Framework of Echocardiographic Analysis System for Clinical Diagnosis

Haowei Kuang, Shiwei Luo, Huaitang Wang

Southern University of Science and Technology
{11910710, 21990005, 11913005}@mail.sustech.edu.cn

Abstract

Heart is the central organ of human beings, heart disease will seriously harm people's health, serious and even life safety will produce great harm. At present, the clinical diagnosis of heart disease mainly uses MRI image, CT image, echocardiography and so on. Echocardiography is the least harmful and lowest cost diagnostic method. However, the ultrasound images collected by the current equipment have a lot of noise, and the images generated are not intuitive heart images. Reading color ultrasound images requires professional doctors to rely on their own experience. To solve this problem, we hope to build a set of complete echocardiogram analysis system, which takes the original ultrasound image as input, automatically completes the ultrasonic image processing, the measurement and extraction of clinical indicators, and completes the comprehensive analysis of echocardiography based on this, and generates a complete echocardiogram analysis report. As a systematic project, what we want to accomplish in this semester is to explore and build its basic framework. In the second stage of innovation practice, we mainly completed the sub-tasks of heart failure diagnosis and super-resolution enhancement.

1. Introduction

1.1. Background

The heart is one of the most important organs in the human body. The role of the heart is to promote the blood flow, to provide adequate blood flow to organs and tissues, to supply oxygen and various nutrients, and take away the final products of metabolism, so that cells maintain normal metabolism and function.

Heart disease mainly includes heart failure, arrhythmia, coronary heart disease, coronary artery abnormalities and so on. Heart disease has the characteristics of high incidence, strong harm, progressive and sudden. According to the latest Report on Cardiovascular Health and Diseases in China 2020, there are 330 million people suffering from cardiovascular diseases in China in 2020, among which the number of heart-related diseases is more than 30 million, and the prevalence rate of the elderly over 70 years old is more than 10%. The incidence rate is still on the rise. Heart disease is extremely harmful. Cardiovascular disease is the leading cause of death in urban and rural areas in China, accounting for more than 40%. And the onset of general heart disease is insidious, the disease develops slowly, there will not be too many symptoms at the initial stage of the disease, so many people miss the best treatment time. And symptoms can erupt suddenly, and if not handled in time, death can result directly. Therefore, doctors recommend regular heart check-ups, especially for the elderly.

Current cardiac function examination methods mainly include electrocardiogram, coronary CT, coronary angiography, ultrasound and so on. The electrocardiogram can only be used to detect some functional abnormalities, unable to see the overall shape. Coronary CT and coronary angiography require injection of contrast agent, harmful to the human body, is not suitable for regular examination. Therefore, non-invasive and high accuracy ultrasound examination has become the focus of our attention.

Ultrasound examination is a medical imaging diagnosis technology based on ultrasound, which can intuitively see the size, structure and pathological lesions of muscles and internal organs. In addition to the patient's basic personal information, the ultrasonic diagnosis report is divided into three parts. The first is the sonogram obtained by machine scanning, the middle is the parameter values manually measured by the doctor on the sonogram, and the last is the diagnostic opinion manually input according to these parameter values and the sonogram.

At present, except for the ultrasound images obtained by the imaging machine, the remaining two parts need to be manually filled one by one according to the knowledge of doctors at the present stage, which causes great trouble to the ultrasound examiners.

1.2. Problem Formalization

For the above reasons, we propose the formalization of our question:

We hope to build a complete set of echocardiographic analysis system, which can automatically complete the processing of ultrasonic images and the measurement and extraction of clinical indicators with the original ultrasound images as input, and complete the comprehensive analysis of echocardiography based on this, and generate a complete echocardiographic analysis report.

To complete the system, it is necessary to carry out specific de-noising on ultrasonic images first, and then extract the required parameters by processing the images on this basis. Finally, knowledge map is constructed according to these parameters to generate diagnostic reports. This is a systematic work that requires the joint participation of many sub-projects. What we want to complete this semester is to build the framework of the complete process of the system.

