

Research Article

Dense Dilated Attentive Network for Automatic Classification of Femur Trochanteric Fracture

Yuxiang Kang,¹ Jie Yu,² Zhipeng Ren,¹ Guokai Zhang ,³ Wen Cao,¹ Yinguang Zhang ,¹ and Qiang Dong ¹

¹Department of Orthopaedics, Tianjin Hospital, Tianjin 300211, China

²Qingdao Central Hospital, Qingdao 266042, China

³School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

Correspondence should be addressed to Yinguang Zhang; ygz_yyy@ieee.org and Qiang Dong; dongqiangtianjin@126.com

Received 12 August 2021; Accepted 11 September 2021; Published 22 October 2021

Academic Editor: Yi-Zhang Jiang

Copyright © 2021 Yuxiang Kang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Automatic classification of femur trochanteric fracture is very valuable in clinical diagnosis practice. However, developing a high classification performance system is still challenging due to the various locations, shapes, and contextual information of the fracture regions. To tackle this challenge, we propose a novel dense dilated attentive (DDA) network for more accurate classification of 31A1/31A2/31A3 fractures from the X-ray images by incorporating a DDA layer. By exploiting this layer, the multiscale, contextual, and attentive features are encoded from different depths of the network and thus improving the feature learning ability of the classification network to gain a better classification performance. To validate the effectiveness of the DDA network, we conduct extensive experiments on the annotated femur trochanteric fracture data samples, and the experimental results demonstrate that the proposed DDA network could achieve competitive classification compared with other methods.

1. Introduction

Femur trochanteric fracture is one of the most commonly occurred fractures among elderly people. Especially, with the rapid growth of the aging population worldwide, the occurrence of this fracture increases rapidly which severely threatens the health of elderly people. Moreover, since this fracture could lead to high mortality rates and dramatically affect the quality of patients' life, effective and timely treatment is essential to relieve the pain of patients during the clinical diagnosis. Currently, the most effective way to diagnose this disease is by utilizing medical imaging such as X-rays or computed tomography (CT) to classify the types of fractures and then applying an appropriate treatment plan based on the corresponding diagnosis result. Typically, the OA/OTA classification criterion has been the most frequent and reliable used method to diagnose the condition of the fracture in the clinical fracture diagnosis. In this criterion, there are three types, e.g., 31A1, 31A2, and 31A3 (as shown in Figure 1), where 31A1 represents the simple pertrochanteric

fracture, 31A2 denotes the multifragmentary pertrochanteric fracture and lateral wall incompetent (≤ 20.5 mm) fracture, and 31A3 is the intertrochanteric (reverse obliquity) fracture [1]. Nevertheless, the conventional diagnosis method inspects patient images slice by slice which is usually tedious and time-consuming for the radiologists, and moreover, with different clinical experiences of radiologists, the final diagnosis result is liable to be empirical and subjective, which may hamper making the follow-up treatment plan. To tackle this challenge, a practicable way is to design a fracture computer-aided system [2–6] that helps the radiologist classify the fracture types automatically. In the past, a considerable number of researches have been proposed, for example, Demir et al. [7] developed a novel exemplar pyramid method for the humerus fracture; it extracted histograms of oriented gradients and local binary pattern features from the input images and then combined them with four conventional classifiers to classify the fractures. Boudissa et al. [8] explored the influence of semiautomatic bone-fragment segmentation on the reproducibility of the

