Cardiac Segmentation and Diagnosis

Xiaorui Zhang (xzhan227@jhu.edu)

Abstract This paper is a final write-up for the cardiac segmentation and diagnosis project. Cardiac MRI is considered the gold standard if cardiac function analysis. However, manual evaluation is expensive and non-reproducible. In this project, we created an automatic pipeline consisting of segmentation and cardiac disease classification to tackle the "Automatic Cardiac Diagnosis Challenge" (ACDC) using the deep learning method. We evaluated the performance of three UNet like models, and the best one was used to participate in the ACDC segmentation challenge. Dynamic and instant features were extracted using the segmentation result, and an ensemble of MLP and random forest classifiers were trained to do the diagnosis challenge. Our model proved its accuracy in the segmentation challenge on the ACDC test set and achieved promising classification results on the training set. Our project won first place in the class competition and demonstrated how small changes in the training model could improve overall performance.

1 Introduction

Analysis of cardiac function is important in clinical cardiology. Because Cardiac MRI (CMR) can discriminate different types of tissues, it is considered as the gold standard of cardiac function analysis through the assessment of the left and right ventricular ejection fraction (EF) and stroke volumes (SV), the left ventricle mass, and myocardium thickness. [1] However, manually evaluating these time series is expensive and non-reproducible; the huge benefits of CMRI are still not exploited in today's clinical routine. Accurate and automatic approaches for simultaneous multi-structure segmentation and computer-assisted diagnosis are thus desirable.

Our team proposed an automatic approach for CMRI scans segmentation and classification of cardiac diseases in this project. Based on the segmentation for each time step of the CMRI, domain-specific features were extracted motivated by the cardiologist's workflow. These features were then used to train an ensemble of MLP and random forest classifiers. Finally, we evaluated our methods for segmentation and diagnosis on the ACDC data set [2] and achieved dice scores of 0.945 (LVC), 0.885 (RVC), and 0.900 (LVM) on the test set (50 cases). We reported a classification accuracy of 74% on the test set.

2 Theory and Background

Convolutional neural networks (CNN) have shown outstanding performance in medical image segmentation [3], where UNet-like architectures are often used [4]. There are two paths in UNet, down-sampling, and up-sampling. The down-sampling path is used to extract and interpret the context (local contextual information), while the up-sampling path is used to enable precise localization (global information). The coarse contextual information captured by the down-sampling path is transferred to the up-sampling path by means of skip connections. In 2017, Fabian et al. used an ensemble of UNet inspired architectures for segmentation of cardiac structures on each time instance of the cardiac cycle and won first place in ACDC 2017 segmentation challenge and found that 2D UNets outperformed 3D UNets with the ACDC dataset [5].

Our project implemented the original 2D UNet model first [4], and its performance was reported. After that, several modifications were made to this model to account for some of its inherent drawbacks. Odena et al. [6]found that the transposed convolution layer used in the original 2D UNet for upsampling can create a characteristic checkerboard-like pattern of varying magnitudes. Alternatives upsampling methods such as nearest-neighbor interpolation or bilinear interpolation could be used to avoid that artifact and speed up the learning.

Deep supervision is a technique that combines segmentation maps created at different points in the network. This idea was used in the original FCN-design by Long et al. [7] to reduce the coarseness of the final segmentation. Chen et al. [8] [9] implemented deep supervision on 3D medical image segmentation by creating multiple segmentation maps at different resolutions and bring to a matching resolution by deconvolution and combined via element-wise summation. Kayalibay et al. [10] found that in a UNet like model, deep supervision can be used to speed up convergence by "encouraging" earlier layers of the network to produce good segmentation results. Moreover, deep supervision could alleviate the class imbalance in our dataset. [8] [9]

Both modifications were made to our original UNet, and the improvements of these modifications were reported.

Computer-assisted diagnosis (CAD) uses texture information to discriminate healthy from pathology tissue. Medrano-Gracia et al. found the major principal modes of shape variation to be associated with known clinical indices of adverse modeling, including heart size, sphericity, and concentricity [11]. Fabian et al. [5] extracted two sets of features from their segmentation results to perform disease classification. All features were designed to quantify the traditional assessment procedures of expert cardiologists by describing static and dynamic properties of the structures of interest.

