Master's Thesis
석사 학위논문

Multi Dimensional LSTM for 3D patch based pancreas segmentation

Jeonghwan Kim(김 정 환 金 廷 環)

Department of
Robotics Engineering

DGIST

2020

Multi Dimensional LSTM for 3D patch based pancreas segmentation

Jeonghwan Kim(김 정 환 金 廷 環)

Department of

Robotics Engineering

DGIST

2020

Multi Dimensional LSTM for 3D patch based pancreas segmentation

Advisor: Professor Sanghyun Park
Co-advisor: Professor DongHye Ye

by

Jeonghwan Kim
Department of Robotics Engineering

DGIST

A thesis submitted to the faculty of DGIST in partial fulfillment of the requirements for the degree of Master of Science in the Department of Robotics Engineering. The study was conducted in accordance with Code of Research Ethics[1]

07. 06. 2020

Approved by

Professor Sanghyun Park   (signature)
(Advisor)

Professor DongHye Ye   (signature)
(Co-Advisor)

---

[1] Declaration of Ethical Conduct in Research: I, as a graduate student of DGIST, hereby declare that I have not committed any acts that may damage the credibility of my research. These include, but are not limited to: falsification, thesis written by someone else, distortion of research findings or plagiarism. I affirm that my thesis contains honest conclusions based on my own careful research under the guidance of my thesis advisor.

Multi Dimensional LSTM for 3D patch based pancreas segmentation

Jeonghwan Kim

Accepted in partial fulfillment of the requirements for the degree of Master of Science.

05. 28. 2020

Head of Committee   Prof. SangHyun Park (signature)

Committee Member   Prof. DongHye Ye   (signature)

Committee Member   Prof. Okkyun Lee   (signature)

ABSTRACT

Compute Tomography (CT) imaging is mostly used to diagnose abdominal disease. Among many organs in abdominal CT, the pancreas is one of important regions to diagnose diabetes, pancreatic cancer, and pancreatitis. Thus, it is important to fin the pancreas area and quantitatively analyze the present of disease. However, it is difficult to segment the pancreas area due to the size and shape variation depending on patients and ambiguous boundaries with surrounding organs. A lot of methods have been proposed for the automatic pancreas segmentation, but the accurate segmentation is still challenging. Recently, deep learning methods have been achieved better performance than conventional machine learning based segmentation methods for many applications. Thus, in this paper, we propose a new pancreas segmentation method using deep learning.

Applying a deep learning algorithm to the 3D CT image segmentation is non-trivial due to a memory limitation. Several deep learning methods address this issue by dividing a 3D medical image into small 3D patches and then applying 3D Convolutional Neural Networks (CNN) on the patches. However, anatomical information cannot be considered in this way. To address these issues, in this paper, we propose a Multi-Dimensional Long Short-Term Memory (MDLSTM) method which can consider the anatomical information using a sequential learning scheme. MDLSTM is compose of a 3D CNN that performs the patch-wise segmentation on local regions and a Long Short-Term Memory (LSTM) that propagates information of adjacent patches. For evaluation, the NIH CT dataset was used to train the MDLSTM model, and Dice-Sorensen Coefficient (DSC), Precision, and Recall were used to compare performance with other 2D and 3D CNN methods. When using the proposed method, we can see the DSC and precision were increased by 2%, 16%, respectively compared to the conventional CNN method.

Keywords: Computed Tomography, Pancreas segmentation, Convolutional Neural Network, Long Short-Term Memory

List of Contents

List of Tables

List of Figures

I. INTRODUCTION

1.1 Introduction

Organ segmentation is a fundamental step in analyzing medical image. Among them, the pancreas provides import information to diagnose diabetes, pancreatic cancer, and pancreatitis. For this reason, it is important to localize the pancreas, but pancreas segmentation is a difficult task because the size and shape of the pancreas is more diverse than other organs and difficult to classify with surrounding organs. Radiologists' manual segmenting of the pancreas is a time-consuming, tedious, and laborious task. In addition, there are differences in the results made by radiologists by hand. So, it is required to automatic segmentation algorithm. As machine learning technology has developed, methods were proposed to extract handcraft features using pre-knowledge and then perform pancreas segmentation [10], [11]. Recently, as hardware performance and deep learning technology have been developed, many researches using deep learning have been conducted, and better performance has been obtained in various applications than previous methods. Among several deep learning algorithms, the convolutional neural network shows good performance for the problem of segmentation, and it can be used in pancreas segmentation. However, despite the good performance of the convolutional neural network, pancreas segmentation is still challenging.

The reasons it is difficult to the pancreas segmentation within a 3D CT image is too large. Therefore, it is necessary to change the size that can be learned for deep learning through preprocessing. Two methods are typically used as a preprocessing. It is extracting 2D slices and small 3D images. When segmentation is performed with 2D images, 3D information is lost, but anatomical information can be used. When segmentation is performed with 3D images, there is a method to reduce the size of the original image or extract patches from the original image for deep learning. There is an advantage of using 3D information, but there is a disadvantage of losing a lot of information while reducing the size, and not using anatomical information when performing segmentation while using patches. For the 2D image segmentation methods, FCN [13], U-Net [5], and DeepLab v3[14] were proposed. For the 3D image segmentation methods, 3D U-Net [6] was proposed which is an extended version of 2D U-Net.