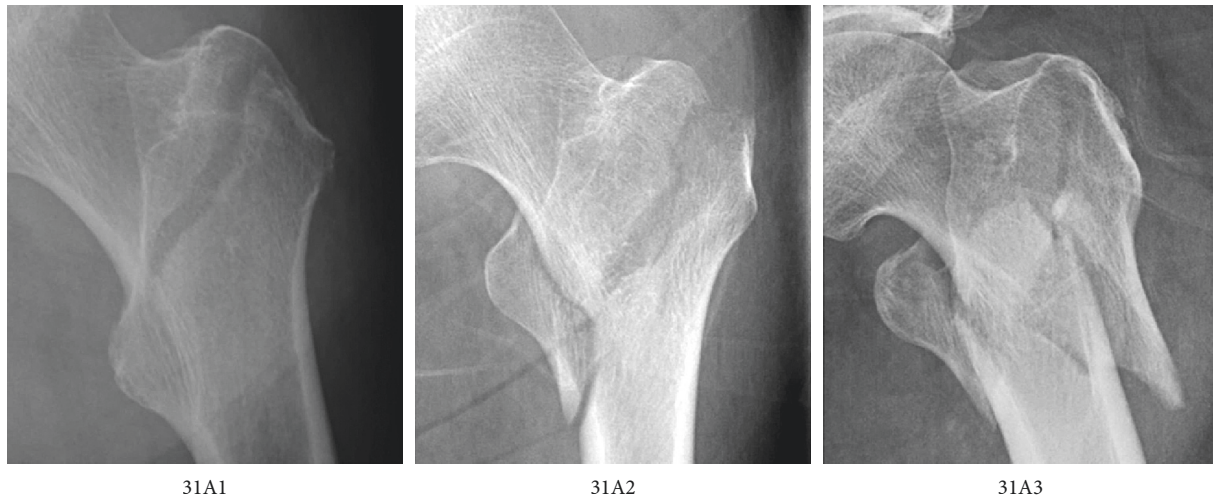


FIGURE 1: The fracture samples of 31A1, 31A2, and 31A3.

fracture classification, and it claimed that with the assistance of this technique, the classification accuracy of the fracture could be effectively improved. Additionally, in [9], the author proposed a 3D classification intertrochanteric fractures system which used the Hausdorff distance-based K-means method to classify the fractures into five types; the experimental results found that the unsupervised K-means method could gain promising classification performance with clinical significance. Cho et al. [10] evaluated the 3D CT images for boosting the diagnosis performance of femur trochanteric fractures, and it is shown that incorporating the CT could efficiently improve the reproducibility of stability for femur trochanteric fractures. Mall et al. [11] utilized different machine learning methods with the gray level cooccurrence matrix (GLCM) to classify the categories of fractures or no fractures; it proved that the proposed method could gain significant improvement on different evaluation metrics. Despite the great success, those methods have achieved in classifying the fracture task; those have deficiencies in capturing the robust and high-level semantic features due to hand-crafted feature predefinition.

Recently, the deep convolution neural network (CNN) has been proved its effectiveness in many computer vision tasks [12–16]. For instance, Lindsey et al. [17] suggested a deep CNN model based on the UNet structure achieving the automatic detection of the wrist fractures, and then it was evaluated on two different datasets; the result demonstrated that the proposed model could boost the clinical diagnosis performance. Then, Krogue et al. [18] labeled 3026 hip fractures and trained them with the DenseNet to achieve the automatic detection of hip fractures. Similarly, in [19], the authors utilized the faster R-CNN [20] to locate and classify the distal radius fractures automatically, and it obtained the mean average precision score of 0.866 at that time. To learn more high-level features, in [21], the authors employed a cropping process with the Inception V3 network to filter the unnecessary parts and thus leading to an improvement of the fracture detection. Besides, in [22], it used the Inception-ResNet faster R-CNN architecture to

construct a wrist fracture detection model and tested it on the unseen dataset, which proved that the designed model could gain high sensitivity and specificity.

Although those methods, especially the CNN ones, have gained promising results on the fracture classification task, an automatic fracture classification model should be simple and stable and provide effective information for the follow-up treatment plan. Specially, the femur trochanteric fractures usually have various locations, shapes, and contextual information in the clinical practice, which make it challenging to achieve a higher classification performance. Moreover, few works have considered the contextual information at different scales which may further limit the capability of the classification models. To tackle those challenges and efficiently improve the ability to learn strong representations from the fracture regions, in this paper, we develop a dense dilated attention (DDA) network to aggregate the multiscale, contextual, and attentive features from the femur trochanteric fracture region. Specially, in our DDA network, we incorporate the dense connection with dilated convolution by utilizing different dilated rates to learn the multiscale representations, and meanwhile, the dense connection could also alleviate the vanishing gradient problem and enable the network to reuse the hierarchical features. Furthermore, a dilated attention (DA) module is designed which encourages the network to encode more contextual and attentive representations automatically. To validate the effectiveness of DDA network, we perform extensive experiments on the femur trochanteric fracture images, and the experimental results show that our proposed DDA network could efficiently improve the classification performance by successfully extracting the discriminative features from the input image.

