Paper Reading No.7

Cardiac MRI Segmentation with Strong Anatomical Guarantees

Sheng Lian

July 2019

1 Brief Paper Intro

- ***Paper ref:*** MICCAI 2019 , http://arxiv.org/abs/1907.02865
- ***Authors:*** See Figure 1.

Nathan Painchaud[1], Youssef Skandarani[1,2], Thierry Judge[1], Olivier Bernard[3], Alain Lalande[2], and Pierre-Marc Jodoin[1]

Department of Computer Science, University of Sherbrooke, Canada[1]
Université de Bourgogne Franche-Comté, Dijon France[2]
Université de Lyon, Lyon, France[3]

Figure 1: authors' brief intro.

- ***Paper summary:*** Authors use an adversarial variational autoencoder (aVAE) for modeling the cardiac's morphological prior, which act as a guide in the post-processing step of cardiac MRI segmentation task.

The aVAE's latent space encodes a smooth manifold on which lies a large spectrum of valid cardiac shapes.

- ***Reading motivation:*** When I first see Figure 2 when browsing `arxiv.org`, I feel like: *That's it!*. Unlike the other prior guided MICCAI 19 paper (paper reading No.5), the post-processing steps presented in this paper are able to properly modify the predicted segmentation mask towards the structure that the cardiac should be.

  Just like HUAWEI P30 pro, maybe the camera can only capture blurred image of the moon, but through AI's '脑补', HUAWEI can recover the details on the moon surface. (See zhihu link).

  When I did research on medical image segmentation, I tried some post-processing approaches, such as CRF. But these approaches '没有灵魂'. They typically involves morphological operators or some connected component analysis to remove small isolated regions, while cannot guarantee the anatomical plausibility.

  But, after I read this paper, I have some confused issues. For example, how to implement the automatically judgement of whether the segmentation result is compliant? (as introduced in section 3.1)

  So, let's see how this paper do.

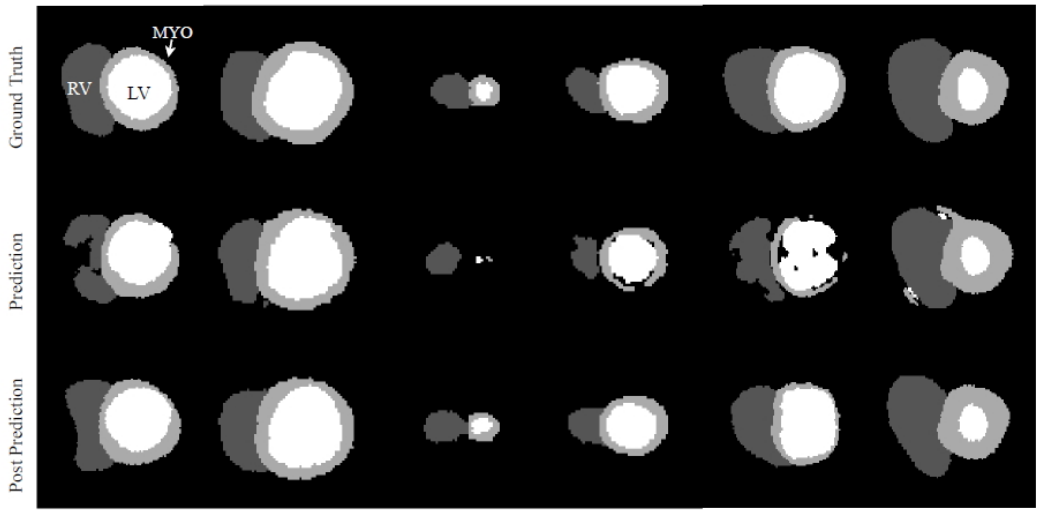

Figure 2: That's it! Groundtruth and erroneous maps before and after our post-processing method.

