ICIC Express Letters
Part B: Applications
Volume **13**, Number **2**, February **2022**

ICIC International ©2022 ISSN 2185-2766
pp. 155–160

AUTOMATIC SEGMENTATION OF FINGER BONE REGIONS FROM CR IMAGES USING IMPROVED U-NET AND MSGVF SNAKES

Kohei Kawagoe[1], Seiichi Murakami[2], Tohru Kamiya[1] and Takatoshi Aoki[2]

[1]Department of Control Engineering
Kyushu Institute of Technology
1-1, Sensui, Tobata, Kitakyushu 804-8550, Japan
kawagoe.kouhei622@mail.kyutech.jp; kamiya@cntl.kyutech.ac.jp

[2]Department of Radiology
University of Occupational and Environmental Health
1-1, Iseigaoka, Yahata-nishi 807-8555, Japan
a-taka@med.uoeh-u.ac.jp

Received June 2021; accepted August 2021

Abstract. *Bone disease including rheumatoid arthritis and osteoporosis is one of serious diseases especially elderly people. Diagnostic imaging using computed radiography (CR) imaging is an effective method for diagnosing the bone disease. However, accurate diagnosis should be made within a limited time, increasing the burden on doctors. Therefore, the goal of this paper is to develop a computer aided diagnosis (CAD) system for automatically diagnosing bone disease from CR images. The proposed method based on an improved U-Net and multiscale gradient vector flow snakes (MSGVF Snakes) extracts each phalange region from the CR image and creates an image for each phalange. In the experiment, the proposed method was applied to CR images of 101 cases, and the performance was evaluated. As a result, our proposed model based on improved U-Net for the segmentation obtained intersection over union (IoU) of 0.939, true positive (TP) of 0.984, and false positive (FP) of 0.068. The initial segmentation accuracy was higher than the conventional approach such as U-Net, Residual U-Net. Final segmentation performance of the phalanges based on MSGVF Snakes are TP of 0.945 and FP of 0.030.*
Keywords: Rheumatoid arthritis, Osteoporosis, Computer aided diagnosis, Convolutional neural network, Multiscale gradient vector flow snakes

1. **Introduction.** At present, Japan became a super-aging society. In Japan, the percentage of people over the age of 65 is increasing. In 2018, it was 28.1%. As a matter of aging, the risk of injury and illness in the elderly has increased, and the number of patients has also increased accordingly. Bone disease is one of the most common diseases in the elderly. Symptoms of bone disease include growth disorders, skeletal deformities, joint disorders, and fractures. Representative cases include rheumatoid arthritis and osteoporosis. Bone disease leads to a decrease in quality of life (QOL) as the condition progresses. In addition, new drugs for rheumatoid arthritis and osteoporosis are being developed that can alleviate symptoms and reduce morbidity. Therefore, early detection and early treatment are important to prevent the disease.

Visual screening based on imaging becomes a routine work for medical diagnosis for early detection. In recent years, its reliability and importance have increased with the improvement of imaging technology such as computed tomography, ultrasound, and magnetic resonance imaging. However, there are problems such as differences in diagnostic results due to the experience of doctors, and the burden on doctors due to the limited

DOI: 10.24507/icicelb.13.02.155

time of diagnosis. Therefore, we propose a computer aided diagnosis (CAD) system that can perform quantitative analysis to solve these problems.

Kajihara et al. [1] proposed a method to extract each phalange image by applying multiscale gradient vector flow (MSGVF) Snakes [2] from computed radiography (CR) images. Hatano et al. [3] proposed an extraction method for extracting phalanx regions using U-Net [4]. Also, there are many image processing algorithms which is related to CNN [5-10]. These methods still have problems with mis-detected phalanges and versatility. Therefore, we propose a method for extracting phalange region from the CR images by combining our previous model with MSGVF Snakes. In this paper, we discuss the effectiveness of the proposed method applied to 101 CR images.

The remainder of this paper is structured as follows. In Section 2, we describe our image processing techniques in detail. Section 3 illustrates the experimental results and Section 4 addresses discussion and conclusion.

2. Image Processing Techniques.

In this paper, we describe a CAD system based on automatic segmentation of bone area, automatic registration, and feature analysis which compose CAD system for quantitative analysis of bone deformation from CR image. Main process includes automatic extraction of region of interest (ROI) based on improved U-Net, and MSGVF Snakes.

