공학석사 학위논문

Effective Algorithm for Left Ventricle Segmentation in MRI

심장 MRI에서 좌심실 세분화를 위한 효과적인 알고리즘

2017 년 12 월

서울대학교 대학원

전기·컴퓨터공학부

이일규

Effective Algorithm for Left Ventricle Segmentation in MRI

지도교수 문병로

이 논문을 공학석사 학위논문으로 제출함
2017 년 12 월

서울대학교 대학원
전기·컴퓨터공학부
이일규

이일규의 석사 학위논문을 인준함
2017 년 12 월

위 원 장 ______ 박 근 수 ______ (인)

부위원장 ______ 문 병 로 ______ (인)

위 원 ______ 김 선 ______ (인)

Abstract

Effective Algorithm for Left Ventricle Segmentation in MRI

Ilkyu Lee
Department of Computer Science & Engineering
College of Engineering
The Graduate School
Seoul National University

The short-axis left ventricle segmentation of Cine MRI is a representative imaging analysis used as a medical care. It is difficult to obtain the same segmentation results when performing image analysis with a human hand. Time and effort are being consumed due to different segmentation results. The algorithm consists of two deep learning models proposed in this paper, it provides saving time and effort and also obtain same segmentation results always. The first model uses the selective search to detect the region of interest from irrespective size of image and obtain the center point of left ventricle by deep learning. The second model consists of applying coordinate transformation to the image and finding the boundary of the endocardium and epicardium by deep learning. The number of patients used was 24, and totally 194 slices of Cine MRI were used.

Among them, 19 were used for training and 5 were used for testing. deep learning for short-axis left ventricle segmentation and other algorithms were used to solve the problem, based on the analysis of the experimental results, we identify the problems and show the possibility.

keywords : Image processing, left ventricle segmentation, artificial neural network, cardiac magnetic resonance imaging
Student Number : 2016-21225

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Before classifying object in image, region of object in image should be specified. The same process has to be done in medical image also. First, related work about the image-processing technology, to generate possible object locations, using the diverse search and using a variety of complementary image portioning such as selective search [1]. Using region proposal networks instead of selective search was makes their algorithm even faster [2]. Even, extends method by adding detects objects in an image while simultaneously generating a high-quality segmentation mask for each instance [3]. As artificial intelligence evolves with image processing, more attempts are being made to apply the techniques to medical image processing. To estimate the frame-wise regional wall thicknesses from cardiac MR sequence, they build a new architecture of network with recurrent neural network [4]. A single neural network architecture, trained end-to-end, can deliver a fully-automated and accurate segmentation of the left ventricle using a stack of MR short-axis images [5]. Also using knowledge of the filed, origin problem is projected into other problem. Mapping the pixels from Cartesian to polar coordinate, and using multiple seed, region-growing method to calculate boundary of epicardium [6]. Conventional image processing methods have been proposed to classify various classes and they are a vast model to be used for medical image processing.

This paper introduces a system that can simplify models using domain knowledge for processing medical images. We introduce selective search which proposes candidates of region, and neural network which solves numerical problem in Chapter 2. Chapter 3 showing the details of the proposed system. Proposed method is composed with two model. First model computes center point of endocardium in the proposed region of interest(ROI) and second model searches the boundary through preprocessing on image from the center point of the previous model. Chapter 4 compares the performance with the result of the various model and other models. Finally, we will discuss the limitations and possibilities of future work related to them. The proposed algorithm combines the technique of image processing with domain knowledge of medical image and compares it with various models. It shows that the improvement of model structure and the combination of domain knowledge are effective for problem solving.

Chapter 2

Preliminaries

2.1 Selective search [1]

Selective search is a heuristic algorithm that combines exhaustive search and segmentation in an image. Three considerations in designing a selective search are:

Capture scale regardless of size using hierarchical structure
Combination of various strategies
Avoid bottlenecks in other processes with fast operations

The algorithm requires create initial regions, there are various way to acquire initial regions. We have created a starting area using graph-based method [7]. The Greedy algorithm is used to extend the regions. Similarity between all regions is calculated, and the process of merging two different regions which have highest similarity. Repeating same process until every region merge as one. The general method is detailed in Algorithm 1.

Algorithm 1: Hierarchical Grouping Algorithm

Input: (colour) image

Output: Set of object location hypotheses L

Obtain initial regions R = $\{r_1, r_2, \cdots, r_n\}$ using [7]

Initialize similarity set $S = \varnothing$

foreach Neighbouring region pair $(r_{i,} r_j)$ **do**

  Calculate similarity $s(r_{i,} r_j)$

  $S = S \bigcup s(r_i, r_j)$

while $S \neq \varnothing$ **do**

  Get highest similarity $s(r_{i,} r_j) = \max(S)$

  Merge corresponding regions $r_t = r_i \bigcup r_j$

  Remove similarities regarding $r_i : S = S \setminus s(r_{i,} r_*)$

  Remove similarities regarding $r_j : S = S \setminus s(r_{*,} r_j)$

  Calculate similarity set S_t between r_t and its neighbours

  $S = S \bigcup S_t$

  $R = R \bigcup R_t$

Extract object location boxes L from all regions in R

The similarity used in the algorithm consists of four functions.

