Inverse Problems and Imaging doi:10.3934/ipi.2020057

AUTOMATIC SEGMENTATION OF THE FEMUR AND TIBIA BONES FROM X-RAY IMAGES BASED ON PURE DILATED RESIDUAL U-NET

Weihao Shen†, Wenbo Xu†, Hongyang Zhang and Zexin Sun

School of Artificial Intelligence, Hebei University of Technology
Tianjin 300401, China

Jianxiong Ma and Xinlong Ma

Tianjin Institute of Orthopaedics, Tianjin Hospital, Tianjin University
Tianjin 300211, China

Shoujun Zhou*

Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
Shenzhen 518055, China

Shijie Guo and Yuanquan Wang*

School of Artificial Intelligence
Hebei Key Laboratory of Robot Perception and Human-Robot Interaction
Hebei University of Technology
Tianjin 300401, China

Abstract. X-ray images of the lower limb bone are the most commonly used imaging modality for clinical studies, and segmentation of the femur and tibia in an X-ray image is helpful for many medical studies such as diagnosis, surgery and treatment. In this paper, we propose a new approach based on pure dilated residual U-Net for the segmentation of the femur and tibia bones. The proposed approach employs dilated convolution completely to increase the receptive field, in this way, we can make full use of the advantages of dilated convolution. We conducted experiments and evaluations on datasets provided by Tianjin hospital. Comparison with the classical U-net and FusionNet, our method has fewer parameters, higher accuracy, and converges more rapidly, which means the high performance of the proposed method.

1. **Introduction.** X-ray images are the most commonly used imaging modality for clinical studies of the lower limb bones, and segmentation of bone structure in x-ray images is of paramount importance for clinical studies. In comparison with the segmentation of other modal images such as MR and CT images, less research has been conducted on the segmentation of x-ray images. Unlike the modalities such as MR and CT that are in slice manner, the bones in pelvis region in x-ray images often overlap with other organs, since the bones are naturally 3D while the x-ray images are just 2D projection of the 3D bones. Therefore, automatic segmentation

2020 *Mathematics Subject Classification.* Primary: 68T07; Secondary: 68T20.

Key words and phrases. Image segmentation, U-Net, dilated convolution, X-ray imaging, femur and tibia bones.

The first author is supported by NSFC grant 61976241.

† These authors contribute equally to this paper.

* Co-Corresponding author: Shoujun Zhou and Yuanquan Wang.

of bone structures in x-ray images is intrinsically difficult.

In the past decade, the deep learning networks have got fantastic results in many image analysis tasks, such as classification, detection, tracking and segmentation. In 2015, Long et al. [21] proposed the fully convolutional network (FCN), which opened up a new field, using end-to-end networks, for semantic segmentation. However, the FCN model usually cannot get high accuracy results. Classical FCN-8s finally reached 90.3% pixel accuracy on the test set of PASCAL VOC2011 by training on a large set. However, 90.3% pixel accuracy is still insufficient to medical image segmentation, because the extraction of some medical image information requires images with high resolution. Amazingly, a novel network architecture, i.e., the U-Net, is proposed by Ronneberger et al. [29] in 2015, which improves the prediction accuracy. The U-Net is composed of an encoder and a decoder. Through a series of tricks like skip-connection, max-pooling, and transpose convolution, the U-Net can collect global information and local information effectively. One can make precise predictions sine that the global information tells us what it is while the local one tells us where it is. Unfortunately, there are several issues in the classical U-Net, which make it unsuitable for X-ray bone image segmentation. For example, since the max-pooling operation requires downsampling the feature map, it will not only increase the receptive field but also shrink the size of outputs. Max-pooling operation first divides feature maps into blocks. Then it selects the maximum value in each block to form the next level of feature maps. Obviously, the information remaining except the maximum value in each block lost irreversibly. The loss of structured information caused by the max-pooling operation is irreversible [32], which means it's difficult for one to get a high-precision prediction using the classical U-Net.

To increase the receptive field without the loss of structure information, Yu et al. [43] proposed a new kind of convolution called dilated convolutions. By rational design, not only can we guarantee the integrity of the information, but also we can increase the receptive field of our network effectively. Hongwei Li et al. [17] have achieved automatic brain structures segmentation using deep residual dilated U-net. But they also used max-pooling operations to increase the receptive field so that they cannot connect each layer freely.

In this paper, we propose a novel network to deal with the task of X-ray lower limb image segmentation. All the max-pooling operations are removed so that we increase the receptive field by taking use of the dilated convolution. Since the size of outputs and inputs of each layer are identical, we can employ skip connections between different layers with different levels of structure information. We called the proposed network pure dilated residual U-Net, PDRU-Net in short. The PDRU-Net owns high-speed of convergence, and accurate predictions on our validation set, about 0.9875 in terms of Dice Coefficient. All of these make it more suitable for the segmentation of the X-ray bone image. We verified our method on the training set of the X-ray lower limb bone images obtained from Tianjin hospital, and we compared our model with the U-Net and FusionNet [28].

In the next section, we summarize the related works on X-ray bone segmentation, and then, the proposed network architecture is introduced. Afterward, we recommend the selection of hyper-parameters in the process of training. We compared our network with the U-Net and FusionNet, and conclusion is drawn at the end.

