Femur segmentation in X-ray image based on improved U-Net

Fan Lianghui¹, Han JunGang¹, Jia Yang¹, Yang Bin²
1. School of Computer, Xi’an University of Posts and Telecommunications, Xian Shanxi 710121, China
2. Xi’an Honghui Hospital, Xi’an 710054, China
592556555@qq.com, School of Computer, Xi’an University of Posts and Telecommunications

Abstract. Segmentation of Femur bone from X-ray images is an indispensable step in computer aided analysis of medical images and orthopaedic examinations. It is more complex than segmentation from CT and MR images, due to some associated less dense tissues that are hard to distinguish from the femur bone in X-ray images. This paper presents an improved method based on U-Net to automatically extract the femurs from hip X-ray images. This method changes the structure of the U-Net network, which can effectively map the non-linear relationship between hip image and femur image, and accurately segment femur image. The paper also added the absolute deviation loss function to improve the segmentation effect. Experimental results show that this method is accurate, robust, and achieves an average dice similarity coefficient of 0.966. The segmentation results are satisfactory.

1. Introduction
Analysis of proximal femur images plays an important role in Musculoskeletal Disorders (MSDs) diagnosis, preoperative planning and surgical planning of orthopedic conditions. Femur segmentation from X-ray image of whole hip bone is the prerequisite condition for bone morphologic feature analysis.

Because of the high noise, low contrast, tissue overlapping with hip bones and the narrow gap between femur and acetabulum[1], femur segmentation is a challenging task. In the previous works, segmentation error always occurs in joint regions [2]. Classical natural image segmentation algorithms apply homogeneous features[2, 3] to segment object regions. Unfortunately, such features are appropriate to natural pictures but not satisfied by X-ray images. For example, Figure 1 is a sample X-ray image of hip joint. Region 1 and 2, region 3 and 4 are with high homogeneity, while 2 and 4 belong to femur, 1 and 3 belong to hipbone. In this circumstance, color segmentation with threshold is ineffective for segmenting femur from the image. The arrows show some blurring regions on the edge of proximal femur. To determine the accurate edge in blurring region is also challengeable.

Most proposed segmentation methods are performed on CT and MRI images [4-11], rather than on X-ray images. Chen et al.[12] developed a model-based approach for automatically extracting femur contours from standard hip X-ray images. The method begins with detecting the positions of the femoral shaft, head and turning points, then a model femur contour is registered piecewise to the X-ray image according to the detected features. Finally, active contours with curvature constraints are applied to accurately identify the femur contours. 81.4% femur contours were successfully segmented from 172 X-ray images. Unfortunately, the proposed model-based method is not fit in fracture or odd
femurs segmentation and is time consuming. Feng et.al. [13] presented an atlas-based approach for automatic femur segmentation in X-ray images, which deforms the aligned atlas to extract detailed femur contours. Test results show that the proposed algorithm is robust and accurate in segmenting femur contours. However, it is extremely difficult to segment femurs with major fractures, i.e., not only the shapes but also the topologies of femur contours can be changed. Behiels et.al. [14] proposes an improved search procedure for Active Shape Model (ASM) to segment femur from X-ray Images. During the shape searching, shape model needs to change smoothly. But the method is sensitive to the morphological parameters (such as translation, rotation and scaling). Dong and Zheng [15] use a 3D statistical model for the femur bone contour extraction, but in their model the existence of occlusion would lead to a failure in extraction the contour around the femur neck area. Chav et.al. [16] proposed a prior shape segmentation method to extracted femur boundary. The method applies subspace to achieve segmentation of rectangular image, but the result of segmentation is not satisfactory. Ouertani et.al. [1] improved hip joint segmentation by simultaneously solving a dual edge-detection problem, but the results differed between the recovered contours and the manual segmentations.

![Figure 1. Sample X-ray Image of hip joint.](image)

In recent years, deep learning has been widely used and often outperforms traditional methods in the field of medical image analysis because of its great modeling capacity, and ability of extraction features. In this paper, we present a method to automatically segment the X-ray image of hip joint based on the improved u-net network, and the method is evaluated on the real X-ray data set. The experimental results show that this method can accurately segment the X-ray image of femur, which is better than the existing method.