The rest of the paper is organized as follows: Section 2 presents the details of our proposed DDA network, and in Section 3, we first introduce the experimental data and evaluation metrics and then show the comparison results of different experiment settings. Finally, an elaborate discussion and conclusion of this paper are given in Section 4.

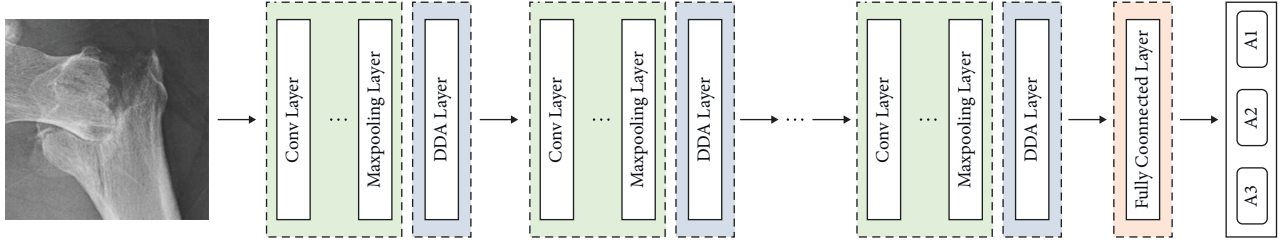


FIGURE 2: The main architecture of our proposed DDA network, where DDA layer is the dense dilated attention module.

2. Methodology

The automatic classification of femur trochanteric fracture is a challenging task due to its complex contextual information and various fracture regions. Hence, improving the network ability to extract multiscale representations, contextual information and intensity details are particularly important for accurate femur trochanteric fracture classification. To address those above challenges, a DDA network is developed for the accurate classification of the fracture categories; in the following subsections, we will provide the detailed descriptions of the network architecture and DDA module.

2.1. Network Architecture. As illustrated in Figure 2, given the X-ray images as the input of the DDA network, it first passes a series of convolution layers, max-pooling layers, and then a DDA module is implemented in the middle of the network to refine the feature representations, which will be described in detail in Subsection 2.3. After that, the final prediction category is output by a fully connected (FC) layer with the softmax activation in an end-to-end manner. The detailed parameters of the network are shown in Table 1. Notably, to preserve more spatial information of the image, we do not use the stride in the convolution layer. Specially, the ReLU activation is used to learn more nonlinear information, and batch normalization layer is utilized after each nonlinear activation to accelerate the convergence of the network.

2.2. Dilated Convolution. In our DDA module, we employ the dilated convolution to enlarge the receptive field without losing feature map resolution. Moreover, as the receptive field increases, it also provides more multiscale contextual features from the input [23]. Specifically, the dilated convolution could be divided into three steps: (1) sampling the input feature map based on the dilated rate; (2) conducting the convolution operation on the sampled values; (3) merging the obtained sampled values to a new feature map. Here, we denote the kernel size of the convolution layer as $k \times k$, and then the output feature map dimension of the traditional convolution layer I_c could be calculated as

$$I_c = \frac{W - k + 2p}{s} + 1, \quad (1)$$

TABLE 1: Parameters setting of the DDA network, where ‘‘Conv’’ denotes the convolution layer, and ‘‘FC’’ represents the fully connected layer.

Layer name	Output size	Filter size	Filter number
Conv1	512×512	3×3	256
Max-pooling	256×256	—	—
Conv2	256×256	3×3	128
Max-pooling	128×128	—	—
DDA layer	128×128	—	—
Conv3	128×128	3×3	64
Max-pooling	64×64	—	—
Conv4	64×64	3×3	64
Max-pooling	32×32	—	—
DDA layer	32×32	—	—
Conv5	32×32	3×3	32
Max-pooling	16×16	—	—
Conv6	16×16	3×3	32
Max-pooling	8×8	—	—
DDA layer	8×8	—	—
FC	Softmax	—	3

where W is the dimension of the input feature map; p and s are the padding size and stride, respectively. For the dilated convolution, its output dimension I_c^d could be defined as

$$I_c^d = \frac{W - k - r + 2p}{s} + 1. \quad (2)$$

Notably, when the stride is set as 1, the receptive field of the dilated convolution layer R_l could be formulated as

$$R_l = R_{l-1} + (k - 1) * r_l, \quad (3)$$

where r_l is the dilated rate of l th layer, and through this operation, the receptive field could increase rapidly. Finally, the dilated convolution layer could be given as

$$y[i] = \sum_m x[i + r \cdot m] \cdot w[m], \quad (4)$$

where $x[i]$, $y[i]$ are the input and output of the i th position, separately; $w[m]$ represents the learnable parameters of m th filter.