2.1. Overview of the proposed method.

The aim of this study is to develop a CAD system for realizing quantitative evaluation of bone diseases. To achieve this purpose, it is necessary to develop three elemental technologies: automatic segmentation, automatic registration, and feature analysis of the phalangeal region. Figure 1 shows an overview of the CAD system constructed in this paper. Extraction of the bone region in the first step is performed for reducing the burden on doctors such as coping with the complexity of positioning processing and positioning of ROI. In addition, since bone disease does not occur in the entire bone, and symptoms appear in only a part at an early stage, image analysis is performed on each finger bone. By applying the process, we will improve the accuracy of diagnosis and reduce the burden on doctors. As the preprocessing steps, original CR image with 10 bits is converted to bitmap image with 8 bits. After that, converted images are enhanced based on histogram equalization.

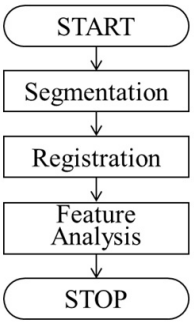


FIGURE 1. The flow of CAD system

2.2. Automatic segmentation of finger bone region.

A convolutional neural network (CNN) is a specific type of artificial neural network (ANN) that uses perceptrons, a machine learning unit algorithm. In this paper, a CNN algorithm based on improved U-Net as an initial segmentation and MSGVF Snakes as final segmentation are used in combination for segmentation of phalanges region. The segmentation targets are the proximal and middle phalanx from the CR images. In addition, we build a residual U-Net [4] as a CNN model for segmentation, and learn and extract bone region from a known CR image.

In this method, we segment the proximal and middle phalange region based on two step processing: an improved CNN as an initial segmentation and MSGVF Snakes techniques for the final segmentation. The whole flow of this method inputs an image and performs extraction of a finger bone image at CNN. Next, noise removal and labeling processing are performed on the obtained extracted image as pre-processing to create each finger bone image. After that, each phalange region is set as an initial position of MSGVF Snakes to extract the final segmentation. This makes it possible to accurately extract only the phalanges.

2.3. **Improved U-Net.** U-Net is a CNN model that combines residual U-Net [6] and nested U-Net [11]. The basic structure is Residual U-Net, and the connectivity of the skip connection part is changed to a nested structure. At this point, Nested U-Net will introduce a convolution block, while the improved U-Net will introduce a residual block. The improved U-Net extracts image features from the input at the encoder and restores the image at the decoder. In this method, full pre-activation is used as residual block. Each block is composed of 3×3 convolution, ReLU, and batch normalization. Therefore, we down sample the feature map by applying 2 slides in the first 3×3 convolution of each residual block as a substitute for maximum pooling. Figure 2 shows the network structure of our improved U-Net.

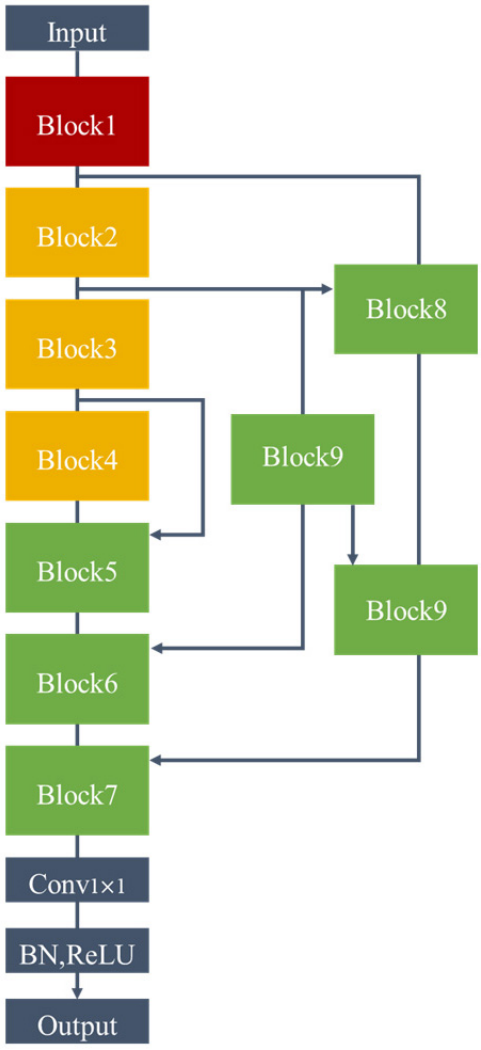
The advantage of the improved U-Net is that the introduction of the residual block in the skip connection part enables the transmission of feature information while suppressing image degradation compared to the residual U-Net. In addition, optimization can be performed easily.