2 Backgrounds

MRI is a common technology for cardiac diseases. In clinical practice, cardiac parameters are usually estimated from the knowledge of the endocardial and epicardial borders of the left ventricle (defined as the cavity (LV) and the myocardium (MYO)) and the endocardial border of the right ventricle (RV) in end-diastolic (ED) and end-systolic (ES) phases. In the last few years, several deep learning segmentation methods (in particular CNNs) have had great success at estimating these clinical parameters. But, unfortunately, these methods still generate anatomically impossible shapes, so these methods are still unfit for day-to-day clinical use.

3 Methods

The framework of the proposed model is shown in 3, which contains three main blocks:

1) an adversarial VAE (aVAE) that learns a 32-dim latent representation of anatomically correct cardiac shapes;

2) an anatomically-constrained data augmentation of the latent vectors;

3) a post-processing VAE which converts erroneous segmentation maps into anatomically plausible ones.

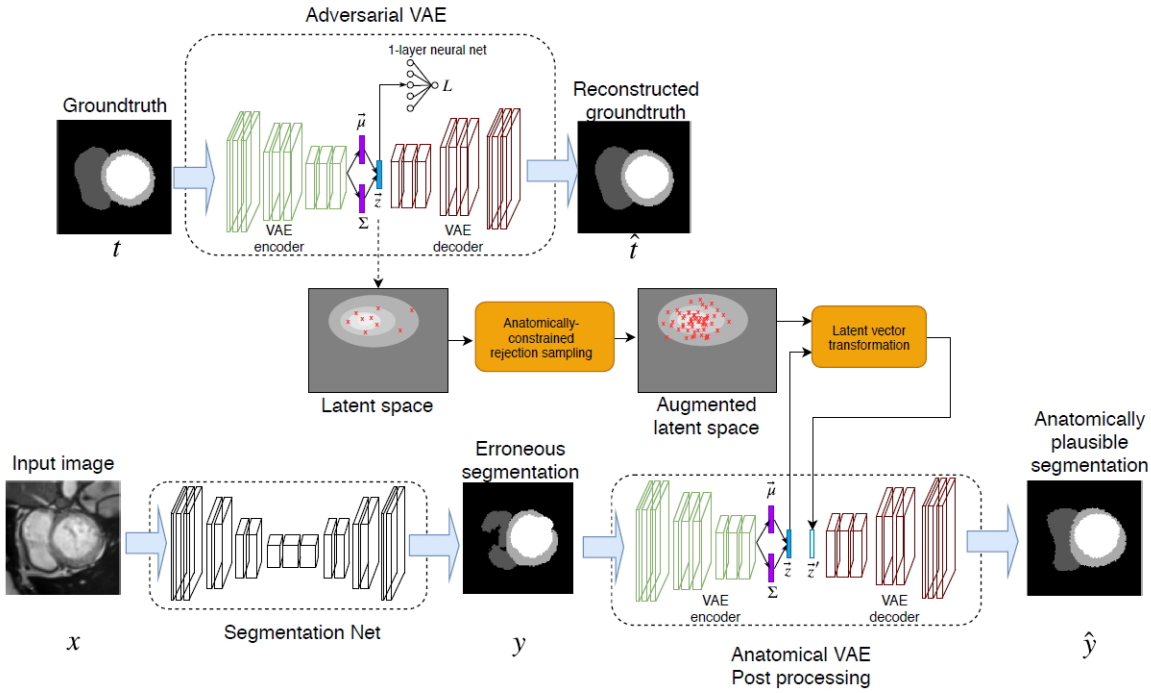

Figure 3: Schematic representation of the method.

3.1 Cardiac MR Images and Anatomical Metrics

First, for getting prior knowledge, authors defined 16 anatomical metrics that will be used to detect incorrect cardiac shapes. Holes, connections, concavities and other issues are considered. This will be used for finding if the latent vectors generated valid in the following steps.

3.2 Adversarial Variational Autoencoder (aVAE)

The structure of autoencoder (AE) and variational autoencoder (VAE) can be briefly summarized in Figure 4 (source: kevin frans' blog.) Also, the detailed VAE recap can be referred in kevin frans' blog.

