A Temporal Subtraction Technique for Phalange CR Image Using CNN

Hikaru Ono [1], Tohru Kamiya [1,*], and Takatoshi Aoki[2]

[1] Department of Mechanical and Control Engineering, Kyushu Institute of Technology, Kitakyushu, Japan
[2] Department of Radiology, University of Occupational and Environmental Health, Kitakyushu, Japan
Email: ono.hikaru240@mail.kyutech.jp (H.O.); kamiya@cntl.kyutech.ac.jp (T.K.); a-taka@med.uoeh-u.ac.jp (T.A.)
*Corresponding author

***Abstract*—X-ray examinations are widely used in the diagnosis of Rheumatoid Arthritis (RA). However, the condition of many phalanges and joints must be evaluated visually, which causes a lack of objectivity due to subjective evaluation by the physician and an increased workload for the physician in reading the images. In this paper, we propose an image analysis method for hand Computed Radiography (CR) images based on the temporal subtraction method to support the diagnosis of rheumatoid arthritis. The proposal method consists of three steps. First, a Convolutional Neural Network (CNN) model for semantic segmentation, which is efficient in terms of computational complexity and accuracy, is proposed to extract phalangeal regions. Second, a geometric-matching CNN with instance-specific optimization is used to align the phalangeal regions. Finally, the current image and the aligned past image are subtracted to visualize the temporal changes. We applied the proposed method to hand CR images and confirmed its effectiveness.**

***Keywords*—Computer Aided Diagnosis (CAD) system, segmentation, image registration, temporal subtraction technique, Rheumatoid Arthritis (RA)**

I. Introduction

Rheumatoid Arthritis (RA) is a chronic inflammatory disease characterized by joint swelling, joint tenderness, and destruction of synovial joints [1]. In the early stages, the joints of the hands and feet are easily affected, and gradually destruction and deformation of joints throughout the body occur. Aggressive treatment and tight monitoring from an early stage are important to control the progression of symptoms [2]. However, the diagnosis of rheumatoid arthritis requires a lot of time and effort due to visual evaluation on X-ray images. of rheumatoid arthritis has required a lot of time and effort due to the need for visual evaluation using X-ray images. To solve these problems, the development of a Computer Aided Diagnosis (CAD) system is expected to provide physicians with the results of computer analysis as a second opinion.

The temporal subtraction technique [3] is one of the image analysis techniques that support the reading of X-ray images. This technique visualizes the presence or absence of temporal changes by generating subtraction images of the past and current images from X-ray image of the same patient. By visualizing temporal changes, it is expected to improve diagnostic accuracy and shorten reading time on visual screening.

In a study of temporal subtraction techniques to aid in the diagnosis of rheumatoid arthritis, Ichikawa *et al.* [4] proposed a CAD system to detect the progression of joint space narrowing. However, visual alignment of the joints is necessary. Kajihara *et al.* [5] also proposed an image segmentation method for phalangeal region using Multi Scale Gradient Vector Flow (MSGVF) Snakes and image registration method based on Salient Region Feature (SRF). Although these methods achieve automatic segmentation and registration, they have issues with accuracy and processing speed.

To analyze the temporal changes in phalanges, it is necessary to develop accurate segmentation methods and correct alignment methods of previous and current images on same subject. Therefore, we propose a segmentation and registration method for phalangeal regions using a Convolutional Neural Network (CNN). In the segmentation step, we propose U-ConvNeXt, which is efficient in terms of computational complexity and accuracy. In the registration step, the phalangeal region images are aligned using geometric matching CNN [6] with instance-specific optimization [7]. Finally, the proposed method is applied to hand CR images and the results and discussion are described. This paper develops an image segmentation methods using deep learning and confirms its usefulness with synthetic data on CT images. It also proposed an accurate image registration method based on geometric CNN based image alignment methods. As a result, we established a technique for correctly detecting of temporal changes from difference images. This technique enables effective enhancement of newly appearing lesions and contributes to assisting radiologist in diagnosis. From the experimental results, segmentation performance and registration accuracy are improved compare with conventional method.

The rest of this work is organized as follows: The fundamental technique for the proposed method based on segmentation and registration are described in Section II. Experimental results and discussions are shown in Section III. Finally, conclusion is presented in Section IV.

