Automated Cardiac Segmentation and Diagnosis

Zhaoshuo Li

Abstract— **This report is submitted to EN.520.633 Medical Image Analysis course. This goal of this project is to segment the four areas of a MRI cardiac image and classify the patients into five disease categories. The segmentation task used a UNet structure with ResNet backbone. The classifier is a random forest classifier which takes in both instant and dynamic image features. The data used for training is the cardiac volumes of 50 patients provided by the course instructor, and the mean dice score of segmentation on test dataset is 0.88. The end-to-end classification accuracy of the random forest classifier is 0.80. The total inference time is 79 seconds for all test data. It has been shown that the proposed approach is very effective for the target tasks, outperforming all the other approaches in the class.**

I. THEORY AND BACKGROUND

Cardiac imaging has been an important CAD tool for heart function monitoring. This is usually done through MRI. The current gold standard is to segment the different region of the heart to analyze, e.g. ejection volume, to diagnose diseases. However, manual segmentation can be labor-intensive and prone to error. In this project, the goal is to automatically segment the 3D MR images into four regions and classify the patient into one of the five categories for diagnosis.

In MR images, the blood pool usually appear brighter than the surrounding structures, However, segmentation is still very challenging due to the overlap intensity distributions, shape variation, and fuzziness [1]. Traditionally, the segmentation task is solved by handcrafting features and proposing various cost functions for optimization. In [2], the authors segmented the LV via graph cut with parametric parameters as prior information. While in [3], graph cut is conditioned with a shape model constructed via a PCA from a set of representative shapes of the RV obtained by manual segmentation. This shape prior is added to the global cost function as a distance loss. In [4], LV is segmented by finding the 3D centre of the cardiac through first harmonic images obtained by fourier transform. The DC component of FT is used to find proper thresholds for LV cavity segmentation. In [5], the cardiac image is first transformed into polar coordinates. The first slice that the LV is detected is set as the reference slice and deformable registration is done to align the slices in the ED and ES phase separately. The average image of all slices are taken and the segmentation is found to be the minimal surface across all slices by minimizing the cost function consisting of gradient, the number of pixels between the two contours not labeled as myocardium, and the variation in the thickness of the myocardium. A different approach of using motion field between ED and ES phase at the basal slice and Hough transformation to find LV is presented [6]. Five Gaussians are fitted to the histogram to represent different classes, and segmentation is performed by minimizing the intensity-based pixel affinity, distance between labels and sizes of the regions. The image is again transformed to polar coordinates and a path is drawn to be homogeneous with high affinity, strong gradient, and smoother transitions along the column.

A statistical modelling approach is taken in [7], which builds 3D active appearance models (AAM) out of Gaussian smoothed local 2D AAMs. To find correspondence of the current volume to the model, iterative matching of local 2D slices and updating of global 3D shape based on the best 80% match is carried out, and the final segmentation can be obtained from correspondence. In [8], an atlas based registration is proposed with initialization using region based registration. The masks in atlas are transformed with overlapping region corrected. Then distance weighting interpolation is applied on the transformation between classes, with fluid deformable registration followed up to fine tune segmentation result.

More recently, deep neural network based approach is used and outperformed most of the previous suggested methods. UNet, a popular segmentation architecture, used encoding-decoding methods to segment an image both on latent features and the original image. In [9], a 5 layer 2D and 3D UNets were trained trained on the publicly available ACDC Challenge dataset. The final segmentaiton result is used as the mean of the two predicted output. In this project, this paper is heavily followed since it produced the best result so far in cardiac imaging.

As for the classification task, not much is reported in the literature. [10] used 14 features with random forest classifier and achieved 91% accuracy on MICCAI 2017 ACDC. The features include patient-based features, e.g. weight, as well as image-based features, e.g. volume change of LV between ED and ES. This paper was used to guide the design of the classification task.

In this project, a final layer-by-layer segmentation with an improved UNet is used. The classifier is trained by extracting image features from the segmented result. In the following sections, the methods used in the project is presented in section II. The result will be discussed in section IV. And the conclusion will be presented in section V.