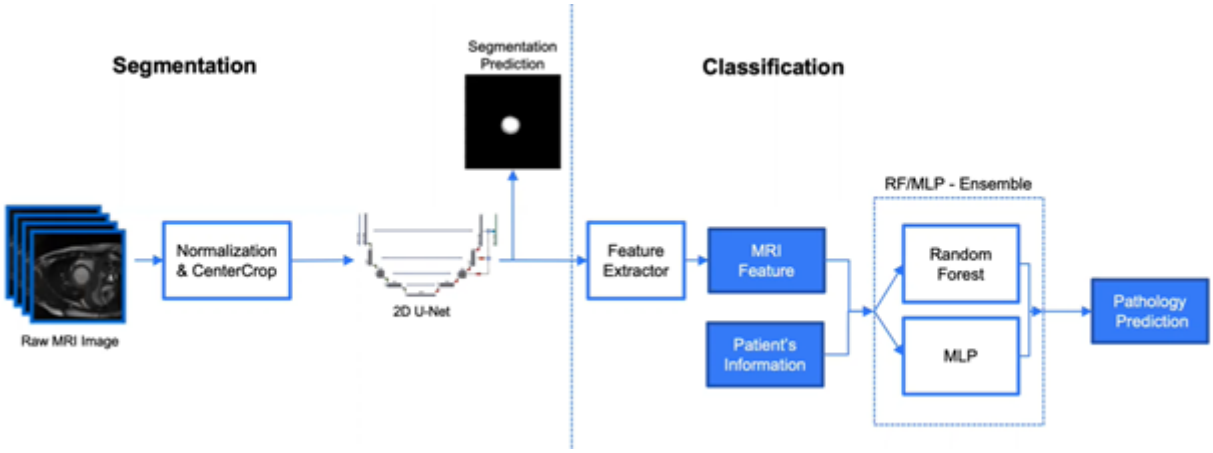

Fig. 1: Project Pipeline

In the project, the modified UNet model was used to perform multi-structure segmentation for each time step of the CMRI. Motivated by the cardiac diagnosis workflow that is used in clinical practice, we extracted both dynamic and instant features from our segmentation results. These features were then used to train an ensemble of MLP and random forest classifiers. (See Figure.1) Finally, we evaluated our methods for segmentation and diagnosis on the ACDC data set.

3 Methods

Dataset The ACDC dataset includes short-axis cine-MRI of 150 patients acquired in daily clinical practice. Each time series is composed of 28 to 40 3D volumes. 150 patients are evenly divided into 5 subgroups: patients with previous myocardial infarction (MINF), dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM), abnormal right ventricle (ARV), and normal(healthy) subjects (NOR). In addition, the height and weight for each patient are also provided. The training dataset is composed of 100 patients, i.e., 20 for each group. There are 50 patients in the test set, i.e., 10 for each group. The ground truth for segmentation and classification is provided for the training set by two independent experts who had to reach a consensus in case of discordance.

Data set analysis:Why choose 2D UNet over 3D? The training model (modified 2D UNet) was selected based on the characteristic of our dataset and the computing power we have. The dataset has high in-plane resolution ranging from 0.49 to 3.69 mm^2 and a low resolution (5-10mm) in the direction of the long axis of the heart. Therefore, most of the pathological information is concentrated in the 2D plane, and the data is relatively sparse along the long axis. In terms of extracting useful features, 2D UNet should be sufficient, and implementing 3D UNet might not significantly improve the segmentation result. Moreover, implementing 3D UNet means we need to face the problem of data scarcity along the long axis. Although this problem could be solved by performing interpolation along the long axis to create more data, we didn't have enough computing resources to do a 3D convolution training that acquires much higher memory and a more powerful GPU. Another potential risk of using 3D UNet is that it is more prone to overfitting since 3D UNet uses the whole 3D volume as input. But for each patient, we only have two 3D volumes that can be used for training, one at ED and one at ES.

