year-old (A-H).

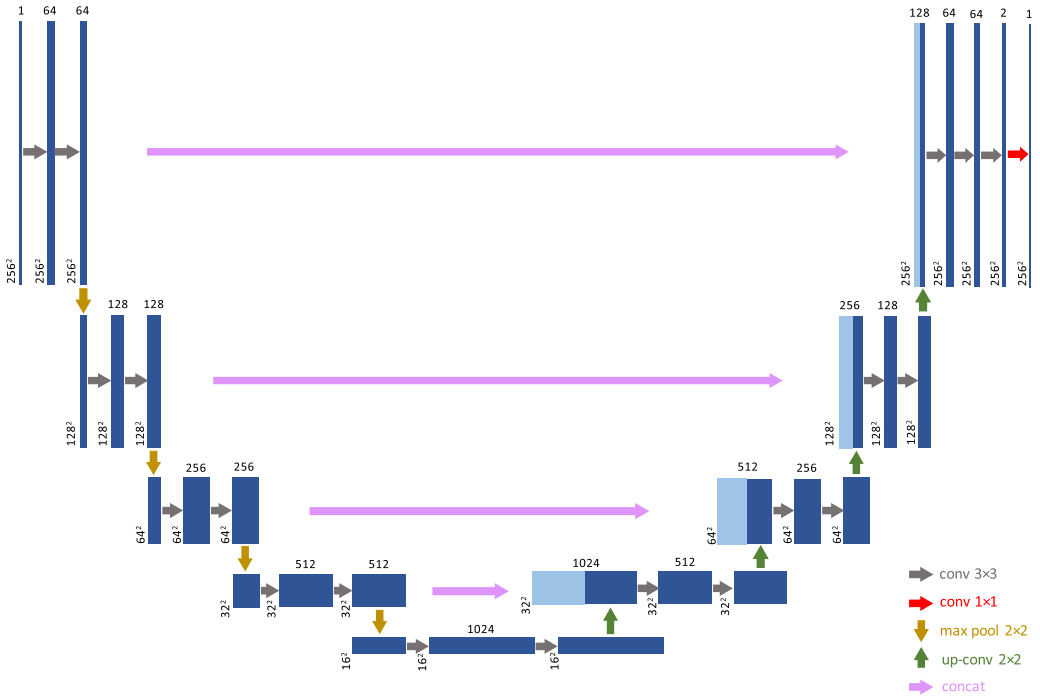


FIGURE 3. U-Net architecture with 4 down-sampling and up-sampling operations (U-Net^4).

segmentation. We compare different numbers of down-sampling and up-sampling operations of U-Net architecture to obtain higher performance in hand bones segmentation. Particularly, we adopt multi-scale block like Inception model to extract the scale-relevant features, which consists of multiple filters with different kernel size. The network then segments the hand bones end-to-end.

A. LIGHTWEIGHT U-NET ARCHITECTURE

The number of down-sampling and up-sampling operations in pyramidal shape networks depends on the specific problem for higher segmentation accuracy. Therefore, we compare different number of down-sampling and up-sampling operations of U-Net architecture to find a specific lightweight network structure for hand bone segmentation. The original U-Net (U-Net^4) we adopted is illustrated in Fig. 3. The size of input image is $256 * 256$. It consists of the repeated application of two 3×3 convolutions (unpadded convolutions), each followed by a rectified linear unit (ReLU) [26] and a 2×2 max pooling operation with stride 2 for down-sampling. Every step in the expansive path consists of an up-sampling of the feature map followed by a 2×2 convolution (''up-convolution'') that halves the number of feature channels, a concatenation with the correspondingly cropped feature map from the contracting path, and two 3×3 convolutions, each followed by a ReLU. At the final layer a 1×1 convolution is used to map each 64-component feature vector to the desired number of classes. We compared different U-Net architectures with different number of down-sampling and up-sampling operations (Fig. 4) and U-Net^{++} [23] for the same hand bone segmentation task and we found the U-Net^2 could obtain optimal performance.

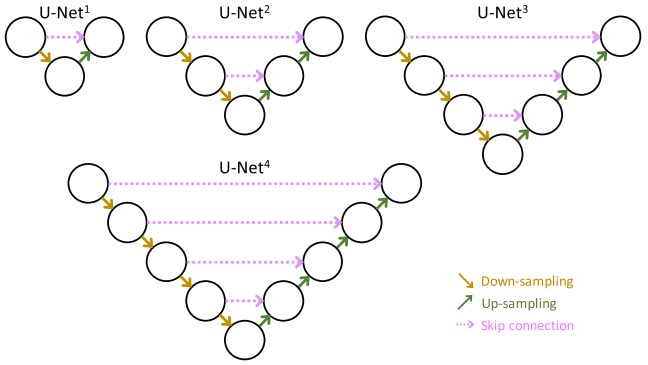


FIGURE 4. U-Net architectures with different number of down-sampling and up-sampling operations.

