

# An approach for fully automatic femoral neck-shaft angle evaluation on radiographs

Cite as: Rev. Sci. Instrum. **91**, 013706 (2020); <https://doi.org/10.1063/1.5089738>

Submitted: 22 January 2019 . Accepted: 23 December 2019 . Published Online: 14 January 2020

Qiang Wei, Jungang Han , Yang Jia, Liyang Zhu, Shuai Zhang, Yufeng Lu, Bin Yang, and Shaojie Tang 



[View Online](#)



[Export Citation](#)



[CrossMark](#)



## Lock-in Amplifiers



Zurich  
Instruments

[Watch the Video](#) 

# An approach for fully automatic femoral neck-shaft angle evaluation on radiographs

Cite as: Rev. Sci. Instrum. 91, 013706 (2020); doi: 10.1063/1.5089738

Submitted: 22 January 2019 • Accepted: 23 December 2019 •

Published Online: 14 January 2020



View Online



Export Citation



CrossMark

Qiang Wei,<sup>1,2</sup> Jungang Han,<sup>1,2</sup> Yang Jia,<sup>1,2</sup> Liyang Zhu,<sup>1,2</sup> Shuai Zhang,<sup>3</sup> Yufeng Lu,<sup>3</sup> Bin Yang,<sup>3,a)</sup> and Shaojie Tang<sup>4,a)</sup>

## AFFILIATIONS

<sup>1</sup>School of Computer, Xi'an University of Posts and Telecommunications, Xi'an, Shaanxi 710121, China

<sup>2</sup>Shaanxi Key Laboratory of Network Data Analysis and Intelligent Processing, Xi'an University of Posts and Telecommunications, Xi'an, Shaanxi 710121, China

<sup>3</sup>Department of Orthopaedics, Xi'an Honghui Hospital, Xi'an, Shaanxi 710054, China

<sup>4</sup>School of Automation, Xi'an University of Posts and Telecommunications, Xi'an, Shaanxi 710121, China

<sup>a)</sup>Authors to whom correspondence should be addressed: [yangbin9013@139.com](mailto:yangbin9013@139.com) and [tangshaojie@xupt.edu.cn](mailto:tangshaojie@xupt.edu.cn)

## ABSTRACT

Femoral neck-shaft angle (NSA) is the angle included by the femoral neck axis (FNA) and the femoral shaft axis (FSA), which is a critical anatomic measurement index for evaluating the biomechanics of the hip joint. Aiming at solving the problem that the physician's manual measurement of the NSA is time consuming and irreproducible, this paper proposes a fully automatic approach for evaluating the femoral NSA on radiographs. We first present an improved deep convolutional generative adversarial network to automatically segment the femoral region of interest on radiographs of the pelvis. Then based on the geometrical characteristic of the femoral shape, the FNA and FSA are fitted, respectively, and thus, the NSA can be evaluated conveniently. The average accuracy of the proposed approach for NSA evaluation is 97.24%, and the average deviation is 2.58° as compared to the measurements manually evaluated by experienced physicians. There is no significant statistical difference ( $P = 0.808$ ) between the manual and automatic measurements, and Pearson's correlation coefficient is 0.904. It is validated that the proposed approach can provide an effective and reliable tool for automatically evaluating the NSA on radiographs.

Published under license by AIP Publishing. <https://doi.org/10.1063/1.5089738>

## I. INTRODUCTION

The femoral neck-shaft angle (NSA), also known as the caput-collum-diaphyseal (CCD) angle or inclination angle, is included by the femoral neck axis (FNA) and the femoral shaft axis (FSA) and is a critical anatomic measurement index for the evaluation of biomechanics of the hip joint.<sup>1</sup> The structure near the femoral neck-shaft bears the torsion force of the femur in the perspective of biomechanics. The NSA plays an important role in gait, which facilitates movement at the hip joint, enabling the limb to swing clear of the pelvis. Kaur *et al.*<sup>2</sup> summarized that the mean values of the NSA range from 122° to 136°, and normal individuals are found from around 110° to almost 150°. What is more, the NSA in children is usually larger than that in adults, and the value is also associated with the other variations such as the gender, region, and

lifestyle factors.<sup>3</sup> Reference 4 indicated that the NSA has implications for hip fracture risk. Mills *et al.*<sup>5</sup> found the rise of the NSA may aggravate the wear and tear of the acetabulum and femoral head, so leading to osteoarthritis. In hip dysplasia<sup>6</sup> and Legg–Calvé–Perthes disease,<sup>7</sup> physicians are required to accurately measure the NSA and select the appropriate treatment plan to achieve expected therapeutic effects. Additionally, the analysis of the NSA in a large population could give us an insight into the associated disease conditions.<sup>8</sup>

The NSA is now commonly measured manually by physicians in practice on the standard AP (anteroposterior) radiographs of the hips with both hips internally rotated.<sup>9</sup> However, the measurement of the NSA on radiograph is much influenced by the experience of physicians, and manual measuring is usually time consuming and irreproducible. Hence, the objective of this study is to propose an

approach for automatically evaluating the NSA on radiographs and demonstrate its performance. The contributions of this paper are summarized as follows:

- (1) This study shows the possibility of a fully automatic measurement for evaluating the NSA on radiographs, and it provides reliable results comparable with physicians' measurements.
- (2) An improved DCGAN (Deep Convolutional Generative Adversarial Network) is proposed to segment femur and femoral head regions from radiographs of the pelvis with a high degree of accuracy. Then, the minimum enclosing circle algorithm is used to determinate the center coordinates of the femoral head.

The method of NSA measurement by physicians is shown in Fig. 1. For exact determination of the center point C of the femoral head, a series of different sizes of circles are attempted with their margins touching the edges of the femoral neck at two points, A and B, and we call A and B the inside point and outside point, respectively. So the FNA is presented by a line from the point C perpendicular to the line AB. Similarly, the three gray lines almost perpendicular to the femoral shaft orientation are used to define the FSA. The angle between the FNA and the FSA is regarded as the NSA.<sup>10</sup>

The remainder of this paper is organized as follows. Section II briefly describes the related work. In Sec. III, we preprocess the data of pelvis radiographs, present a network of improved DCGAN to segment femur and femoral head regions, and make the fitting of the FNA and FSA. The experiments and results are discussed in Sec. IV. In Sec. V, we summarize our findings and conclude with our expectations in the future work.