2.3. Dense Dilated Attention Layer. As illustrated in Figure 3, the aim of our dense dilated attention (DDA) layer is to learn more contextual and multiscale features across different layers to leverage the classification performance of the DDA network [24]. Note that the shallow layers usually contain the position information of the input, while the deep layers have high-level

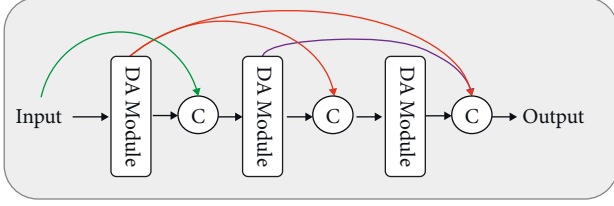


FIGURE 3: The structure of the dense dilated attention layer.

semantic representations. Therefore, combining those features across different layers could enhance the discrimination capability on fracture regions. Additionally, in order to guide the network's focus on the most salient regions from different receptive fields, a DA module is integrated into the DDA layer. Specially, taking the previous input feature map from the previous layer, it first passes through the DA module to learn the attentive and contextual features, and then the obtained ones are concatenated with the previous inputs as the input of the next layer. Note that the DDA layer mainly contains three DA modules with dilated rates of $\{1, 2, 3\}$, respectively, and its detailed structure is shown in Figure 4.

Mathematically, we denote the input of each DA module as $F_D^r \in \mathbb{R}^{W \times H \times C}$ with dilated rate of r , where $W \times H \times C$ represents the width, height, and channel numbers of F_D^r . Then, we adopt three 1×1 convolution layers to transform the F_D^r to three embeddings $\Phi \in \mathbb{R}^{W \times H \times \hat{C}}$, $\eta \in \mathbb{R}^{W \times H \times \hat{C}}$, $\tau \in \mathbb{R}^{W \times H \times \hat{C}}$, separately:

$$\phi = W_\phi(F_D^r), \quad \eta = W_\eta(F_D^r), \quad \tau = W_\tau(F_D^r), \quad (5)$$

where $W_\phi(\cdot)$, $W_\eta(\cdot)$, and $W_\tau(\cdot)$ denote the corresponding convolution operation; \hat{C} represents the channel number of those embeddings. After that, ϕ , η , and τ are flattened to the dimension of $\hat{C} \times HW$. To gain the contextual relation of F_D^r , a matrix multiplication between ϕ and η is applied, which can be given as

$$M = \phi^T \times \eta, \quad (6)$$

where $M \in \mathbb{R}^{HW \times HW}$ is the similarity matrix. Next, a softmax activation S is employed to normalize M to the interval of $[0, 1]$, and it could be formulated as

$$\tilde{M} = S(M). \quad (7)$$

Then, the attentive feature map $A \in \mathbb{R}^{HW \times HW}$ is gained by multiplying the \tilde{M} with τ , and it could be formulated as

$$A = \tilde{M} \times \tau^T. \quad (8)$$

Therefore, the final output of the DA module F_{DA}^r is defined as

$$F_{DA}^r = F_D^r + W_\delta(A), \quad (9)$$

where $W_\delta(\cdot)$ is the 1×1 convolution operation. By adopting the hierarchical DA modules with the dense connection, the DDA layer could not only extract the multiscale features from different receptive fields but also learn the attentive and contextual information from the input.