In this paper, we propose a Multi-Dimensional Long Short-Term Memory (MDLSTM) to solve the problem when using 3D U-Net. MDSLTM is composed of 3D U-Net and Convolutional LSTM (CLSTM). The principle of operation is to use the results from 3D U-Net as input to MDLSTM to learn the relationship with surrounding patches. For the CT datasets used in the experiment, the NIH pancreas dataset was used. As a result of the proposed

method, it was confirmed that the DSC was improved by 2% and the Precision was improved by 16% compared to the conventional 3D U-Net.
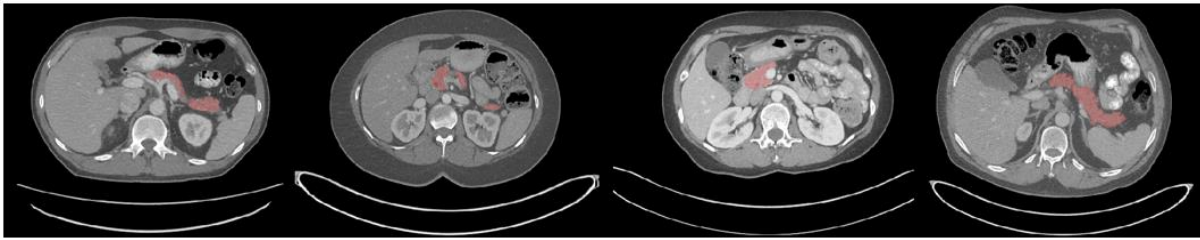

Figure 1. Axial view CT images from different subjects.

1.2 Related works

The result of the pancreas segmentation for the two basic methods (2D U-Net, 3D U-Net) shows significantly lower performance than those of other organs. Deep learning model is difficult to learn how to find pancreas area because the pancreas has different sizes and shapes for each patient. Recently, many previous studies have been conducted to solve this problem.

H. R. Roth et al. [1] proposed a binary classification method for voxel using small images in axial, coronal, and sagittal spaces that share a point in a 3D image. It is a method that performs segmenting without using the Fully convolutional Network (FCN), the performance is not good, but it uses axial, coronal, and sagittal spatial information and suggests the coarse-to-fine stage method. J. Cai et al. [2] performs pancreas segmentation using Convolutional Long Short-Term Memory (CLSTM) that learns information between the result of 2D U-Net, and refinement of segmentation using CLSTM. It helps to reduce the disadvantage of not using 3D spatial information. However, if a wrong prediction is generated in the previous slice, a problem occurs that incorrectly predicts until the next slice. Also, since only the information of the previous images is used, the information of the next images cannot be used. Hao Li et al. [3] to use 3D information, the number of seven slices are stacked as channels and used for 2D U-Net. Besides, they proposed a structure that learns continuous information between each 2D image using Bidirectional LSTM (BDLSTM). The structure of using 7 slices as channels and using it for 2D U-Net does not predict for each image. It shows the results considering the overall information of the 7 slices. Since the appearance between the stacked slices is not similar, detailed consideration of each image is not possible, and the performance is reduced.

Zhuotun Zhu, et al. [8] used a coarse-to-fine method with using 3D patches. In the coarse stage, input patches are randomly extracted from the entire image to find the approximate location of the pancreas. In the fine stage, bounding boxes are created using the results of the coarse stage, and patches are randomly extracted from the

bounding boxes for 3D U-Net. Because the unnecessary patch for learning is not used in fine stage, it is possible to obtain a better segmentation result of the coarse stage, but there is s disadvantage in that anatomical information cannot be used. Ningning Zhao, et al. [9] also using the coarse-to-fine method, but the entire image of the abdominal CT is down-sampled and used as the input of the coarse stage. A bounding box is generated based on the results from the coarse stage, and then up sampled to be used as input for the fine stage to obtain the final result. However, when learning a 3D U-Net in the coarse stage, there is a problem in that the pancreas is too small in the original image, and the number of data is not enough to learn well.

The previous studies have a common point that the research was conducted to learn anatomical information and 3D spatial information in the abdominal CT image. It attempted to supplement for the inability to use 3D spatial information in 2D U-Net, and to supplement the inability to use anatomical information in 3D U-Net. However, a method of using CLSTM in 2D U-Net does not use all 3D spatial information. This is because the characteristics of the other axis cannot be considered because information of the 2D image extracted based on one axis is propagated. In addition, using a coarse-to-fine step in 3D U-Net does not communicate with each other in the coarse stage and fine stage, the anatomical information of the coarse stage does not propagate to the fine stage. In this paper, to solve the inability to use anatomical information in 3D U-net, LSTM method is used to improve performance by delivering information about surrounding patches.