Manuscript received March 4, 2024; revised April 9, 2024; accepted May 10, 2024; published July 23, 2024.

II. Methods

In this section we provide segmentation and registration methods to generate temporal subtraction images of the phalangeal region.

A. U-ConvNeXt

It is necessary to extract the phalangeal region from the hand CR image for registration. The extraction of phalangeal regions is automated using CNN. We propose a new CNN model called U-ConvNeXt for automatic extraction of phalangeal regions. U-ConvNeXt extends U-Net [8] and uses the ConvNeXt block [9] as an encoder to expand the receptive field and reduce computational complexity.

An overview of the U-ConvNeXt architecture is presented in Fig. 1. First, the input image is reduced to 1/2 the image size by the stem block shown in Fig. 2. Our stem block differs from U-Net in that the first convolutional layer is changed to stride = 2 to reduce computational complexity. The feature map output from the stem block is input to "Stage 1", which consists of a down sampling block and a ConvNeXt blocks. The down sampling block consists of a 3×3 convolutional layer with stride = 2 and BN (Batch Normalization) to reduce the size of the feature map. The ConvNeXt block consists of a 7×7 depthwise convolution layer and two 1×1 convolution layers, as shown in Fig. 3, to expand the receptive field and reduce computational complexity. This process is repeated four times, from "Stage 1" to "Stage 4". The number of ConvNeXt blocks in each stage is [2,2,6,2], respectively. The decoder is built by replacing the activation function in the double convolution block of U-Net from Rectified Linear Unit (ReLU) to Gaussian Error Linear Unit (GELU) [10], with the other layers kept the same.

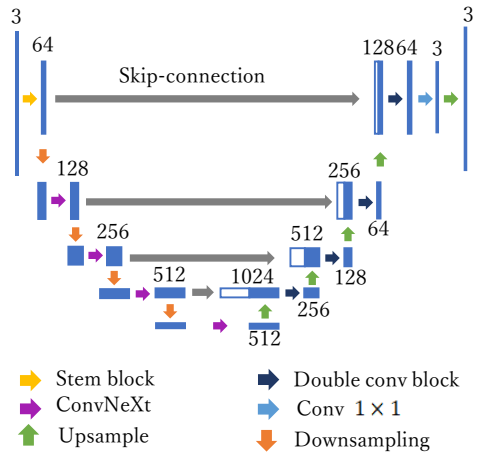


Fig. 1. An overview of the U-ConvNeXt architecture.

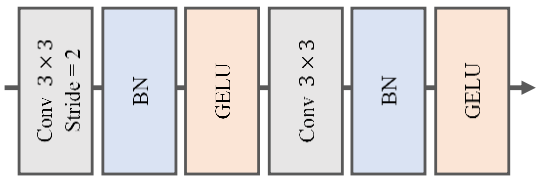


Fig. 2. Architecture of the stem block. "Conv" indicates convolution and "BN" indicates batch normalization.

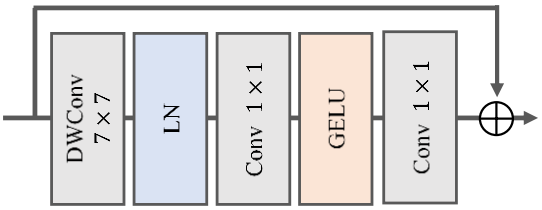


Fig. 3. Architecture of the ConvNeXt block. "DWConv" indicates depthwise convolution and "LN" indicates layer normalization.

B. Generation of Phalangeal Region Images

The phalangeal region image is generated based on the segmentation map output from the CNN. First, the area of each region is calculated. Regions with small areas are considered as noise, they are deleted, and only the phalangeal regions are selected. Next, the region corresponding to the phalangeal region of the segmentation map is extracted from the CR image. The extracted region is placed in the center of the image to generate a phalangeal region image.