Data preprocessing: We first resampled all volumes to 1.25 X 1.25 X Z_{origin} mm per voxel to address the varying spatial resolution problem in our dataset. Second, the intensity of every image was normalized, so they all have zero means and unit variances. Various data augmentation techniques were used to help train a well-generalized model on limited data. The methods implemented including random crop, horizontal/vertical flip. The region of interest was selected using prior knowledge of knowing the heart is at the center of each slice in this dataset. A 270 by 270 pixels center crop was implemented first to filter out the region of interest, then a 256 by 256 pixels random crop was performed within ROI to generate a training image.

UNet Modifications: In this project, we decided to implement the original UNet first, evaluate its performance, understand its pros and cons, and then try to modify it to perform better in the challenge. Upsampling is a crucial step in UNet as precise localization is achieved during up-sampling by combining the contextual information from the contracting path. However, one drawback of the original UNet actually comes from upsampling. The transposed convolution used in this expansive path can create a checkerboard-like pattern of varying magnitudes when processing 2D images. The checkerboard artifacts undoubtedly harmed our segmentation. Therefore the first change we made to the model was switching the transposed convolution to bilinear upsampling followed by a one by one convolution layer. Here we had a couple of choices when choosing the interpolation method. For instance, we could have used nearest neighbor or higher-order B splines. We used bilinear mainly because it's quick and smoother

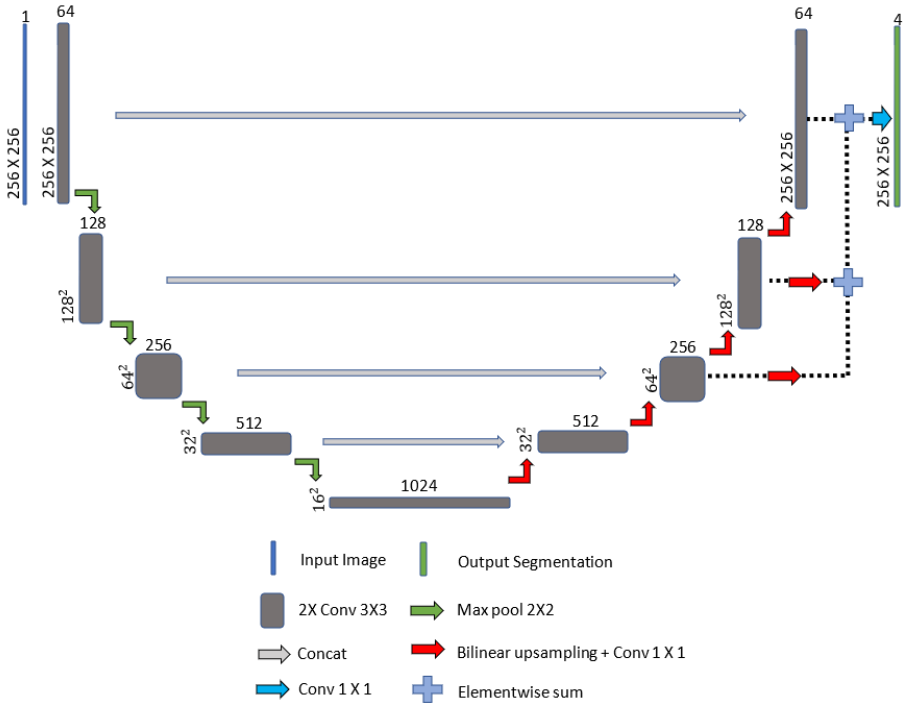

Fig. 2: Final network architecture

than the nearest neighbor. We hope this change could eliminate the artifacts and speed up the training.
Another common yet critical problem that calls for modifications of our model is class imbalance. In our dataset, voxels are dominated by background. Especially in this project, we only have one background, but three foreground classes. Class imbalance is an overwhelming issue. Element-wise sum and upsampling are used in deep supervision to combine segmentation maps at different resolutions to a matching resolution. As mentioned in the background section, this technique can speed up the convergence, improve the discriminative ability of our network, and alleviate class imbalance.