2. **Related works.** X-ray bone image segmentation plays an important role in the diagnosis of orthopedic diseases. However, because of the high noise, low contrast, overlapping and imaging pose variability, segmentation of the femur and its adjacent bones is a challenging task, and almost all the general medical image segmentation methods have been applied to the task on hand, for example, the proposed method in [3] integrates an entropy-based segmentation method with an adaptive thresholding-based contour tracing to generate the bone-contour of an X-ray image. In [31], the canny edge detector is employed to extract the femur bone automatically. Both of the two methods are sensitive to noise and cannot make use of prior information of the object.

The active contour, or snake model [15, 26, 45], is one kind of popular methods for the bone segmentation from X-ray images. The snake model is a controllable continuous spline curve, which can deform iteratively under the action of internal and external forces to match the object contour in the image. Jiang et al. [15] proposed a segmentation method for fractured bones in X-ray images using active contour model with prior shape information as global constraint. In [26], the gradient vector flow (GVF), a typical external force for active contours, was employed in an automated fracture detection system to obtain accurate femur contour. Zhang et al. [45] proposed a model named gradient vector flow over manifold(GVFOM) for image segmentation.

Since the bones have a tendency towards some average shape, the statistical shape model (SSM) is also employed for this task. In [42], a hierarchical multi-object SSM representing joint structure was employed to segment the femur and pelvis from 3D data automatically, both shape and pose variations are embedded in a combined pelvis and femur SSM. Tack et al. [37] proposed a segmentation method employing convolutional neural networks in combination with statistical shape models. Segmentation methods based on SSM perform well on bone segmentation when sufficient training samples are available.

The active shape model (ASM) is a widely used statistical analysis algorithm, which has higher robustness and is less vulnerable to noise. Smith et al. [33] proposed the Spline-ASM method that is a novel extension of the standard ASM algorithm by using cubic spline interpolation to maintain curved contours. Wu et al. [39] proposed a framework for the bone segmentation and 3D visualization of pelvic CT images, where a registered active shape model (RASM) with novel initialization was incorporated in the segmentation method to extract pelvic bone tissues.

Classification-based methods are also employed for bone structure segmentation. Zhang et al. [44] presented an automatic knee cartilage segmentation system, where the support vector machine(SVM) is employed for classification. An interactive X-ray bone segmentation method was proposed in [34] where the mean-shift algorithm is combined with a region growing algorithm based on the maximal similarity between regions. Öztürk and Albayrak [25] proposed a voxel-classification-driven region-growing method, where an automatic method to select seed points and an effective vicinity-correlated sparse subsampling strategy were employed to improve the performance of automatic segmentation of knee joint cartilage in high-field magnetic resonance images.

Graph-based segmentation methods work well in muti-object segmentation tasks, and are widely employed in bone segmentation. Ababneh et al. [1] employed a graph-cut based method to segment knee magnetic resonance images, the method is automatic and can separate bone from highly similar adjacent structures

accurately.

Atlas-based approaches are also popular for bone segmentation, the basic idea of which is to label the input images according to the atlas that contain anatomical information. In [8], an atlas-based approach is employed for automatic segmentation of femurs in x-ray images. The segmentation algorithm consists of two stages: global alignment and local refinement. The global alignment method was used to register the whole atlas to the image under joint constraints followed by a local refinement method to extract detailed contours. Results show that the proposed method is robust and accurate. Carballido-Gamio et al. proposed an automatic segmentation method based on multi-atlas [4], which enabled accurate and reproducible quantitative analyses of the proximal femur.

Recently, the convolutional neural network (CNN) has shown great performance in several image processing tasks [16, 36, 18, 2, 5, 38]. Compared to traditional hand-craft features, one of the main advantages of CNNs is that convolutional filters are learned from the training image data via an automatic optimization process. Ronneberger et al. [29] proposed the U-Net based on CNN, which is widely used for semantic segmentation due to its high accuracy, and many variants of the U-Net [6, 40, 11, 14, 24, 23] were proposed subsequently. 3D U-Net [6] is a simple extension of U-Net, which was applied to 3D image segmentation. 3D U-Net retains the original excellent features of U-Net, replacing 2D convolution with 3D convolution.The network solves the awkward situation that 3D images are sliced into the model one by one for training, and greatly improves the training efficiency. A model named Res-UNet [40] was employed to segment retinal image. In the model structure of Res-UNet, residual connection was introduced in every submodule of U-Net to accelerate convergence and improve the accuracy of deep model. A weighted attention mechanism was added to the model, which makes the model learn more features to distinguish vascular and non vascular pixels. Guan et al. [11] replaced every submodule of U-Net with dense connection [14], which means that the output of a certain layer in the submodule is taken as a part of the input of subsequent layers, and the input of a certain layer comes from the combination of the output of previous layers, this kind of model structure makes more effective use of the features of each layer, and also reduces the number of parameters to a certain extent. Attention UNet [24] introduced attention mechanism in UNet. With the increase of training times, the attention weight obtained in this image tends to get a large value in the target organ area and a small value in the background area, which is helpful to improve the accuracy of image segmentation. In [23], 3D convolution was employed in a model named V-Net to realize the end-to-end segmentation of 3D image, and a novel loss function based on Dice coefficient was introduced to deal with the strong imbalance between foreground voxels and background voxels.