C. Image Registration

Geometric-matching CNN [6] is used to align the past and current images. Because the images in this paper were taken with equipment and hands in the same position in the past and current, rigid transformation shown in Eq. (1) is used for registration.

$$R = \begin{pmatrix} \cos\theta & -\sin\theta & T_x \\ \sin\theta & \cos\theta & T_y \\ 0 & 0 & 1 \end{pmatrix} \quad (1)$$

Since it is necessary to estimate the three parameters (θ, T_x, T_y) in a rigid transformation, the number of output dimensions in the Geometric-matching CNN is changed to three.

An overview of the proposed registration method is shown in Fig. 4. The geometric transformation parameters estimated by the CNN can be interpreted as simply an approximation or initialization to the optimal deformation [7]. Therefore, after initial alignment by Geometric-matching CNN, the weight parameters of the CNN model can be improved by using instance-specific optimization [7]. The geometrical transformation parameters re-estimated by the CNN are applied to the initially aligned phalangeal region images and optimized to minimize the mean squared error in pixel values between image pairs.

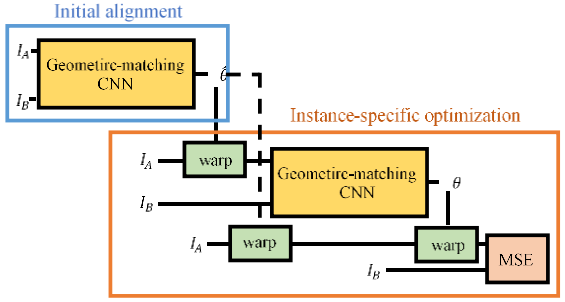


Fig. 4. Overview of the registration method.

D. Generation of Subtraction Images

After the registration process, a subtraction image of the past and current images is generated to highlight temporal changes. Since the image density of CR images differs from case to case due to differences in imaging conditions, the linear gradation process [11] shown in Eq. (2) is applied before generating the subtraction image.

$$y = \left(\frac{\sigma}{\sigma'}\right)x + \left\{\mu - \left(\frac{\sigma}{\sigma'}\right)\mu'\right\} \quad (2)$$

where, x is the pixel value of the image, y is the pixel value after correction, μ is the mean of pixel values of the base image, σ is the standard deviation of pixel values of the base image, μ' is the mean of pixel values of the corrected image, and σ' is the standard deviation of pixel values of the corrected image, respectively.

III. Experimental Results

We evaluated the performance of our proposed method for segmentation and registration to confirm its effectiveness. Information on the experimental equipment is shown in Table I.

TABLE I. Experimental Equipment

Equipment	Information
OS	Ubuntu 20.04.4 LTS
CPU	Intel® Core™ i7-6850K CPU@3.6GHz
RAM	32GB
GPU	NVIDIA GeForce RTX 2080 SUPER
VRAM	8GB

A. Evaluation of Segmentation Performance

In this paper, U-ConvNeXt was applied to hand CR images for segmentation of the phalangeal regions. The data set consists of 202 one-hand-only images cropped to a size of 512×512 from the hand CR images of 101 cases. We evaluated with 5-fold cross validation due to the number of data. IoU (Intersection over Union) and mIoU (mean Intersection over Union) shown in Eq. (3) and (4) were used as evaluation metrics.

$$\mathrm{IoU} = \frac{|A \cap B|}{|A \cup B|} \quad (3)$$

$$\mathrm{mIoU} = \frac{1}{L}\sum_{l=1}^{L} \mathrm{IoU}_l \quad (4)$$

where A is the ground truth, B is the prediction, L is the number of classes, IoU_l is the IoU in class l.

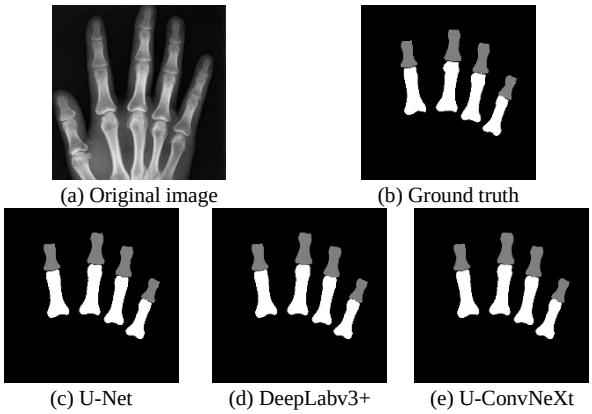
We compared U-ConvNeXt with U-Net [8] and DeepLabv3+ [12]. Table II shows the segmentation accuracy and Table III shows the memory usage and processing speed for each model. Fig. 5 also shows an example of the segmentation results for each model. In the figure, the gray region represents the media phalanges and the white region represents the proximal phalanges. The mIoU of each model was 95.30% for U-Net, 95.15% for DeepLabv3+, and 95.21% for U-ConvNeXt.