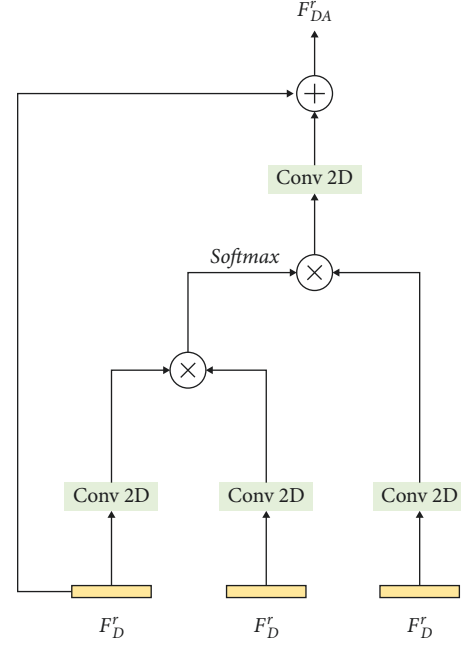


FIGURE 4: The structure of the DA module.

2.4. Training Loss Function. Denote the output feature map from the fully connected layer as $\{x_1, x_2, x_3\}$ and the corresponding label as $\{l_1, l_2, l_3\}$. To gain the predicted scores of each class, we apply a softmax activation function $\sigma(\cdot)$, which could be given as

$$\sigma(x)_i = \frac{e^{x_i}}{\sum_{n=1}^3 e^{x_n}}, \quad \text{for } i = \{1, 2, 3\}. \quad (10)$$

Additionally, to optimize the network, we use the binary cross-entropy \mathcal{L}_{bce} as the loss function, which could be formulated as

$$\mathcal{L}_{\text{bce}} = \sum_{j=1}^N g_j \log p_j + \sum_{j=1}^N (1 - g_j) \log(1 - p_j), \quad (11)$$

where N is the number of the data samples, and $p_j \in [0, 1]$, $g_j \in \{l_1, l_2, l_3\}$ are the predicted probability and corresponding true label.

3. Experiment

In this section, we conduct extensive experiments to validate the effectiveness of the DDA network. Specially, we first introduce the experimental data, implementation details, and evaluation metrics. Then, we compare the experiments with different amounts of data samples. Next, an ablation analysis of the DDA layer and DA module is explored to validate their efficiencies for this classification task. Finally, we reimplement some other fracture classification methods to compare them with our proposed DDA network.

3.1. Dataset. The total number of the experimental dataset is 390, and it consists of three categories 31A1, 31A2, and 31A3 with the amount of 117, 125, and 128, respectively. The mean

age of the patients is 65, the maximum age is 91, and the minimum age is 26. All the categories of the experimental data are annotated by three traumatic orthopedic specialists with more than 15 years of experience. Notably, the final category of data is based on the AO/OTA criterion. Since the initial resolution of the image is 1417×1772 , we crop the region of interest (ROI) with the maximum bounding-box, and then we resize those ROIs to 512×512 before input them into the DDA network to accelerate the training process of the network.

3.2. Implementation Details. In our experiments, since the initial resolution of the image is large, we first resize the image input to 512×512 . Moreover, to alleviate the overfitting problem, data augmentation is utilized to generate more data samples, and it includes random rotation, flip, and contrast. The whole network is implemented by the PyTorch deep learning library, and it is optimized by Adam. The initial learning rate is set as 0.001 and it decreases by 0.1 after 10 epochs. To accelerate the training process, we use the NVIDIA GeForce RTX 2070 Graphics Card, and the batch-size is set as 2.

3.3. Evaluation Metrics. To evaluate the performance of the proposed DDA network, we apply four evaluation metrics; here, we denote the true positive, false positive, true negative, and false negative as TP, FP, TN, and FN. Then, the accuracy which calculates the correct prediction among the total numbers of samples could be calculated as

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (12)$$

The sensitivity measures the ratio of correct TP prediction to the whole number of true positive samples. It could be formulated as

$$\text{sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (13)$$

The specificity calculates the ratio of correct TN to the whole numbers of false positive samples, and it is given as

$$\text{specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (14)$$

The receiver operating characteristic (ROC) curve is the most used graphical plot to measure the performance of the classifier, and the area under curve (AUC) is the score to measure the classification performance in which the higher score indicates the better distinguishing performance.

3.4. Data Sample Analysis of DDA Network. In this section, we first explore the influence of different data samples on the classification performance of the DDA network. Here, we divide our training data to 20%, 40%, 60%, 80%, and 100%, while the amount of the testing data samples is unchanged. The comparison result is shown in Figure 5, and it can be concluded that the more the data samples, the better performance the network would gain.