Recent works [46, 35, 20] apply deep neural network model based on U-Net to medical image segmentation. Zhou et al. [46] proposed a new architecture named D-UNet for chronic stroke lesion segmentation, which combines 2D and 3D convolution innovatively in the encoding stage. Sun et al. [35] proposed a novel attention-guided dense- upsampling network named AUNet for accurate breast mass segmentation. Liu et al. [20] proposed a multi-scale deep fusion network named MSDF-Net for stroke lesion segmentation.

Modified U-Net was used in bone segmentation such as works in [37, 7, 27]. Deniz et al. [7] extended the U-Net into three dimensions for volumetric segmentation using three-dimensional convolution, up-convolution and max-pooling layers,

and concatenated dilated convolutions with different dilation rates to improve the segmentation performing. Tack et al. [37] proposed a knee menisci segmentation method that combines U-Net with statistical shape models (SSMs). Pries et al. [27] proposed a method called deep morphing, where a variant of the U-net called FusionNet [28] was combined with a point distribution model (PDM) to detect the proximal femur with prior knowledge. Generative adversarial networks(GANs) [10, 47, 41, 22, 30, 12, 19] were widely used in semantic segmentation. In [9], a method based on GAN was used to segment femur in X-ray images,in which an adaptive contours estimation approach and an absolute deviation loss function was employed to optimize the networks.

3. **The proposed PDRU-Net.** The proposed model is an end-to-end trainable network for automatic segmentation of the femur and tibia bones from X-ray image. The network is called pure dilated residual U-Net(PDRU-Net in short) which is based on the U-Net, it means that our PDRU-Net consists of an encoder path to extract global and local information and a decoding path to merge the above information and make dense predictions. Figure 1 represents the structure of the proposed network. The PDRU-Net exploits the inherent advantages of the U-Net skip connections [29], residual learning [13] and dilated convolutions [43] to capture rich context information and offer robust bone structure segmentation. Figure 2 shows the diagram of two key modules of our network. The network architecture starts with a standard block that consists of two consecutive convolutional layers with 40 filters of size 3×3. Each residual block consists of two consecutive convolutional layers with 40 filters of size 3×3 and a residual skip connection that is able to avoid gradient vanishing. It is worth mentioning that a batch normalization operation and a Rectified Linear Unit (ReLU) activation function follow successively after each convolution layer.

In contrast to the classical U-Net, to avoid information loss caused by the downsampling operation, we remove all the max-pooling layers so that we also removed all the upsampling layers. Therefore, the output and input size of each layer are of size 512×512. Because of this priority of our network, we can add the skip connection structure between any two blocks freely to accelerate the speed of convergence further. Unlike the classical U-Net that achieves skip connection by concatenation, we achieve all the skip connections in our network by addition. In addition, as shown in Figure 1, our network has nested skip connections that are able to avoid gradient vanishing help structure information flow within and across levels in the network.

In the task of X-ray lower limb bone image segmentation, each bone takes about a quarter of the picture, i.e., an image of size 256×256 since the whole image is of size 512×512. In order to achieve the segmentation goal of each target bone, we need to enable the receptive field of our network to cover each target bone at least. Considering that in the real X-ray original image, the position of the target bones is not neatly located in the center of the image due to deformity of bones or different patient postures, we need to further expend the receptive field of our network to make our method applicable to all different cases of real original images. As mentioned earlier, the dilated convolution [43] is able to increase the receptive field without the loss of structure information caused by the downsampling operation. Due to this advantage, our network used dilated convolution to increase the receptive field. As shown in Figure 1, we increase the dilated rate in the encoding

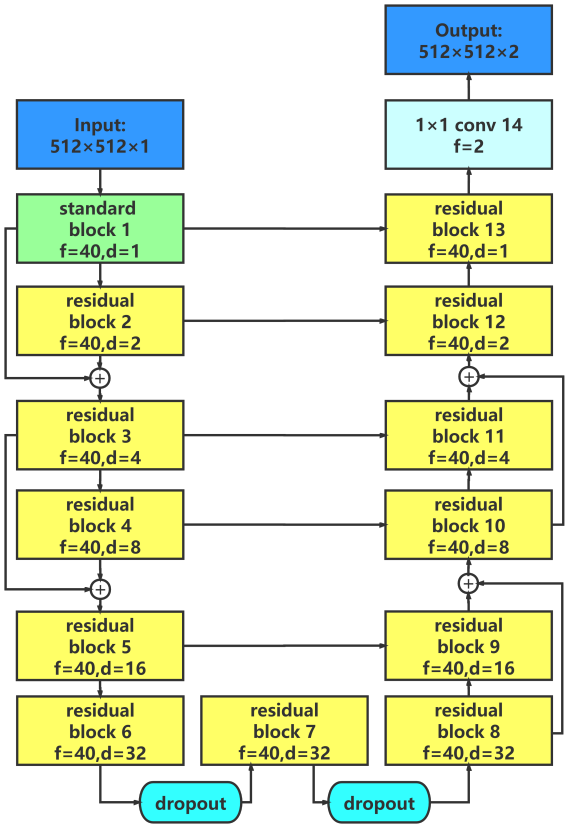

FIGURE 1. The Architecture of PDR U-Net. f represents the number of filters.d represents the dilated rate.The keep rate of dropout is 0.7