II. METHODS

A. Data

There are 50 patients data provided for training in this project. Along with the ED and ES volumes, the labels of background, left ventricle (LV), mycardium and right ventricle (RV) are given. There are 5 disease groups in this

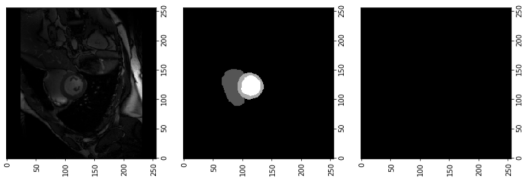


Fig. 1. A failure case for vanilla UNet with 5 layers. Left: input, Middle: ground truth label, Right: prediction.

dataset, MINF, DCM, HCM, RV and NOR. The information about the patient, including the weight, height, disease type are given.

B. Segmentation

According the report of [9], cardiac MRI suffers from poor resolution in the out-of-plane direction (so called Z axis) while having fine resolution in plane (so called XY plane). Therefore, the 3D UNet structure does not improve the result by a significant margin. Moreover, due to the increase of the parameters, the network becomes harder to train and memory limitation becomes a issue. Therefore, only 2D UNet is used. During early experiments, it has been found that shallow UNet is very prone to overfit on the intensity feature since in most of the cardiac MRI, the LV region appears brighter. Therefore, when the contrast is low, the network will fail (Figure 1).

To mitigate this problem, an improved UNet called AlbuNet [11] is used. The same 5-layer UNet with more powerful encoding ability is used by incorporating the ResNet-34 backbone (Figure 2). The residual blocks from the ResNet can help with gradient flow during the training process. The input size of the AlbuNet is chosen to be 256×256 to reduce the memory burden, which also matches well with the MRI image sizes.

As for data pre-processing, a per-slice normalization is used since there are two different machines used for acquiring the image data (Equation 1).

$$I^*_{x,y} = \frac{I_{x,y}}{I_{max}} \tag{1}$$

To avoid the change of spatial property of the cardiac image, instead of re-sizing the images to the target size, the image is padded first if it is smaller than the target shape and then randomly cropped. To avoid overfitting, the images are augmented by randomly disturbing the brightness and contrast, as well as spatial transformations such as flip and affine. In the end, the images are roughly zero-meaned such that the gradient flow will not have the zig-zag issue (Equation 2).

$$I^{**}_{x,y} = \frac{I^*_{x,y} - 0.5}{0.5} \tag{2}$$

C. Classification

Since the gold standard approach of diagnosis by clinicians is to analyze the image features such as ejection fraction to determine if the patient's heart malfunctions, the cardiac images were not used to for the classification task directly. Instead, image features are extracted and used to train the classifier. The random forest has proven to be a powerful classifier [10], though prone to overfitting. The classifier is firstly trained on the ground truth data on 62 selected features first, including instant, dynamic features of cardiac as well as patient specific information (Table I). All features are used to train $1,000$ trees first to determine the best 10 features to use, in order to avoid overfitting. Following up, 500 trees are trained further on the 10 features for inference (Figure 3).

TABLE I

FEATURES FOR INITIAL CLASSIFICATION TRAINING

Instant Feature	Dynamic Feature	Patient Specific
Circularity Mean	Ejection fraction	Weight
Circumference Max	Volume ratio myocardium/LV	
Minor axis length Mean	Volume ratio RV/LV	
Minor axis length Max		
Minor axis length Std		
Volume/Body Suface		

The final selected features are:

- Mean circularity of Myocardium ED
- Mean thickness of RV ES
- Volume of RV ES
- Volume/surface of RV ES
- Max thickness of LV ES
- Mean thickness of LV ES
- Mean circumference of LV ES
- Volume of LV ES
- Volume/surface of LV ES
- Ejection fraction of LV
- RV/LV volume ratio ED
- RV/LV volume ratio ES
- Myocardium/LV volume ratio ED
- Myocardium/LV volume ratio ES

The classifier is then fine-tuned based on the predicted segmentation to improve the stability of diagnosis due to the segmentation error from the AlbuNet.

III. EXPERIMENT AND RESULTS

A. Segmentation

The hyper-parameters used for training are of batch size 1 to avoid memory overloading. Initial learning rate of 0.001 and decay of 0.95 per epoch. Adam optimizer is used for weight update. Total train time is 150 epcohes. Weighted dice loss is used due to the imbalance of the pixel numbers appeared in the cardiac images and also the relative appearance variation. Experimentally, the best weight of dice loss is set to $0.1905, 0.9524, 1.9048$ and 0.9524.