Color similarity

Each color channel is normalized using a histogram. Each Ri can obtain a color histogram $C_i = \{c_i^1, \cdots, c_i^n\}$. The histogram intersection between regions is calculated and defined as color similarity:

$$s_{color}(r_i, r_j) = \sum_{k=1}^{n} \min(c_i^k, c_j^k)$$

Texture similarity

A histogram is obtained using a similar method to SIFT [8]. SIFT itself is very effective for texture recoding. The selective search takes a gaussian derivative value in each color channel in eight directions and obtains its histogram. The texture similarity is defined as intersection of the histograms of the regions:

$$s_{texture}(r_i, r_j) = \sum_{k=1}^{n} \min(t_i^k, t_j^k)$$

Size similarity

Size similarity is designed to merge smaller region first. This similarity is defined that regions are composed of similar size through the algorithm:

$$s_{size}(r_i, r_j) = 1 - \frac{size(r_i) + size(r_j)}{size(im)}$$

$size(im)$ is number of pixel of whole image.

Fill similarity

The fill similarity is a similarity degree expressing positional relationship between two regions. If R_i has a positional structure including R_j, it merges the two regions in preference, and if R_i and R_j are barely in contact, they are not merged. BB_{ij} tells the size of tight bounding box including R_i and R_j. Filling similarity is as follows:

$$s_{fill}(r_i, r_j) = 1 - \frac{size(BB_{ij}) - size(r_i) - size(r_j)}{size(im)}$$

By combining the above four similarities, final similarity function becomes as follows:

$$s(r_i, r_j) = a_1 s_{color}(r_{i,} r_j) + a_2 s_{texture}(r_{i,} r_j) + \\ a_3 s_{size}(r_{i,} r_j) + a_4 s_{fill}(r_{i,} r_j)$$

Changing the value of a_i can add weight to each similarity and expand the area based on the above equation

2.2 Artificial Neural network

Artificial neural networks are statistical learning algorithms inspired by the neural network structure of biology in machine learning and cognitive science. An artificial neural network is a model in which artificial nodes composed of networks of synapses change the weight of synapses through learning and have problem solving ability. The network consists of input layer, hidden layer and output layer, the structure of which is shown in Figure 2.1. CNN and RNN have been studied to solve certain problems in artificial neural networks.

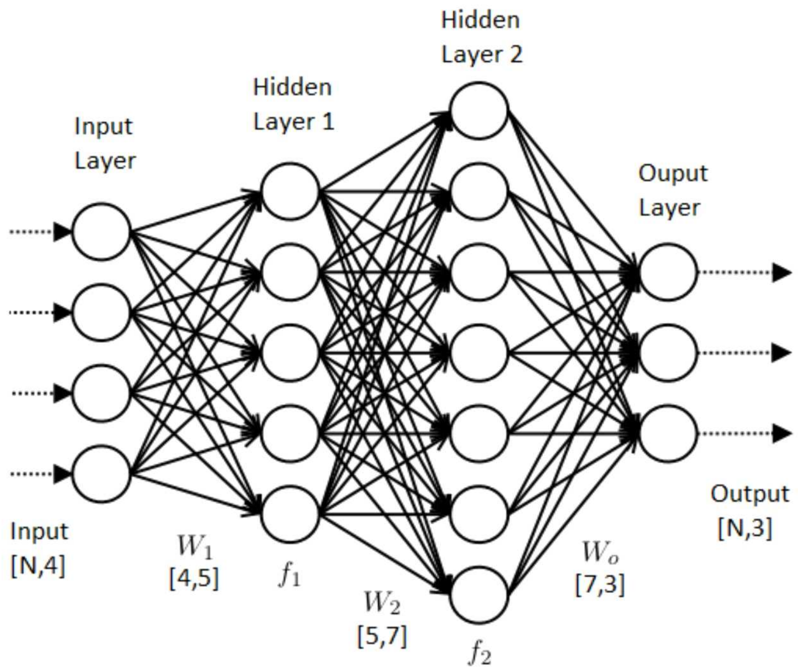


Figure 2.1: General architecture of artificial neural network

Convolutional Neural Netwrok(CNN)

CNN, which shows a powerful effect in image processing, is a modified form of artificial neural network. Architect of model designed to share the weight of weight between layers, it has a translation invariance property. CNN is widely used as an image classification algorithm, which means that the network is learning filters that are manually created in the image processing field previously. People were free from prior knowledge of video filters and their efforts to design them.

Recurrent Neural Netwrok(RNN)

RNN has a recursive structure that allows the pattern to be remembered and used continuously at the input of the sequence. However, the cyclic structure generated a vanishing gradient and it was difficult to learn weights. LSTM [9] has been proposed and is one of the most widely used deep learning methods for natural language processing. The LSTM structure has a forget gate, input gate and output gate. These gate are more advantageous for storing past patterns. This can help you to look at the context of the entire input rather than the local information. The structure of RNN and LSTM is shown in Figure 2.2. RNN has various structures according to the type of input and output. Typical types are shown in Figure 2.3.