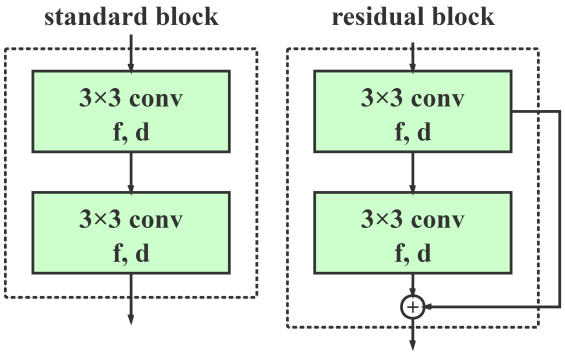

FIGURE 2. The details of the standard block and residual block. f represents the number of filters. d represents the dilated rate

path layer by layer. According to the calculation formula of the receptive field proposed in [43], the result of the receptive field in the last layer of encoding path, i.e., residual block 8, is 509, which is sufficient to cover each target bone from different original images of size 512×512. The receptive field of each block in the encoding path of the proposed network is shown in Table 1.

Our network can also reduce the number of training parameters effectively.

TABLE 1. The receptive field of each block in the encoding path of PDRU-Net.

Block Type	Convolutional Layer	Receptive Field
standard block 1	conv1_1	1-1+1×2+1=3
dilated rate = 1	conv1_2	3-1+1×2+1=5
residual block 2	conv2_1	5-1+2×2+1=9
dilated rate = 2	conv2_2	9-1+2×2+1=13
residual block 3	conv3_1	13-1+4×2+1=21
dilated rate = 4	conv3_2	21-1+4×2+1=29
residual block 4	conv4_1	29-1+8×2+1=45
dilated rate = 8	conv4_2	45-1+8×2+1=61
residual block 5	conv5_1	61-1+16×2+1=93
dilated rate = 16	conv5_2	93-1+16×2+1=125
residual block 6	conv6_1	125-1+32×2+1=189
dilated rate = 32	conv6_2	189-1+32×2+1=253
residual block 7	conv7_1	253-1+32×2+1=317
dilated rate = 32	conv7_2	317-1+32×2+1=381
residual block 8	conv8_1	381-1+32×2+1=445
dilated rate = 32	conv8_2	445-1+32×2+1=509

In the classical U-Net, in order to reduce the loss of structure information caused by downsampling, the number of feature map channels is doubled after each max-pooling layer, which make the training parameters increasing exponentially. Our network uses dilated convolution to increase receptive field without loss of information, so the number of convolution filters is set to 40 and do not need to double it after each layer. Finally, our PDRU-Net only has about 0.36M parameters to train, while the classical U-Net has about 33M parameters.

In summary, the PDRU-Net solves the problem mentioned above successfully, i.e., we can avoid the training parameters increasing exponentially when the size of our input image is large. We can also keep the details of the image better than classical U-Net. With the skip connection between blocks with the different receptive field, our network architecture can convergent more rapidly.

4. **Experimental results.**

4.1. **Data preprocessing and augmentation.**

The X-ray bone images are obtained from Tianjin Hospital, China, and there are 100 X-ray images, there is bone malformation for each image to some degree. We acquire our label set by marking the original images carefully by an orthopedics specialist. The LabelMe software [49] is employed to draw the outlines of the tibia and the femur, there are 112 points marked on the femur and 104 points on the tibia respectively, one example

is shown in Figure 3 .After that, we do data augmentation to avoid overfitting of our network. To do data augmentation, we have to merge the images and the corresponding labels at first, by making input image occupy the red channel, while femur label occupies the green channel and tibia label occupies the blue channel. Then, by rotating the images in the range of -45°∼+45°, with shifting and scaling, we can expand one sample to 50 ones. Then, we split the merged image into the input image and label image. The whole process of data augmentation is demonstrated in Figure 4 .We use random sampling method to divide the whole data set into training set, validation set and test set. At last, we get 5000 samples in total, which consists of 4400 training samples, 300 validation samples, and 300 test samples.

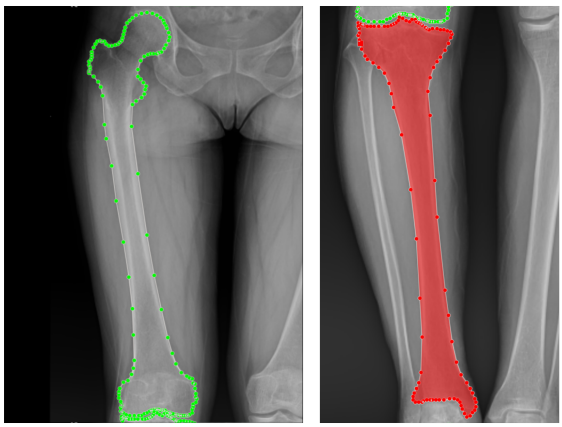

FIGURE 3. the illustration on the left is an unfilled femur label and on the right is a filled tibia label