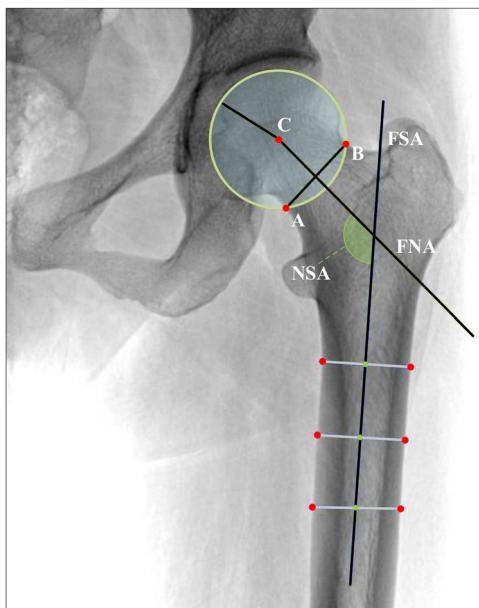


FIG. 1. Measuring method of the NSA by physicians on an AP radiograph.

## II. RELATED WORK

The NSA plays a critical role in the research of femoral geometry. It is also related to the other hip parameters, including femoral head offset and femoral neck length, and it is commonly used in presurgical planning of the proximal femur. To obtain accurate angles, Anderson and Trinkaus<sup>11</sup> took direct measurements with the dry bones of the femur, and Miller *et al.*<sup>12</sup> acquired precise measurements of the NSA by using ultrasound. Cerveri *et al.*<sup>13</sup> used 3D CT (computed tomography) reconstructions to derive the angles. These methodologies could produce precise measurements. However, this would not be routinely available or far from the everyday assessment of clinical situations. Additionally, He *et al.*<sup>14</sup> presented a method based on the 3D multiline shape of the femur to automatically evaluate the NSA from CT data of the hip joint. The method needs a manual extraction of the femur contour, verification data are also less, and the NSA deviation still needs to be further discussed. Teo *et al.*<sup>15</sup> also proposed a formula which explored the relationship between an AP and a 45° oblique radiograph with the NSA measurements. That study was limited by the small sample size, and intraobserver and interobserver reliabilities have not been evaluated.

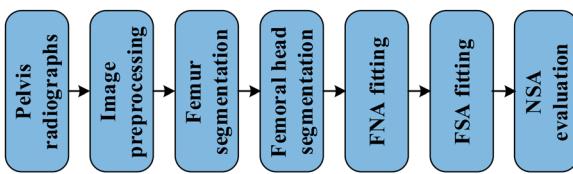
CT is considered precise and reproducible in the measurement of the NSA. Due to a higher cost and deep radiation exposure, its use is usually restricted to individual complex cases and revision operations. Radiograph examination in two dimensions is frequently performed at most of the institutes, and it is also cost-effective and safe as compared to CT and MRI (Magnetic Resonance Imaging). Hence, it can be an effective tool for measurement of the NSA, and so we select available radiographs for our study. Although there are a variety of methods for the measurement of the NSA on radiographs in the current publishing literature studies, there still exists no standard method. The important reasons for this situation are that NSA measurement on radiographs is sensitive to hip rotation and the definitions of the FNA and FSA varied significantly in the publications.<sup>16</sup> Recently, Bizdikian *et al.*<sup>17</sup> compared the performance of several measurement methods on radiographs and suggested that the most reliable and valid method should be considered a consistent measurement method for the NSA on radiographs. According to physicians' advice, we choose a measurement method which is coincident with the method in Ref. 17 mentioned above and facilitates the implementation of the automatic algorithm. The details of the method are depicted in Fig. 1. We note that the approaches for automatically measuring the NSA on radiographs are really rare, and the automatic measurement of the NSA on radiographs is of great significance.

## III. MATERIALS AND METHODOLOGY

This paper proposes a fully automatic approach for evaluating the NSA on radiographs. The approach consists of the preprocessing stage, femur and femoral head segmentation stages, FNA fitting stage, FSA fitting stage, and NSA evaluation, as depicted in Fig. 2.

### A. Materials and preprocessing

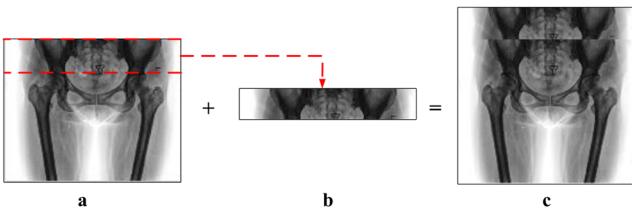
We use 80 standard AP pelvis radiographs acquired from Xi'an Honghui Hospital, Xi'an, China. These samples were chosen in a quasirandom way in medical practices. The radiographs with the

**FIG. 2.** Algorithmic flow chart of the proposed approach.

following problems were excluded from the dataset: the radiographs that were not clear enough to be distinguished by physicians, the radiographs with obvious fractures, bone defects, and bone tumors in the proximal femur, and those radiographs with the previous hip surgery. All the radiographs were taken using standard protocols for AP pelvic radiographs. The patients were required to keep an internal rotation of  $10^\circ$ – $15^\circ$  of the feet with the supine position. Such factors as gender and age are not considered in our dataset. Although Ref. 3 reported that these demographic factors could affect the value of the NSA, the objective of our study is to propose an automatic approach for evaluating the NSA, instead of investigating the effects of these factors on the value of the NSA.

To increase the validity of this study, three experienced physicians were employed for the measurements of the NSA. They were blinded to each other's findings. The measurements were performed on the radiographs using the PACS (Picture Archiving and Communication System). Two weeks later, the same procedure was repeated by these physicians for the second time for the measurements of the NSA, and we take the mean of these two values as the same observer's measurements. Notably, our study will focus on the difference between manual and automatic methods, excluding experiment deviation caused by nonstandard positions, and less attention is paid to the intraobserver and interobserver correlation coefficients between different physicians. Finally, we chose the mean of measurements by three physicians as the golden standard of the NSA measurements for our study.

