Automatic detection of spine in CT image by U-Net

Mikoto Kamata[1)] Kunihiko Fukushima[2)] Hayaru Shouno[3)] Isao Hayashi[4)] and Masayuki Kikuchi[5)]

[1)]School of Computer Science, Tokyo University of Technology [2)]Fuzzy Logic Systems Institute [3)]Graduate School of Informatics and Engineering, University of Electro-Communications [4)]Faculty of Informatics, Kansai University
E-mail: mikoto.kamata@gmail.com

Abstract

In this study, a neural network called U-Net (Convolutional Networks for Biomedical Image Segmentation) [1] is used to automatically extract spinal parts from CT images of patients. This neural network has a U-shaped structure as its name, and it is possible to perform highly accurate learning with a small number of images. The U-Net is composed of convolution layers, pooling layers, and upsampling layers. In this study, unlike general U-Net, we introduced the mechanism of drop out during learning, and we use hyperbolic tangent function instead of sigmoid function at the stage of the layer of final convolution. As a result, we got whole spine pattern in the accuracy of above 99%.

1. Introduction

Recently the rise of deep learning has been activating AI field, bringing many examples of applications in various industrial fields. On the other hand, there are still many manual work as well. It is a challenge making these jobs more efficient. As a specific example, in the medical field, there is a case where a 3D mesh model of the spine is created from the CT images of the patient, but at this time, the spinal part and the non-spatial part have to be divided for each CT image. Although this mission is done by hand at present, it is very time-consuming and labor-intensive, so the task of automating this task is a problem. In this research, as a preliminary step to solve this problem, we aim to automatically extract spinal parts in CT images. Specifically, a neural network called U-Net [1] is used. As a previous study, there is one that addresses similar problems using FCN (Fully Convolutional Network) [2], but extracts only a part of the spine. This study aimed to extract the whole spine, contrary to the previous study. Since U-Net is a neural network considered for segmentation of medical images, it was expected to achieve highly accurate segmentation of the spine.

2. Analysis method

Experiments in this study were carried out using two spine data (1066 sheets of 512 × 512) published by SpineWeb [4]. This data does not contain the information about the region of the entire spine. Therefore, we created "teacher image" containing only white spine pattern on the empty (black) background using 3DSlicer which is open source software for visualization and analysis of medical images. After that, we prepared a program for learning and testing. At this time, in order to improve learning speed, normalization is performed beforehand. Specifically, following conversion was adopted for each input CT image:

$$i_{new}(x) = \frac{i(x) - i_{avr} - 3\sigma}{6\sigma} \quad (1)$$

where $i(x)$, $i_{new}(x)$, i_{avr} are the input pixel value at position x in 2D coordinate before conversion, the one after conversion, and the average of pixel values across entire image respectively, and sigma is the standard deviation of all pixels on that image.

This is somewhat different from the usual normalization method, but there are two reasons to use such a normalization method in this research. Firstly, when normalization is performed by a general method for CT images, the contrast of the input image is lowered. The second is to flexibly deal with the difference in CT value. Generally, the CT values tend to differ greatly among different CT scanners. The conventional methods cannot absorb such changes in CT values. Therefore, it is hard to handle when input images with different CT values are used for learning and testing. Contrary to the conventional methods, the proposed method can overcome such changes in CT values.

This study used 1066 CT image of two spines. Specifically, 559 CT images of the first spine are first shuffled, then divided into two subsets, 279 for learning and 280 for test, and perform learning and testing. After that, all 507 CT images of the second spine are used for testing. The batch size is 1, and the number of epochs is determined automatically; the learning process stops when learning rate is no longer improved four times. After that, correct rate and loss rate are calculated. The correct rate is obtained by averaging pixel-wise matching results (1: matched, 0: unmatched) between teacher-images and output images of U-Net, and the loss rate is the reciprocal thereof. Finally, predicted patterns of unlearned spine were supplied into 3DSlicer and 3D mesh model was generated by it. We checked the shape of obtained 3D mesh model whether it represented whole spine or not.