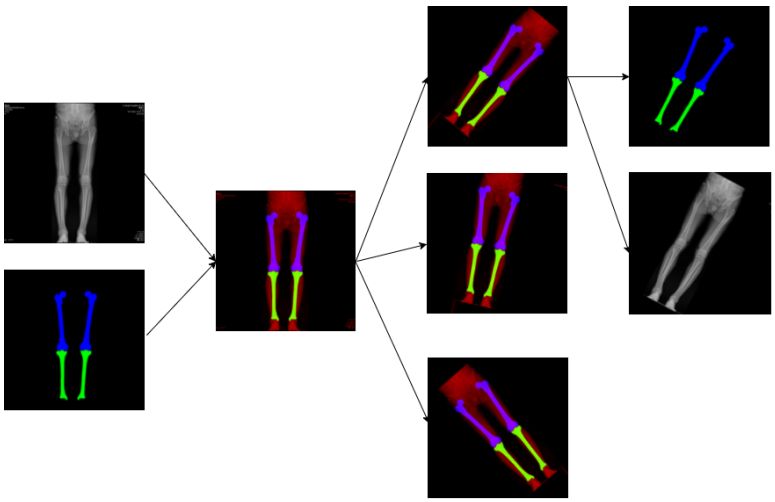

FIGURE 4. The whole process of data augmentation

4.2. **Experimental setup.** In this section, we implement the proposed deep network and make comparison with the classical U-Net [29] and FusionNet [28]. All of our program are implemented using an open-source deep learning library Keras [48], and we choose TensorFlow as the backend deep learning engine. As a high-level API, Keras can help researchers to implement their idea quickly and efficiently. Our experiments are mainly conducted on one workstation equipped with one NVIDIA GTX GeForce 1080Ti GPU with 11 GB video memory, 32 GB main memory, and an Intel i7 CPU.

During training stage, we choose the same training set and experimental setting for all approaches, i.e., PDRU-Net,U-Net and FusionNet. We use the training samples mentioned in Section 4.1 as the training set. We maximize the Dice Coefficient (DC) to optimize our network, which is shown in Eq (1), where P represents the prediction label, while G represents the ground truth label. We set smooth=1 to avoid P and G both are 0 simultaneously. The networks are optimized using Adam gradient descent algorithm, and we set $\beta 1$=0.9, $\beta 2$=0.99, and batch-size=1. We set the learning rate to 0.0004 and the training epoch number to 30. During testing stage, we evaluate the networks on the test samples mentioned in Section 4.1 and compare the three networks in terms of convergence rate, number of parameters and prediction accuracy.

$$DC = \frac{|P| \bigcap |G| \times 2 + smooth}{|P| + |G| + smooth} \tag{1}$$

After about 15 hours of training, our network has run 30 epochs, reaching the Dice Coefficient of 97.26% on the test set, which is a relatively high level result in the segmentation task. Figure 5 shows the change of loss value on training set and validation set during the train process. In terms of the number of model parameters, our PDRU-Net only has about 0.36M parameters, while the number of parameters of classical U-Net and FusionNet are about 33M and 78M, respectively. With few parameters, our network can avoid overfitting effectively.

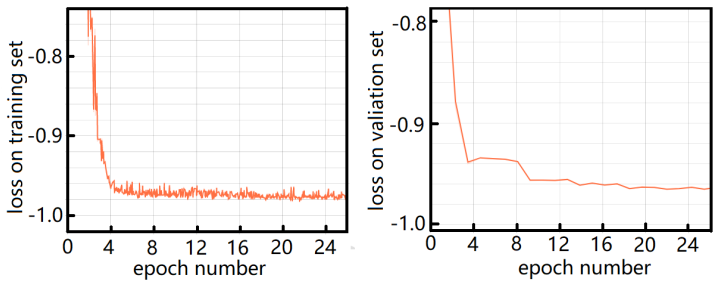

FIGURE 5. The caption on the left is the loss of PDRU-Net on the training set, and that on the right is the loss of PDRU-Net on the validation set

To better evaluate the performance of PDRU-Net, we employ the U-Net and FusionNet for comparison. We test our network architecture on the data set provided by Tianjin Hospital, and all training details of U-Net and FusionNet are the same as that of PDRU-Net. Figure 6 shows the segmentation results of the three

models in the first three training epochs, and the results show that the convergence speed of our network is significantly higher than that of U-Net and FusionNet.

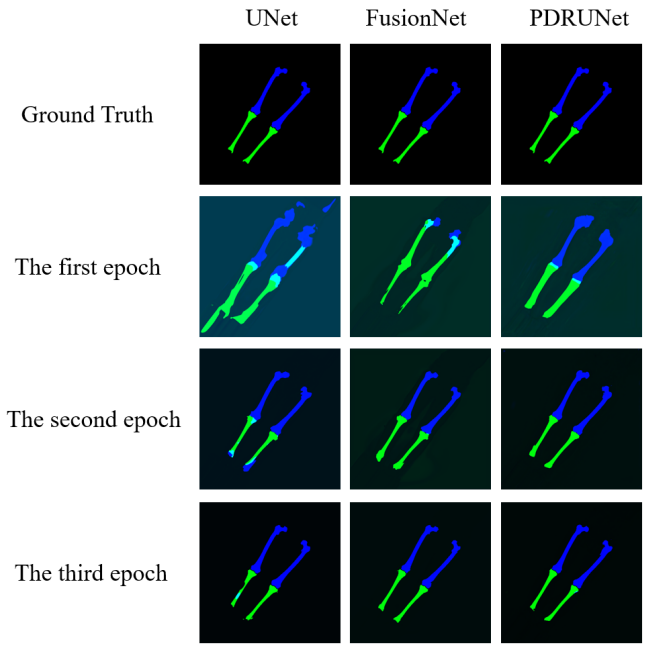

FIGURE 6. Segmentation results of the first three training epochs

The segmentation results of several samples on the test set are shown in Figure 7, the segmentation results show that our network architecture outperforms U-Net evidently and performs slightly better than the FusionNet. To further compare the performance of these three models, we select two common semantic segmentation metrics: Pixel Accuracy (PA) and Dice Coefficient (DC). The Pixel Accuracy is defined as follows:

$$PA = \frac{\sum_{i=0}^{k} p_{ii}}{\sum_{i=0}^{k} \sum_{j=0}^{k} p_{ij}} \tag{2}$$

where p_{ij} represents the number of pixels of class i yet predicted to class j, where there are k different classes. p_{ii} represents the number of pixels of class i and predicted to class i, which means the number of correctly predicted pixels. The calculation formula of DC is shown before in Eq (1). The experimental results on test set are presented in Table 2, from which one can see that the DC of our network is 97.3%, which is significantly better than those of the U-Net and FusionNet, i.e., 91.8% and 94.4%, respectively. Our PDRU-Net performs better than the U-Net and FusionNet in Pixel Accuracy. In order to further compare the performance of these three models, we add three commonly used segmentation indicators (recall, precision and F1 score) to table 2. The precision of the U-Net and FusionNet is very high, but their recall value is very low, only 83.9% and 87.7% respectively. Our model achieves a good balance between recall and precision, both recall and

precision of our model are very high. F1 score is a balance measure of recall and precision, and the F1 score of our network is higher than that of the U-Net and FusionNet.

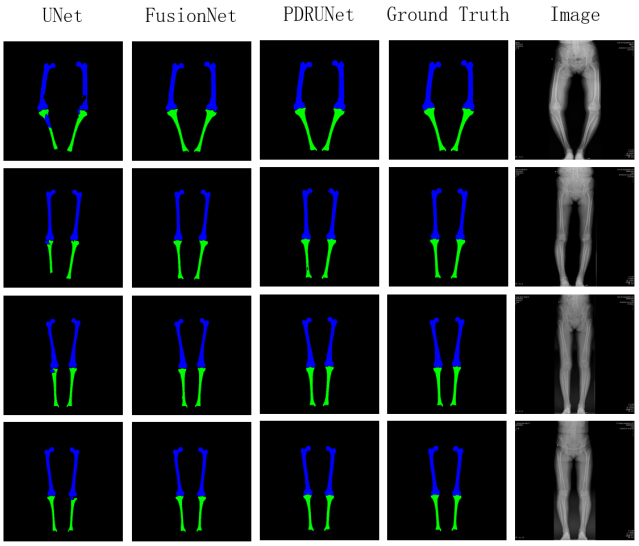

FIGURE 7. The first three columns show the segmentation results of U-Net, FusionNet and PDR U-Net respectively.The third and fourth columns show the corresponding ground truth images and input images respectively

TABLE 2. Comparison of the PDRU-Net,U-Net and FusionNet.

model	parameters	Dice Coefficient	Pixel Accuracy	Recall	Precision	F1 score
U-Net	∼33M	0.918	0.943	0.839	0.987	0.907
FusionNet	∼78M	0.944	0.969	0.877	**0.997**	0.933
PDRU-Net	∼0.36M	**0.973**	**0.987**	**0.953**	0.976	**0.964**

5. **Conclusion.** In this paper, we introduced a novel deep neural network for automatic segmentation of the femur and tibia bones from X-ray images. The proposed network architecture is called pure dilated residual U-Net (PDRU-Net), which is a novel extension of the classical U-Net. We increase the receptive field by pure dilated convolution operation to avoid information loss caused by the max-pooling layers in the U-Net. In addition, our network can add the skip-connection structure between any two blocks freely, this property can not only accelerate the speed of convergence but also help information flow across levels in the network. Thanks to data augmentation and few parameters, out network can effectively avoid overfitting. We evaluated the PDRU-Net on test set and compared the segmentation results with the U-Net and FusionNet, and the results showed that the PDRU-Net outperforms those methods in two common semantic segmentation metrics.

Acknowledgments. This work was supported in part by the National Science Foundation Program of China (NSFC) under Grants 61976241, 61871173, 81827805, in part by the National High Technology R&D Program of China under Grant No. 2015AA043203, in part by the National key R&D program of china under Grant No. 2018YFA0704102.

REFERENCES

[1] S. Y. Ababneh, J. W. Prescott and M. N. Gurcan, Automatic graph-cut based segmentation of bones from knee magnetic resonance images for osteoarthritis research, *Medical Image Anal.*, **15** (2011), 438–448.

[2] V. Badrinarayanan, A. Kendall and R. Cipolla, Segnet: A deep convolutional encoder-decoder architecture for image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.*, **39** (2017), 2481–2495.

[3] O. Bandyopadhyay, A. Biswas and B. B. Bhattacharya, Long-bone fracture detection in digital x-ray images based on digital-geometric techniques, *Comput. Methods Programs Biomed.*, **123** (2016), 2–14.

[4] J. Carballido-Gamio, et al., Automatic multi-parametric quantification of the proximal femur with quantitative computed tomography, *Quantitative Imaging in Medicine and Surgery*, **5** (2015), 552–568.