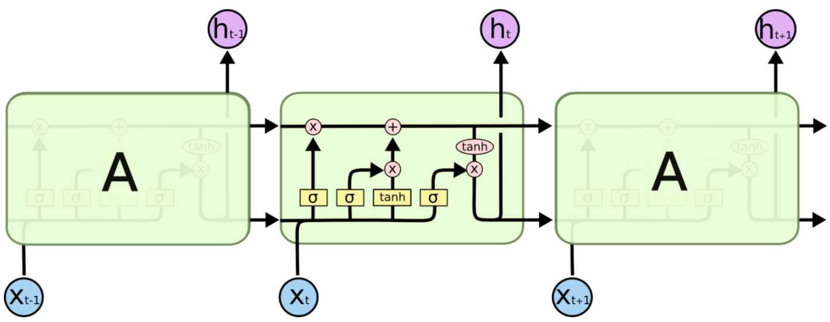


Figure 2.2: Design of LSTM cell

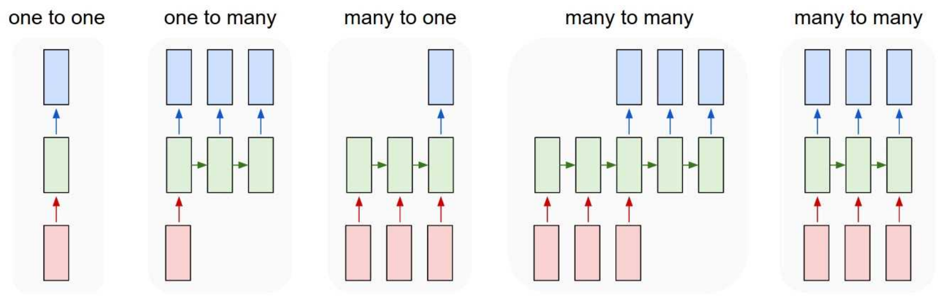


Figure 2.3: Typical architecture of RNN models

Chapter 3

The Proposed System

3.1 Formulation of the Problem

Understanding and isolating organs in cine MRI is an important indicator when diagnosing a patient's disease. In the case of myocardial disease, it is necessary to distinguish the myocardium mask and look inside. The myocardium mask is hard to make the same boundary subject to experts. So developing an application that always produces the same results for the same input can be a good tool. To solve this problem, the proposed system consists of two models. First, we search for ROI containing LV from cine MRI, and detect the endocardial and epicardial boundary in searched region from previous model. Our methods don't have any post processing after obtaining result from model, cause we want to observe effects of each proposed methods.

3.2 ROI detection model

This ROI detection model extracts a candidate region from input of cine MRI which has irrelevant size. A deep learning model is used to regress center point of the endocardial region. The structure of the model is the same as Table 3.1.

Table 3.1: Architecture of ROI detection model

Input: 140 * 140 image

Output: (x, y) coordinate of center point.

Layer	In	Weights	Pool	Out
CN1	140*140	5*5*64	2*2	70*70*64
CN2	70*70*64	5*5*128	2*2	35*35*128
CN3	35*35*128	5*5*256	2*2	18*18*256
FC1	18*18*256	18*18*256*512	-	512
FC2	512	512*512	-	512
Output	512	512*2	-	2

We cropped an image of 70 pixels from the center of the proposed region given from a selective search and used it as input to the deep learning model. In order to balance training data, using intersection over union(IoU) which is define as follow:

$$IoU(area_1, area_2) = \frac{area_1 \bigcap area_2}{area_1 \bigcup area_2}$$

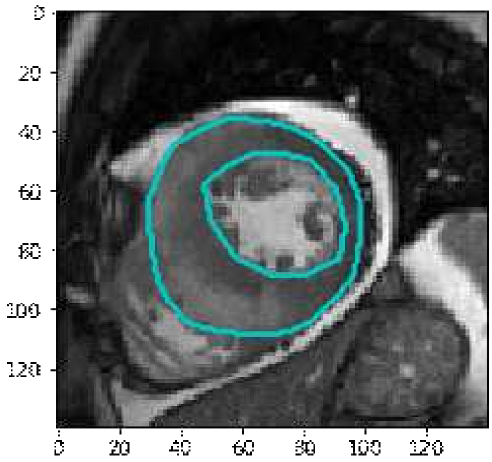
IoU of the proposed region and the ground truth is more than 55%, it is used as a positive dataset, and if it is less than 15%, it is used as a negative dataset. Data with values between 15% and 55% are not used [2]. In order to increase the amount of data to be learned, a selective search may be used to increase the candidate region as input and to apply some error to the center point of the ground truth. In addition, we used a method of rotating image or changing the brightness of image. Using this model, we can obtain a regression of the center of endocardium from the test input image and this mean error is 19.76 and standard deviation is 13.36. By modeling the obtained results as a unimodal Gaussian, one center point can be obtained. The mean center error is 5.43 and standard deviation is 3.67. The loss function used in learning is set as follows:

$$Loss = (coordinate_{gt} - coordinate_{calc})^2$$

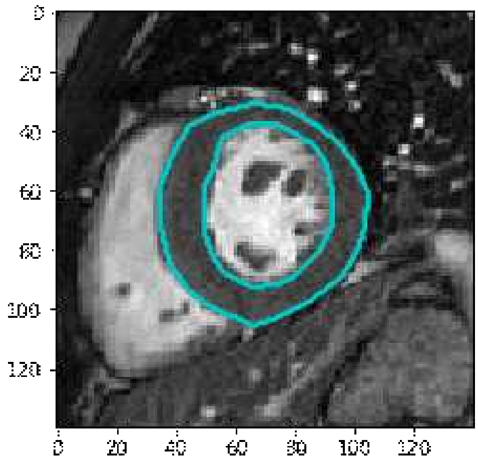
Learning rate is 10^{-4} to the adamoptimizer[10] to optimize. Pass this center coordinate to the next model.