2.4. **MSGVF Snakes.** To segment an ROI from an image based on MSGVF Snakes [2], it is necessary to input the initial control points by user. To realize automation as a CAD system, boundary tracking is performed on the rough extraction area, and contours are sampled at regular intervals. The sampling point is given as an initial control point, and a detailed area extraction of the phalanges is performed by MSGVF Snakes.

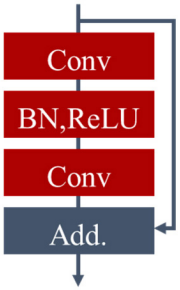
As the processing flow, first we set the scale λ_{M-1} and calculate the edge map for the smoothed image according to the scale. Then, based on the obtained edge map, the control points are moved in the direction of the vector field until the active contour converges. After convergence, the scale value is updated to the value obtained by subtracting 1.

The control points after convergence are given again as the initial control points, and the same process is repeated until the scale λ_0 is reached. The scale λ_0 indicates the original image, and is possible to perform local extraction gradually from the global extraction that is less susceptible to noise by repeating from the large scale λ_{M-1} to the scale λ_0. In addition, B-spline [12] interpolation is used as a control point interpolation method. Control points can be represented smoothly by using the B-spline interpolation.

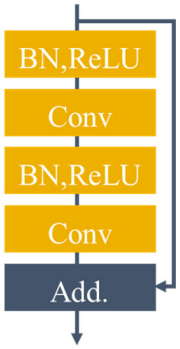
MSGVF is a method of expressing a vector field using an edge map of an image roughened by scale based on scale space theory. Conventional GVF is susceptible to noise and is characterized in that it forms a local vector field. MSGVF broadens the range of vectors concentrated on edges by using an edge map obtained from a roughened image as compared to the edge map in the original image. Therefore, a vector field can be formed globally. Due to this feature, even when the initial contour in Snakes is apart from the contour of the extraction target, it converges to the target contour. In addition, since the rough image is used, the region extraction can be performed more accurately because it is less susceptible to noise. The convergence behavior in Snakes is established by minimizing the energy function which consists of internal and external forces. If the position of parameterized Snakes is expressed as $r(s) = (x(s), y(s))$, the general Snakes energy function can be expressed by Equation (1). Here, E_{int} represents internal energy, and E_{ext} represents external energy. Internal energy is defined to keep the dynamic contour smooth. External energy is used to derive an active contour towards the image features



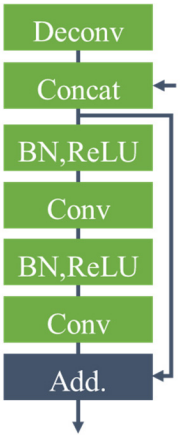
(a) Network structure



(b) Block1



(c) Block2 ∼ Block4 (Full pre-activation)



(d) Block5 ∼ Block7

FIGURE 2. Network structure of improved U-Net: Conv: 3×3 convolution, Deconv: 2×2 deconvolution, BN: Batch normalization, Concat: Concatnation, Add.: Addition

of the region of interest. The method using the vector field obtained by MSGVF as this external energy is MSGVF Snakes.

$$E_{snake} = \int_0^1 [E_{int}(r(s)) + E_{ext}(R(S))]ds \tag{1}$$

The formulas for E_{int} and E_{ext} are shown below.

$$E_{int} = \frac{\left(\alpha|r_s(s)|^2 + \beta|r_{ss}(s)|^2\right)}{2} \tag{2}$$

$$E_{ext} = \frac{u_\lambda(x,y) \times u_\lambda(\bar{x},\bar{y}) + v_\lambda(x,y) \times v_\lambda(\bar{x},\bar{y})}{\left((u_\lambda(x,y))^2 + (v_\lambda(x,y))^2 \times \left(u_\lambda(\bar{x},\bar{y})\right)^2 + (v_\lambda(\bar{x},\bar{y}))^2\right)} \tag{3}$$

Equation (2) describes the commonly used curve broadening. Equation (3) shows the inner product of the vectors, taking advantage of the feature that the facing vector near the edge is in the opposite direction. $(\bar{x},\bar{y})$ in Equation (3) indicates the control points after movement, and is the direction of MSGVF at the position (x,y) of the movement amount. The equations for $\bar{x}$ and $\bar{y}$ are shown below.

$$\bar{x} = x + \eta u(x,y) \tag{4}$$

$$\bar{y} = y + \eta v(x,y) \tag{5}$$

α and β in Equation (2) and η in Equations (4) and (5) are positive constants. E_{int} works to reduce closed curves, especially when the active contour is set outside the extraction target.