Here, authors implemented an adversarial VAE (aVAE) [2] which forces the latent space to be as linear as possible. The structure of aVAE is displayed in the upper of Figure 3, where a single-layer neural network is used to predict the slice index of the input image x given its latent vector z using a regression loss. Since the regression's gradient signal propagates through the encoder, it forces it to learn a more linear (and thus less convoluted) latent space. This step is **so tricky** that it makes the latent space smoother and makes interpolated values more anatomically plausible.

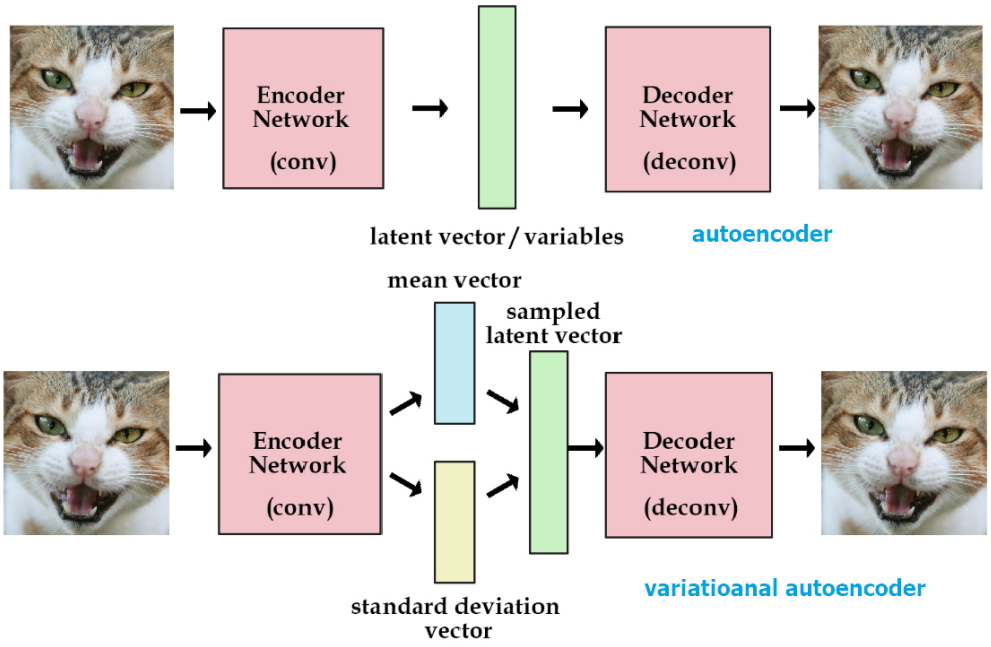

Figure 4: structure of autoencoder (AE) and variational autoencoder (VAE). (source: kevin frans' blog.)

3.3 Anatomically-Constrained Data Augmentation

After 3.2, we got a serials of latent vectors that is anatomically correct (from true data). But the number is too few to densely populate the 32d manifold of anatomically correct latent vectors.

So, authors increase the number of anatomically correct latent vectors with a rejection sampling (RS) method. This operation allows us to generate 4 million latent vectors which all have a valid cardiac shape.

3.4 Latent Vector Transformation

This paper uses a post-processing VAE (at the bottom right of Figure 3 to convert erroneous segmentation maps into anatomically valid segmen-

tations. This pp-VAE has the same architecture and weight as the aVAE mentioned above.

The goal is to transform the latent vector z of an erroneous cardiac shape to a similar but anatomically valid latent vector z'. This transformation can be summarized as follows:

$$\boldsymbol{z}' = \arg\min_{\boldsymbol{z}'} ||\boldsymbol{z} - \boldsymbol{z}'||^2, \quad s.t.\ \mathbb{1}\left(\mathrm{dec}(\boldsymbol{z}')\right) = 1$$

Said otherwise, the goal is to find the anatomically valid latent vector that is the closest to z, which can be simplified to

$$\alpha = \arg\min_{\alpha} |\alpha|, \quad s.t.\ \mathbb{1}\left(\boldsymbol{z} + \alpha\boldsymbol{\delta}_{z'}\right) = 1$$

4 Experiment

The method is trained and tested on 2017 ACDC dataset [1] which contains cine-MR images of 150 patients, 100 for training and 50 for testing. As is introduced above, this method can be act as post-processing step and accommodate any segmentation method. So, authors test this method on ACDC dataset by accommodate it with 11 excellent segmentation method. Figure 5 shows the anatomically invalid results before and after the proposed method (the right part *Nearest Neighbors* seems make no sense?). We can find that after the proposed method, the number of anatomically invalid results reduced dramatically.