3.3 Boundary detection model

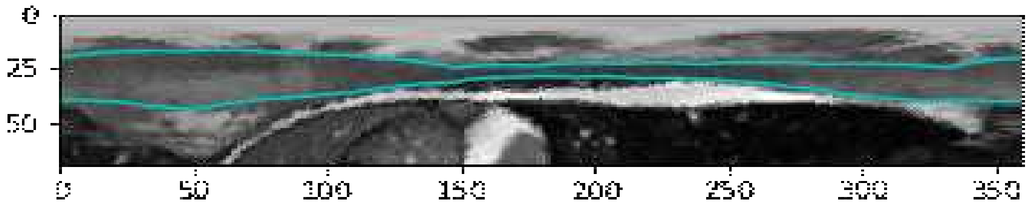
Boundary detection model transforms cine MRI image from Cartesian coordinate to polar coordinate. See Figure 3.1, with transformed input image, target boundaries are more intuitively observed. To regress the height of the endocardial and epicaridal boundary, we use a deep learning model with the same structure will shown in following Chapter 5. Add extra recurrent cells after ROI detection model. The transformed image has a structure of circulation, so that the entire context can be understood using recurrent cells.



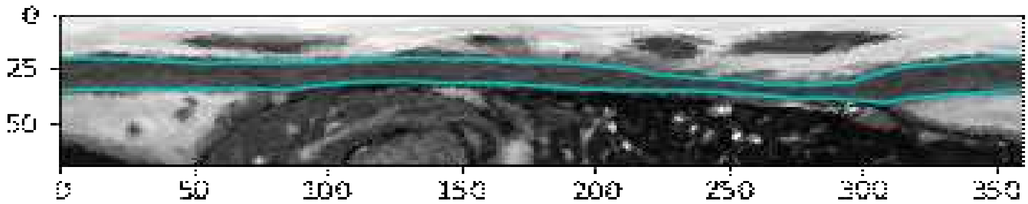
<a> Input 1



<b> Input 2



<c> Transformed 1



<d> Transformed 2

Figure 3.1: Origin image and transformed image

In detail about image transforming, cine MRI takes a circular image with a radius of 70 pixels from the center point calculated in the previous model. Transforming a circular image in a Cartesian coordinate system into a polar coordinate system allows x and y to be transformed into images with angles and distances. The size of the transformed image is 70 in height and 360 in length. The maximum value of boundary data is 54.193, so that we set the height as 70. If you want to obtain a mask for the entire image, you have to consider both x and y coordinate. After transforming, each vertical line of image has 2 height of boundary. We does not know how many boundaries to consider before. We need to find exactly 720 values. To regress value of heights, use imput image size as 70 pixels length of height and 10 pixles length of width. Giving width as s 10 is to calculate more context from input image. We assume that every boundary heights can't be specify with 1 vertical line image. As in the previous model, to increase the amount of data to be learned, we used any point inside endocardium boundary. When the center point change, transformed image have different height respectively. We used a method of changing brightness of image and rotate it also. The final layer of the model obtains endocardial height and epicardial height. Since the last result is not a mask, we must return the results back to the Cartesian coordinate system. To train the model, using adamoptimizer with learning rate value as 10^{-4}. The output layer is designed reflect that the epicardial height should always be greater than endocardial height. Loss function for training model is defined as follows:

$$Loss = height_{gt} - height_{calc}$$

In related studies to find the mask of myocardium, post-processing is applied after obtaining the values for the boundary from the model. For examples, Deletes the boundary value with a negative height, or ignores the case where the endocardium height is greater than the epicardium height. And also, taking moving average of heights to smoothen the boundary from models. Our works have willing to show how each architecture affects on result, so we don’t apply any of post-processing.

Chapter 4
Experiment

4.1 Data of Subjects

The data used in the experiment were de-identified and provided by Department of Radiology, Samsung Hospital. The total number of patients was 23, and about 8 trainable MRI slices were obtained for each patient. In machine learning, data is divided into a training set and a test set to verify the algorithm. 9 patients who received the contrast agent and 4 patients who did not receive the contrast agent were used to study the model. The other patients used as test set to analyze the performance of the algorithm. To avoid model overfitting, 10% of the training set was used as a validation set.

4.2 Experimental Setup

To compare performance of proposed model and other, we build three more model. CNN model with Cartesian coordinate image (CNN-C), NN model with polar coordinate image(NN-P), CNN model with polar coordinate image(CNN-P) and our proposed method(CRNN-P) are

used to compare each performance. Architecture of respective model is shown Table 4.1 and Table 42. Dice index and IoU index are used to measure similarity between produced mask and ground truth.
Dice index are defined as follows:

$$Dice(A,B) = \frac{2TP}{2TP + FP + FN} \ , given \begin{cases} TP: True\ Positive \\ FP: False\ Positive \\ FN: False\ Negative \end{cases}$$