B. MULTI-SCALE NETWORK ARCHITECTURE

We designed a multi-scale convolutional neural network (msCNN) based on the U-Net[2] architecture to learn the scale-relevant density maps from hand bones with different sizes (Fig. 5)[24]. The first convolution layer is a traditional convolutional layer with 9×9 kernel size to remap the image feature. Multi-Scale Blob (MSB) is employed in the network, which is an Inception-like model (Fig. 6) consisting of multiple filters with different kernel size (including 3×3, 5×5, 7×7 and 9×9). ReLU is applied after each convolution layer working as the activation function of previous convolutional layers except the last one [26]. Detailed parameter settings are listed in Table 1. Moreover, to evaluate the segmentation performance with MSB, the MSBs adopted in this network were replaced by single kernel sizes (including 3×3, 5×5, 7×7 and 9×9) and the segmentation results of 4 single kernel size networks were tested separately. The energy function is computed by a sigmoid activation function over the final feature map combined with the dice coefficient loss function. The dice coefficient is defined as

$$\frac{2 * S_1 * S_2}{S_1 + S_2}$$

where S_1 is the segmentation result and S_2 is the groundtruth.

C. EVALUATION METRICS

The segmentation results have been evaluated using the dice coefficient, IoU, sensitivity and the specificity. The IoU is defined as $(S_1 \cap S_2)/(S_1 \cup S_2)$ where S_1 is the segmentation result and S_2 is the groundtruth. In addition, sensitivity is used to evaluate the number of TP and FN that is

$$\text{sensitivity} = \frac{TP}{TP + FN},$$

and specificity is defined as

$$\text{specificity} = \frac{TN}{TN + FP},$$

in which TP, FP, TN and FN denote the true positive, false positive, true negative and false negative measurements, respectively.

TABLE 1. The architecture and parameters of lightweight U-Net architecture multi-scale convolutional network.

Part	Type	Filters	Filter Size	Output Shape
conv	Conv	64	9×9	(256, 256, 64)
	ReLU	-	-	
msb	MSB Conv	4 × 16	(9/7/5/3) × (9/7/5/3)	(256, 256, 64)
	ReLU	-	-	
max pool	MAX Pool	-	2 × 2	(128, 128, 64)
msb	MSB Conv	4 × 32	(9/7/5/3) × (9/7/5/3)	(128, 128, 128)
	ReLU	-	-	
	MSB Conv	4 × 32	(9/7/5/3) × (9/7/5/3)	
	ReLU	-	-	
max pool	MAX Pool	-	2 × 2	(64, 64, 128)
msb	MSB Conv	4 × 64	(9/7/5/3) × (9/7/5/3)	(64, 64, 256)
	ReLU	-	-	
	MSB Conv	4 × 64	(9/7/5/3) × (9/7/5/3)	
	ReLU	-	-	
up-conv	UpSample	-	2 × 2	(128, 128, 128)
	Conv	128	2 × 2	
	ReLU	-	-	
conv	Conv	128	3 × 3	
	ReLU	-	-	
	Conv	128	3 × 3	
	ReLU	-	-	
up-conv	UpSample	-	2 × 2	(256, 256, 64)
	Conv	64	2 × 2	
	ReLU	-	-	
conv	Conv	64	3 × 3	
	ReLU	-	-	
	Conv	64	3 × 3	
	ReLU	-	-	
	Conv	2	3 × 3	(256, 256, 2)
	ReLU	-	-	
conv	Conv	1	1 × 1	(256, 256, 1)

Moreover, to evaluate the detail segmentation accuracy of small bones of the hand, the detection accuracy (DACC) were calculated as

$$\text{DACC} = \frac{N_s}{N_g},$$

where Ns is the number of small bones of the hand segmented with the network and Ng is the number of small bones of the hand in groundtruth. The segmented bones are not connected to any adjacent bones. The small bones of the hand are located in three ROIs: Phalangeal ROIs; Carpal ROIs and Ulnar & Radius ROIs (Fig. 1). Specifically:

Phalangeal ROIs include: Distal phalangeal epiphysis (Dis); Middle phalangeal epiphysis (Mid) and Proximal phalangeal epiphysis (Pro);

Carpal ROIs include: Hamate (Ham); Capitate (Cap); Triquetral (Tri); Lunate (Lun); Trapezoid (Tra1); Trapezium (Tra2) and Scaphoid (Sca);

Ulnar & Radius ROIs include: Distal ulnar epiphysis (Uln) and Distal radius epiphysis (Rad).