Submissions	Original	VAE	Nearest Neighbors		
			w/o RS	w/ RS	Dicho
Zotti-2	55	16	0	0	0
Khened	55	16	0	0	0
Baumgartner	79	17	0	0	0
Zotti	82	15	0	0	0
Grinias	89	12	0	0	0
Isensee	128	21	0	0	0
Roh	287	40	0	0	0
Wolterink	324	42	0	0	0
Jain	185	28	0	0	0
Yang	572	182	0	0	0
ACNN	139	41	0	0	0

Figure 5: Number of anatomically invalid segmentation results on the ACDC test set for 11 segmentation methods with and without our post-processing methods.

Fig 6 shows the Average Dice index and Hausdorff distance comparison. We can find that RS augmentation and dischotomic search can improve the segmentation results in metrics. But, the overall performance is decreased in Dice and Hausdorff after adopting the proposed method.

Submissions	Original	VAE	Nearest Neighbors		
			w/o RS	w/ RS	Dicho
Zotti-2	.913/9.7	.910/10.1	.899/14.4	.909/11.0	.910/10.1
Khened	.915/11.3	.912/12.3	.894/15.2	.909/12.7	.912/10.9
Baumgartner	.914/10.5	.911/11.2	.889/18.2	.907/12.6	.910/10.6
Zotti	.910/9.7	.907/10.9	.878/19.6	.903/12.6	.907/11.0
Grinias	.835/15.9	.833/19.3	.752/32.5	.825/16.9	.833/15.8
Isensee	.926/9.1	.923/10.7	.881/18.4	.917/11.2	.923/9.2
Roh	.891/12.2	.887/14.6	.756/32.2	.874/15.1	.887/12.8
Wolterink	.907/10.8	.903/13.0	.752/32.8	.887/13.5	.903/11.0
Jain	.891/12.2	.886/12.6	.820/31.9	.878/14.2	.886/11.6
Yang	.800/27.5	.752/21.7	.455/29.7	.722/11.5	.752/10.2
ACNN	.892/12.3	.886/26.2	.885/12.0	.885/12.2	.889/13.1

Figure 6: Average Dice index and Hausdorff distance (in mm) for the ACDC test set with and without our post-processing method.

5 My Thoughts

1) When the first time I saw this paper, I was excited.

2) A post-processing VAE which converts anatomically invalid cardiac shapes into close but correct shapes is proposed, and this idea is novel to me.

3) The 16 anatomical metrics in 3.1 confused me. As introduced, these metrics are not differentiable and are not included in the loss. So, how did they calculated? Are they implemented automatically? If it is manually, then it doesn't make sense. If it is automatically, how to ensure that the premise (regional division) of these judgement conditions is established?

4) The traditional metrics of medical segmentation is worth discussing. If this paper's results are solid, after adopting the post-VAE, the anatomically incorrect results are warped towards the closest anatomically viable shape, but the DICE and Hausdorff distance are decreased. So, when we do research on medical image segmentation, is the maintaining of anatomically structure more important, or the evaluation metrics?

References

[1] Olivier Bernard, Alain Lalande, Clement Zotti, Frederick Cervenansky, Xin Yang, Pheng-Ann Heng, Irem Cetin, Karim Lekadir, Oscar Camara, Miguel Angel Gonzalez Ballester, et al. Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: Is the problem solved? *IEEE transactions on medical imaging*, 37(11):2514–2525, 2018.

[2] Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, Ian Goodfellow, and Brendan Frey. Adversarial autoencoders. *arXiv preprint arXiv:1511.05644*, 2015.