II. METHOD

This section introduces the 3D patch-based pancreas segmentation using MDCLSTM. The overall structure of MDCLSTM can be seen in Figure 2. The MDCLSTM method consists of two stage. The first stage extracts small patches from the original image and create a feature map of each patch using 3D U-Net. The second stage is to deliver feature maps generated in 3D U-Net to adjacent patches using MDCLSTM and perform pancreas segmentation. Using the proposed method, anatomical information can be used by viewing a wider area compared using only 3D U-Net while performing pancreas segmentation.
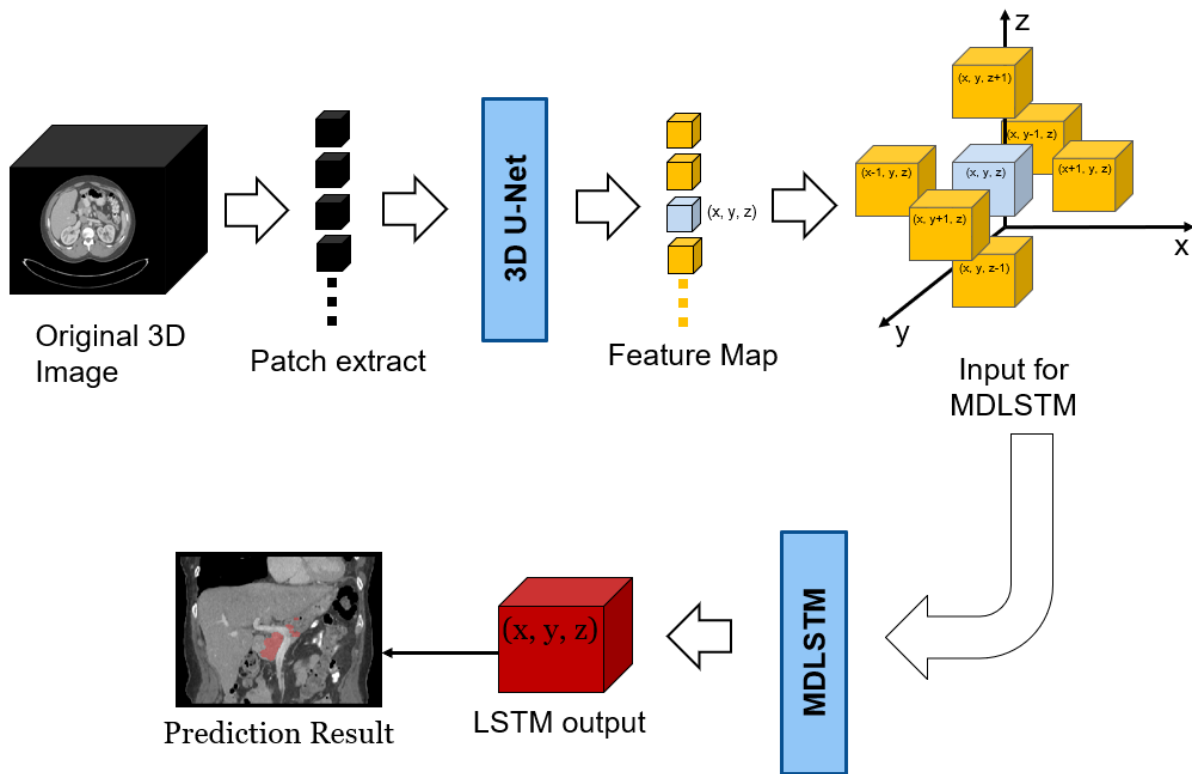

Figure 2. Overall structure of MDCLSTM

2.1 U-Net

U-Net consists of encoding and decoding stages. In the encoding stage, there are CNN composed of 3x3 kernels, Relu function, and 2x2 max pooling function. U-Net can be separated into the Encoding and Decoding stages. In the Encoding stage, there are 3x3 kernel CNN, Relu functions, and 2x2 max pooling functions. In this study, a batch normalization technique is additionally used to improve learning stability and acceleration. In the Decoding stage, there are Transposed CNN which makes the larger image using a 3x3 kernel (Up-convolution), CNN using a 3x3 kernel, Relu function. In the final stage, the final segmentation result is obtained through CNN using a 1x1 kernel. In addition, the feature map generated for each depth in the Encoding stage can be propagated to each depth in the Decoding stage, thereby reducing information lost in the Encoding stage (Figure 3). In this study, we

compared the performance of the proposed method using 2D U-Net and 3D U-net. The comparison method used DSC, Precision, and Recall.
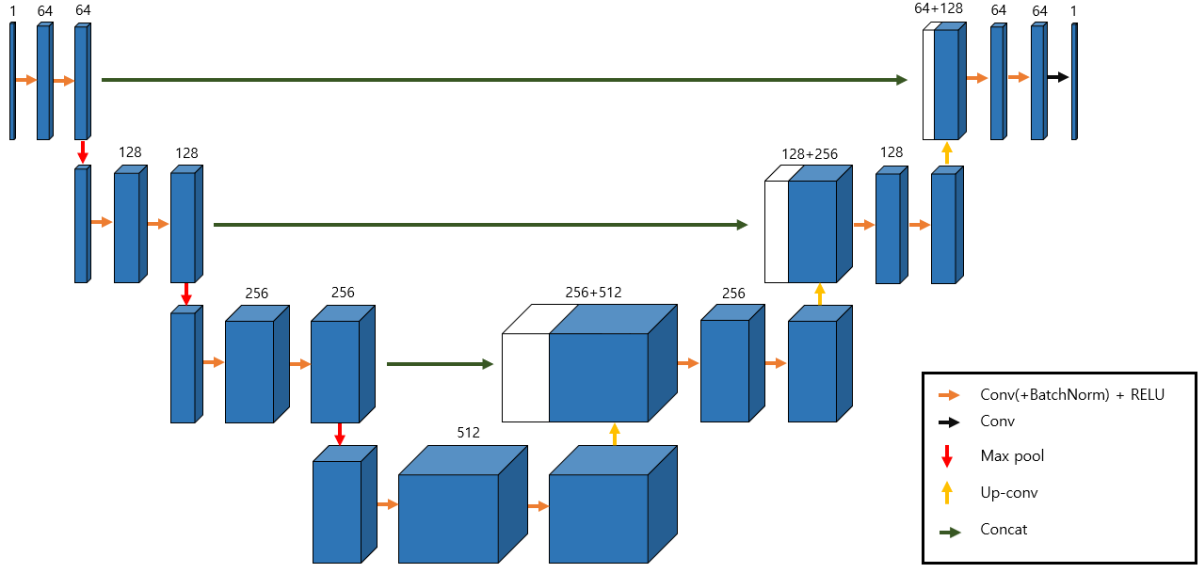