3. **Experimental Results.** In this paper, we evaluate the performance of the phalange image extraction method combining the improved U-Net and MSGVF Snakes. In the experiment, the performance was evaluated by comparing the image obtained by the proposed method and the image obtained by the conventional method with the correct image. The results were applied to CR images of 101 cases which is obtained on visual screening.

The evaluation method used intersection over union (IoU), true positive (TP) and false positive (FP). In Equations (6) to (8), A is the correct answer area, and B is the extraction result by the proposed method.

$$\text{IoU} = \frac{A \cap B}{A \cup B} \tag{6}$$

$$\text{TP} = \frac{A \cap B}{A} \tag{7}$$

$$\text{FP} = \frac{B - A \cap B}{B} \tag{8}$$

In this study, we obtained the segmentation results shown in Table 1 for CR images. In our improved U-Net, we achieved IoU = 0.939, TP = 0.984, and FP = 0.068 from 101 CR cases. In addition, the results of final segmentation performance based on MSGVF Snakes were TP = 0.945 and FP = 0.030. From Table 1, the proposed method has improved on the extraction performance. In this study, an extraction method combining improved U-Net and the MSGVF Snakes method was implemented. In our improved U-Net, the phalange region was globally extracted, and the obtained image was used as the initial position to locally extract the phalange region using the MSGVF Snakes.

Table 1. Experimental results

	IoU	TP	FP
Improved U-Net	0.939	0.984	0.068
Conventional method [2]	0.909	0.971	0.077
Final segmentation	–	0.945	0.030
Conventional method [1]	–	0.929	0.059

(a) Improve U-Net (b) MSGVF Snakes

FIGURE 3. Experimental result

4. **Conclusions.** We developed a method for extracting phalanges by combining improved U-Net and the MSGVF Snakes method, and evaluated its performance using an evaluation formula and processing time. As a result, we obtained the performance with IoU = 0.939, TP = 0.984, and FP = 0.068. The final segmentation of the phalanges by MSGVF Snakes showed TP = 0.945 and FP = 0.030. As the future work, we need further improvement for the automatic extraction of the phalanges region and segment the phalanges region.

REFERENCES

[1] S. Kajihara et al., Automatic segmentation of phalanges regions on CR images based on MSGVF Snakes, *International Conference on Control, Automation and Systems*, pp.1290-1293, 2014.

[2] J. Tang, Vessel boundary tracking for intravital microscopy via multiscale gradient vector flow snakes, *IEEE Trans. Biomed.*, vol.51, no.2, pp.316-324, 2004.

[3] K. Hatano et al., Detection of phalange region based on U-Net, *International Conference on Control, Automation and Systems*, pp.1338-1342, 2018.

[4] O. Ronneberger, P. Fischer and T. Brox, U-Net: Convolutional networks for biomedical image segmentation, *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp.234-241, 2015.

[5] A. Krizhevsky, I. Sutskever and G. E. Hinton, ImageNet classification with deep convolutional neural networks, *Adv. Neural Inf. Process. Syst.*, vol.60, no.6, pp.84-90, DOI: 10.1145/3065386, 2012.

[6] Z. Zhang, Q. Liu and Y. Wang, Road extraction by deep residual U-Net, *IEEE Geoscience and Remote Sensing Letters*, vol.15, no.5, pp.749-753, 2018.

[7] K. He, X. Zhang, S. Ren and J. Sun, Deep residual learning for image recognition, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.770-778, DOI: 10.1109/CVPR.2016.90, 2016.

[8] K. He, X. Zhang, S. Ren and J. Sun, Identity mappings in deep residual networks, in *Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science*, B. Leibe, J. Matas, N. Sebe and M. Welling (eds.), Cham, Springer, DOI: 10.1007/978-3-319-46493-0_38, 2016.

[9] S. Ioffe and C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, *arXiv.org*, arXiv: 1502.03167v3, 2015.

[10] J.-Y. Zhao, J. Gong, S.-T. Ma and Z.-M. Lu, Curvature gray feature decomposition based finger vein recognition with an improved convolutional neural network, *International Journal of Innovative Computing, Information and Control*, vol.16, no.1, pp.77-90, 2020.

[11] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh and J. Liangc, UNet++: A nested U-Net architecture for medical image segmentation, in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. DLMIA 2018, ML-CDS 2018. Lecture Notes in Computer Science*, D. Stoyanov et al. (eds.), Cham, Springer, DOI: 10.1007/978-3-030-00889-5_1, 2018.

[12] M. Unser, A. Aldroubi and M. Eden, B-spline signal processing: Part I – Theory, *IEEE Trans. Signal Processing*, vol.41, no.2, pp.821-832, 1993.