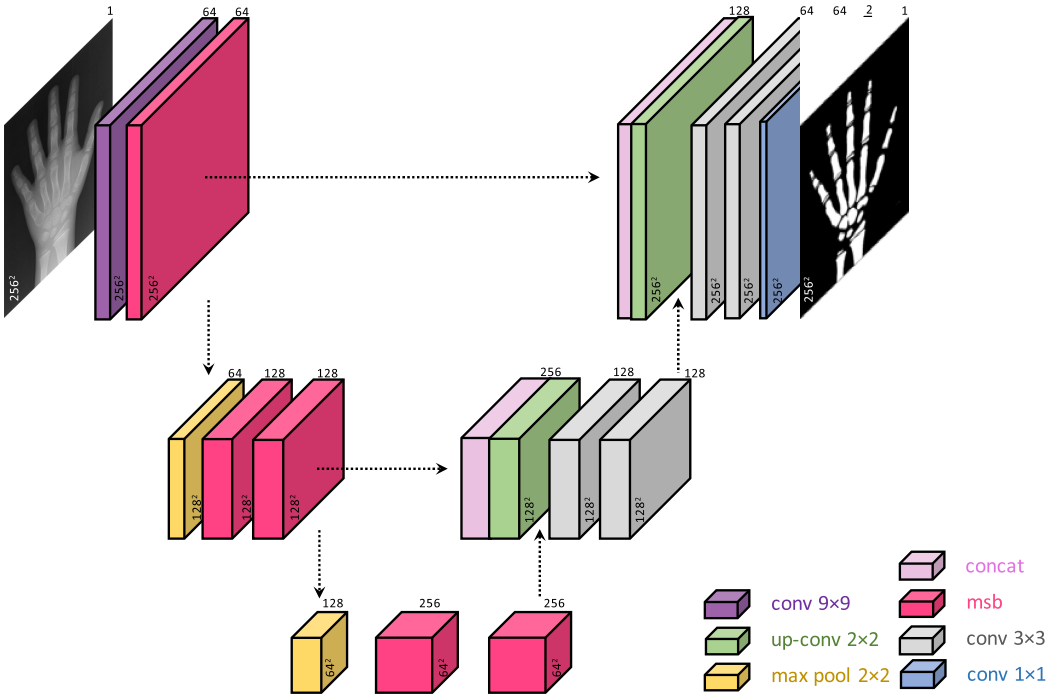


FIGURE 5. Lightweight U-Net architecture multi-scale convolutional network based on the U-Net^2 for pediatric hand bone segmentation in X-ray image.

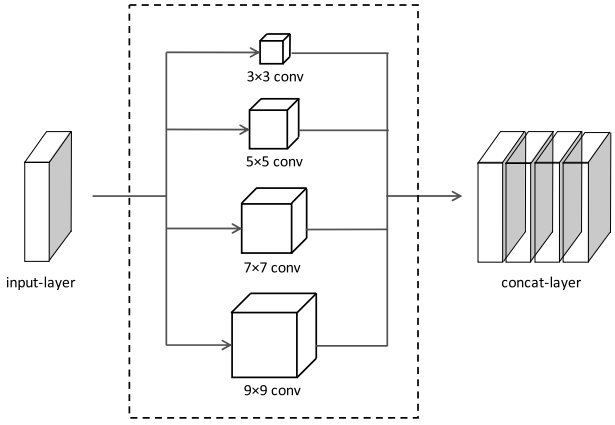


FIGURE 6. Multi-scale block with different kernel size.

III. EXPERIMENTAL RESULTS

A. DATASET

The assessment of the segmentation accuracy of the proposed network described in the previous section, was carried out on the Digital Hand Atlas Database System [1], a public and comprehensive X-ray dataset for automated skeletal bone age benchmarking. The dataset contains 1391 left-hand X-ray scans of children of age up to 18 years old, divided by gender and race. Each X-ray scan comes with two bone age values, provided by two expert radiologists. To evaluate the segmentation performance of small bones of the hand, the images of children more than 7 years old are excluded in our experiment, cause the small bones of the hand of children start to merge together since 8 years old. Finally, 429 X-ray images of children from birth to 7 years of age were employed and we randomly split the dataset into training (252 images), validation (89 images), and test (88 images) sets, without patient overlap.

B. TRAINING DETAILS

The input images and their corresponding segmentation maps are used to train the network with the stochastic gradient descent implementation. Data augmentation [27] was performed including image rotation within $-20°$ to $20°$, image translation within the ratio 0-0.2, image scaling within the ratio 0-0.2, image horizontal flip, and image brightness shifting in the whole image within the ratio 0.8-1.2. The data augmentation operation is performed by Keras Application Programming Interface [28]. The experiments are tested on Keras and trained on a NVIDIA GeForce GTX 1080Ti GPU (11GB) with a 64 GB RAM. The training time(s/epoch) is 7 and running time(s/img) is 0.0287.