Final network architecture: Combining the two modifications we mentioned above, our final model that was used for the challenge is depicted in figure2. All three models mentioned above were trained for 100 epochs using the ADAM solver and a pixel-wise categorical cross-entropy loss. The initial learning rate of 5×10^{-4} was decayed by 0.94 per epoch. As mentioned in preprocessing section, training examples were generated as a random crop of 256 by 256 pixels taken from a randomly selected slice in the collection of all ED and ES slices of all patients in the training set. The initial feature maps are 48, and the network is trained with a batch size of 10. Each feature extraction block (shown in gray in fig 2) consists of two zero-padded 3X3 convolutions, followed by batch normalization and ReLU activation. The initial feature maps are doubled with each of the four max pooling layers and halved with each of the four upsampling operations. Deep supervision is implemented using the last two lower resolution segmentation results. First, the segmentation map with the lowest resolution is upsampled with bilinear interpolation to have the same size as the second-lowest resolution segmentation map. The sum of the two is then upsampled and added to the highest-resolution segmentation map in the same way.

Classification: Both dynamic and instant features were extracted. The features were selected to quantify the traditional clinical assessment procedures by describing static and dynamic properties. All features we used are listed in Figure 3.

Instant Features	**RV**	**MYO**	**LV**
max thickness*		X	
min thickness*		X	
std thickness*		X	
mean thickness*		X	
std thickness of LVM between LVC and RVC*			
mean thickness of LVM between LVC and RVC*			
mean circularity*	X	X	
max circumference*	X	X	
mean circumference*	X	X	
patient weight			
patient height			
Dynamic volume feature	**RV**	**MYO**	**LV**
Vmax	X	X	X
Vmin	X	X	X
dynamic ejection fraction	X	X	X
volume median	X	X	X
volume kurtosis	X	X	X
volume skewness	X	X	X
volume standard deviation	X	X	X

Fig. 3: Features Table

To extract the dynamic volume features throughout

the entire cardiac cycle. We first used the trained segmentation model to predict anatomical structures in all time steps of CMRI. Then the features in Fig.3 were calculated for each patient using the segmentation result. These features were used to train an ensemble of multilayer perceptrons (MLP) and a random forest for pathology classification. The MLP is trained for 1000 epochs using the ADAM solver with an initial learning rate of 5 e-4. The random forest was trained with 1000 trees. During resting, the softmax outputs of all MLPs were averaged and are combined sequentially with the random forest output to obtain the final ensemble prediction.

4 Experiments and Results:

Results of our modifications: By switching transposed convolution in the original UNet to bilinear upsampling, we reduced the cross-entropy loss on the training set by 5.45%. The dice scores of the validation set were improved by 1.5689%, 0.3208%, 1.358% for LV, RV, and MYO, respectively. By adding the deep supervision to our modified UNet model, the cross-entropy loss on the training set was further reduced by 4.48%. The dice scores of the validation set were further improved by 0.7%, 1.77%, 0.67% for LV, RV, and MYO, respectively. We noticed that the dice score of RV is more sensitive to the change of deep supervision than the change of bilinear upsampling. This result may imply the insertion of coarse segmentation results especially helped the classification of RV. Using our final network architecture, we achieved individual dice scores of 0.961 for LV, 0.943 for RV, and 0.9193 for MYO. (Detailed results can be found in table2). Based on the segmentation result, geometric features were extracted and utilized by an ensemble classifier to predict the diagnosis, yielding promising outcomes for the training dataset. We received a 93% accuracy for the training set.