The initial radiographs were stored in a DICOM format and with a size of  $2688 \times 2208$  pixels (width × height). The femur should be first segmented from the radiograph of the pelvis for automatically evaluating the NSA, and convolutional neural network models in deep learning are applied to the task of femur segmentation in our study. In order to balance the calculation of neural networks and retain enough image information, the initial DICOM format is converted to the JPG format after desensitization, and the image is cropped to  $512 \times 512$  pixels. Figure 3 shows the image preprocessing stage,

**FIG. 3.** Preprocessing of the femur image: (a) raw femur image, (b) cropped part of the femur image, and (c) processed femur image.

we keep equal scale on the two dimensions of samples to ensure the accuracy of NSA calculation. As shown in Fig. 3, image (b), the upper part of the initial image (a), is combined with the image (a) and so produces the square of (c).

The entire image dataset is divided into two parts. 50 radiographs are used to train the neural networks and augmented by image rotation, horizontal flip, translation, and other transformations. Also, other 30 radiographs are used as the test dataset. A sample for training neural networks is given in Fig. 4. Each sample consists of a pelvis image, a femur image, and a femoral head image labeled manually.

## B. Femur segmentation based on improved DCGAN

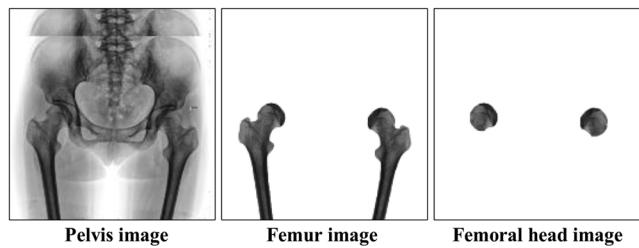
Inspired by zero-sum game theory, Goodfellow *et al.*<sup>18</sup> creatively proposed the GAN (Generative Adversarial Network) in 2014. In recent years, the GAN has achieved a great success in the tasks of image and text generation. The GAN commonly consists of a generator network (G) and a discriminator network (D). G could generate images adapted to the distribution of real data by the input random noise, while D estimates the probability that the input image is real or synthetic. They are alternatively optimized to improve their respective performances. When D cannot distinguish the generated image and real image, it is considered that G is well trained and has learned the distribution of the input data.

As an improvement of the GAN, the DCGAN (Deep Convolutional Generative Adversarial Network) adds the convolutional layers into the generator and discriminator models so that the models achieve a better performance and the subsequent works have been improved.<sup>19</sup> Based on the DCGAN, we present an improved DCGAN in this paper. Additionally, for applying to the segmentation tasks of femur regions, we take the idea of the CGAN (Conditional Generative Adversarial Network),<sup>20</sup> using input images instead of the random noise. Hence, the input of the generator in our model is the pelvis image, and the output is the corresponding femur; the improved network architecture is shown in Fig. 5.

In this paper, we adopt an improved DCGAN with the identical network structures to segment femur and femoral head regions successively from preprocessed radiographs. Since the femur is the first part to be segmented, we focus on the segmentation of the femur region for illustrating the model of the improved DCGAN.

### 1. Loss function

In the original GAN, G could generate images from the noise vector, and D estimates the source of input images. However, we

**FIG. 4.** A sample for training neural networks.

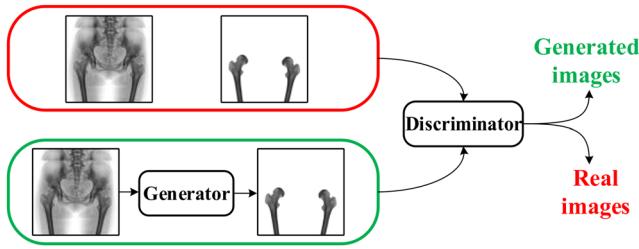


FIG. 5. Improved DCGAN architecture.

need to apply the model to segment the femur from the pelvis image. Instead of mapping the noise vector to the segmentation result, the input pelvis image is mapped to the segmented femur image. The distinction between segmentation and labeled images usually can be measured by the loss function based on pixels in the field of semantic segmentation. Therefore, the Dice is added to original loss function to improve the accuracy of network segmentation,

$$\text{Dice} = \frac{2|X \cap Y|}{|X| + |Y|}, \quad (1)$$

where  $X$  is our segmentation result,  $Y$  is the physicians' manual segmentation result, and a higher Dice means a better segmentation performance. In this paper,  $L_{\text{Dice}} = 1 - \text{Dice}$  is selected as the penalty function, and so it is minimized. Finally, the loss function of the generator model is defined as

$$L_G = \sum_i \log(D(x_i, G(x_i))) + \left(1 - \sum_i \frac{2|G(x_i) \cap y_i|}{|G(x_i)| + |y_i|}\right),$$

which is equivalent to

$$L_G = \sum_i \log(D(x_i, G(x_i))) - \sum_i \frac{2|G(x_i) \cap y_i|}{|G(x_i)| + |y_i|}, \quad (2)$$

where  $x_i$  represents the pixel of the pelvis image, and  $y_i$  and  $G(x_i)$  represent the pixels of the femur image by manual segmentation and generator model, respectively.

Discriminator model distinguishes whether the input image is real or synthetic, and its loss function is defined as

$$L_D = \sum_i \log(D(x_i, y_i)) + \sum_i \log(1 - D(x_i, G(x_i))). \quad (3)$$

Therefore, the whole optimization process of the improved DCGAN can be described as

$$\begin{aligned} \arg \min_G \max_D L_{\text{DCGAN}}(G, D) &= E_{x,y \sim P_{\text{data}(x,y)}} [\log D(x, y)] \\ &\quad + E_{x \sim P_{\text{data}(x)}} [\log(1 - D(x, G(x)))] \\ &\quad - E_{x,y \sim P_{\text{data}(x,y)}} \left[ \frac{2|G(x) \cap y|}{|G(x)| + |y|} \right], \end{aligned} \quad (4)$$

where  $x$  and  $y$  represent the pelvis image and femur image, respectively.