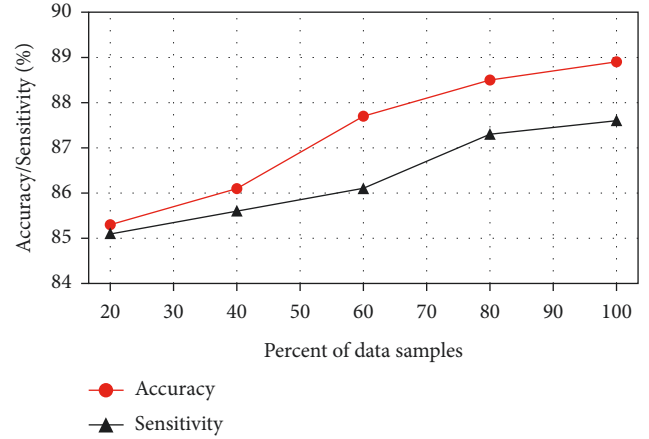


FIGURE 5: The accuracy/sensitivity of different percent of data samples.

That could be suggested that with more data samples, the network can extract the image features more sufficiently and effectively.

3.5. The Effectiveness of DDA Layer. The aim of our DDA layer is to learn the multiscale, attentive, and contextual information from the input image. Therefore, in this section, we conduct experiments w/o the DDA layer to explore its effects on the final classification performance. Moreover, since different depths of layers contain discriminative representations, therefore we also test the effectiveness of DDA layer on different depth locations. As reported in Table 2, DDA-1, DDA-2, and DDA-3 represent the depth location of locating DDA layer, in which the smaller value denotes the shallower depth location of the DDA network, and the DDA-W and DDA-O denote the network with or without the DDA layer. From the result, we observe that the best performance is achieved by the “DDA-W,” which can be explained that with the proposed DDA layer, the network is able to learn more high-level representations and then boost the classification performance.

3.6. Impact of DA Module. Different from the conventional dense connection, we develop a DA module that can guide the network to capture more attentive information with self-attention from different receptive fields. To validate the effectiveness of the proposed DA module, we compare three different network settings: with DA module; without DA module; with DA module (dilated rate as 1); with DA module (dilated rate as 2); with DA module (dilated rate as 3). The comparison result is shown in Table 3; the experimental result demonstrates that adopting the designed DA module could efficiently improve the classification performance compared with the setting without DA module. Moreover, with different dilate rates, the network tends to gain different performance; however, the best one is achieved by combining those three dilate rate settings, with the accuracy, sensitivity, specificity, and AUC of 88.9%, 87.6%, 85.9%, and 0.97, respectively.

TABLE 2: The effectiveness of DDA layer.

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
DDA-1	88.0	86.0	84.3	0.91
DDA-2	88.1	86.3	85.1	0.92
DDA-3	88.5	87.1	85.3	0.94
DDA-O	87.3	85.3	82.1	0.85
DDA-W	88.9	87.6	85.9	0.97

TABLE 3: The comparison results of different DA settings, where “DA-W” and “DA-O” denote the network w/o DA module, and “DA($r=1,2,3$)” represents the DA module with dilate rates of 1, 2, and 3, respectively.

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
DA ($r=1$)	88.2	87.2	85.1	0.93
DA ($r=2$)	88.1	86.8	84.7	0.91
DA ($r=3$)	88.3	87.0	85.2	0.92
DA-O	87.4	85.9	82.7	0.88
DA-W	88.9	87.6	85.9	0.97

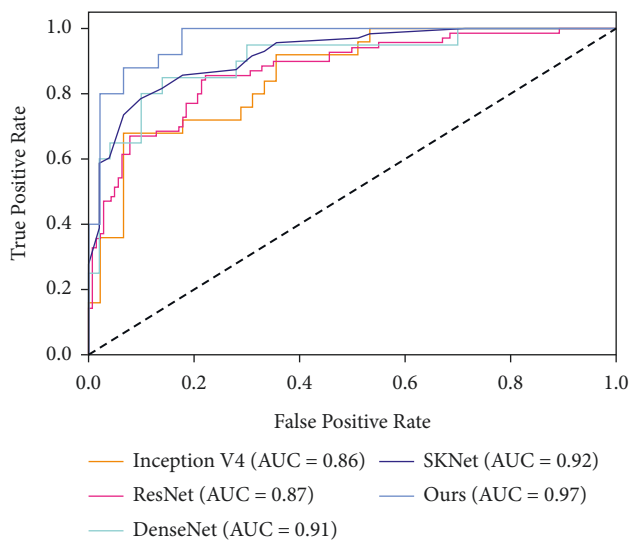


FIGURE 6: The AUC scores of different methods.