C. COMPARISON OF SEGMENTATION PERFORMANCE WITH DIFFERENT U-NET BASED NETWORK ARCHITECTURES

We compare the segmentation results for hand bone segmentation with different U-Net architectures with different number of down-sampling and up-sampling operations and U-Net^{++}. Fig. 7 shows typical segmentation results (the redundant parts of the segmentation are circled in blue, the neglected parts are circled in yellow and the red arrows point

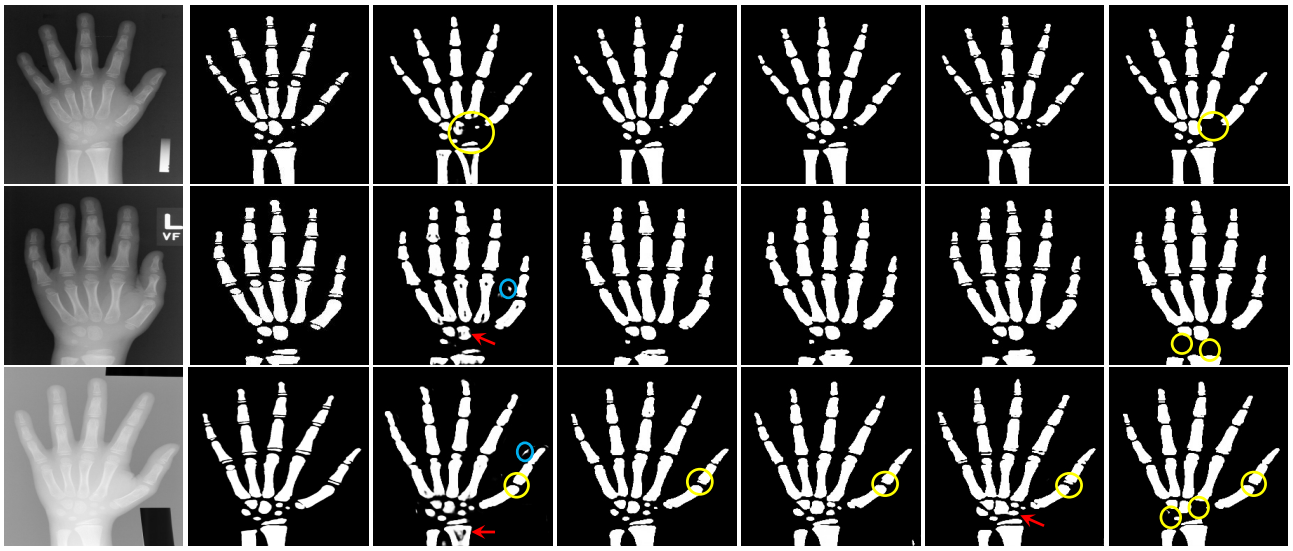


FIGURE 7. Examples of hand bone segmentation results of different U-Net based network architectures. The redundant parts of the segmentation are circled in blue, the neglected parts are circled in yellow and the red arrows point to the false alarms.

TABLE 2. Segmentation results of different U-Net based networks.

	U-Net^1	U-Net^2	U-Net^3	U-Net^4	U-Net^{++}
Dice (%)	89.0±5.3	**93.1±2.4**	92.8±2.1	92.9±2.4	92.9±2.3
IoU (%)	88.4±4.8	**92.5±2.3**	92.1±2.0	92.3±2.3	92.2±2.3
Sensitivity (%)	91.7±4.4	94.7±1.9	**94.9±2.1**	94.3±2.2	94.9±2.5
Specificity (%)	97.6±2.5	98.6±0.6	98.4±0.5	**98.6±0.5**	98.4±0.5
Params	0.4M	2.0M	7.7M	31.0M	36.2M

TABLE 3. Segmentation results of modified U-Net^2 with different kernel sizes.

	3×3	5×5	7×7	9×9	msb
Dice (%)	**93.1±2.4**	92.5±2.3	92.2±2.1	91.7±2.4	92.9±2.3
IoU (%)	92.5±2.3	91.8±2.2	91.5±2.0	91.0±2.2	**92.9±2.3**
Sensitivity (%)	94.7±1.9	92.9±1.8	93.7±1.7	92.9±1.8	**94.9±2.5**
Specificity (%)	**98.6±0.6**	**98.6±0.6**	**98.6±0.6**	98.4±0.7	98.4±0.5
Params	2.0M	4.9M	9.2M	14.9M	6.5M

to the false alarms). It shows that U-Net^2, U-Net^3 and Net^4 achieves better performance than the other networks. U-Net^1 is not robust enough against small input variance with only 1 down-sampling step, and U-Net^{++} fail to segment some small bones.

Concluded from Table 2, except the U-Net^1, the other networks get similar segmentation performance by the measurement of dice, IoU, sensitivity, specificity, respectively. In addition, U-Net^2 achieves the same segmentation performance with much less parameters and higher computational efficiency than other networks.