2. The Network

CNN has achieved superior performance on visual object recognition and image classification [17], and the resulting U-Net network has made outstanding achievements in image segmentation [18]. In view of this, we improved the U-Net network model based on practical problems. The network improvements are described in detail below.

2.1. The Structure of Network

In the structure of U-Net network, there are two stages of convolution and deconvolution. In the convolution stage, the pooling layer is following two successive convolution layers. In this stage, the network retains the feature of the pooling layer. In the deconvolution phase, the feature of deconvolution layer is connected the feature of the corresponding pooled layer that previously retained for deconvolution.

The improvement of the network is removing the pooling layer and adds the batch normalized layer. Instead of cutting out the features extracted from the convolution layer, we merge the feature that have same dimension from convolution. The pooling layer is used to reduce the dimension of the output results by down sampling the input features. As reduces the output feature dimension, it discards the image's global information about the location that is importance for segmentation task. Therefore, we removed the pooling layer from the network for retaining the global information of the image. After removing the pooling layer in the network, the network training speed is slowed down because of the increase of data dimension. We add batch normalization layer to improve the convergence speed of the
network, and also avoid the influence of abnormal values in the image. In the improved network, the features of the convolution layer and the corresponding deconvolution layer have the same dimension. So, we can directly merge those features that have same dimension without cutting operation. The network structure is shown in figure 2.

In the shallow convolution layer, the global position information of pixel points is clear. In the deep convolution layer, the classification of pixels is accurate. These two kinds of information are both critical for medical image segmentation. In the U-Net network, skip to fuse can make use of both information to improve the accuracy of image segmentation. We continued this technique. The network proposed in this paper has 8 convolution layers and 8 deconvolution layers. The number of convolution layer filters is 64, 128, 256, 512, 512, 512, 512, and 512 respectively. The size of filter is fixed at 4×4 and set the strides value as 2. And the activation function of those filter is LReLu. The deconvolution layer is symmetric with the convolution layer. And the number of filters is 512, 512, 512, 512, 256, 128, and 64 respectively. The parameter setting is the same as the convolution layer. Except for the final output layer, the activation function of filter in deconvolution use ReLu function. The Tanh function is used as the activation function in the final output of the network. The batch normalization layer is added after each layer of convolution and deconvolution. In order to increase the generalization capability of the network, we add dropout[19] layer after the first three layers of deconvolution and set dropout keep a probability of 0.5.

Another difference in our network is that what the network outputs is not a binary image of the femur, but the original femur image. As shown in the figure 2.

2.2. Training

In the original U-Net network, a new weighted cross entropy loss function was defined according to the problem of cell segmentation. In our femur segmentation task, we use the original cross entropy loss function because the femur has a relatively fixed form.

\[
L_n(x, z) = -\sum_{i=1}^{d} x_i \log z_i + (1 - x_i) \log(1 - z_i)
\]

(1)

Where x and z stand for the input hip image and the corresponding ground-truth femur image respectively. Their vector length is d.

We also add the absolute loss error function, which is used as the loss function of the network together with the cross-entropy loss function. The loss function of the network is shown in eq (2).

\[
L_x = \lambda_1 L_n(x, z) + \lambda_2 \sum_{i=1}^{d} |x_i - z_i|
\]

(2)

Where \(\lambda_1\) and \(\lambda_2\) are used to balance the loss.
In practice, we set the minibatch size, the value of $\lambda_1$ and the value of $\lambda_2$ are set as 1, 1 and 100, respectively. The learning rate of the Adam optimizer is fixed in $2 \times 10^{-4}$, and parameter $\beta_1$ and $\beta_2$ are set as 0.5 and 0.999.

3. Experiment

3.1. Datasets and data pre-processing

The hip image dataset includes 50 DICOM X-ray images collected from the Xi’an Honghui Hospital. Data Masking is done to remove the sensitive information about the patient. The DICOM images are with a size of 2,688 × 2,208 (width × height) and a display window of [0, 4,096]. In order to balance the network computation and retain sufficient image information, the DICOM images were converted into PNG format and the image size was changed accordingly. Each image is transformed into a square with 512 × 512 pixels by cutting and stitching operations. Taking the femur segmentation image as an example, in Fig. 3, image (b) is a part of image (a), while image (c) is the combination of images (a) and (b).

An example of the hip and femur image is given in Figure 4. Each sample consists of a hip image and a femur image labeled manually.