## 2. Network framework of improved DCCGAN

U-NET model<sup>21</sup> is employed for the generator due to its excellent performance in medical image segmentation. As we know, the low-level features of the image, such as the edges, can be extracted

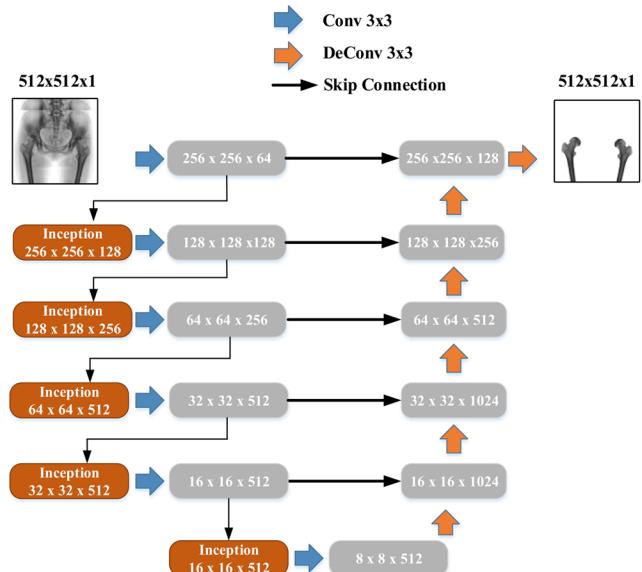


FIG. 6. Network structure of the generator model.

from the shallow convolutional layers. We take the idea of the U-NET model, in order to transfer these low-level features to deconvolutional layers by the skip connections. The network structure of the generator model is shown in Fig. 6, whose input and output are the pelvis image and corresponding femur image, respectively. Moreover, there are some inception modules in the generator model, which could enhance the depth and width of networks and extract the deep semantic features of images.<sup>22</sup> The specific structure is shown in Fig. 7. Batch normalization (BN) is also employed to accelerate the training process and enhance the independence of different layers.<sup>23</sup> The leaky-ReLU (leaky rectified linear unit) as the nonlinear activation function is applied in the encoding layers, while the decoding layers take the ReLU (rectified linear unit), and the tanh activation function is employed in the last layer of the model to generate the femur image.

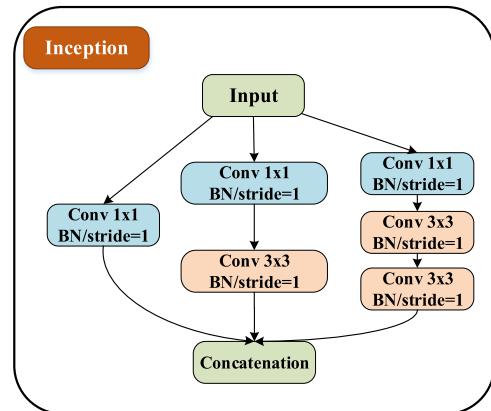


FIG. 7. Framework of the inception module.

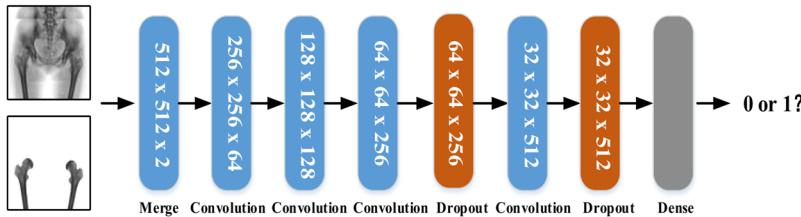


FIG. 8. Network structure of the discriminator model.

The numbers of filters in the encoding layers of G are set as 64, 128, 256, 512, 512, and 512, respectively. The decoding process is a mirror process of the encoding stage, which has the same parameter configuration as the encoding process. In order to prevent the networks from overfitting and improve the generalization performance of the model, we add the dropout layers after the deconvolutional layers of G, and the parameters are set to 0.5.

Traditional convolutional neural networks commonly adopt the pooling layers to achieve the scaling of feature maps, which will lose part of features. In this paper, we use the convolutional layers with stride instead of pooling layers to reduce the feature maps in the G and D models. D extracts the features of the femur images from different sources as the evidence of discrimination by convolutional neural networks. The network structure of D is shown in Fig. 8. The numbers of filters are 64, 128, 256, and 512, respectively. The leaky-ReLU is used as the nonlinear activation function after the convolutional layers.

### C. Femoral neck axis fitting

In this study, we first automatically segment the regions of the femur and femoral head by using the improved DCGAN from the radiographs of the pelvis. Then, the FNA is automatically fitted according to the measurement method in Ref. 10. The left and right femurs are processed, respectively, which is conducive to the image operations and facilitates the statistics of the mathematical characteristics of left or right NSAs. The FNA fitting includes the following steps:

- (1) Femoral head areas are usually viewed as a round shape on radiographs, whose important parameters are the center point and radius. In this paper, we need to determine the center point of the femoral head so as to fit the FNA. Hence, the minimum enclosing circle algorithm is used.<sup>24</sup> The algorithm is specially designed for the minimum enclosing circle problem, computing the smallest circle that contains every point in a given point set in the plane. We use the implementation of this algorithm in OpenCV that is an open sourced computer vision library and obtain the center point of the femoral head region. In Fig. 9(d), the red circle is the smallest fitted circle and point C is the center of the femoral head.
- (2) The femoral head region in Fig. 9(c) is used as a mask image, the femur image in Fig. 9(b) is considered the marker image for the matrix mask operation, and the result is shown in Fig. 9(e). Note that, in the fitting procedure of the femoral head center point C, the arc intersects the femoral neck contour with the inside and outside points. According to this location relationship, we label all candidate points, shown as

blue points in the margin of the femur in Fig. 9(e), whose distances to the point C are equal to the radius of the femoral head. The last two farthest points of these candidates are selected as the inside point A and outside point B, and we join the two points to form the red line AB in Fig. 9(e).

A line from the point C perpendicular to the line AB is considered the femoral neck axis, as shown in Fig. 9(f). The formula for determining the FNA is as follows:

$$y = y_c - \frac{(x_a - x_b)}{(y_a - y_b)}(x - x_c), \quad (5)$$

where  $(x_a, y_a)$  and  $(x_b, y_b)$  represent the coordinates of the points A and B, respectively, while  $(x_c, y_c)$  is the coordinate of the point C.