[5] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy and A. L. Yuille, Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, *IEEE Trans. Pattern Anal. Mach. Intell.*, **40** (2018), 834–848.

[6] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox and O. Ronneberger, 3d u-net: Learning dense volumetric segmentation from sparse annotation, in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016 - 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II*, Lecture Notes in Computer Science, 9901, 2016, 424–432.

[7] C. M. Deniz, S. Hallyburton, A. Welbeck, S. Honig, K. Cho and G. Chang, Segmentation of the proximal femur from MR images using deep convolutional neural networks, *Sci. Rep.*, **8** (2018), 16485.

[8] F. Ding, W. K. Leow and T. S. Howe, Automatic segmentation of femur bones in anterior-posterior pelvis x-ray images, in *Computer Analysis of Images and Patterns, 12th International Conference, CAIP 2007, Vienna, Austria, August 27-29, 2007, Proceedings* (eds. W. G. Kropatsch, M. Kampel and A. Hanbury), Lecture Notes in Computer Science, 4673, Springer, 2007, 205–212.

[9] L.-H. Fan, J.-G. Han, Y. Jia, C. Zhao and B. Yang, Segmentation of femurs in x-ray image with generative adversarial networks, *DEStech Transactions on Engineering and Technology Research*, 289–295.

[10] I. J. Goodfellow, et al., Generative adversarial nets, in *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada* (eds. Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence and K. Q. Weinberger), 2014, 2672–2680, URL `http://papers.nips.cc/paper/5423-generative-adversarial-nets`.

[11] S. Guan, A. A. Khan, S. Sikdar and P. V. Chitnis, Fully dense unet for 2d sparse photoacoustic tomography artifact removal, preprint, `arXiv:1808.10848`.

[12] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin and A. C. Courville, Improved training of wasserstein gans, in *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA* (eds. I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan and R. Garnett), 2017, 5767–5777, URL `http://papers.nips.cc/paper/7159-improved-training-of-wasserstein-gans`.

[13] K. He, X. Zhang, S. Ren and J. Sun, Deep residual learning for image recognition, in *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, IEEE Computer Society, 2016, 770–778.

[14] G. Huang, Z. Liu, L. van der Maaten and K. Q. Weinberger, Densely connected convolutional networks, in *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, IEEE Computer Society, 2017, 2261–2269.

[15] R. Jiang, J. Meng and P. Babyn, X-ray image segmentation using active contour model with global constraints, *2007 IEEE Symposium on Computational Intelligence in Image and Signal Processing*, (2007), 240–245.
[16] A. Krizhevsky, I. Sutskever and G. E. Hinton, Imagenet classification with deep convolutional neural networks, in *Advances in Neural Information Processing Systems*, 2012, 1097–1105.
[17] H. Li, A. Zhygallo and B. H. Menze, Automatic brain structures segmentation using deep residual dilated u-net, in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries - 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part I* (eds. A. Crimi, S. Bakas, H. J. Kuijf, F. Keyvan, M. Reyes and T. van Walsum), Lecture Notes in Computer Science, 11383, Springer, 2018, 385–393.
[18] M. Lin, Q. Chen and S. Yan, Network in network, in *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings* (eds. Y. Bengio and Y. LeCun), preprint, arXiv:1312.4400.
[19] M. Liu, T. Breuel and J. Kautz, Unsupervised image-to-image translation networks, in *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA* (eds. I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan and R. Garnett), 2017, 700–708, URL http://papers.nips.cc/paper/6672-unsupervised-image-to-image-translation-networks.
[20] X. Liu, et al., Msdf-net: Multi-scale deep fusion network for stroke lesion segmentation, *IEEE Access*, **7** (2019), 178486–178495.
[21] J. Long, E. Shelhamer and T. Darrell, Fully convolutional networks for semantic segmentation, in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, IEEE Computer Society, 2015, 3431–3440.
[22] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang and S. P. Smolley, Least squares generative adversarial networks, in *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, IEEE Computer Society, 2017, 2813–2821.
[23] F. Milletari, N. Navab and S. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in *Fourth International Conference on 3D Vision, 3DV 2016, Stanford, CA, USA, October 25-28, 2016*, IEEE Computer Society, 2016, 565–571.
[24] O. Oktay, et al., Attention u-net: Learning where to look for the pancreas, preprint, arXiv:1804.03999.
[25] C. N. Öztürk and S. Albayrak, Automatic segmentation of cartilage in high-field magnetic resonance images of the knee joint with an improved voxel-classification-driven region-growing algorithm using vicinity-correlated subsampling, *Comp. Bio. Med.*, **72** (2016), 90–107.
[26] T. T. Peng, et al., Detection of femur fractures in x-ray images, *Master of Science Thesis, National University of Singapore*.
[27] A. Pries, P. J. Schreier, A. Lamm, S. Pede and J. Schmidt, Deep morphing: Detecting bone structures in fluoroscopic x-ray images with prior knowledge, preprint, arXiv:1808.04441.
[28] T. M. Quan, D. G. C. Hildebrand and W. Jeong, Fusionnet: A deep fully residual convolutional neural network for image segmentation in connectomics, preprint, arXiv:1612.05360.
[29] O. Ronneberger, P. Fischer and T. Brox, U-net: Convolutional networks for biomedical image segmentation, in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015-18th International Conference Munich, Germany, October 5-9, 2015, Proceedings, Part III* (eds. N. Navab, J. Hornegger, W. M. W. III and A. F. Frangi), Lecture Notes in Computer Science, 9351, Springer, 2015, 234–241.
[30] T. Salimans, I. J. Goodfellow, W. Zaremba, V. Cheung, A. Radford and X. Chen, Improved techniques for training gans, in *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain* (eds. D. D. Lee, M. Sugiyama, U. von Luxburg, I. Guyon and R. Garnett), 2016, 2226–2234, URL http://papers.nips.cc/paper/6125-improved-techniques-for-training-gans.
[31] P. Santhoshini, R. Tamilselvi and R. Sivakumar, Automatic segmentation of femur bone features and analysis of osteoporosis, *Lecture Notes on Software Engineering*, 194–198.
[32] K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition, in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings* (eds. Y. Bengio and Y. LeCun), preprint, arXiv:1409.1556.