3.2. Qualitative result

The results of femur segmentation of improved U-Net and original U-Net are showed in Figure 5. It can be clearly seen from Figure 5 that the improved U-Net can accurately segment the femur from the hip joint, while the original U-Net network segmentation results are not clear, and the contour of the femur is incomplete.

![Figure 3. Femur image pre-processing](image)

![Figure 4. Hip image (left) and its manually segmented femur (right)](image)
3.3. Qualitative result

Table 1 summarizes Dice coefficient, Hausdorff distance, recall, specificity and precision measured in pixel for 17 different subjects. Dice coefficient is a similarity measure related to the Jaccard index. The formula is defined as follows:

$$s = \frac{2 |X \cap Y|}{|X| + |Y|}$$  \hspace{1cm} (3)

where $X$ stands for the femur pixel set of segmentation result from the output of $G$, and the $Y$ indicates the femur pixel set of ground truth image. The larger the Dice coefficient is, the more similar the two images are.

The Hausdorff distance measures the dissimilarity of two shapes. As shown in eq (4)

$$d_h(X,Y) = \max\{\sup_{x,y} d(x,y), \inf_{x,y} d(x,y)\}$$  \hspace{1cm} (4)

where sup and inf stands for the supremum and the infimum respectively, $X$ and $Y$ denote the contours set of the segmentation result and the ground truth respectively. The smaller the distance, the closer the match of two images, the better the segmentation results.

We also calculate the three evaluation criteria of image segmentation in this paper. The three criteria are calculated on the pixel level based on the segmented image and the ground truth image. TP stands for the pixel number of to correctly segmented femur. FN stands for the pixel number of that the femur is incorrectly segmented as background. FP stands for the pixel number of the background incorrectly segmented as femur, and TN stands for the pixel number of correctly segmented background. The three definitions are as follows:

$$Recall = \frac{TP}{TP + FN}$$  \hspace{1cm} (5)

$$Specificity = \frac{TN}{FP + TN}$$  \hspace{1cm} (6)

$$Precision = \frac{TP}{TP + FP}$$  \hspace{1cm} (7)

Recall represents the ratio of the number of pixels belong to correctly segmented femur to the number of pixels in ground truth. Specificity represents the ratio of the number of pixels belong to correctly
segmented background to the number of pixels belong to the whole segmented background. Precision represents the ratio of the number of pixels belong to correctly segmented femur to the number of pixels in the whole segmented femur.

Table 1. Quantitative segmentation results

| ID | Dice  | Hausdorff | Recall  | Specificity | Precision |
|----|-------|-----------|---------|-------------|-----------|
| 1  | 0.9684 | 9.00      | 0.9675  | 0.9595      | 0.9249    |
| 2  | 0.9751 | 39.02     | 0.9355  | 0.9399      | 0.8981    |
| 3  | 0.9850 | 36.02     | 0.9576  | 0.9414      | 0.9074    |
| 4  | 0.9787 | 8.25      | 0.9213  | 0.9390      | 0.9365    |
| 5  | 0.9760 | 13.00     | 0.9373  | 0.9598      | 0.9483    |
| 6  | 0.9661 | 9.07      | 0.9183  | 0.9198      | 0.9281    |
| 7  | 0.9722 | 45.64     | 0.9491  | 0.9287      | 0.9222    |
| 8  | 0.9592 | 48.05     | 0.8536  | 0.9386      | 0.9304    |
| 9  | 0.9686 | 29.08     | 0.9291  | 0.9596      | 0.9362    |
| 10 | 0.9776 | 40.81     | 0.9247  | 0.9192      | 0.9331    |
| 11 | 0.9586 | 5.00      | 0.9419  | 0.9294      | 0.9445    |
| 12 | 0.9483 | 20.21     | 0.9388  | 0.9493      | 0.9139    |
| 13 | 0.9682 | 2.83      | 0.9500  | 0.9390      | 0.9497    |
| 14 | 0.9483 | 9.49      | 0.9293  | 0.9393      | 0.9339    |
| 15 | 0.9677 | 7.00      | 0.9417  | 0.9287      | 0.9124    |
| 16 | 0.9567 | 7.00      | 0.9579  | 0.9484      | 0.9214    |
| 17 | 0.9481 | 2.83      | 0.9497  | 0.9193      | 0.9147    |
| average | 0.9660 | 19.55   | 0.94    | 0.9388      | 0.9268 |

The average Dice coefficient reached 0.966, and the average Hausdorff distance is 19.55, indicating a high similarity between femur segmentation results and ground true. The value of three indexes: Recall, Specificity and Precision are very high, proving that the method is robust and accurate. From the data in the table 1, we can conclude that our method has an impressive effect on the segmentation of the femur.