1.3. Difficulty Analysis

There are many difficulties in completing this system:

1) Unlike radiological images, which are stable and easy to model, ultrasonic images are unstable due to manual manipulation.
2) The dimension of ultrasonic image is too much, and many parameters need multidimensional comprehensive analysis, which is difficult to carry out.
3) Public data sets on ultrasound images are extremely scarce.
4) Heart structure is complex, different structures have high similarity, it is very difficult to accurately segment.
5) A great deal of prior knowledge is required.

2. Related Work

2.1. Echocardiographic Analysis

Echocardiography interpretation and guidelines rely heavily on use of quantitative measures. Image processing techniques with underlying machine learning algorithms have shown promise for rapid identification of structures and quantification of related parameters. Assessment of left ventricular volume and function was one of the first applications of artificial intelligence to minimise error and reduce operator subjectivity [2, 3, 5, 23]. Methods have evolved so that, recently, Knackstedt et al. demonstrated that left ventricular ejection fraction and longitudinal strain could be analysed in approximately 8 s using machine learning methods[14] . Within 3D echocardiography, random forest models to identify borders have been shown to provide an accurate identification of left and right ventricular cavities so that derived left and right ventricular volumes are comparable to those measured by cardiac magnetic resonance [4, 15, 25, 29]. Furthermore, machine learning has been shown to aid in the assessment of valvular heart disease, for example, mitral valve disease [6, 7]. Automated assessments of 3D transoesophageal echocardiograms of the mitral valve provided more reproducible and consistent quantitative assessment of the mitral valve annulus size and its morphology than human interpretation [11, 17]. An extensive work also has been done in the field of aortic valve segmentation for planning transthoracic aortic valve implantation procedure[20, 21, 22]. But these studies are currently only looking at specific parts of echocardiography, with the aim of directly diagnosing disease based on echocardiography.

2.2. Diagnostic Report Generation

Automatically medical report generation is a significant and difficult task, which need to interpret and summarize the insights gained from medical images such as radiography or biopsy samples [9]. It needs accurate abnormality detection and state-of-art detailed image caption generation. It can be treated as the image captioning models' application to the medical domain. There are some existing studies to automatically generate medical reports or annotations for medical images[9, 10, 12, 16, 24]. Most existing studies are based on two public medical image-caption pair datasets.

The first is ImageCLEFcaption [9, 10]. ImgeCLEFcaption is an evaluation campaign about medical concept detection and report generation. The organization provides a large scale dataset of 184,000 image-caption pairs. The second dataset is the public IU X-Ray dataset [8], which contains 7,470 pairs of radiology images and reports. There are several other studies which adopt paragraph generation models to generate medical reports for X-Ray images. Shin and et al[24]. firstly present a deep learning model to efficiently detect abnormalities from an image and annotate its contexts. Jing and et al. [12] propose to leverage CNN to detect tags, and then combine the hierarchical recurrent network and attention mechanism to generate detailed medical reports.

However, among the methods of more than 40 relevant papers so far, all of them are aimed at radiological images without exception, and the data set is also limited. So far, we have not seen any methods related to the diagnostic report generation of ultrasonic images, especially echocardiography.

3. Dataset

Due to the lack of information in the original image dataset, we use EchoNet Dynamic dataset[18] (Figure 2) for development. The EchoNet-Dynamic dataset contains 10,030 echocardiography videos, spanning the range of typical echocardiography lab imaging acquisition conditions, with corresponding labeled measurements including ejection fraction, left ventricular volume at endsystole and enddiastole, and human expert tracings of the left ventricle as an aid for studying machine learning approaches to evaluate cardiac function. A standard full resting echocardiogram study consists of a series of videos and images visualizing the heart from different angles, positions, and image acquisition techniques. The dataset contains 10,030 apical-4-chamber echocardiography videos from individuals who underwent imaging between 2016 and 2018 as part of routine clinical care at Stanford University Hospital. Each video was cropped and masked to remove text and information outside of the scanning sector. The resulting images were then downsampled by cubic interpolation into standardized 112x112 pixel videos. In addition to the video itself, each study is linked to clinical measurements and calculations obtained by a registered sonographer and verified by a level 3 echocardiographer in the standard clinical workflow. A