TABLE II. Comparison of Segmentation Accuracy

Model	Media(%)	Proximal(%)	*mIoU*
U-Net [8]	94.89	95.71	95.30
DeepLabv3+ [12]	94.65	95.66	95.15
U-ConvNeXt	94.76	95.66	95.21

TABLE III. Comparison of Memory Usage and Processing Speed

Model	Memory	Speed (iteration /sec) Train/Test
U-Net [8]	7.42	1.70/12.30
DeepLabv3+ [12]	5.11	2.66/12.67
U-ConvNeXt	4.63	3.79/17.68



(a) Original image (b) Ground truth (c) U-Net (d) DeepLabv3+ (e) U-ConvNeXt

Fig. 5. Example of segmentation results.

B. Evaluation of Registration Performance

In this paper, the proposed method is applied to 240×240 pixels phalangeal region images to evaluate the registration performance. Due to the small number of phalangeal region images, we used the Pascal VOC 2011 dataset [13] for training the Geometrical-matching CNN.

Registration accuracy is evaluated by the overlap of the phalangeal regions obtained by segmentation. As evaluation metrics, we use TP, FP, and Dice:

$$\mathrm{TP} = \frac{|A \cap B|}{|A|} \quad (5)$$

$$\mathrm{FP} = \frac{|B| - |A \cap B|}{|A|} \quad (6)$$

$$\mathrm{Dice} = \frac{2|A \cap B|}{|A| + |B|} \quad (7)$$

where, A is the area of the target image and B is the area of the aligned image respectively.

C. Experimental Results on Synthetic Data

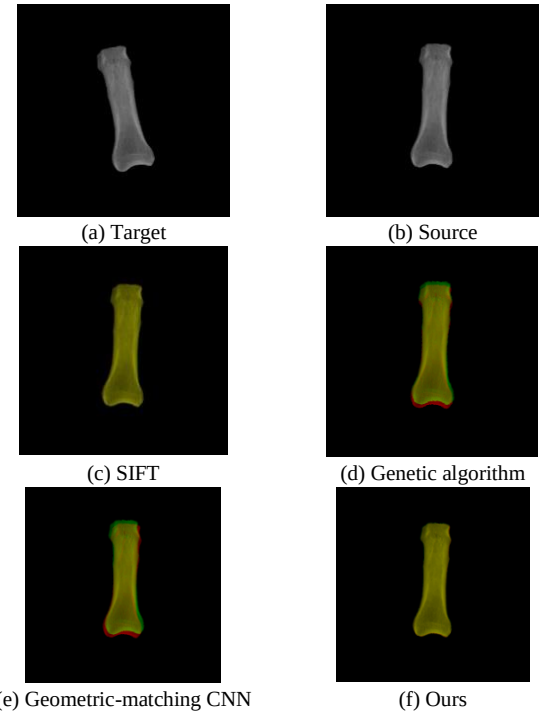
Real data consisting of past and current images is difficult to accurately evaluate the registration accuracy due to bone deformation and other factors. Therefore, we used 560 pairs of synthetic data. The synthetic data were created by randomly adding rotations in the range of −15° to 15° and translations in the range of −10 to 10 pixels to the phalangeal region images.

In this paper, the proposed method is compared with Scale-Invariant Feature Transform (SIFT) [14], Genetic Algorithm (GA) [15], and Geometric-matching CNN [6]. The parameters of the GA were determined based on Ref. [15] and the number of generations was set to 200. The proposed method used gradient descent for 50 iterations on each test pair. Table IV shows the experimental results of registration on synthetic data. However, SIFT lacked the feature points necessary to calculate the rigid transformation parameters in the four

image pairs. The proposed method achieved a Dice score of 99.04%. Fig. 6 also shows an example of result images. Fig. 6(c) to (d) are composites of the target image in red and the aligned image in green, with the area where both overlap represented in yellow.