3. Results

Fig. 3.1 is a graph plotting the correct rate for each 1 epoch. In Fig. 3.1, the blue line graph "acc" shows the predicted result of the unlearned part of the spine whose counterpart is used for learning, and the orange line graph "val_acc" shows the prediction result of the unlearned spine.

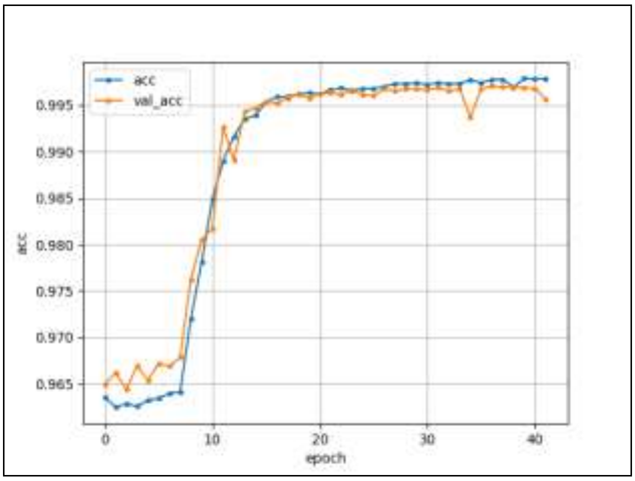

Fig 3.1: Correct percentage of spine extraction results

As shown in Fig. 3.1, even in predicting unlearned spinal cords, U-Net could learn with almost the same high precision as in predicting the learned spinal cord.

Fig. 3.3 show the result of extracting the spinal image of unlearned part, whose counterpart is used for learning, and Fig. 3.4 are the results of extracting the unlearned spine. Fig. 3.3 and Fig. 3.4 are composed of three rows and five columns, respectively. From the top, the first row is raw data, the second row is the teacher image, the third row is the prediction result.

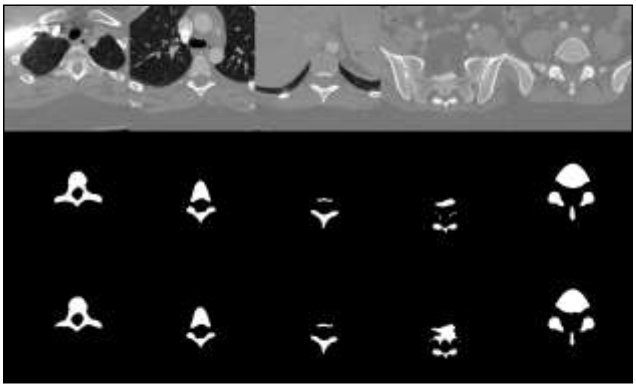
Fig 3.3: Results of extraction of unlearned part of learned spine (first row: raw data, second row: teacher image, third row: prediction result)

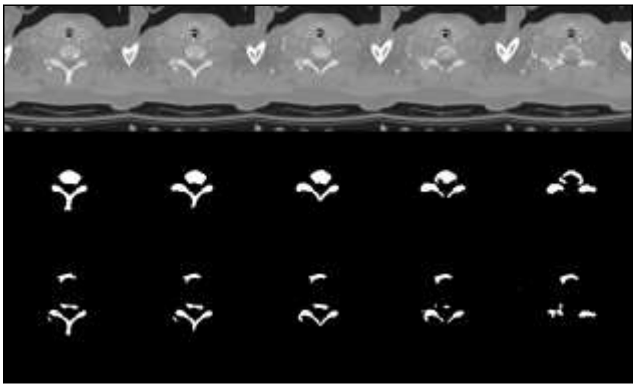
Fig 3.4: Result of extraction of unlearned spine (first row: raw data, second row: teacher image, third row: prediction result)

As shown in Fig. 3.3, we can appreciate that the network could generate almost identical approximation patterns of spine as the teacher images. Also, in Figure 3.4, we can find somewhat defective responses in the predicted image comparing to the teacher image, nevertheless there are correct responses to some extent.