### D. Femoral shaft axis fitting

The proximal part of the femur has a complex profile involving the femoral head and the great trochanter. Meanwhile, the FSA could be fitted separately by the femoral shaft part. In order to reduce the interference of the proximal part to the FSA fitting, we just use the distal part of femur images. In the FSA fitting process, physicians need to manually draw three horizontal lines almost perpendicular to the femoral shaft orientation. Our approach can automatically detect several lines as the green lines shown in Fig. 10(a), so as to

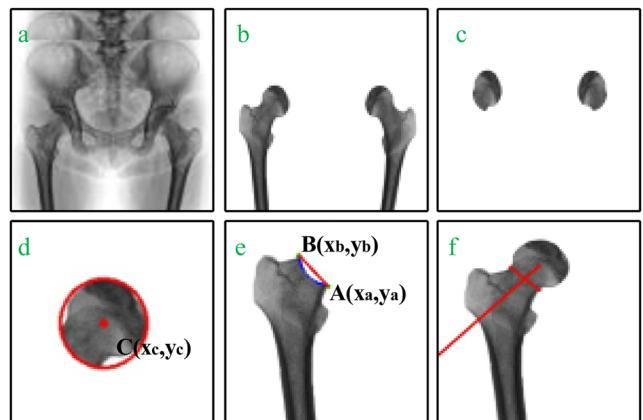
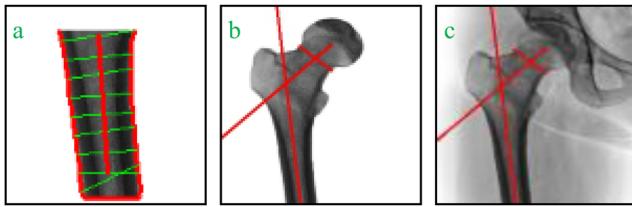


FIG. 9. Process of unilateral FNA fitting: (a) preprocessed image of the pelvis; (b) femur region segmented by the improved DCGAN; (c) femoral head area segmented; (d) the smallest fitting circle; (e) calculated inside point and outside point of the neck; (f) result of left FNA fitting.



**FIG. 10.** Process of unilateral FSA fitting: (a) FSA fitting, (b) NSA on the femur, and (c) NSA on the radiograph.

better fit the FSA. The details of the fitting procedure are listed as follows:

- (1) The Canny detection is the commonly used operator to detect the edges of the image,<sup>25</sup> and it is used to identify the edges of the femoral shaft in this work. It can be seen from Fig. 10(a) that both sides of the femoral shaft are surrounded by the red lines, and the orientation of the lines could be considered the orientation of the femoral shaft. Then, we need to make several lines perpendicular to the femoral shaft orientation to define the FSA. However, femoral shaft contours seem not absolutely straight, and all horizontal lines perpendicular to the entire femoral shaft orientation could not be solved once. So we divide the whole contour of the femoral shaft into several small segments which could be approximately viewed as straight lines, and then we take the perpendicular lines of these segments passing through their end points. These perpendicular lines are the green lines shown in Fig. 10(a), which are perpendicular to the femoral shaft orientation on the whole.
- (2) A few segments mentioned above are likely to have obvious gaps with the straight line, so their perpendicular lines may not be vertical to the entire femoral shaft contour. In order to make the fitted FSA reflecting the femoral shaft orientation better, we need to shift some perpendicular lines. First,

all perpendicular lines should be compared with each other and the cosine similarity is calculated by (6). Considering that the errors could exist in the fitting process, any two lines with a cosine similarity greater than 0.99 can be viewed as being parallel to each other in our study. Then, we take the perpendicular line group which includes the largest amount of parallel lines, and their midpoints are connected in a straight line, which presents the FSA as the red line shown passing through the femoral shaft in Fig. 10(a). The cosine similarity equation is

$$\cos(\theta) = \frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{\|\mathbf{v}_1\| \|\mathbf{v}_2\|}, \quad \cos(\theta) = 1 \quad \text{if } \cos(\theta) \geq 0.99, \quad (6)$$

where  $\mathbf{v}_1$  and  $\mathbf{v}_2$  represent the vectors of any two vertical lines.

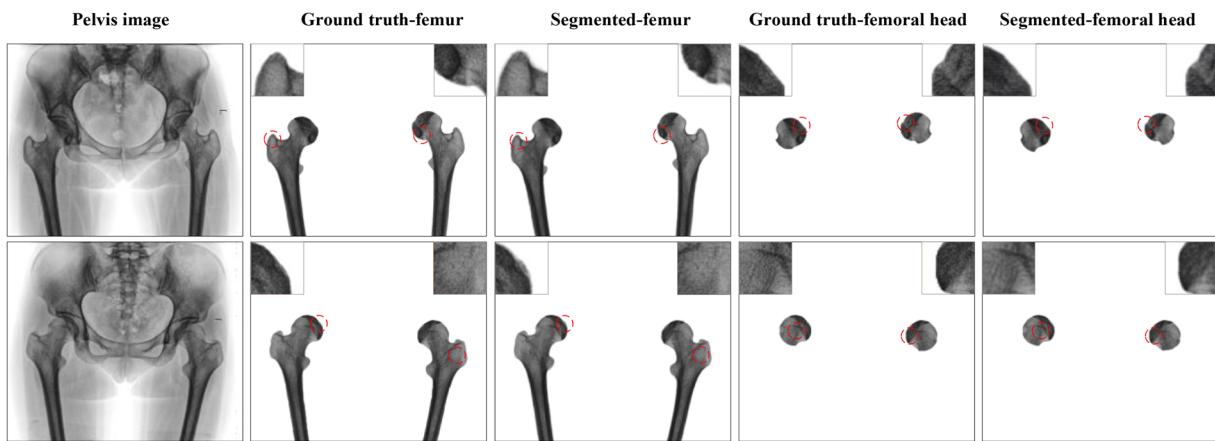
In the previous subsections, we fit the FNA and FSA based on geometrical characteristic of the femoral shape with the knowledge on image processing and deep learning. After obtaining the orientations of both the FNA and the FSA, the NSA could be conveniently evaluated from them by using (6). Figures 10(b) and 10(c) present the NSA measurement in the femur and original radiograph of the pelvis, respectively.