3.7. Comparison with Other Methods. To further evaluate the performance of the DDA network, in this section, we compare the results with different classification methods. As shown in Figure 6, we first compare our method with some baseline classification methods: Inception V4 [25], ResNet [26], DenseNet [27], and SKNet [28]. Note that we reimplement those methods and all the parameter settings are based on the default values. From the result, we observe that our proposed network could gain the highest AUC score of 0.97, which proves the effectiveness of our proposed network. Furthermore, we also report the comparison result with some other fracture methods; despite some of them are not with the same classification task, we reimplement those methods on the same dataset. As shown in Table 4, our method gains the best performance among all the evaluation metrics.

TABLE 4: The comparison results with other methods.

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
Exemplar [7]	87.5	85.7	82.0	0.80
K-means [9]	87.3	86.1	82.3	0.79
GLCM [11]	87.2	85.9	82.5	0.83
Inception V4 [25]	87.8	86.1	82.3	0.86
ResNet [26]	87.5	86.7	81.9	0.87
DenseNet [27]	87.2	86.3	83.4	0.91
SKNet [28]	87.1	86.8	83.6	0.92
Ours	88.9	87.6	85.9	0.97

4. Conclusion

In this paper, a DDA network is designed to achieve the classification of femur trochanteric fracture from X-ray images automatically. Since the fracture usually comes with various locations, shapes, and contextual information in the clinical practice, a novel DDA layer is developed which can automatically extract the multiscale, contextual, and attentive features to enhance the feature learning ability for achieving a more accurate classification performance. Extensive experiments on the annotated femur trochanteric fracture demonstrate that the proposed DDA network could gain competitive performance on this classification task. In future work, we will extend our work to different fracture classification tasks and collect more data samples to make the model more robust.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was sponsored by the Science and Technology Talent Cultivation Project of Tianjin Health Commission (KJ20215) and Research Project of Tianjin Sports Bureau (21DY014).

References

- [1] E. Meinberg, J. Agel, C. Roberts, M. Karam, and J. Kellam, “Fracture and dislocation classification compendium-2018,” *Journal of Orthopaedic Trauma*, vol. 32, no. 1, pp. S1–S10, 2018.
- [2] W. Li, Z. Wang, L. Zhang, L. Qiao, and D. Shen, “Remodeling pearson’s correlation for functional brain network estimation and autism spectrum disorder identification,” *Frontiers in Neuroinformatics*, vol. 11, p. 55, 2017.
- [3] X. Xu, W. Li, M. Tao et al., “Effective and accurate diagnosis of subjective cognitive decline based on functional connection