Moreover, the DACC of small bones of the hand were shown in Fig. 8 for evaluating the detail segmentation accuracy of different networks. In total, DACC of 12 different types of hand bones were calculated with 5 methods and the red column indicated the results of U-Net^2. The DACC of U-Net^2 was higher than other networks in 9 types of hand bones and was the second high in the rest 3 types (Pro, Rad and Tra1).

D. COMPARISON OF SEGMENTATION PERFORMANCE OF MODIFIED U-NET^2 WITH DIFFERENT KERNEL SIZES

The improvement of segmentation performance with msCNN was evaluated replacing the MSBs with different single kernel sizes based on U-Net^2 (Fig. 5). Fig. 9 shows typical segmentation results. It shows that msCNN achieves better

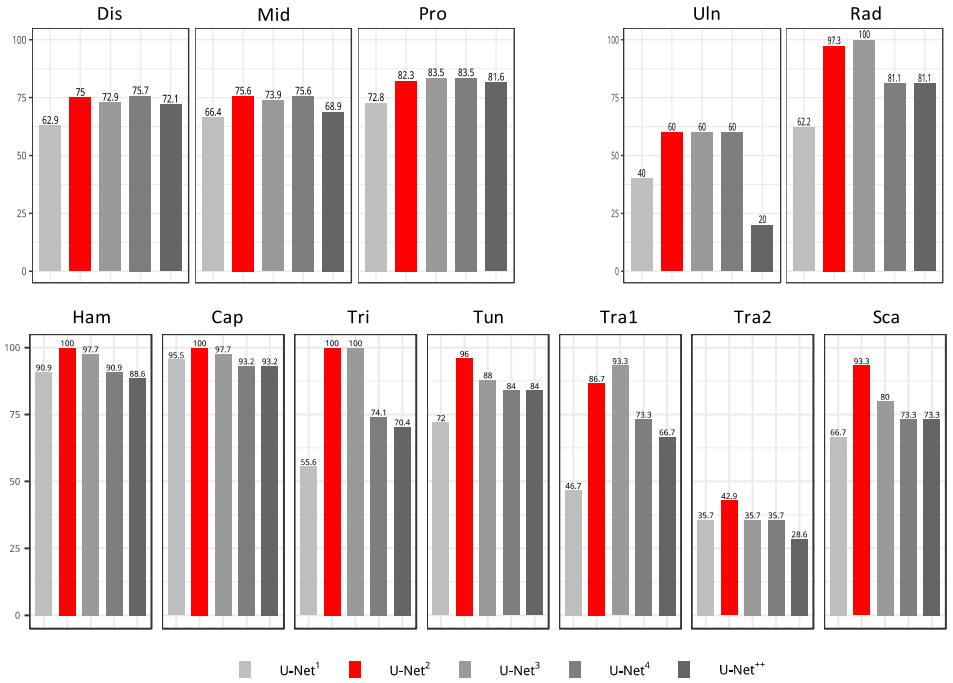


FIGURE 8. The detection accuracy (DACC) of small bones of the hand of different U-Net based network architectures.