## IV. RESULTS AND DISCUSSION

### A. Segmentation results

In the experiment, the Adam optimization method is used for gradient descent ( $\beta_1 = 0.5$ ,  $\beta_2 = 0.999$ ). The learning rate is  $lr = 0.0002$ , batch-size = 1, and the number of training epochs is 300. The rate in leaky-ReLU is 0.2. The entire training procedure takes 5 h and 23 min with NVIDIA DGX-1, which is equipped with 8 T V100 GPUs, 2 Intel Xeon processors, and 512 GB memory.

Figure 11 shows the segmentation results by using the networks on femur and femoral head regions. The ground truth and segmented result are referred to as the images manually labeled by the physicians and those automatically segmented by the improved



**FIG. 11.** Segmentation results by using the proposed networks.

**TABLE I.** Network comparison results on femur images.

Method	Dice	Precision	Recall	F-measure
U-NET	0.9387	0.9373	0.9309	0.9341
U-NET++	0.9511	0.9450	0.9388	0.9418
DCGAN + U-NET	0.9605	0.9518	0.9447	0.9482
Proposed	0.9762	0.9549	0.9622	0.9585

**TABLE II.** Network comparison results on femoral head images.

Method	Dice	Precision	Recall	F-measure
U-NET	0.9367	0.9435	0.9388	0.9412
U-NET++	0.9522	0.9536	0.9474	0.9504
DCGAN + U-NET	0.9654	0.9713	0.9718	0.9715
Proposed	0.9786	0.9854	0.9841	0.9847

DCGAN, respectively. We can see that there is a high similarity between the ground truths and the segmented images, especially for the details of the images, as shown by the amplified red circle parts in different individuals in Fig. 11. In order to evaluate the performance of the proposed approach, we employ the Dice coefficient, precision, recall, and F-measure,

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (7)$$

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (8)$$

$$\text{F-measure} = 2 \times \text{Precision} \times \frac{\text{Recall}}{\text{Precision} + \text{Recall}}, \quad (9)$$

where TP represents the number of target region pixels correctly segmented, TN is the number of background pixels correctly segmented, FP stands for the number of background pixels incorrectly segmented into target regions, and FN stands for the number of target region pixels wrongly labeled as background pixels.

In order to verify the superiority of the improved DCGAN, we compare the segmentation effects of the four network structures in Tables I and II with 30 test images: (1) using the U-NET model alone, (2) adding the inception module into U-NET (U-NET++), (3) using U-NET as a generator of the DCGAN

(DCGAN + U-NET), and (4) using U-NET++ as a generator of the DCGAN (proposed).

It can be observed from Tables I and II that the proposed networks can achieve an excellent segmentation performance as compared to other networks. The inception module enhances the depth and width of networks, increasing the adaptability of the networks to multiple scales, and hence, the U-NET++ model has a better segmentation performance compared to the U-NET model alone. What is more, we combine the U-NET++ model with the DCGAN and achieve a great performance in the segmentations of the femur and femoral head. The high Dice coefficient indicates that our segmentation results are highly similar to the ground truth. Moreover, the metrics of recall, precision, and F-measure are also high, which validates the robustness and accuracy of the proposed approach and demonstrates that the proposed networks could provide a relatively accurate segmentation for the task of NSA evaluation.

## B. NSA measurement results

The proposed approach is tested with 60 femurs. Figure 12 shows the four examples of NSAs acquired by our approach. The longer red lines present the FNA and FSA, respectively. Meanwhile, Fig. 12 gives the NSA measurements of the individuals.

Comparing the NSAs automatically evaluated by the proposed approach (NSA<sub>a</sub>) with those manually measured by physicians (NSA<sub>m</sub>), we obtain the average absolute deviation as 2.58° and the average accuracy as 97.24% with the test of 60 femurs. The accuracy index used here is defined as

$$\text{Accuracy} = \left( 1 - \left| \frac{\text{NSA}_m - \text{NSA}_a}{\text{NSA}_m} \right| \right) \times 100\%. \quad (10)$$

Figure 13(a) illustrates the manually measured left NSA vs the automatically evaluated left NSA (MLNSA vs ALNSA) on 30 femurs. The average absolute deviation is 2.65°, and the average accuracy is 97.14%. Figure 13(b) illustrates the manually measured right NSA vs the automatically evaluated right NSA (MRNSA vs ARNSA) on 30 femurs. The average absolute deviation is 2.32°, and the average accuracy is 97.35%. Wilson *et al.*<sup>26</sup> showed that there is a difference among the NSA observers in the PACS, which is limited to ±6°. Thus, the proposed approach for the fully automatic evaluation of the NSA shows a better performance.

We also carry out the statistical analysis of the physicians' manual measurements and our automatic evaluations by the SPSS 22 software and list the statistical results in Table III.