[33] R. Smith, Segmentation and fracture detection in x-ray images for traumatic pelvic injury.
[34] C. Stolojescu-Crisan and S. Holban, An interactive x-ray image segmentation technique for bone extraction, in *International Work-Conference on Bioinformatics and Biomedical Engineering, IWBBIO 2014, Granada, Spain, April 7-9, 2014* (eds. I. Rojas and F. M. O. Guzman), Copicentro Editorial, 2014, 1164–1171, URL http://iwbbio.ugr.es/2014/papers/IWBBIO_2014_paper_121.pdf.
[35] H. Sun, et al., Aunet: Attention-guided dense-upsampling networks for breast mass segmentation in whole mammograms, *Phys. Med. Biol.*, **65** (2020), 055005.
[36] C. Szegedy, et al., Going deeper with convolutions, in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, IEEE Computer Society, 2015, 1–9.
[37] A. Tack, A. Mukhopadhyay and S. Zachow, Knee menisci segmentation using convolutional neural networks: Data from the osteoarthritis initiative, *Osteoarthritis and Cartilage*, **26** (2018), 680–688.
[38] W. Wang, Y. Wang, Y. Wu, T. Lin, S. Li and B. Chen, Quantification of full left ventricular metrics via deep regression learning with contour-guidance, *IEEE Access*, **7** (2019), 47918–47928.
[39] J. Wu, A. Belle, R. H. Hargraves, C. Cockrell, Y. Tang and K. Najarian, Bone segmentation and 3d visualization of CT images for traumatic pelvic injuries, *Int. J. Imaging Syst. Technol.*, **24** (2014), 29–38.
[40] X. Xiao, S. Lian, Z. Luo and S. Li, Weighted res-unet for high-quality retina vessel segmentation, in *2018 9th International Conference on Information Technology in Medicine and Education (ITME)*, 2018, 327–331.
[41] Y. Xue, T. Xu, H. Zhang, L. R. Long and X. Huang, Segan: Adversarial network with multi-scale L 1 loss for medical image segmentation, *Neuroinformatics*, **16** (2018), 383–392.
[42] F. Yokota, T. Okada, M. Takao, N. Sugano, Y. Tada and Y. Sato, Automated segmentation of the femur and pelvis from 3d CT data of diseased hip using hierarchical statistical shape model of joint structure, in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2009, 12th International Conference, London, UK, September 20-24, 2009, Proceedings, Part II* (eds. G. Yang, D. J. Hawkes, D. Rueckert, J. A. Noble and C. J. Taylor), Lecture Notes in Computer Science, 5762, Springer, 2009, 811–818.
[43] F. Yu and V. Koltun, Multi-scale context aggregation by dilated convolutions, in *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings* (eds. Y. Bengio and Y. LeCun), preprint, arXiv:1511.07122.
[44] K. Zhang, W. Lu and P. Marziliano, Automatic knee cartilage segmentation from multi-contrast mr images using support vector machine classification with spatial dependencies, *Magnetic Resonance Imaging*, **31** (2013), 1731–1743.
[45] Z. Zhang, C. Duan, T. Lin, S. Zhou, Y. Wang and X. Gao, GVFOM: A novel external force for active contour based image segmentation, *Inf. Sci.*, **506** (2020), 1–18.
[46] Y. Zhou, W. Huang, P. Dong, Y. Xia and S. Wang, D-unet: A dimension-fusion U shape network for chronic stroke lesion segmentation, preprint, arXiv:1908.05104.
[47] J. Zhu, T. Park, P. Isola and A. A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, IEEE Computer Society, 2017, 2242–2251.
[48] Keras: Deep learning library for theano and tensorflow, https://github.com/keras-team/keras, 2015.
[49] Lableme, http://labelme.csail.mit.edu/Release3.0/.

Received December 2019; revised May 2020.

E-mail address: shenweihao1997@163.com
E-mail address: xuwenbo2019@ia.ac.cn
E-mail address: 2465860483@qq.com
E-mail address: primaluce0@gmail.com
E-mail address: mjx969@163.com
E-mail address: maxinlong8686@sina.com
E-mail address: sj.zhou@siat.ac.cn

E-mail address: guoshijie@hebut.edu.cn
E-mail address: wangyuanquan@scse.hebut.edu.cn