Figure 3. U-Net architecture

2.2 Convolutional Long Short-Term Memory (CLSTM)

The Neural Network (NN) structure has a disadvantage in that the size of input information must be a constant value. To alleviate this problem, Recurrent Neural Network (RNN) method is proposed. RNN can be used without restrictions on the size of input data and it is suitable to use for process continuous data. Long Short-Term Memory (LSTM) is an improved method of RNN that can learn long-term and short-term information. LSTM is composed of input gate, forget gate, and output gate. The output of LSTM is hidden state and cell state. To use continuous information in 3D images, the Convolutional LSTM (CLSTM) method has been proposed. By using this technique, input data from LSTM can be processed in images.

$$i_t = \sigma(W_{xi} * x_t + W_{hi} * h_{t-1} + b_i) \tag{1}$$

$$f_t = \sigma(W_{xf} * x_t + W_{hf} * h_{t-1} + b_f) \tag{2}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot tanh(W_{xc} * x_t + W_{hi} * h_{t-1} + b_i) \tag{3}$$

$$o_t = \sigma(W_{xo} * x_t + W_{ho} * h_{t-1} + b_o) \tag{4}$$

$$h_t = o_t \odot tanh(c_t) \tag{5}$$

$\sigma()$ and tanh() mean sigmoid function and hyperbolic tangent function, respectively. * operation means CNN and $\odot$ means element-wise product. i_t f_t, o_t is input gate, forget gate, and output gate at time 't' respectively. x_t means input data, and W, are parameter of CLSTM. h_t and c_t means hidden state and cell state, respectively. x_t passes through the CNN and is used as the input of each gate along with the hidden state

of the previous time. At the forget gate, the decision as to whether to use previous information is calculated using hidden state and input data (2), (3). The input gate decides how to add the calculated value from the hidden state and input data to cell state (1), (3). The output gate calculates how much the cell state value will be transferred to the hidden state (5).
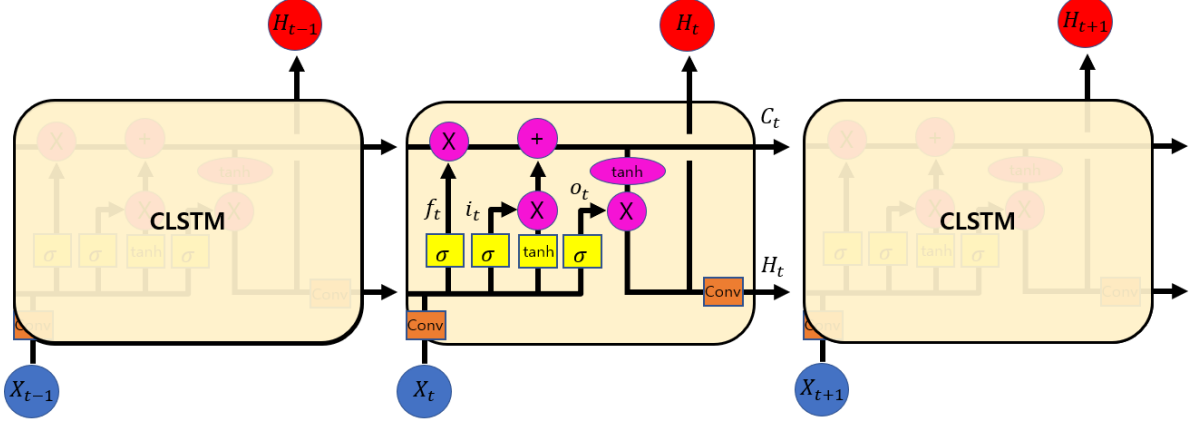

Figure 4. CLSTM architecture

2.3 Multi-Dimensional Convolutional Long Short-Term Memory (MDCLSTM)

The 3D patch -based pancreas segmentation method has the disadvantage of not knowing the overall information of the image. In the case of an abdominal CT image, cannot use the anatomical information is the major cause of performance degradation because the overall organ structure is similar for each patient. To solve this problem, we proposed MDCLSTM method that can perform the 3D patch-based pancreas segmentation by using the information of surrounding patches. 3D patch cannot apply CLSTM method to set one axis like 2D slice. Because if set based on one axis and apply CLSTM, the information of the other axis cannot be used. Therefore, this study intends to use a network that can use the information on various directions in 3D using MDCLSTM (Figure 5). In this study, MDCLSTM using the information in 6 directions, and the total number of 6 CLSTM are used. The composition of MDCLSTM is as follows.