**FIG. 12.** Four examples of NSAs automatically evaluated by the proposed approach.

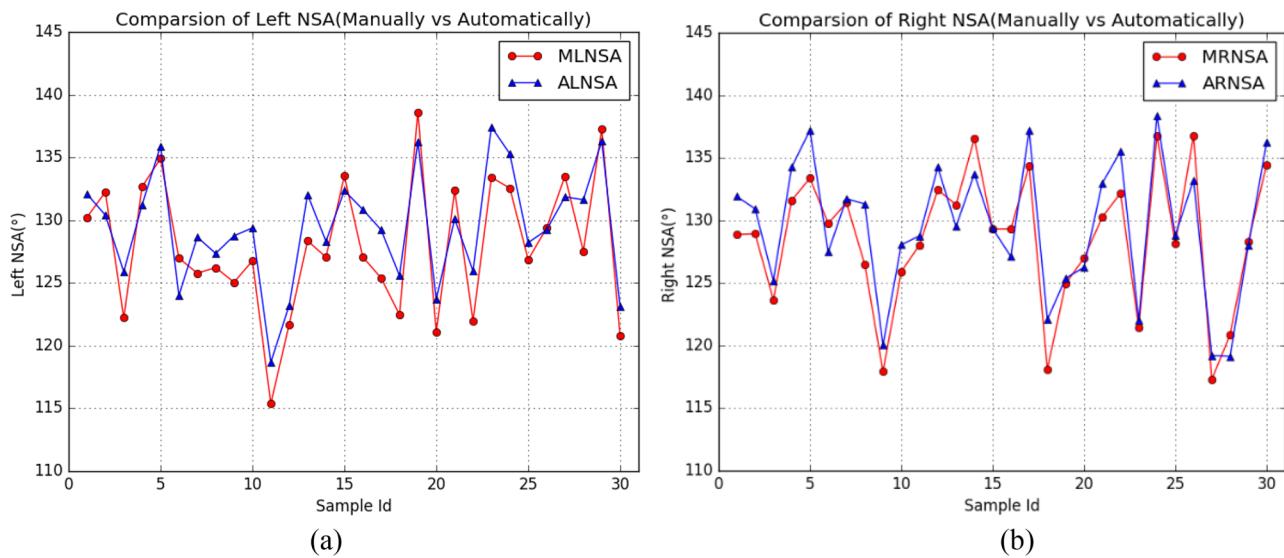


FIG. 13. Comparison of NSA manual measurements with automatic measurements: (a) left NSAs and (b) right NSAs.

**TABLE III.** Analysis of the NSA from manual measurements and automatic evaluations.

Method	Hip joint number	Maximum	Minimum	Mean	Standard deviation
Manual measurements	60	138.78	115.38	128.98	6.11
Automatic evaluations	60	137.41	117.67	129.51	5.59
Pathak <i>et al.</i> <sup>27</sup>	220	138	118.8	128.60	...
Le Bras <i>et al.</i> <sup>28</sup>	25	136.9	112.3	126.5	7.2

Table III lists the main statistics characteristics, such as the maximum, minimum, and mean. In our study, the mean of the NSA with the manual measurement is  $128.98^\circ$  (range  $115.38^\circ$ – $138.78^\circ$ ), and the mean with the automatic evaluation is  $129.51^\circ$  (range  $117.67^\circ$ – $137.41^\circ$ ). Due to a random collection of the dataset and a small amount of the sample, the data present a large dispersion degree, which could explain that the standard deviation in manual or automatic measurements is wide, as shown in Table III, and the values are closer to the standard deviation in Ref. 28. In subsequent automatic measurements, the standard deviation may change as the samples increase. The mean of the NSA with the automatic evaluation is also closer to the mean in Ref. 27; thus, there is a consistent tendency for a larger amount of measurements of the NSA. Gilligan *et al.*<sup>3</sup> indicated that the mean of the NSA for modern humans is  $127^\circ$  involving multiple parts of the world, and this value of the NSA could be greatly affected by several factors; we do not investigate the effects, and we merely propose an automatic evaluation for the measurements of batch NSAs in our study.

The data are further subject to a series of statistical analyses to assess the stability of the proposed approach. A one-way

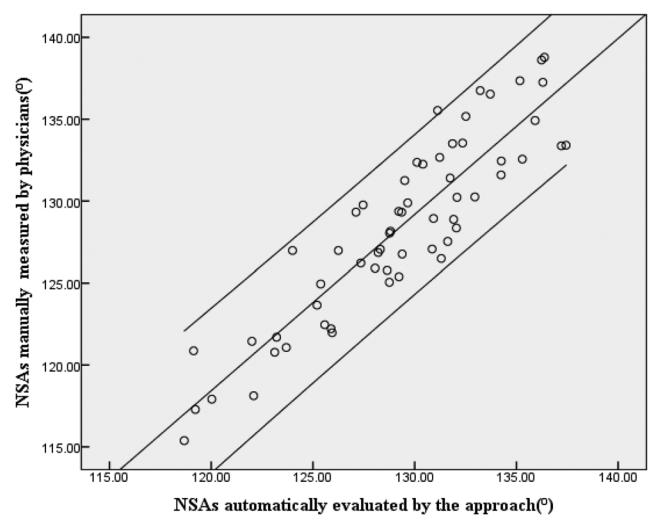


FIG. 14. Pearson's correlation analysis of the manual measurements and automatic evaluations (95% confidence interval marked).

analysis of variance is first performed on NSA measurements from different methods, and the significance difference  $\alpha = 0.05$  is set for all statistical analyses in our study. The experimental results show that there is no statistical difference between the physicians' manual measurements and the automatic evaluations by the proposed approach ( $P = 0.808$ ). In addition, Pearson's correlation coefficient  $r$  between them is 0.904, and Fig. 14 shows the analyzed result with the regression lines of 95% confidence intervals. Almost all of the scatters are within the 95% confidence intervals, which indicates that there exists a high correlation between the two methods. Reference 26 demonstrated that Pearson's correlation coefficient of intraobserver and interobserver consistency assessments was 0.98 and 0.86, respectively, with NSA observers in the PACS. Pearson's correlation coefficient between the manual and automatic methods is 0.904 (0.80  $< r < 1.00$ ) in this paper, indicating that the proposed approach is relatively stable in the automatic evaluation of NSAs.

## V. CONCLUSIONS

The objective of this study is to develop a fully automatic approach to evaluate the NSA on radiographs. In the proposed approach, we improve the DCGAN and apply it to the tasks of segmenting the femur and femoral head regions. Based on the research on femoral geometry with the knowledge on image processing, the FNA and FSA are fitted so that the NSA can be automatically evaluated.

We use the minimum enclosing circle algorithm to detect the center point of the femoral head region, which simplifies the problem solving. Compared with manual measurements, the results acquired by the proposed approach show an accuracy of 97.24% and an average deviation of  $2.58^\circ$  in the test data. Meanwhile, it is found that there is no significant statistical difference between manual and automatic results. Furthermore, the improved DCGAN can also be used for the segmentation of other bone tissues, and it is a general model for segmenting the medical image. In future work, we need to validate the proposed approach on more radiograph datasets and embed it into a robust tool.

## ACKNOWLEDGMENTS

This work was supported in part by the Shaanxi Key Laboratory of Network Data Intelligent Processing (Grant No. XUPT-KLND 210802), by the project for the innovation and entrepreneurship in the Xi'an University of Posts and Telecommunications (Grant No. 2018SC-03), by the Key Lab of Computer Networks and Information Integration (Southeastern University), Ministry of Education, China (Grant No. K93-9-2017-03), by the Department of Education Shaanxi Province (Grant Nos. 15JK1673 and 18JK0722), and by the Shaanxi Provincial Natural Science Foundation of China (Grant No. 2016JM8034).