Figure 3.5 shows a screenshot of 3D slicer loaded using the results of extracting unlearned spine.

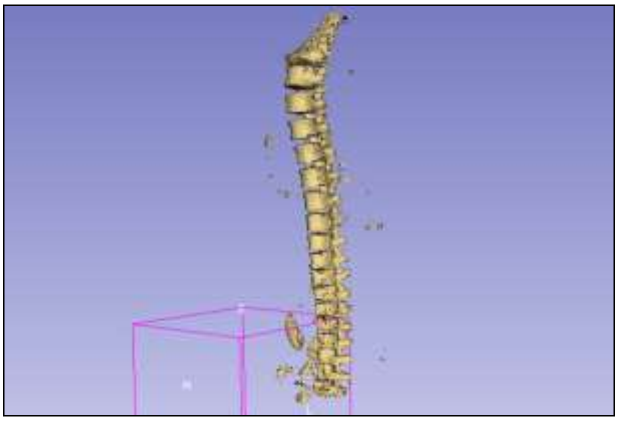

Fig 3.5: 3D mesh model generated by 3DSlicer constructed from prediction results of unlearned spine

As shown in Figure 3.5, we can see that the 3D mesh model of the spine has been constructed successfully. On the other hand, we can find some scattered factors other than the spine in the 3D mesh model.

4. Discussion

When U-Net is used as it is, overfitting is easy to occur, and accuracy of about 80% can be obtained at most. In this research, in order to improve the accuracy of learning, in addition to using dropout to prevent overfitting, the output function at the last convolution was changed to hyperbolic tangent function. Also, original normalization method for input CT images was introduced as a pre-process to clarify tissues in images. As a result, it is predicted that it was able to obtain a correct rate of 99%. Also, since the data for learning was only one vertebra, it can be thought that higher accuracy can be obtained by learning more spines. The reason why scattered factors other than the spine were extracted is because there is a tissue with a CT value close to the CT value of the spine in the CT image. It is thought that this problem can be solved by improving the accuracy of learning, and how to improve the accuracy is the future problem. It is also expected that it will be possible to increase usefulness by enabling spine classification such as adult/child, female/male, fracture/intact, and so on.

5. Conclusions

This study focused on automatic extraction of spinal parts in CT images using deep learning as a foothold for automatically generating spinal mesh models from CT images of patients. Specifically, we used a network model called U-Net, which is a U-shaped neural network aiming at segmentation of medical images. It can increase the discrimination rate with a small number of sheets. In addition, It has also the ability to overcome the weak points for conventional CNN such as identifying both the local features of the object and the overall position information on the original image or making the boundaries between the objects clear. U-Net constructed in this research is somewhat different from normal U-Net, it performs dropout to avoid overfitting, and uses hyperbolic tangent function instead of sigmoid function at the stage of final convolution. Also, normalization was done beforehand to make clear the tissues in image. As a result of experiments using 279 CT images for learning and 787 CT images for testing, both the predicted result of the learned spine and the predicted result of the unlearned spinal cord were more than 99% correct rate. Also, when 3DSlicer imported prediction results of unlearned spine, 3D mesh model of the spine could be constructed, but it was found that it extracted irrelevant factors other than the spine. This is due to the existence of other tissues having similar CT values to the spine in the CT image.

References

[1] Olaf Ronneberger, Philipp Fischer, Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", Medical Image Computing and Computer-Assisted Intervention, pp.234-241, 2015.

[2] Image classification from spinal CT images using depth learning, http://suga.sd.soft.iwate-pu.ac.jp/images/sotsuron/PDF/0312014012_20180109145155_.pdf, (accessed 2017.1.13).

[3] 3DSlicer, https://www.slicer.org/, (accessed 2017.12.1).

[4] SpineWeb, http://spineweb.digitalimaginggroup.ca/spineweb/index.php?action=home, (accessed 2017.1.1)