$$h_t^n = H^n(W_{hh^n} * h_{t-1}^n + W_{ch^n} * c_{t-1}^n + W_{\hat{y}h^n} * \hat{y}_t + b_{h^n}) \tag{6}$$

$$\tilde{y} = \sum_{n=1}^{6}(W_{h^n\hat{y}} * h_t^n) + b_{\hat{y}} \tag{7}$$

$\hat{y}_t$ In this study, MDCLSTM using the information in 6 directions, and the total number of 6 CLSTM are used. The composition of MDCLSTM is as follows. The composition of MDCLSTM is as follows means the result of 3D U-Net, and H^n means hidden state from n direction CLSTM (6). $W_{\hat{Q}h^n}$ and b_{h^n} mean learning parameters of CLSTM, and MDCLSTM uses the CNN for CLSTM results in 6 direction to generate the final $\tilde{y}$ (6). $W_{h^n\hat{y}}$ and $b_{\hat{y}}$ mean learning parameters of the CNN to obtain the result (7).
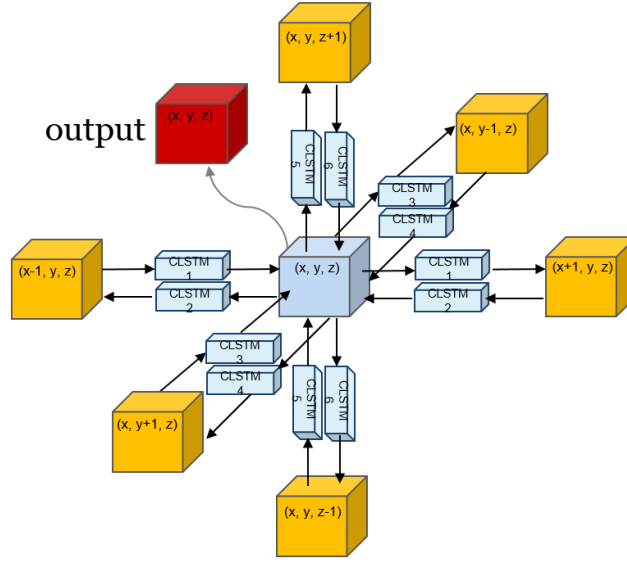
Figure 5. MDCLSTM architecture

2.4 Forward and Backward propagation for the MDCLSTM

The entire network structure of MDCLSTM is differentiated. Therefore, end-to-end learning is possible, and the entire network can be trained at once using forward-propagation and backward-propagation. Binary Cross Entropy (BCE) loss function and Dice-Sorensen Coefficient (DSC loss function were used as the loss function to optimize the network.

$$L_1 = -\frac{1}{N}\sum_{i=1}^{N} y_i log(h(x_i;\theta)) + (1 - y_i)log(1 - h(x_i;\theta)) \tag{8}$$

$$L_2 = 1 - \frac{2 \times |Y \cap \tilde{Y}|}{|Y| + |\tilde{Y}|} \tag{9}$$

$$\text{L} = \lambda L_1 + (1 - \lambda)L_2 \tag{10}$$

Equation (8) means BCE loss function, equation (9) mean DSC loss function. In equation (10), the final loss value is calculated by adjusting the ratio of two loss function values using the λ value. MDCLSTM learning parameters are learned to minimize the loss value.

III. Experimental result

3.1 Data set

To verify the performance of the proposed method, 82 abdominal CT images from The National Institutes of Health Clinical Center were used. The size of Ct images is $512 \times 512 \times D, D \in [181, 466]$ and the thickness of the images is $1.5 - 2.5$ mm. The experiment consisted of 4-fold cross-validation and 82 patient data were randomly distributed.

3.2 Implementation details

Pytorch was used to implement the proposed network in this paper, and learning was performed on NVIDIA TITAN Xp (12 GB memory). The optimization algorithm used the Adam optimization method, and the learning rate was set to an initial value of 1e-3 and multiplied by 1e-1 for every 15 epochs out of a total of 40 epochs to gradually decrease the ratio. DSC loss function and BCCE loss function were used as the loss function, and when experimented while adjusting the λ value in equation (10), the value of $\lambda = 0.5$ showed the best performance.

The 2D U-Net synthesized the results of the axial, coronal, and sagittal axes to derive the final result, and the size of the input slice was $512 \times 512, 256 \times 192, 192 \times 192$ respectively. When learning 3D U-Net, 32x32x32 patch was used, and the result of each patch was placed in the proper place on the original CT to get the result. The 3D U-Net used in MDCLSTM uses the same structure as the comparison method, and the size of the input patch is set the same. When performing pancreas segmentation of a specific patch in MDCLSTM, create a feature map for each patch using 3D U-Nets, and use the feature maps as inputs of 6 CLSTM.