TABLE IV. RESULTS ON SYNTHETIC DATA (GM CNN SHOWS GEOMETRIC-MATCHING CNN)

Method	TP (%)	FP (%)	Dice (%)
SIFT	95.17	4.86	95.15
GA [15]	95.59	4.52	95.54
GM CNN [6]	93.66	6.39	93.64
Ours	99.08	1.00	99.04



(a) Target (b) Source (c) SIFT (d) Genetic algorithm (e) Geometric-matching CNN (f) Ours

Fig. 6. Example of result image on synthetic data.

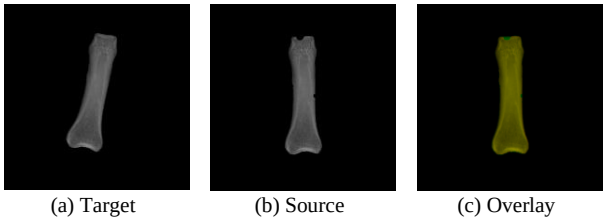
D. Results on Synthetic Data with Pseudo Lesions

Bone erosions and deformities due to rheumatoid arthritis may be present in the actual phalangeal region image pairs. Therefore, we created pseudo-lesion images with 1%, 3%, and 5% of the phalangeal region missing. Table V shows the experimental results on synthetic data with pseudo lesions in the proposed method.

TABLE V. RESULTS ON SYNTHETIC DATA WITH PSEUDO LESIONS

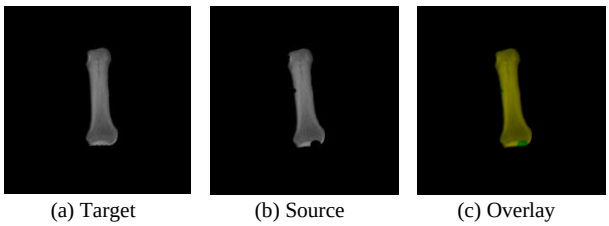
Missing rate (%)	TP (%)	FP (%)
1	99.15	1.79
3	99.11	3.78
5	99.07	5.90

For images with missing regions, the closer the FP is to the missing rate, indicates a better result. Examples of the results images at each missing rates are shown in Fig. 7 to Fig. 9. It can be seen that 1%, 3%, and 5% of missing rates of source images are correctly represented in all overlay images.



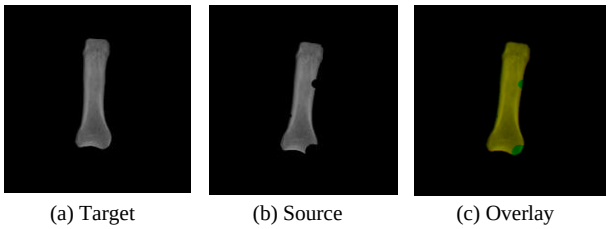
(a) Target (b) Source (c) Overlay

Fig. 7. Example with 1% missing rate.



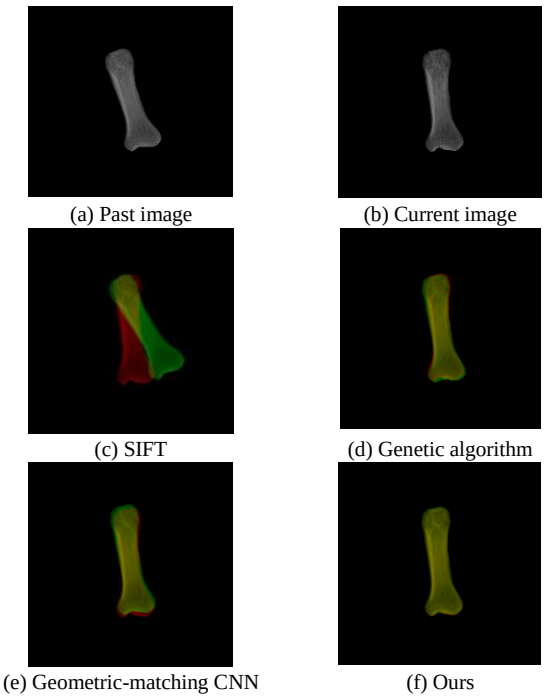
(a) Target (b) Source (c) Overlay

Fig. 8. Example with 3% missing rate.