Pathology	*Dice LV*	*Dice RV*	*Dice MYO*
DCM	0.977	0.944	0.913
HCM	0.937	0.926	0.936
MINF	0.969	0.930	0.914
NOR	0.964	0.955	0.928
RV	0.958	0.961	0.905

Fig. 4: Training set result

5 Discussion:

We achieve dice scores of 0.945(LV), 0.885(RV), 0.90(LVM) on the ACDC test set, which earned us first place in the segmentation part of the project one competition. We ranked 4th in the classification challenge. The fact that our great segmentation results didn't yield a promising classification on the test set is a little frustrating, but after listening to other groups' presentations, we realized there are two main problems in our prediction model. The first and foremost problem is overfitting. Group 4, who won first place in the prediction competition, had fewer extracted features than us but used a ten folds cross-validation. The use of cross-validation had a huge impact in this sense of avoiding overfitting. We believe by adding cross-validation and drop out to our model, decreasing the layers in the MLP, our model can be boosted significantly. The second problem is extracting inaccurate features such as the myocardium's thickness. Extracting myocardium thickness from segmentation results at apical and basal slice can be really challenging, and the algorithm we implemented (a canny edge detector that outlines the myocardium and calculates the minimum distance between the outlines) can give us false results at these slices. An example is given in Fig.5. A more robust feature extractor or simply dumping the inaccurate features can help to improve the classification as well.

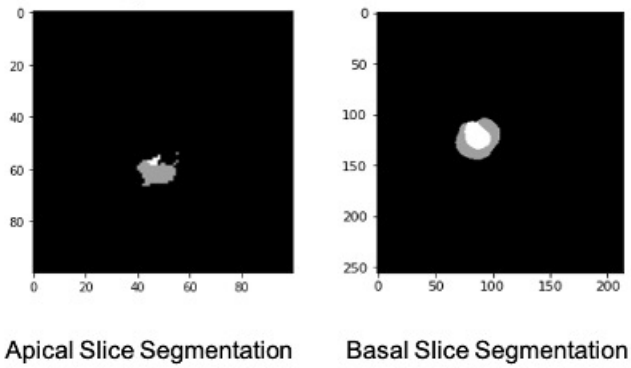

Fig. 5: Problematic segmentation results

6 Team Member Contributions:

Generally speaking, Wenkai and I did the first project. Angelina and Chenyu did the second project. In this project, Wenkai and I worked together throughout the project. Because this is our first time implementing deep learning, we always discuss together, learn from the internet together, and help each other out whenever we encountered problems. More specifically, I was in charge of designing and training the model for segmentation, extracting dynamic features, and Wenkai was in charge of data preprocessing, extracting instant features, and training the classifier. Here I want to give a huge shoutout to my teammate. We made this project happen together.

References

[1] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, X. Yang, P.-A. Heng, I. Cetin, K. Lekadir, O. Camara, M. A. G. Ballester, *et al.*, "Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved?," *IEEE transactions on medical imaging*, vol. 37, no. 11, pp. 2514–2525, 2018.

[2] J. Pierre-Marc, L. Alain, and B. Olivier, "Automated cardiac diagnosis challenge description." `https://acdc.creatis.insa-lyon.fr/description/`.

[3] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.

[5] 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*, pp. 120–129, Springer, 2017.

[6] A. Odena, V. Dumoulin, and C. Olah, "Deconvolution and checkerboard artifacts," *Distill*, 2016.

[7] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440, 2015.

[8] H. Chen, Q. Dou, L. Yu, and P.-A. Heng, "Voxresnet: Deep voxelwise residual networks for volumetric brain segmentation," *arXiv preprint arXiv:1608.05895*, 2016.

[9] Q. Dou, H. Chen, Y. Jin, L. Yu, J. Qin, and P.-A. Heng, "3d deeply supervised network for automatic liver segmentation from ct volumes," in *International conference on medical image computing and computer-assisted intervention*, pp. 149–157, Springer, 2016.

[10] B. Kayalibay, G. Jensen, and P. van der Smagt, "Cnn-based segmentation of medical imaging data," *arXiv preprint arXiv:1701.03056*, 2017.

[11] P. Medrano-Gracia, B. R. Cowan, B. Ambale-Venkatesh, D. A. Bluemke, J. Eng, J. P. Finn, C. G. Fonseca, J. A. Lima, A. Suinesiaputra, and A. A. Young, "Left ventricular shape variation in asymptomatic populations: the multi-ethnic study of atherosclerosis," *Journal of Cardiovascular Magnetic Resonance*, vol. 16, no. 1, pp. 1–10, 2014.