3.3 Evaluation metrics

To verify the performance of the proposed method, DSC, Precision, and Recall were used. DSC is a method of calculating the similarity between the actual pancreatic region and the created pancreatic region by the trained model (11). Precision is the ratio of the actual pancreatic region to the created pancreatic region by the trained model (12). Recall is the ratio of the created pancreatic region by the trained model to the actual pancreatic region (13). In the below equation, TP, FP, and FN mean True Positive, False Positive, and False Negative, respectively.

$$\text{DSC} = \frac{2 \times |Y \cap \tilde{Y}|}{|Y| + |\tilde{Y}|} \tag{11}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{12}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{13}$$

3.4 Qualitative and Quantitative Analysis

In the abdominal CT image, the intensity value varies from each patient and often varies within the same organ. Because of the characteristics of these abdominal CT images, it is important to use anatomical information to perform pancreas segmentation. 2D U-Net views the entire image on the axial, coronal, and sagittal axes, it can use the anatomical information to perform pancreas segmentation. When using anatomical information, CNN can learn about the overall position and shape of organs. Therefore, the error rate of predicting a non-pancreas organ as a pancreas is small and we can see the precision value is high. 3D U-Net has the advantage of using spatial information in the abdominal CT image compared to 2D U-Net, but it cannot use anatomical information because it performs for a single patch. If a patch of a non-pancreas organ has characteristics like the pancreas, it is often mistaken, and the precision value is significantly lower than that of the 2D U-Net. Using MDCLSTM, it is possible to reduce the error rate of FP using anatomical information while propagating information of adjacent patches, and the precision, DSC value increases by 16%, 2%, respectively (Table 1).

	DSC (%)	Precision (%)	Recall (%)
2D U-net	79.40(±7.98)	85.43(±6.07)	75.96(±7.98)
2D U-net + BDCLSTM	**79.99(±7.92)**	85.58(±5.91)	76.45(±12.63)
3D U-net	73.04(±8.47)	69.23(±10.6)	**78.85(±10.45)**
3D U-net + MDCLSTM	75.29(±7.89)	**85.86(±5.87)**	68.31(±11.36)

Table 1. Segmentation result of the previous segmentation method and purposed segmentation method.

Figure 6 shows the quantitative results of the comparison methods and proposed method. In the case of 2D U-Net because it uses anatomical information, there are low errors that predict non-pancreas organs as the pancreas. However, since only 2D information is used, a problem arises that it cannot perform well pancreas segmentation for real pancreas area where can predict well using spatial information. In the case of 3D U-Net because it does not use a wide area of anatomical information, it can be seen from Figure 6 that it often makes false predictions about the non-pancreas area. When using MDCLSTM method, we can see disappear the cases of false predictions, and it is also solved if it is not predicted as a pancreas in the middle of the predicted pancreas
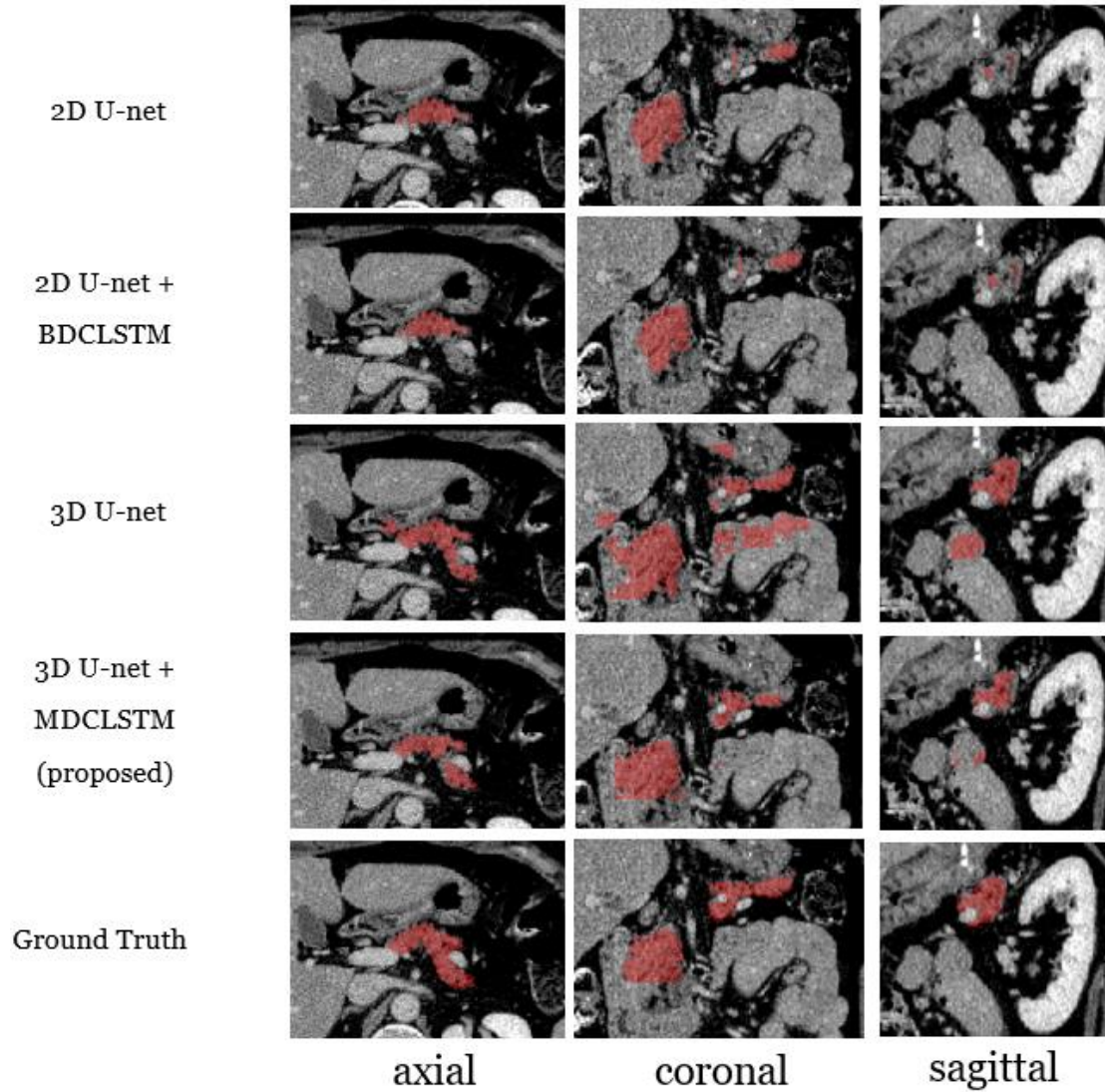
Figure 6. Compare proposed method and previous methods.