(a) Target (b) Source (c) Overlay

Fig. 9. Example with 5% missing rate.



(a) Past image (b) Current image (c) SIFT (d) Genetic algorithm (e) Geometric-matching CNN (f) Ours

Fig. 10. Example of result image on real data.

E. Results on Real Data

Real data consisting of past and current images of the same subject were used to evaluate the registration accuracy. In this paper, 144 pairs of phalangeal region images were created by extracting the media and proximal phalanges from hand CR images of 9 cases. The comparison method is the same as in the experiment on synthetic data. Table VI shows the experimental results on

real data. However, SIFT lacked the feature points necessary to calculate the rigid transformation parameters in the one image pairs. The proposed method achieved a Dice score of 97.98%. Fig. 10 also shows an example of result images.

TABLE VI. RESULTS ON REAL DATA

Method	TP (%)	FP (%)	Dice (%)
SIFT	89.46	9.73	89.82
GA [15]	94.93	4.34	95.27
GM CNN [6]	93.58	5.58	93.96
Ours	97.61	1.61	97.98

F. Running Time

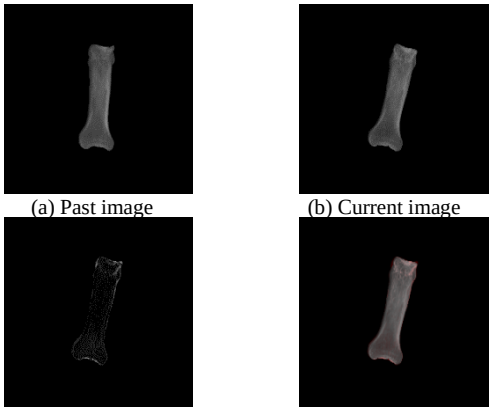
The runtime required for each method to align a pair of images is shown in Table VI. Although the SIFT and GA methods can be implemented on GPUs, they are not implementation in the library used in this experiment and the programs and algorithms need to be improved. The number of generations in GA is 200 and the number of instance-specific optimization in the proposed method is 50 iterations. Table VII shows the comparison of the running time. Proposed method has increased processing times on CPU compared to SIFT and GM CNN, however has faster than GA. Also, the use of GPU provides sufficient processing time.

TABLE VII. COMPARISON OF RUNNING TIME

Method	CPU (s)	GPU (s)
SIFT	0.031	-
GA [15]	16.33	-
GM CNN [6]	0.238	0.011
Ours	12.48	0.633

G. Subtraction Image

We generated subtraction images by aligning the real data with the proposed method. Fig. 11 shows an example of subtraction images. The subtraction image shows a tinted shadow in the phalangeal region, indicating that the temporal changes between the past and current images can be visualized.



(a) Past image (b) Current image

(c) Subtraction (d) Highlighting image

Fig. 11. Example of subtraction image.

H. Discussion

Compared to U-Net, our proposed U-ConvNeXt has 0.09% lower accuracy, but 38% less memory usage, 2.2 times faster training speed, and 1.4 times faster prediction speed. Therefore, U-ConvNeXt has a better tradeoff between accuracy and computational complexity.

In registration, the proposed method achieved the highest accuracy on synthetic and real data by applying instance-specific optimization. In addition, the proposed method performs fast initial alignment by CNN before iterative process by instance-specific optimization, so the run time per image pair is as fast as about 0.6 s on a GPU. The evaluation using synthetic data with pseudo lesions shows that the proposed method performs well in each missing ratio, indicating that the proposed method is effective even in real images with bone deformities such as bone erosion. The accuracy of the real data for the proposed method was lower than that of the synthetic data. This may be because of bone deformation or segmentation accuracy. Therefore, further improvement in segmentation accuracy is needed.