## REFERENCES

- <sup>1</sup>D. Wright, C. Whyne, M. Hardisty, H. Kreder, and O. Lubovsky, "Functional and anatomic orientation of the femoral head," *Clin. Orthop. Relat. Res.* **469**, 2583–2589 (2011).
- <sup>2</sup>P. Kaur, S. Mathew, and U. George, "A study of neck shaft angle in the North-West Indian population on radiographs," *Int. J. Basic Appl. Med. Sci.* **3**, 9–15 (2013).
- <sup>3</sup>I. Gilligan, S. Chandraphak, and P. Mahakkanukrauh, "Femoral neck-shaft angle in humans: Variation relating to climate, clothing, lifestyle, sex, age and side," *J. Anat.* **223**, 133–151 (2013).
- <sup>4</sup>C. Ripamonti, L. Lisi, and M. Avella, "Femoral neck shaft angle width is associated with hip-fracture risk in males but not independently of femoral neck bone density," *Br. J. Radiol.* **87**, 20130358 (2014).
- <sup>5</sup>H. J. Mills, J. G. Horne, and G. L. Purdie, "The relationship between proximal femoral anatomy and osteoarthritis of the hip," *Clin. Orthop. Relat. Res.* **288**, 205–208 (1993).
- <sup>6</sup>J. R. Davids, T. W. Gibson, L. I. Pugh, and J. W. Hardin, "Proximal femoral geometry before and after varus rotational osteotomy in children with cerebral palsy and neuromuscular hip dysplasia," *J. Pediatr. Orthop.* **33**, 182–189 (2013).
- <sup>7</sup>K. W. Park, C. A. Rejuso, R. A. N. Garcia, T. W. Kim, and H. R. Song, "Extent of physeal involvement in Legg–Calvé–Perthes disease," *Int. Orthop.* **38**, 2303–2308 (2014).
- <sup>8</sup>V. Sharma, K. Kumar, V. Kalia, and P. K. Soni, "Evaluation of femoral neck-shaft angle in subHimalayan population of North West India using digital radiography and dry bone measurements," *J. Sci. Soc.* **45**, 3–7 (2018).
- <sup>9</sup>J. Robin *et al.*, "Proximal femoral geometry in cerebral palsy: A population-based cross-sectional study," *J. Bone Jt. Surg., Br. Vol.* **90**, 1372–1379 (2008).
- <sup>10</sup>S. Waldt and K. Woertler, *Measurements and Classifications in Musculoskeletal Radiology* (Georg Thieme, Germany, 2014), pp. 13–15.
- <sup>11</sup>J. Y. Anderson and E. Trinkaus, "Patterns of sexual, bilateral and interpopulational variation in human femoral neck-shaft angles," *J. Anat.* **192**, 279–285 (1998).
- <sup>12</sup>F. Miller, Y. Liang, M. Merlo, and H. T. Harcke, "Measuring anteversion and femoral neck-shaft angle in cerebral palsy," *Dev. Med. Child Neurol.* **39**, 113–118 (1997).
- <sup>13</sup>P. Cerveri *et al.*, "Automated method for computing the morphological and clinical parameters of the proximal femur using heuristic modeling techniques," *Ann. Biomed. Eng.* **38**, 1752–1766 (2010).
- <sup>14</sup>S. He, H. Liu, and Y. Tang, "A method of calculating the neck-shaft angle based on 3D multiline shape of femur," *Chin. J. Biomed. Eng.* **32**, 109–113 (2013).
- <sup>15</sup>P. C. Teo, A. Y. Mohd Kassim, and Y. Thevarajan, "A 45-degree radiographic method for measuring the neck shaft angle and anteversion of the femur: A pilot study," *J. Orthop. Surg.* **21**, 340–346 (2013).
- <sup>16</sup>C. K. Boese *et al.*, "The femoral neck-shaft angle on plain radiographs: A systematic review," *Skeletal Radiol.* **45**, 19–28 (2016).
- <sup>17</sup>A. J. Bizdikian *et al.*, "Validity and reliability of different techniques of neck-shaft angle measurement," *Clin. Radiol.* **73**, 984.e1–984.e9 (2018).
- <sup>18</sup>I. Goodfellow *et al.*, "Generative adversarial nets," in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2014), Vol. 2, pp. 2672–2680.
- <sup>19</sup>A. Radford, I. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *Computer Science* (2015).
- <sup>20</sup>M. Mirza and S. Osindero, "Conditional generative adversarial nets," in *Computer Science* (2014), pp. 2672–2680.
- <sup>21</sup>O. Ronneberger, P. Fischer, and T. Brox, "U-NET: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, November 2015* (Springer, 2015), pp. 234–241.
- <sup>22</sup>C. Szegedy *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, 2015* (IEEE, 2015), pp. 1–9.
- <sup>23</sup>S. Ioffe, and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning, Lille, France, 2015*, 2015, pp. 448–456.
- <sup>24</sup>S. Xu, R. M. Freund, and J. Sun, "Solution methodologies for the smallest enclosing circle problem," *Comput. Optim. Appl.* **25**, 283–292 (2003).

<sup>25</sup>P. Bao, L. Zhang, and X. Wu, “Canny edge detection enhancement by scale multiplication,” *IEEE Trans. Pattern Anal. Mach. Intell.* **27**, 1485–1490 (2005).

<sup>26</sup>J. D. Wilson, W. Eardley, S. Odak, and A. Jenniings, “To what degree is digital imaging reliable? Validation of femoral neck shaft angle measurement in the era of picture archiving and communication systems,” *Br. J. Radiol.* **84**, 375–379 (2011).

<sup>27</sup>S. K. Pathak *et al.*, “Evaluation of femoral neck shaft angle on plain radiographs and its clinical implications,” *Int. J. Res. Orthop.* **2**, 383–386 (2016).

<sup>28</sup>A. Le Bras *et al.*, “3D reconstruction of the proximal femur with low-dose digital stereoradiography,” *Comput. Aided Surg.* **9**, 51–57 (2004).