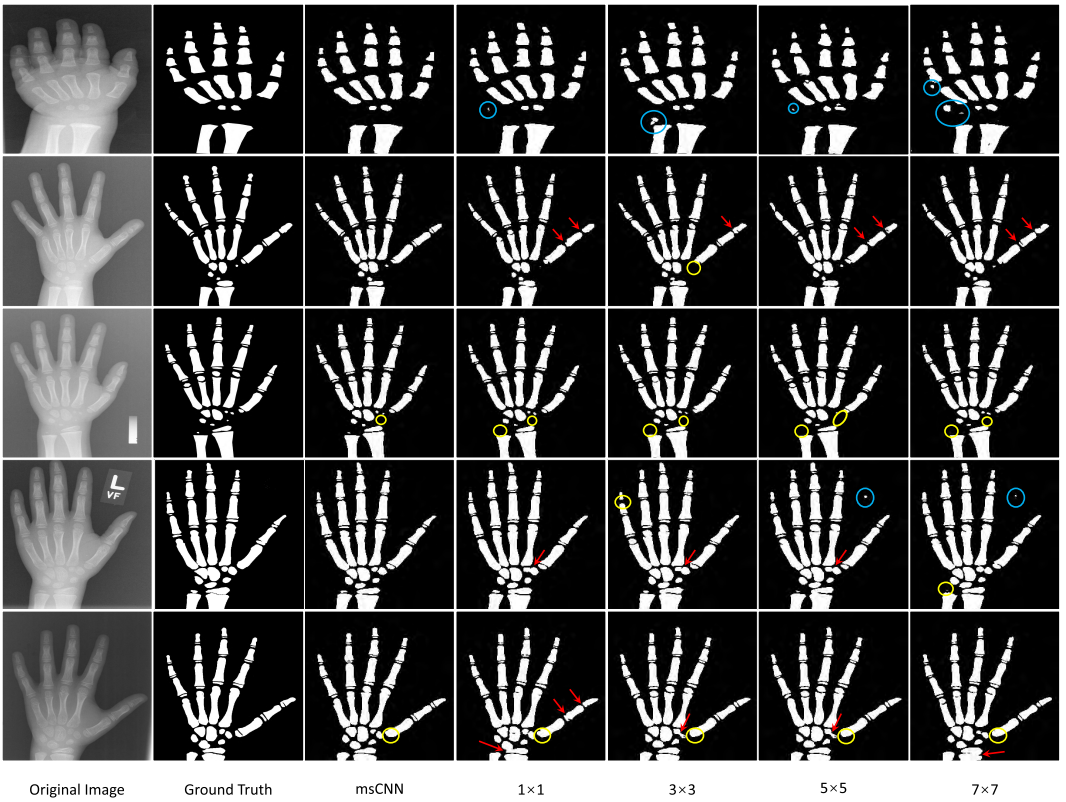


FIGURE 9. Examples of hand bone segmentation results of modified U-Net^2 with different kernel sizes. The redundant parts of the segmentation are circled in blue, the neglected parts are circled in yellow and the red arrows point to the false alarms.

performance than networks with single kernel size. From Table 3, CNN with single kernel size or MSB all get similar segmentation performance by the measurement of dice, IoU, sensitivity, specificity, respectively.

The DACC of small bones of the hand were shown in Fig. 10 with different networks. DACC of 12 different types of hand bones were calculated with 5 methods and the red column indicated the results of msCNN. The DACC of msCNN

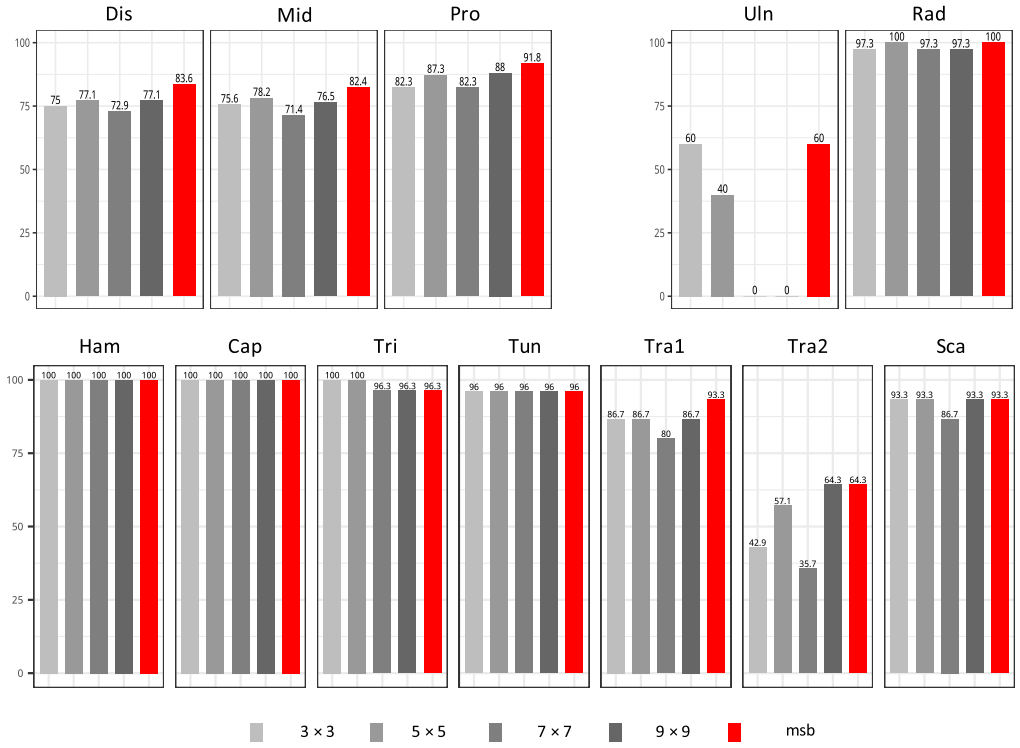


FIGURE 10. The detection accuracy (DACC) of small bones of the hand of modified U-Net^2 with different kernel sizes.

was higher than other networks, particularly with Phalangeal ROIs (Dis; Mid; Pro).

IV. DISCUSSION

There have been lots of automatic BAA studies [1], [2], [6], [8]–[11]. However, in the clinical setting, not only the bone age but also the specific descriptions of hand bones features are needed in medical records. To assess the bone age, a set of ROIs of the hand and wrist joints are required to be analyzed according to the G&P method as well as the TW method. To provide more specific diagnostic message besides bone age, we proposed a lightweight U-Net architecture multi-scale convolutional network for pediatric hand bone segmentation in X-ray image. In our experiment over the Digital Hand Atlas Database System, this method has achieved promising segmentation results, especially for segmentation of small bones of the hand.