IV. CONCLUSION

In this paper, we proposed a U-ConvNeXt for automatic extraction of phalangeal regions and image registration method based on Geometric-matching CNN for detection of temporal changes which is obtained difference time series.

The temporal changes were visualized on the subtraction images generated by the proposed method. This result indicates that the temporal subtraction images generated by the proposed method may be useful in the diagnosis and evaluation of rheumatoid arthritis.

As a future work, it is necessary to verify the effectiveness of the proposed method as a diagnostic support system by applying it to a large number of rheumatoid arthritis cases. Furthermore, state-of-the-art deep learning-based medical image segmentation techniques and non-rigid image registration methods have been proposed. We will be improved to further increase accuracy of the segmentation.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Hikaru Ono conducted the research and wrote the initial manuscript and Tohru Kamiya supervised the project, provided suggestion and recommendations along the way. Takatoshi Aoki gives medical advice and revised manuscripts. All authors had approved the final version.

ACKNOWLEDGMENT

We would like to thank Prof. S. Murakami of Junshin Gakuen University for useful discussions.

REFERENCES

[1] D. Aletaha, T. Neogi, A. J. Silman *et al*., "Rheumatoid arthritis classification criteria: An American college of rheumatology/European league against rheumatism collaborative initiative," *Arthritis Rheum*, vol. 62, no. 9, pp. 2569–2581, 2010.

[2] G. R. Burmester and J. E. Pope, "Novel treatment strategies in rheumatoid arthritis," *The Lancet*, vol. 389, 10086, pp. 2338–2348, 2017.

[3] A. Kano, K. Doi, H. MacMahon *et al*., "Digital image subtraction of temporally sequential chest images for detection of interval change," *Medical Physics*, vol. 21, no. 3, pp. 453–461, 1994.

[4] S. Ichikawa, T. Kamishima, K. Sutherland *et al*., "Radiographic quantifications of joint space narrowing progression by computer-based approach using temporal subtraction in rheumatoid wrist," *The British Journal of Radiology*, vol. 89, no. 1057, 2016.

[5] S. Kajihara, S. Murakami, J. K. Tan *et al*., "Identify rheumatoid arthritis and osteoporosis from phalange CR images based on image registration and ANN," *ICIC Express Letters*, vol. 10, no. 10, pp. 2435–2440, 2016.

[6] I. Rocco, R. Arandjelovic, and J. Sivic, "Convolutional neural network architecture for geometric matching," in *Proc. Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 6148–6157.

[7] G. Balakrishnan, A. Zhao, M. R. Sabuncu *et al*., "VoxelMorph: A learning framework for deformable medical image registration," *IEEE Transactions on Medical Imaging*, vol. 38, no. 8, pp. 1788–1800, 2019.

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015, pp. 234–241.

[9] Z. Liu, H. Mao, C. Y. Wu *et al*., "A ConvNet for the 2020s," in *Proc. Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 11976–11986.

[10] D. Hendrycks and K. Gimpel, "Gaussian Error Linear Units (GELUs)," arXiv preprint, arXiv:1606.08415, 2016.

[11] S. Tachinaga, T. Ishida, H. Isoda *et al*., "Development of temporal subtraction technique on MR images of brain," *Japanese Journal of Imaging and Information Sciences in Medicine*, vol. 31, no. 3, pp. 47–53, 2014. (in Japanese)

[12] L. C. Chen, Y. Zhu, G. Papandreou *et al*., "Encoder-decoder with Atrous separable convolution for semantic image segmentation," in *Proc. European Conference on Computer Vision (ECCV)*, 2018, pp. 801–818.

[13] M. Everingham, L. van Gool, C. K. I. Williams *et al*. (2011). The PASCAL Visual Object Classes Challenge 2011 (VOC2011) Results. [Online]. Available: http://www.pascal-network.org/challenges/VOC/voc2011/workshop/index.html

[14] D. G. Lowe, "Distinctive image features form scale-invariant key points," *International Journal of Computer Vision*, vol. 60, pp. 91–110, 2004.

[15] K. Kawagoe, S. Murakami, H. Lu *et al*., "Registration of phalange region from CR images based on genetic algorithm," in *Proc. International Conference on Control, Automation and Systems (ICCAS)*, 2018, pp. 1464–1467.