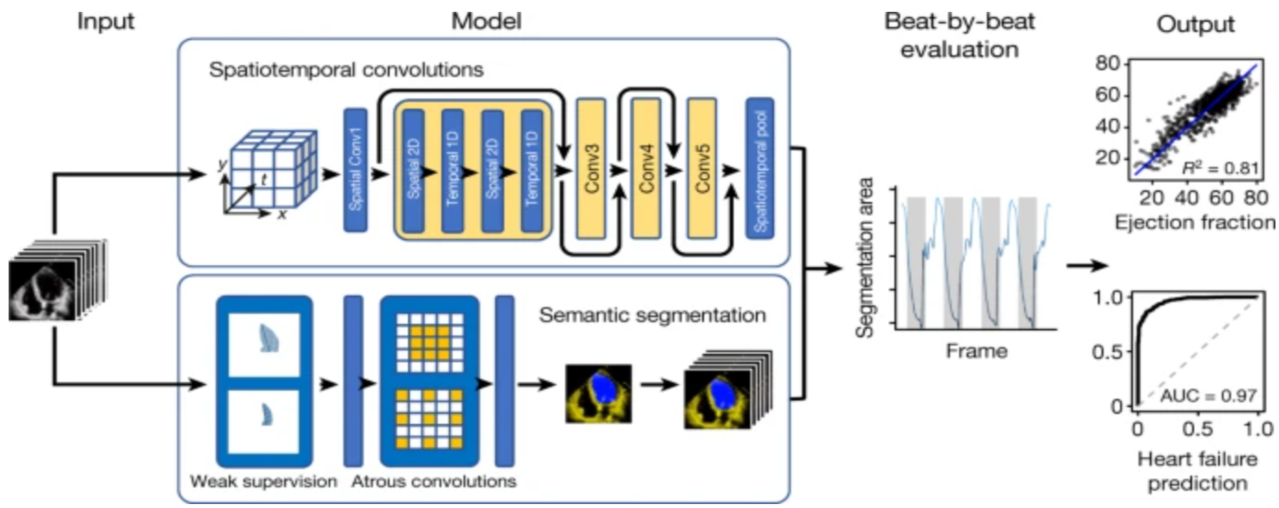

Figure 1. [18]For each patient, EchoNet-Dynamic uses standard apical four-chamber view echocardiogram videos as input. The model first predicts the ejection fraction for each cardiac cycle using spatiotemporal convolutions with residual connections and generates frame-level semantic segmentations of the left ventricle using weak supervision from expert human tracings. These outputs are combined to create beat-to-beat predictions of the ejection fraction and to predict the presence of heart failure with reduced ejection fraction. AUC, area under the curve

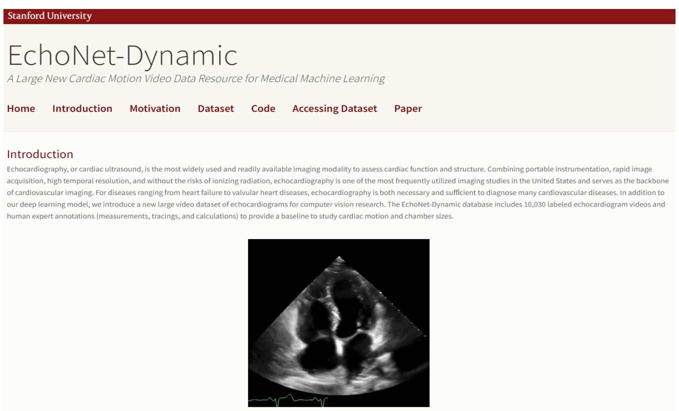

Figure 2. EchoNet-Dynamic Dataset[18]