- and graph theory view,” *Frontiers in Neuroscience*, vol. 14, 2020.
- [4] P. Shan, F. Yang, H. Qi et al., “Alteration of MDM2 by the small molecule YF438 exerts antitumor effects in triple-negative breast c,” *Cancer Research*, vol. 81, no. 15, pp. 4027–4040, 2021.
 - [5] D. Gong, L. Wu, J. Zhang et al., “Detection of colorectal adenomas with a real-time computer-aided system (ENDOANGEL): a randomised controlled study,” *The Lancet Gastroenterology & Hepatology*, vol. 5, no. 4, pp. 352–361, 2020.
 - [6] Mo Zhang, B. Dong, and Q. Li, “Deep active contour network for medical image segmentation,” in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, Lima, Peru, October 2020.
 - [7] S. Demir, S. Key, T. Tuncer, and S. Dogan, “An exemplar pyramid feature extraction based humerus fracture classification method,” *Medical Hypotheses*, vol. 140, Article ID 109663, 2020.
 - [8] M. Boudissa, B. Orfeuvre, M. Chabanas, and J. Tonetti, “Does semi-automatic bone-fragment segmentation improve the reproducibility of the Letournel acetabular fracture classification?” *Orthopaedics and Traumatology: Surgery & Research*, vol. 103, no. 5, pp. 633–638, 2017.
 - [9] J. Li, S. Tang, H. Zhang et al., “Clustering of morphological fracture lines for identifying intertrochanteric fracture classification with Hausdorff distance-based K-means approach,” *Injury*, vol. 50, no. 4, pp. 939–949, 2019.
 - [10] Y.-C. Cho, P.-Y. Lee, C.-H. Lee, C.-H. Chen, and Y.-M. Lin, “Three-dimensional CT improves the reproducibility of stability evaluation for intertrochanteric fractures,” *Orthopaedic Surgery*, vol. 10, no. 3, pp. 212–217, 2018.
 - [11] P. K. Mall, P. K. Singh, and D. Yadav, “GlcM based feature extraction and medical x-ray image classification using machine learning techniques,” in *Proceedings of the 2019 IEEE Conference on Information and Communication Technology*, pp. 1–6, IEEE, Allahabad, India, December 2019.
 - [12] J. Wang, Y. Yang, J. Mao, Z. Huang, C. Huang, and W. Xu, “Cnn-rnn: a unified framework for multi-label image classification,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, IEEE, Las Vegas, NV, USA, June 2016.
 - [13] H. Lee and H. Kwon, “Going deeper with contextual CNN for hyperspectral image classification,” *IEEE Transactions on Image Processing*, vol. 26, no. 10, pp. 4843–4855, 2017.
 - [14] Y. Sun, B. Xue, M. Zhang, G. G. Yen, and J. Lv, “Automatically designing CNN architectures using the genetic algorithm for image classification,” *IEEE transactions on cybernetics*, vol. 50, no. 9, pp. 3840–3854, 2020.
 - [15] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, “Medical image classification with convolutional neural network,” in *Proceedings of the 2014 13th international conference on control automation robotics vision (ICARCV)*, IEEE, Singapore, December 2014.
 - [16] S. S. Yadav and S. M. Jadhav, “Deep convolutional neural network based medical image classification for disease diagnosis,” *Journal of Big Data*, vol. 6, no. 1, pp. 1–18, 2019.
 - [17] R. Lindsey, A. Daluiski, S. Chopra et al., “Deep neural network improves fracture detection by clinicians,” *Proceedings of the National Academy of Sciences*, vol. 115, no. 45, 11591 pages, 2018.
 - [18] J. D. Krogue, K. V. Cheng, K. M. Hwang et al., “Automatic hip fracture identification and functional subclassification with deep learning,” *Radiology: Artificial Intelligence*, vol. 2, no. 2, Article ID e190023, 2020.
 - [19] E. Yahalomi, M. Chernofsky, and M. Werman, “Detection of distal radius fractures trained by a small set of X-ray images and Faster R-CNN,” *Advances in Intelligent Systems and Computing*, in *Proceedings of the Intelligent Computing-Proceedings of the Computing Conference*, pp. 971–981, Springer, Cham, London, UK, July 2019.
 - [20] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: towards real-time object detection with region proposal networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2016.
 - [21] M. D. V. Thurston, T. Mackinnon, and D. H. Kim, “Fracture detection with artificial intelligence: improved accuracy with region of interest focusing,” *European Congress of Radiology-ECR*, vol. 2018, 2018.
 - [22] Y. L. Thian, Y. Li, P. Jagmohan, D. Sia, V. E. Y. Chan, and R. T. Tan, “Convolutional neural networks for automated fracture detection and localization on wrist radiographs,” *Radiology: Artificial Intelligence*, vol. 1, no. 1, Article ID e180001, 2019.
 - [23] L.-C. 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 Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2017.
 - [24] X. Wang, R. Girshick, and A. Gupta, “Non-local neural networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7794–7803, Salt Lake City, UT, USA, June 2018.
 - [25] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, “Inception-v4, inception-resnet and the impact of residual connections on learning,” in *Proceedings of the Thirty-first AAAI conference on artificial intelligence*, San Francisco CA USA, February 2017.
 - [26] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Las Vegas, NV, USA, June 2016.
 - [27] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Honolulu, HI, USA, July 2017.
 - [28] X. Li, W. Wang, X. Hu, and J. Yang, “Selective kernel networks,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 510–519, Long Beach, CA, USA, June 2019.