As the children grow up from birth to 7 years of age, the small bones of the hand start to appear and become larger and larger. After 8 years old, they start to merge together. As a result, the appearance of small bones of the hand, especially small ossification centers in the carpal bones and epiphyses of tubular bones, are vital for bone age assessment, especially for 0-7 years old children [4]. Therefore, precise segmentation of small bones of the hand is more vital in clinical setting. Because the traditional evaluation metrics such as IoU and dice coefficient could only evaluate the segmentation performance from the perspective of the entire image, we adopted the DACC for assessment, which reflects the number of the small bones detected with different methods. As the results show, the network we proposed achieves higher performance in segmentation of small bones of the hand.

Ronneberger *et al.* [18] introduced the U-Net and this architecture performs well for biomedical image segmentation such as cell [18], brain tumor [20] kidney [29] segmenting. This structure contains serval down-sampling and up-sampling steps, providing higher level features such as the location information. And the concatenated layers with normal resolution provide the detailed local appearance of structures. For small bones of the hand segmentation task, detailed features are more important than higher level features relatively. In our experiment, we compared different U-Net architectures and U-Net^2 and U-Net^3 achieve higher detail segmentation performance than other networks. Therefore, we choose U-Net^2 which can achieve satisfying results with a small number of training parameters. The compact structure also allows our model to be converged in a short time during training.

GoogLeNet [25] is another widely used network first proposed in 2014 and won the classification task of ILSVRC2014. An inception module was proposed in the network with multi-scale filters, which could extract more features from different scales. As hand bones become larger with the growth in children (Fig. 2), same sized kernels combination may not counter hand bone scale variations. Considering the improvement of the inception module in

GoogLeNet and the elegant architecture of U-Net, we combine both advantages and use different scaled filters in U-Net^2 without changing the depth of the network. In our experiment, U-Net^2 with MSBs obtain higher DACC than other single kernel size networks, particularly with Phalangeal ROIs (Dis; Mid; Pro). Hand bones in Phalangeal ROIs are relatively smaller than others and close to adjacent bones, making it difficult to obtain precise segmentation. The results indicate promising performance of our method in hand bones segmenting, especially for small bones of the hand.

V. CONCLUSION

We propose a lightweight U-Net architecture multi-scale convolutional network for pediatric hand bone segmentation in X-ray image and it has achieved promising segmentation results, especially for segmentation of small bones of the hand.

REFERENCES

[1] A. Gertych, A. Zhang, J. Sayre, S. Pospiech-Kurkowska, and H. K. Huang, "Bone age assessment of children using a digital hand atlas," *Comput. Med. Imag. Graph.*, vol. 31, nos. 4–5, pp. 322–331, 2007.

[2] V. Gilsanz and O. Ratib, *Hand Bone Age: A Digital Atlas of Skeletal Maturity*. Heidelberg, Germany: Springer, 2005.

[3] S. Ritz-Timme, C. Cattaneo, M. J. Collins, E. R. Waite, H. W. Schütz, H.-J. Kaatsch, and H. I. M. Borrman, "Age estimation: The state of the art in relation to the specific demands of forensic practise," *Int. J. Legal Med.*, vol. 113, no. 3, pp. 129–136, 2000.

[4] W. W. Greulich and S. I. Pyle, "Radiographic atlas of skeletal development of the hand and wrist," *Amer. J. Med. Sci.*, vol. 238, no. 3, p. 393, 1959.

[5] J. M. Tanner, R. H. Whitehouse, N. Cameron, W. A. Marshall, M. J. R. Healy, and H. Goldstein, *Assessment of Skeletal Maturity and Prediction of Adult Height (TW2 Method)*. New York, NY, USA: Academic, 1975.

[6] V. D. Sanctis, S. D. Maio, A. T. Soliman, G. Raiola, R. Elalaily, and G. Millimaggi, "Hand X-ray in pediatric endocrinology: Skeletal age assessment and beyond," *Indian J. Endocrinol. Metabolism*, vol. 18, pp. S63–S71, Nov. 2014.

[7] D. G. King, D. M. Steventon, M. P. O'Sullivan, A. M. Cook, V. P. L. Hornsby, I. G. Jefferson, and P. R. King, "Reproducibility of bone ages when performed by radiology registrars: An audit of tanner and whitehouse II versus Greulich and Pyle methods," *Brit. Inst. Radiol.*, vol. 67, no. 801, pp. 848–851, 1994.

[8] E. Pietka, A. Gertych, S. Pospiech, F. Cao, H. K. Huang, and V. Gilsanz, "Computer-assisted bone age assessment: Image preprocessing and epiphyseal/metaphyseal ROI extraction," *IEEE Trans. Med. Imag.*, vol. 20, no. 8, pp. 715–729, Aug. 2001.