Table 4.1: Architecture of CNN-C and NN-P

CNN-C Model				
Input: 140 * 140, **Output**: 140 * 140 image				
Layer	In	Weights	Pool	Out
CN1	140*140	3*3*1*32	2*2	70*70*32
CN2	70*70*32	5*5*32*64	2*2	35*35*64
CN3	35*35*64	5*5*64*128	2*2	18*18*128
CN4	18*18*128	5*5*128*256	2*2	9*9*256
FC1	9*9*256	9*9*256*1024	-	1024
FC2	1024	1024*1024	-	1024
Out	1024	1024*140*140	-	140*140
NN-P Model				
Input: 70 * 10, **Output**: endocardial height, epicardial height				
Layer	In	Weights	Pool	Out
FC1	70*10	70*10*1024	-	1024
FC2	1024	1024*1024	-	1024
Out	1024	1024*2	-	2

Table 4.2: Architecture of CNN-P, CRNN-P

CNN-P Model				
Input: 70 * 10, **Output**: endocardial height, epicardial height				
Layer	**In**	**Weights**	**Pool**	**Out**
CN1	70*10	3*3*1*32	2*2	35*5*32
CN2	35*5*32	5*5*32*64	2*2	18*3*64
CN3	18*3*64	5*5*64*128	2*2	9*2*128
CN4	9*2*128	5*5*128*256	2*2	5*1*256
FC1	5*1*256	5*1*256*1024	-	1024
FC2	1024	1024*1024	-	1024
Out	1024	1024*2	-	2
CRNN-P Model				
Input: 70 * 10 * 360, **Output**: endocardial height, epicardial height				
Layer	**In**	**Weights**	**Pool**	**Out**
CNN-P	70*10*360	-	-	2*360
RNN	2*360	LSTM 1024 cells	-	2

Table 4.1 and 4.2 are showing the number of weight, however these values don't add up number of bias weights. Each number of bias weights are same as last number of output layer.

4.3 Comparison of Results

To compare the performance between models, see table 4.3. CNN-C was created to compare performance without transformation on input image. NN-P which has the most basic architecture designed for showing benchmark performance with transformed input. NN-P, CNN-P and CRNN-P, which using transformed input data, are showing better performance than the CNN-C. CNN-P has the best performance of all models. CRNN-P, which was expected to show better performance, is shwoing a worse performance than CNN-P. Increasing data set with applying augmentation shows performance of each model are improved.

Table 4.3: Dice index and IoU index of models

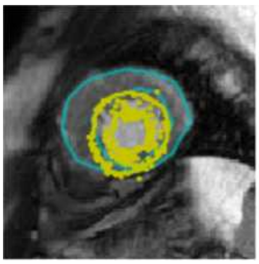
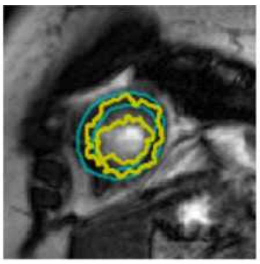
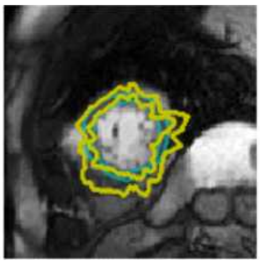
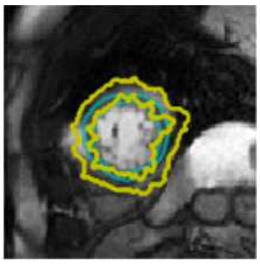
	After Augmentation		Before Augmentation	
	Dice Index	IoU Index	Dice Index	IoU Index
CNN-C	0.580 (0.160)	0.425 (0.148)	0.567 (0.174)	0.416 (0.164)
NN-P	0.670 (0.093)	0.511 (0.103)	0.573 (0.135)	0.413 (0.128)
CNN-P	**0.771 (0.092)**	**0.635 (0.111)**	**0.634 (0.144)**	**0.478 (0.137)**
CRNN-P	0.665 (0.111)	0.508 (0.123)	0.551 (0.159)	0.396 (0.147)

※ Mean (Standard deviation)

4.4 Generated Boundary

The cardiac MRI segmentation differs in difficulty depending on the height of the image. It is largely divided into basal ventricle, middle ventricle and apical ventricle. Figure 4.1 is showing masks generated from basal type of input image. Most of endocardium shape in basal type are left-sheared, because bloods are flowing out from left ventricle. Even doing manual segmentation, there are no clear standard that it's hard to make perfect mask. Figure 4.2 and Figure 4.3 show the masks generated from middle type and apical type, respectively. As shown in figures, CNN-C learns to discriminate each pixel as a mask by pixel-to-pixel approach rather than regressing mask boundaries height, so that it can be seen that sporadic errors occur. Compared to the case of NN-P, NN-P does not show good performance, but unlike CNN-C, the boundary of mask are clearly found and presented. NN-P does not look around the image because it only looks at the given pixel as an input, it shows a microscopic result. On the other hand, CNN-P uses a feature of surrounding pixels to show a reasonable mask. CRNN-P was expected to show the closest correct answer than CNN-P models, but it doesn't. In the evaluation of the boundaries, dice index and IoU index decreased as show in Table 4.4. Looking at the naked eye, the curved surface appears to be smooth, but more errors are occurred in calculating the heights of the boundary. Worst case of mask is shown in Table 4.4 with dice index with ground truth. Every worst case happen in apical level. Apical level MRI is generally dark and it is difficult to grasp the structure of other organs, so the errors are occurs.