The validation result of the trained network is 0.88 for RV, 0.88 for myocardium and 0.95 for LV. Compared to [9], the RV and myocardium did not perform as good by a small margin. The example of the best and worst prediction on the validation data is shown in Figure 5.

On the test dataset, the final score of the trained network is 0.85 for RV, 0.87 for myocardium and 0.93 for LV. The

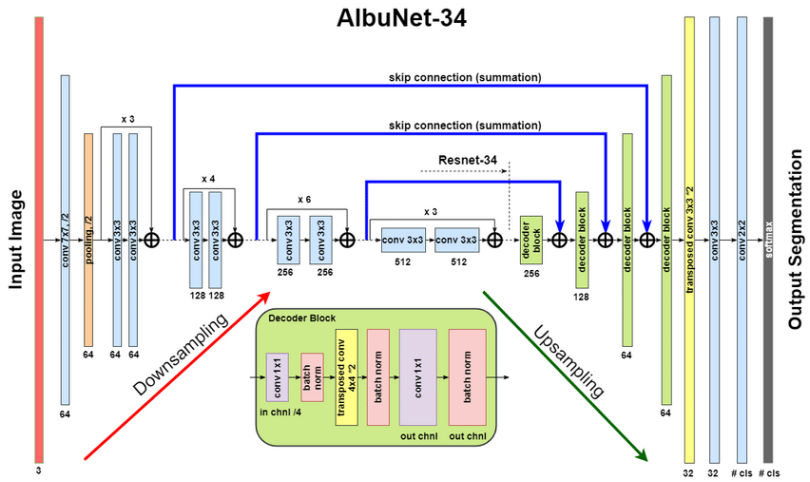


Fig. 2. AlbuNet structure.

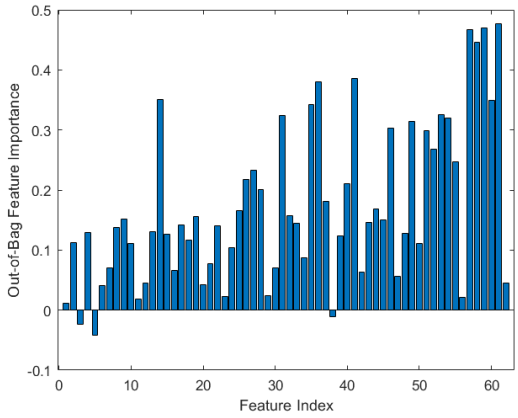


Fig. 3. Feature selection for random forest classifier. The horizontal axis represents the index of different features used for training. The vertical axis represents the feature importance normalized to 0-1. Top 10 features were selected.

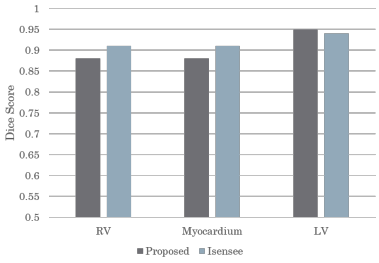
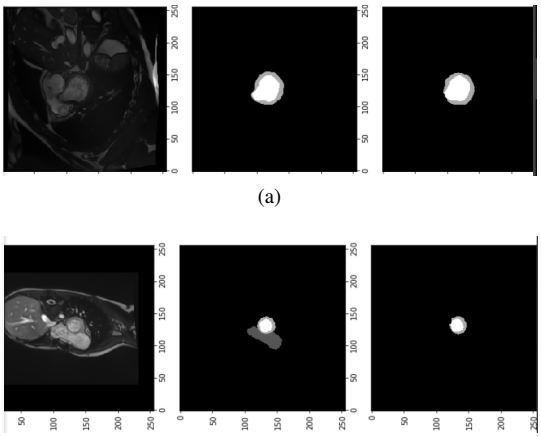


Fig. 4. Comparison of validation dice score with [9].

closest runner up result is 0.7 for RV, 0.66 for myocardium and 0.85 for LV. It can be seen that the proposed method outperformed other team's methods by a significant amount of margin.

B. Classification

The final classification accuracy on validation dateset is 90%, with the confusion matrix shown in Figure 6. The



(a)

(b)

Fig. 5. Best and worst prediction on validation data. (a) is best and (b) is worst.