4. Conclusion

In this paper, we present an improved U-Net for segmentation femur. By comparing the segmentation results, we can conclude that the improved method in this paper is effective. In many adverse conditions, such as extraneous straight lines caused by the analogue imaging process, blurry boundary, femoral deformities, the proposed method can accurately segment femur image. The experimental results show that this method can effectively segment the femur in the original hip image. The next step is to apply this method to the segmentation of other X-ray bone images.

References

[1] Ouertani, F. Simultaneous extraction of two adjacent bony structures in x-ray images: Application to hip joint segmentation. in International Conference on Image Processing. 2015.
[2] Bischof, S. Seeded Region Growing. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2002. 16(6): p. 641-647.
[3] Otsu, N., A Threshold Selection Method from Gray-Level Histograms. IEEE Transactions on Systems Man & Cybernetics, 2007. 9(1): p. 62-66.
[4] Chu, C., et al., MASCG: Multi-Atlas Segmentation Constrained Graph method for accurate segmentation of hip CT images. Medical Image Analysis, 2015. 26(1): p. 173-184.
[5] Cheng, Y., et al., Automatic segmentation technique for acetabulum and femoral head in CT images. Pattern Recognition, 2013. 46(11): p. 2969-2984.
[6] Seim, H., et al. Automatic segmentation of the pelvic bones from CT data based on a statistical
shape model. in Eurographics Conference on Visual Computing for Biomedicine. 2008.
[7] Chandra, S.S., et al., Focused shape models for hip joint segmentation in 3D magnetic
resonance images. Medical Image Analysis, 2014. 18(3): p. 567-578.
[8] Xia, Y., Automated segmentation and quantitative analysis of the hip joint from magnetic
resonance images. Bone, 2015.
[9] Xia, Y., et al. Automated MR Hip Bone Segmentation. in International Conference on Digital
Image Computing Techniques and Applications. 2012.
[10] Schmid, J., J. Kim, and N. Magnenat-Thalmann, Robust statistical shape models for MRI bone
segmentation in presence of small field of view. Medical Image Analysis, 2011. 15(1): p.
155-168.
[11] Li, W., et al., Human Hip Joint Cartilage: MRI Quantitative Thickness and Volume
Measurements Discriminating Acetabulum and Femoral Head. IEEE Trans Biomed Eng,
2008. 55(12): p. 2731-2740.
[12] Chen, Y., et al., Automatic Extraction of Femur Contours from Hip X-Ray Images. 2005:
Springer Berlin Heidelberg. 200-209.
[13] Feng, D., W.K. Leow, and T.S. Howe. Automatic segmentation of femur bones in anterior-
posterior pelvis x-ray images. in International Conference on Computer Analysis of Images
and Patterns. 2008.
[14] Behiels, G., et al., Active Shape Model-Based Segmentation of Digital X-ray Images. Lecture
Notes in Computer Science, 1999. 1679: p. 128-137.
[15] Xiao, D., M.A.G. Ballester, and G. Zheng, Automatic Extraction of Femur Contours from
Calibrated X-Ray Images using Statistical Information. Journal of Multimedia, 2007. 2(5): p.
55-55.
[16] Chav, R. and C. Kauffmann, Method for fast and accurate segmentation processing from prior
shape: application to femoral head segmentation on x-ray images. 2009. 72594Y-72594Y-8.
[17] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahrourdy, B. Shuai, T. Liu, X. Wang, L. Wang, G. Wang,
J. Cai, T. Chen, Recent Advances in Convolutional Neural Networks, Computer Vision and
Pattern Recognition (2015).
[18] Ronneberger, O., P. Fischer, and T. Brox. U-Net: Convolutional Networks for Biomedical
Image Segmentation. in International Conference on Medical Image Computing and
Computer-Assisted Intervention. 2015.
[19] Hinton, G.E., et al., Improving neural networks by preventing co-adaptation of feature detectors.
Computer Science, 2012. 3(4): p. págs. 212-223.