central metric of cardiac function is the left ventricular ejection fraction, which is used to diagnose cardiomyopathy, assess eligibility for certain chemotherapies, and determine indication for medical devices. The ejection fraction is expressed as a percentage and is the ratio of left ventricular end systolic volume (ESV) and left ventricular end diastolic volume (EDV) determined by (EDV - ESV) / EDV. Besides, in this dataset, for each video, the left ventricle is traced at the endocardial border at two separate time points representing end-systole and end-diastole. Each tracing is used to estimate ventricular volume by integration of ventricular area over the length of the major axis of the ventricle. The expert tracings are represented by a collection of paired coordinates corresponding to each human tracing. The first pair of coordinates represent the length and direction of the long axis of the left ventricle, and subsequent coordinate pairs represent short axis linear distances starting from the apex of the heart to the mitral apparatus. Each coordinate pair is also listed with a video file name and frame number to identify the representative frame from which the tracings match.

4. Our Work

According to the Suggestions provided by the teachers in the first defense, we based on the established plan of the first stage of defense, the first order of the diagnosis of heart failure is the starting point, and the four cavity section video data is used in the heart, and the ultrasonic echocardiography of the diagnostic task of heart failure is completed. In addition, in order to benefit the doctor's observation and subjective judgment of raw data, we also completed the hyperresolution enhancement of the heart hypervideo data. In the end, we integrated these work together, and we initially generated our echocardiography report Figure 3.

4.1. Cardiac failure

Cardiac failure, refers to the inability of the heart to be able to reduce the amount of blood in the circulation vein, causing the blood of the venous system to be filled with blood, and the disease of the heart circulatory disorder. The use of echocardiogram hyperdiagnosis cardiac failure is mainly done by calculating the EF. He needs the doctor to measure the total volume and shrinkage of the left ventricle in the left ventricle of the heart, and calculate the percentage of the volume of the final volume of the ventricular diastolic phase by calculating the amount of output per stroke. Then there was a problem of cardiac failure based on the EF. Because the diagnostic index of cardiac failure is simple, it is easier to complete the whole process. And it is helpful to complete of the project, and so on, so we choose the diagnosis of cardiac failure as the first step in the analysis of the ultrasonic echocardiography. In the case of computer assisted diagnosis of cardiac failure, we need to divide the starting point of the center of the body cycle, and then the measurement of the EF, and then predict the prediction according to the previous results. To complete the computer aided diagnosis of cardiac failure, we need to divide the left ventricle and then calculate the EF, and then predict the results based on the previous results.

4.2. Diagnostic model for cardiac failure

Our work is based on the Echonet-Dynamic [18] model published in Nature. EchoNet-Dynamic has three key components (Figure 1). First, we constructed a CNN model with atrous convolutions for frame-level semantic segmentation of the left ventricle. The technique of atrous convolutions enables the model to capture larger patterns and has previously been shown to perform well on non-medical imaging datasets [27]. The standard human clinical workflow for estimating the ejection fraction requires manual segmentation of the left ventricle during end systole and end diastole. We generalize these labels in a weak supervision approach with atrous convolutions to generate frame-level semantic segmentation throughout the cardiac cycle in a 1:1 pairing with frames from the original video. The automatic segmentation is used to identify ventricular contractions and provides a clinician-interpretable intermediary that mimics the clinical workflow.

Second, we trained a CNN model with residual connections and spatiotemporal convolutions across frames to predict the ejection fraction. In contrast to previous CNN architectures for machine learning of medical images, our approach integrates spatial as well as temporal information in our network convolutions[19, 27, 28]. Spatiotemporal convolutions, which incorporate spatial information in two dimensions as well as temporal information in the third dimension, have previously been used in non-medical video-classification tasks[27, 28]. However, this approach has not previously been used for medical data given the relative scarcity of labelled medical videos. We additionally performed a model architecture search to identify the optimal base architecture (Figure 1).