Table 4.4 : Worst case of each model and dice index

CNN-C	NN-P	CNN-P	CRNN-P
			
0.136	0.484	0.448	0.369

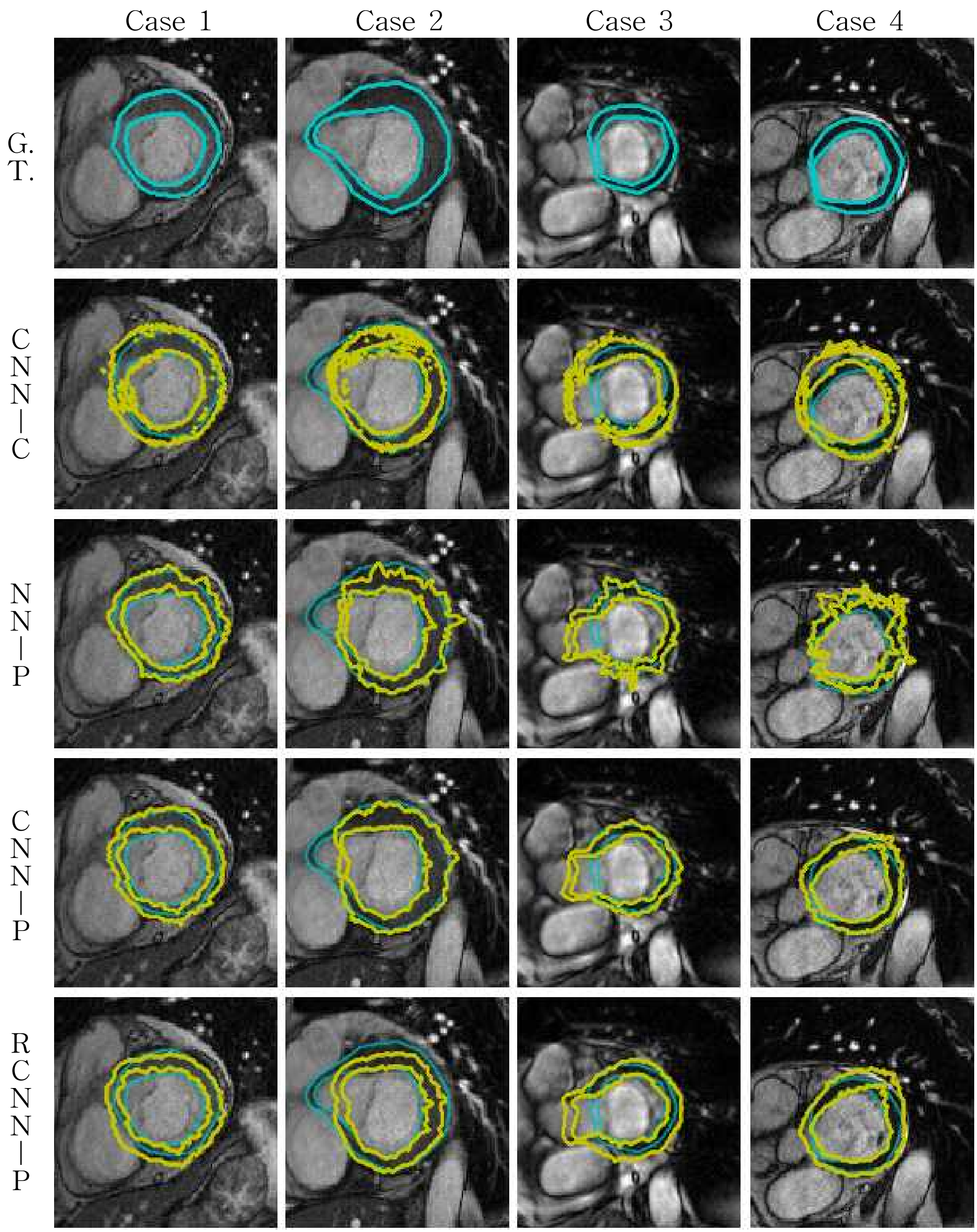


Figure 4.1: Ground truth and generated mask at basal

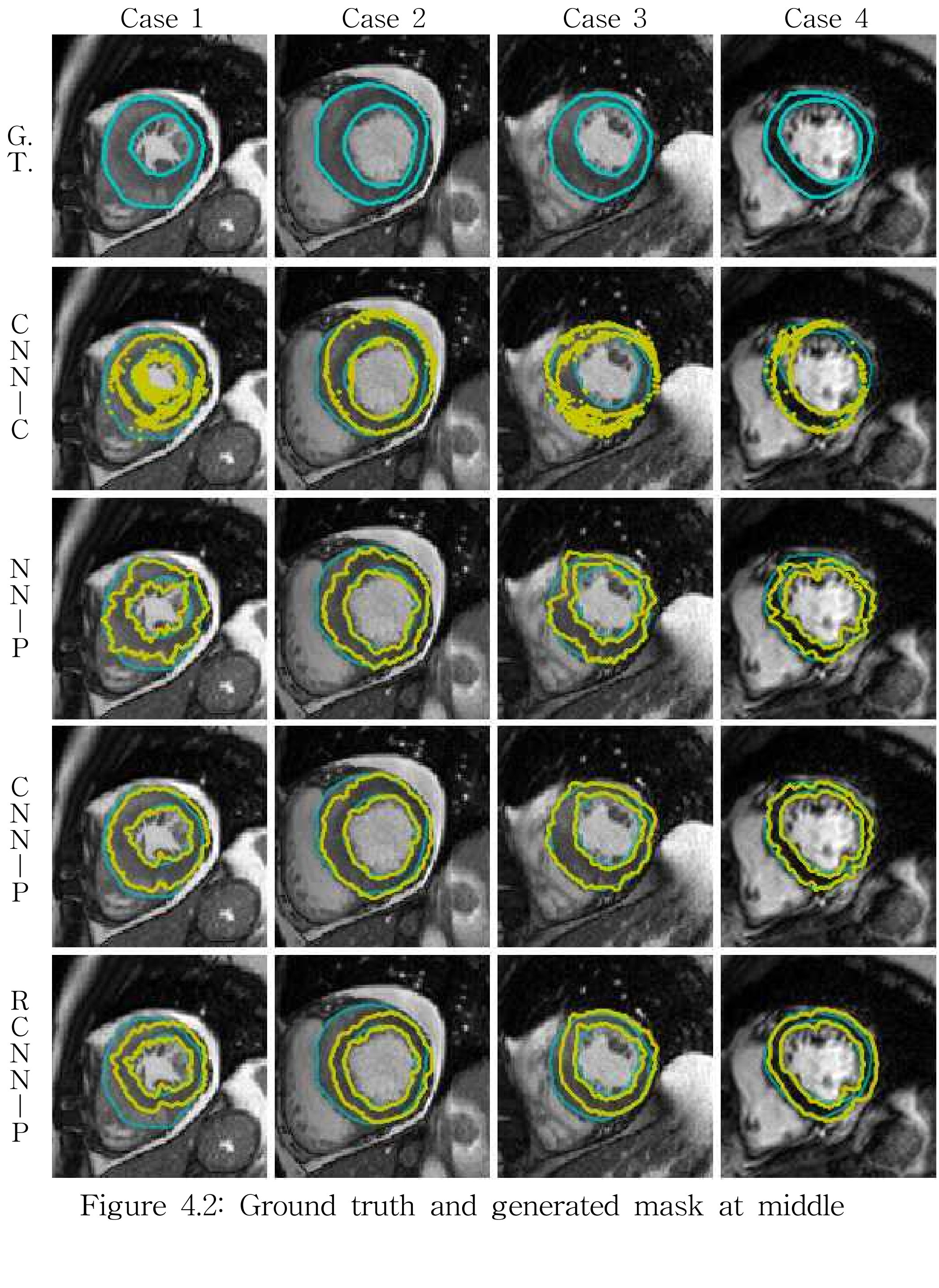


Figure 4.2: Ground truth and generated mask at middle

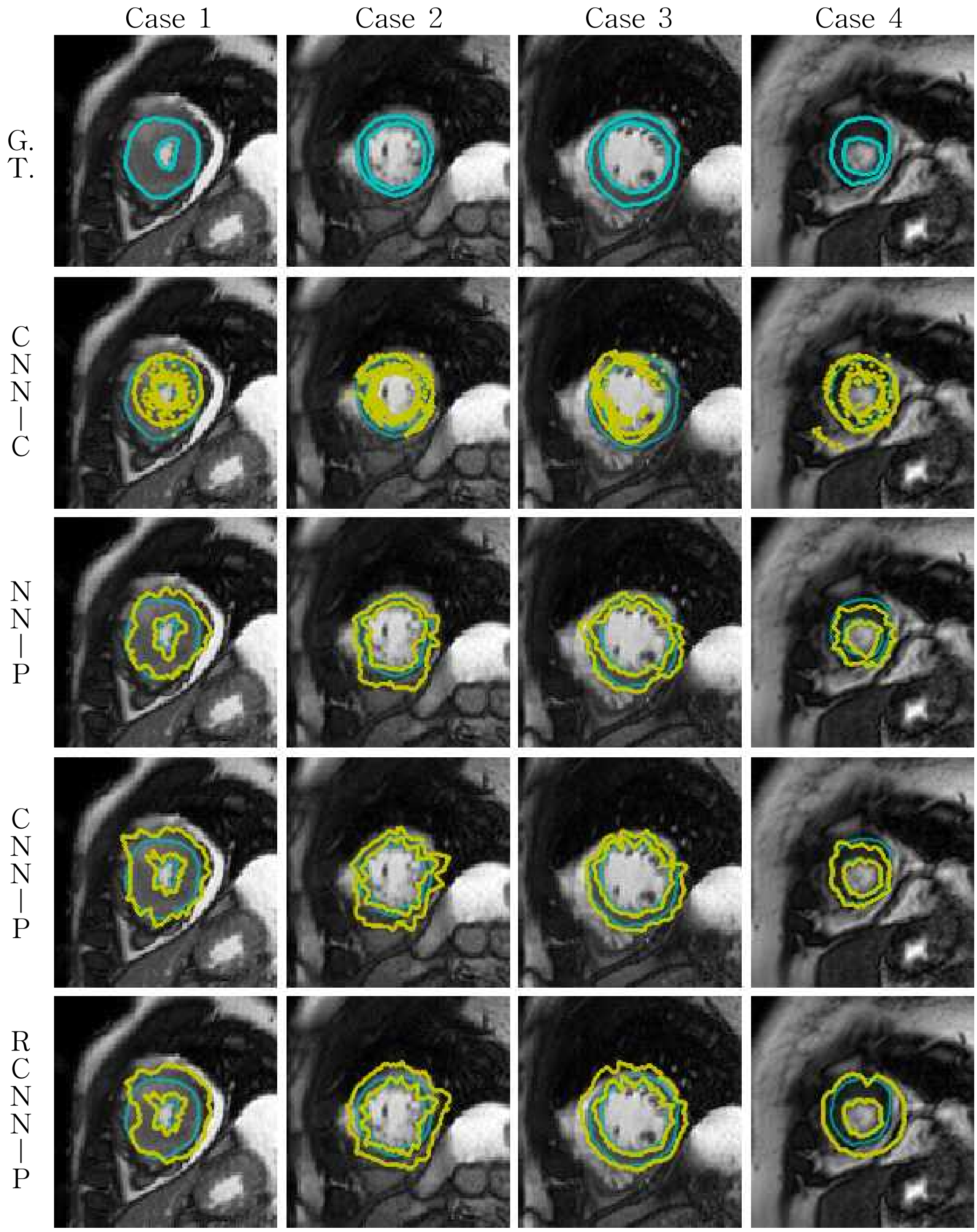


Figure 4.3: Ground truth and generated mask at apical

Chapter 5
Conclusion

Quantifying myocardium mask from cine MRI is widely studied due to diagnosis cardial diseases. Experts could not always make the same mask with same image. This problem makes diagnosis difficult and cross-validates multiple opinions. In this research, a novel approach for applying these problem has been proposed. The proposed CRNN-P system can produce results without human interactions in the whole process and always shows the same results. This can lead to applications that save a lot of time for professionals. Compare CNN-C with NN-P, transforming coordinate of input data is effective to solving problem. The results of NN-P and CNN-P show that the added convolutional layer is effective for the results. On the other hand, the results of CNN-P and CRNN-P show that the added structure does not always have a positive effect. As a result, dice index and IoU index are increased by 32% and 49% on