Table 2 shows the results of 3D U-Net with different sizes of patches used as input. In 3D U-Net, if the patch size were changed from 32 to 48, the performance would be better, but in MDCLSTM, the performance would be lower when the patch size was changed from 32 to 48. As the size of the patch was increased, the amount of information propagated was excessively increased, and the performance was deteriorated. In addition, further research on the optimal information propagation method is needed. We can see the DSC value is the best when the size of the input patch is 48 in 3D U-Net, but the precision value is best when MDCLSTM is added to 3D U-Net. High precision means there is a false-positive, and it can reduce the leading wrong information to the radiologist.

	DSC (%)	Precision (%)	Recall (%)
3D U-net Patch:32	73.04(±8.47)	69.23(±10.6)	78.85(±10.45)
3D U-net Patch:48	**80.97(±6.48)**	80.90(±6.85)	**82.04(±9.75)**
3D U-net + MDCLSTM (Patch:32)	75.29(±7.89)	**85.86(±5.87)**	68.31(±11.36)
3D U-net + MDCLSTM (Patch:48)	67.60(±12.83)	78.99(±11.27)	61.19(±16.03)

Table 2. Segmentation result of each patch size.

It shows the best performance when using a patch size of 48 in 3D U-Net, but over-segmentation occurs compare with the patch size is used as 32 in MDCLSTM (Figure 7). When using a patch size of 48 in 3D U-Net, pancreas segmentation is performed while looking at a larger area than using a patch size of 32, which greatly reduces the misprediction of non-pancreas areas, but the problem of predicting non-pancreas areas as pancreas remains.
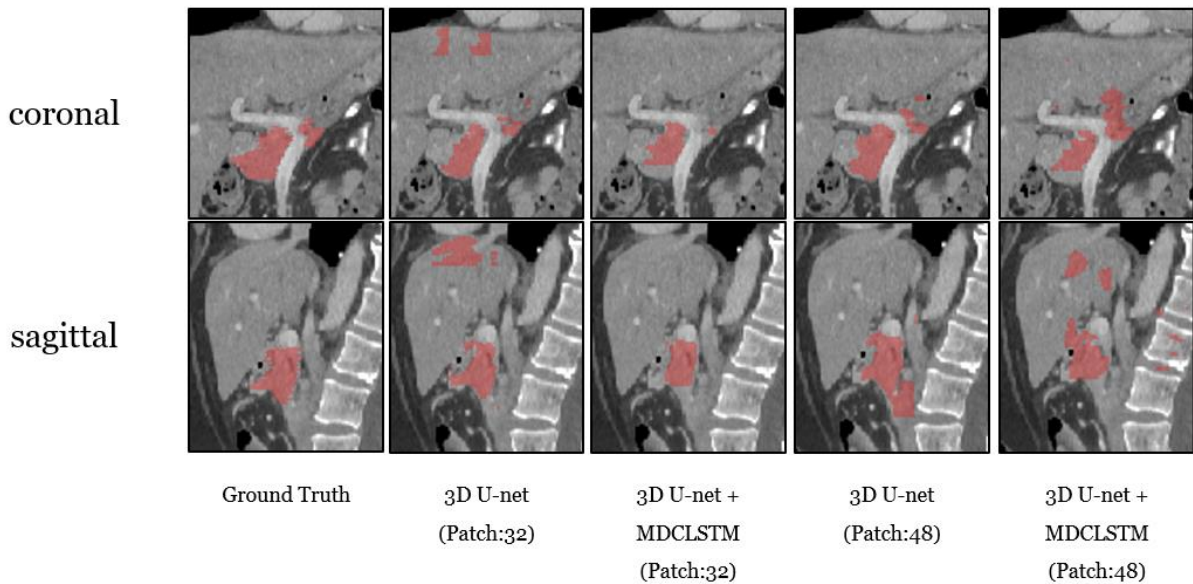

Figure 7. Result of each patch size.

VI. Conclusion

The pancreas is difficult to perform segmentation because size and shape are varies compared to other organs. To perform pancreas segmentation well, it is important to use spatial and anatomical information. However, existing methods do not use this information well, so new method is needed. In this paper, we proposed a method for performing pancreas segmentation using MDCLSTM. When using the proposed method, the problem of not using anatomical information in the 3D U-Net is solved through information propagate between patches. Also, when performing the pancreas segmentation using 3D U-Net, it is possible to reduce the problem of predicting the pancreas in a non-pancreas area and reduce error that giving false information to radiologists. Therefore, the proposed method is suitable for medical images and can be easily applied to other 3D images.

References

[1] Holger R. Roth et al. "DeepOrgan:Multi-level Deep Convolutional Networks for Automated Pancreas Segmentation" *Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2015*, pp. 556-564.