Finally, we make video-level predictions of the ejection fraction for beat-to-beat estimations of cardiac function. Given that variation in cardiac function can be caused by changes in loading conditions as well as heart rate in a variety of cardiac conditions, it is recommended to perform estimations of the ejection fraction for up to five cardiac cycles; however, this is not always done in clinical practice given the tedious and laborious nature of the calculation. Our model identifies each cardiac cycle, generates a clip of 32 frames and averages clip-level estimates of the ejection fraction for each beat as test-time augmentation. Finally, based on relevant clinical knowledge, we established a knowledge map for the use of EF in the diagnosis of heart failure in Asian patients, which guided our prediction

4.3. Parameter Extraction

Combined with the situation of the dataset we have and the recommendations of the doctor in the group, we have finished extraction of the following parameters, the length, diameter, area, volume, and ejection fraction of the left ventricle at the end of diastolic and systolic stages. Based on the existing segmentation results, We also finished the extraction of end diastolic long diameter, end diastolic transverse diameter, end systolic long diameter, and end systolic transverse diameter.

The algorithm we used is EchoNet, an end-to-end deep learning approach for labelling of the left ventricle and estimation of the ejection fraction from input echocardiogram videos alone. It first performs frame-level semantic segmentation of the left ventricle with weakly supervised learning from clinical expert labelling. Then, a three-dimensional convolutional neural network (CNN) with residual connections predicts clip-level ejection fraction from the native echocardiogram videos. Finally, the segmentations results are combined with clip-level predictions to produce beat-to-beat evaluation of the ejection fraction (Figure 5).

4.4. Evaluation Metrics

For the test dataset from Stanford Medicine that was not previously seen during model training, in the original paper, the prediction of the ejection fraction by EchoNet-Dynamic had a mean absolute error of 4.1annotations by human experts, and our result had a mean absolute error of 3.90squared error of 5.17paper, our self-trained model achieved the highly similar results. This is well within the range of typical measurement variation between different clinicians, which is usually described as interobserver variation and can be as high as 13.9of less than 50under the curve of 0.97. In addition, we performed re-evaluation of the videos by blinded clinicians in cases in which the prediction of the ejection fraction by EchoNet-Dynamic diverged the most from the original human annotation. Many of these videos had inaccurate initial human labels (in 43image quality, or arrhythmias and variations in the heart rate.

4.5. Ultrasonic image enhancement

Due to the low resolution and significant noise of existing echocardiographic data, it can interfere with the clinician's observation. In order to solve this problem, we carried out work related to ultrasound image enhancement. Based on the preliminary research and pre-experimental results, we finally selected the Real-ESRGAN model (Figure 3) for ultrasound image enhancement, which is by far the most effective model for image and video super-resolution reconstruction. Our input is a low-resolution cardiac ultrasound image, and the output is a high-resolution image obtained by network reconstruction.

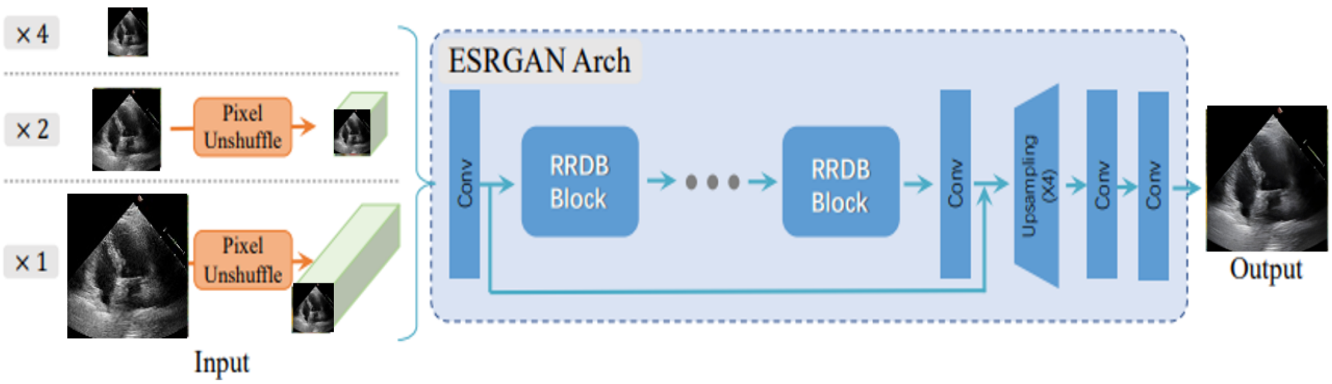

Figure 3. Real-ESRGAN [30]