RV class had the worst accuracy probably due to the bad segmentation of the RV area.

The accuracy on the test data is 80%, outperformed the runner up by 2%.

IV. DISCUSSION

Deep learning has been shown to be a powerful feature extractor that can be used for segmentation purpose. And in this project, both the winner and the runner up used the UNet-like architecture. This strongly indicates that given enough data, the best approach of automated segmentation is deep learning. The best performance of the traditional approaches to segment the cardiac image come from Team 9, who carefully handcrafted all features needed for segmentation. His approach included registration of the test images to training image. The intensity based binarization and morphological operations are used to find the potential regions for LV, RV and myocardium. With prior collection of the probability distribution of the location, the intensity features and the probability features are multiplied to find the best seed. Region growing is used by sampling 10 pixels

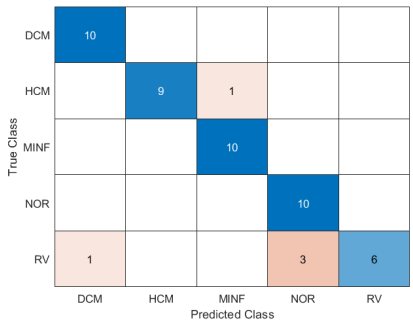


Fig. 6. Confusion matrix.

of the seed region, with a tolerance that minimizes the difference between adjacent tolerance result. Finally, shape smoothing is used to make the prediction better. This pipeline produced relatively good result, while unable to match the deep learning approach.

V. CONCLUSIONS

This report presents the proposed approach for cardiac segmentation. The proposed approach achieved a dice score of 0.88 and 80% accuracy on the diagnosis. The inference time is 79 s and the proposed approach outperformed other ones by a significant margin. The code is made open-source available at this link.

REFERENCES

[1] C. Petitjean and J.-N. Dacher, "A review of segmentation methods in short axis cardiac mr images," *Medical image analysis*, vol. 15, no. 2, pp. 169–184, 2011.

[2] J. Zhu-Jacquot and R. Zabih, "Segmentation of the left ventricle in cardiac mr images using graph cuts with parametric shape priors," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2008, pp. 521–524.

[3] D. Grosgeorge, C. Petitjean, J.-N. Dacher, and S. Ruan, "Graph cut segmentation with a statistical shape model in cardiac mri," *Computer Vision and Image Understanding*, vol. 117, no. 9, pp. 1027–1035, 2013.

[4] X. Lin, B. R. Cowan, and A. A. Young, "Automated detection of left ventricle in 4d mr images: experience from a large study," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2006, pp. 728–735.

[5] M.-P. Jolly, H. Xue, L. Grady, and J. Guehring, "Combining registration and minimum surfaces for the segmentation of the left ventricle in cardiac cine mr images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2009, pp. 910–918.

[6] A. Pednekar, U. Kurkure, R. Muthupillai, S. Flamm, and I. A. Kakadiaris, "Automated left ventricular segmentation in cardiac mri," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 7, pp. 1425–1428, 2006.

[7] S. Zambal, J. Hladvka, and K. Bhler, "Improving segmentation of the left ventricle using a two-component statistical model," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2006, pp. 151–158.

[8] X. Zhuang, D. J. Hawkes, W. R. Crum, R. Boubertakh, S. Uribe, D. Atkinson, P. Batchelor, T. Schaeffter, R. Razavi, and D. L. Hill, "Robust registration between cardiac mri images and atlas for segmentation propagation," in *Medical Imaging 2008: Image Processing*, vol. 6914. International Society for Optics and Photonics, 2008, p. 691408.

[9] F. Isensee, P. F. Jaeger, P. M. Full, I. Wolf, S. Engelhardt, and K. H. Maier-Hein, "Automatic cardiac disease assessment on cine-mri via time-series segmentation and domain specific features," in *International workshop on statistical atlases and computational models of the heart*. Springer, 2017, pp. 120–129.

[10] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum, "Automatic segmentation and disease classification using cardiac cine mr images," in *International Workshop on Statistical Atlases and Computational Models of the Heart*. Springer, 2017, pp. 101–110.

[11] A. A. Shvets, A. Rakhlin, A. A. Kalinin, and V. I. Iglovikov, "Automatic instrument segmentation in robot-assisted surgery using deep learning," in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2018, pp. 624–628.