average. Our proposed system consists of two models, so it requires train model one by one. Redesign model with adding more layer will allow to develope end-to-end model. As an application that serves as a diagnostic aid, it need to improve precision and recall. Using segmented results with late gadolinium enhancement(LGE) MRI, we can expect to estimate the volume and location of lesions in the future.

Bibliography

[1] Uijlings, J. R., Van De Sande, K. E., Gevers, T., & Smeulders, A. W. (2013). Selective search for object recognition. International journal of computer vision, 104(2), 154-171.

[2] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems (pp. 91-99).

[3] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. arXiv preprint arXiv:1703.06870.

[4] Xue, W., Nachum, I. B., Pandey, S., Warrington, J., Leung, S., & Li, S. (2017, June). Direct Estimation of Regional Wall Thicknesses via Residual Recurrent Neural Network. In International Conference on Information Processing in Medical Imaging (pp. 505-516). Springer, Cham.

[5] Poudel, R. P., Lamata, P., & Montana, G. (2016, October). Recurrent fully convolutional neural networks for multi-slice mri cardiac segmentation. In International Workshop on Reconstruction and Analysis of Moving Body Organs (pp. 83-94). Springer, Cham.

[6] Lu, Y. L., Connelly, K. A., Dick, A. J., Wright, G. A., & Radau, P. E. (2013). Automatic functional analysis of left ventricle in cardiac cine MRI. Quantitative imaging in medicine and surgery, 3(4), 200.

[7] Felzenszwalb, P. F., & Huttenlocher, D. P. (2004). Efficient graph-based image segmentation. International journal of computer vision, 59(2), 167-181.

[8] Liu, C., Sharan, L., Adelson, E. H., & Rosenholtz, R. (2010, June). Exploring features in a bayesian framework for material recognition. In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on (pp. 239-246). IEEE.

[9] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.

[10] Kingma, D., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

요약

Cine MRI의 단축 좌심실 세분화는 진료의 보조도구로 사용되는 대표적인 영상 분석이다. 이러한 영상 분석은 사람의 손으로 이루어짐에 따라 항상 동일한 세분화 결과를 얻기가 어렵다. 서로 다른 세분화 결과로 인해 시간과 노력이 소비되고 있다. 이 논문에서 제시하는 두 개의 딥러닝 모델로 이루어진 알고리즘으로 시간과 노력을 아끼고, 또한 항상 동일한 세분화 결과를 얻을 수 있다. 첫 번째 모델은 셀렉티브 서치를 이용해 크기와 무관한 입력 영상에서 관심 영역을 탐색하고, 딥러닝으로 좌심실의 중심 좌표를 얻는 하는 모델이다. 두 번째 모델은 탐색된 영상에 좌표 변환을 적용하고 딥러닝으로 심장 내막과 외막의 경계를 찾아내는 모델로 이루어져 있다. 사용된 환자의 수는 24명이고, 총 194장의 Cine MRI 단면을 사용하였다. 그 중 19명은 트레이닝 셋으로 사용하였고, 5명은 테스트 셋으로 사용하였다. 단축 좌심실 세분화를 위한 딥러닝과 다른 알고리즘을 조합을 가지고 문제를 접근하였고, 그 실험 결과 및 분석을 통해 발생하는 문제점을 파악하고 가능성을 제시한다.

keywords : 영상 처리, 좌심실 세부화, 인공신경망, 심장 자기공명영상
Student Number : 2016-21225