[9] D. Giordano, R. Leonardi, F. Maiorana, G. Scarciofalo, and C. Spampinato, "Epiphysis and metaphysis extraction and classification by adaptive thresholding and DoG filtering for automated skeletal bone age analysis," in *Proc. 29th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2007, pp. 6551–6556.

[10] X. Ren, T. Li, X. Yang, S. Wang, S. Ahmad, L. Xiang, S. R. Stone, L. Li, Y. Zhan, D. Shen, and Q. Wang, "Regression convolutional neural network for automated pediatric bone age assessment from hand radiograph," *IEEE J. Biomed. Health Inform.*, to be published.

[11] C. Spampinato, S. Palazzo, D. Giordano, M. Aldinucci, and R. Leonardi, "Deep learning for automated skeletal bone age assessment in X-ray images," *Med. Image Anal.*, vol. 36, pp. 41–51, 2017.

[12] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.

[14] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[15] A. M. Ali, A. A. Farag, and A. S. El-Baz, "Graph cuts framework for kidney segmentation with prior shape constraints," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Heidelberg, Germany: Springer, 2007, pp. 384–392.

[16] Z. Wang, K. K. Bhatia, B. Glocker, A. Marvao, T. Dawes, K. Misawa, K. Mori, and D. Rueckert, "Geodesic patch-based segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Heidelberg, Germany: Springer, 2014, pp. 666–673.

[17] H. R. Roth, L. Lu, A. Farag, H.-C. Shin, J. Liu, E. B. Turkbey, R. M. Summers, "DeepOrgan: Multi-level deep convolutional networks for automated pancreas segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Heidelberg, Germany: Springer, 2015, pp. 556–564.

[18] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Heidelberg, Germany: Springer, 2015, pp. 234–241.

[19] K. Kamnitsas, C. Ledig, V. F. J. Newcombe, N. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Med. Image Anal.*, vol. 36, pp. 61–78, Feb. 2017.

[20] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, "Automatic brain tumor detection and segmentation using U-net based fully convolutional networks," in *Proc. Annu. Conf. Med. Image Understand. Anal.* Heidelberg, Germany: Springer, 2017, pp. 506–517.

[21] Y. Chen, Z. Cao, C. Cao, J. Yang, and J. Zhang, "A modified U-net for brain MR image segmentation," in *Proc. Int. Conf. Cloud Comput. Secur.* Springer, 2018, pp. 233–242.

[22] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers Tiramisu: Fully convolutional DenseNets for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1175–1183.

[23] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-net architecture for medical image segmentation," in *Proc. Int. Workshop Deep Learn. Med. Image Anal. Int. Workshop Multimodal Learn. Clin. Decis. Support*. Heidelberg, Germany: Springer, 2018, pp. 3–11.

[24] L. Zeng, X. Xu, B. Cai, S. Qiu, and T. Zhang, "Multi-scale convolutional neural networks for crowd counting," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 465–469.

[25] C. Szegedy, X. Xu, B. Cai, S. Qiu, and T. Zhang, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[26] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Int. Conf. Mach. Learn.*, 2010, pp. 807–814.

[27] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," 2017, *arXiv:1712.04621*. [Online]. Available: https://arxiv.org/abs/1511.08977

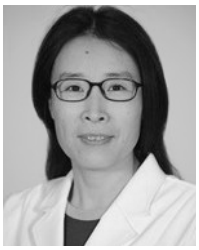
[28] F. Chollet. (2015). *Keras*. [Online]. Available: https://github.com/fchollet/keras

[29] Ö.Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-net: Learning dense volumetric segmentation from sparse annotation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Heidelberg, Germany: Springer, 2016, pp. 424–432.

LIAN DING received the B.Sc. degree from the University of Electronic Science and Technology of China, in 2014. He is currently pursuing the Ph.D. degree with Peking University. He is currently a Researcher with the Academy for Advanced Interdisciplinary Studies, Peking University. His research interest includes medical image processing and applications using machine learning.

KAI ZHAO received the Ph.D. degree from the Peking University of Medical Imaging and Nuclear Medicine, in 2015. He is currently a Radiologist with the Peking University First Hospital. His research interests include functional magnetic resonance imaging and medical image processing.

XIAODONG ZHANG received the Ph.D. degree in biomechanics and medical engineering from Peking University, China, in 2012. He is currently a Medical Physicist with the Department of Radiology, Peking University First Hospital. His research interests include medical image post-processing and MR Imaging.



XIAOYING WANG received the M.D. degree in clinical medicine from Peking University, China, in 1999. She is currently a Professor of radiology with the Peking University First Hospital. Her research interests include medical image analysis and MR Imaging.

JUE ZHANG received the Ph.D. degree in engineering mechanics from Peking University, China, in 2003, where he is currently an Associate Professor with the College of Engineering. His research interests include medical signals & image analysis and MR Imaging.