[2] Jinzheng Cai et al. "Improving Deep Pancreas Segmentation in CT and MRI Images via Recurrent Neural Contextual Learning and Direct Loss Function" *Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2017*, pp. 674-682.

[3] Hao Li, Jun Li, et al. "MDS-Net: A Model-Driven Stack-Based Fully Convolutional Network for Pancreas Segmentation" *arXiv:1903.00832, 2019*

[4] Ningning Zhao et al. "Fully Automated Pancreas Segmentation with Two-stage 3D convolutional Neural Networks" *Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2019*, pp. 674-682.

[5] Olaf Ronneberger, Philipp Fischer, et al. "U-Net: Convolutional Networks for Biomedical Image Segmentation" *Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2015*, pp.234-241.

[6] Ozgun Cicek, Ahmed Abdulkadir, et al. "3D U-Net: Learning Dense Volumetric Segmentation from Sparse annotation" *Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2016*, pp.424-432.

[7] Xingjian Shi, Zhourong Chen, et al. "Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting" *Neural Information Processing Systems (NIPS) 2015*

[8] Zhuoton Zhu, Yingda Xia, et al. "A 3D Coarse-to-Fine Framework for Volumetric Medical Image Segmentation" *arXiv 2017*.

[9] Ningning Zhao, Nuo Tong, et al. "Fully Automated Pancreas Segmentation with Two-stage 3D convolutional Neural Networks", *Medical Image Computing and Computer Assisted Intervention (MICCAI) 2019*, pp. 201-209

[10] Marius Erdt, Matthias Kirschner, et al. "Automatic pancreas segmentation in constant enhanced CT data using spatial anatomy and texture descriptors" *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, Chicago, IL, 2011, pp. 2076-2082

[11] Akinobu Shimizu, Tatsuya Kimoto, et al. "Automated Pancreas Segmentation From Three-Dimensional Constrast-Enhanced Computed Tomography" *International Journal of computer assisted radiology and surgery*, 5(1), 2010, pp. 85-98

[12] Alex Graves, Santiago Fernandez, et al. "Multi-Dimensional Recurrent Neural Networks", *Computing Research Repository (CORR).* 2007, pp. 549-558

[13] Jonathan Long, Evan Shelhamer, et al. "Fully convolutional networks for semantic segmentation" *2015 IEEE Conference on Computer Vision and pattern Recognition (CVPR),* Boston, MA, 2015,

pp.3431-3440

[14] Liang-Chieh Chen, George Papandreou, et al. "Rethinking Atrous Convolution for Semantic Image Segmentation" *arXiv:1706.05587,* 2017

요 약 문

다차원 장기-단기 기억장치를 이용한 3차원 패치기반 췌장 영역화

복부 컴퓨터 단층 촬영 영상은 영상의학자들이 환자를 진단할 때 많이 사용하는 자료이다. 그중 췌장은 당뇨병, 췌장암, 췌장염에 대한 중요한 정보를 가지고 있기 때문에 췌장 영역을 찾아 어떤 질병이 있는지 확인하는 일은 중요하다. 하지만 췌장은 크기와 모양이 환자마다 다양하며 주변 장기와의 분류가 쉽지 않다. 컴퓨터를 사용한 췌장 자동 영역화에 대해 많은 연구들이 나왔지만 여전히 높은 정확도를 기대하기는 어려운 상황이다. 최근 딥러닝을 사용한 연구가 많이 진행되고 있으며 이전의 머신 러닝기법보다 좋은 성능을 보여준다. 특히 합성곱 신경망을 사용한 영상처리 분야에서 좋은 성능을 보여주고 있으며 본 논문에서도 합성곱 신경망을 사용한 췌장 영역화 기법을 제안한다. 3차원 CT 영상은 딥러닝 학습에 사용하기엔 너무 크기 때문에 전처리를 통해 학습에 사용할 수 있는 작은 영상들로 추출을 해야 한다. 3차원 CT 영상을 작은 패치로 추출하여 딥러닝에 사용하면 패치 내에서 세부적인 공간적 정보를 사용하면서 췌장 영역화가 수행되는 장점이 있지만 주변 정보들을 알지 못하기 때문에 해부학적 정보를 사용하지 못하는 문제가 있다. 본 논문에서는 3차원 패치를 사용한 췌장 영역화에서 나타나는 문제를 보완하기 위해 3차원 합성곱 신경망과, 장기-단기 기억장치 네트워크를 사용하여 췌장을 찾아내는 기법인 3차원 패치 기반 다차원 장기-단기 기억장치 네트워크를 제안한다. 제안하는 기법은 3차원 합성곱 신경망을 기반으로 패치에서의 췌장영역을 찾아내며, 장기-단기 기억장치 네트워크를 사용하여 주변 패치들의 정보를 습득해 해부학적 정보를 학습할 수 있다. 제안하는 모델을 학습하기 위해 NIH에서 제공해주는 복부 컴퓨터 단층 촬영 데이터를 사용했으며, 기존의 기법들과 성능 비교를 위해 Dice-Sorensen Coefficient(DSC), Precision, Recall을 사용했다. 제안하는 기법을 사용하면 기존의 기법보다 DSC는 2%, precision은 16% 상향했음을 확인할 수 있었다.

핵심어: 컴퓨터 단층 촬영, 췌장 영역화, 합성곱 신경망, 다차원 장기-단기 기억장치