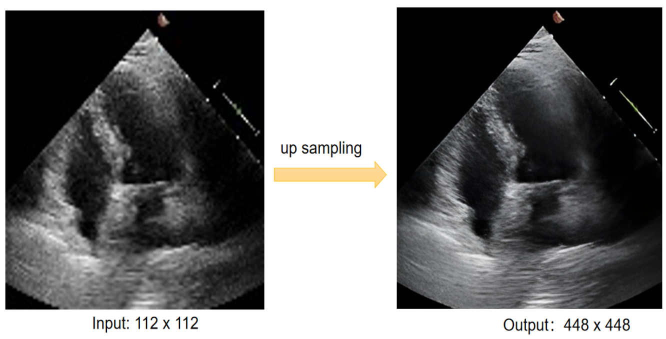

Figure 4. Comparison between input and output

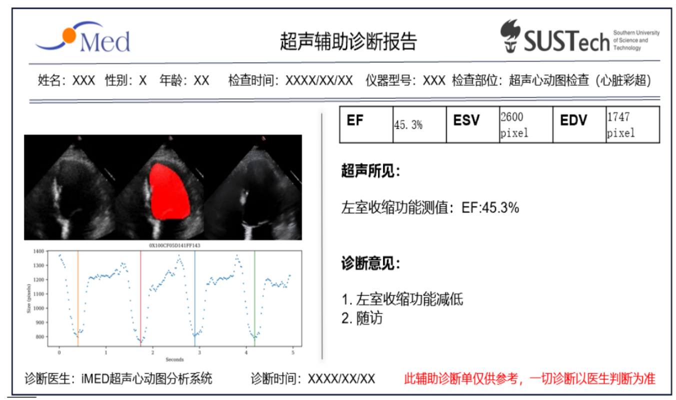

Figure 5. Auxiliary diagnostic report for heart failure

4.6. Datasets and Implementation

Due to the lack of high-definition echocardiographic video data, we adopt DIV2K [1], Flickr2K [26] and OutdoorSceneTraining [31] datasets for training. The training HR patch size is set to 256. Adam optimizer is selected in the training process. Real-ESRNet is finetuned from ESRGAN for faster convergence. RealESRGAN is trained with a combination of L1 loss, perceptual loss and GAN loss, with weights 1, 1, 0.1, respectively. It uses the conv1, ...conv5 feature maps (with weights 0.1, 0.1, 1, 1, 1) before activation in the pre-trained VGG19 network [13] as the perceptual loss.

4.7. Report Generation

According to the results of hyperfraction, segmentation and ejection fraction obtained above, we generated an auxiliary diagnostic report for heart failure (Figure 5). The report includes patient information, ultrasound images, left ventricular segmentation results, hyperfraction results, cardiac parameters, and diagnostic advice.

We use the knowledge of the Left-Handed Doctor Medical Knowledge Map and the Si Zhi Chinese Knowledge Map to interpret the conclusions of our supplementary diagnostic reports for patients' initial understanding of their conditions and to give them some reference advice.

4.8. Knowledge Graph Construction

We have already established a preliminary knowledge graph (Figure 6), and we will continue to expand the content of the knowledge graph in the future. The final knowledge graph will contain not only the parameters that we have calculated, but also many parameters that are helpful in generating diagnostic opinions that we have not calculated, and whose calculations will be completed by others in the future.

4.9. Platform Building and Function Embedding

Finally, we have built an analysis platform [7] for echocardiography, integrating the functions we have completed, such as ultrasonic image processing, left ventricular segmentation, parameter extraction and diagnostic report generation. At the same time, the platform will allow others to add more features to improve the echocardiographic analysis system.

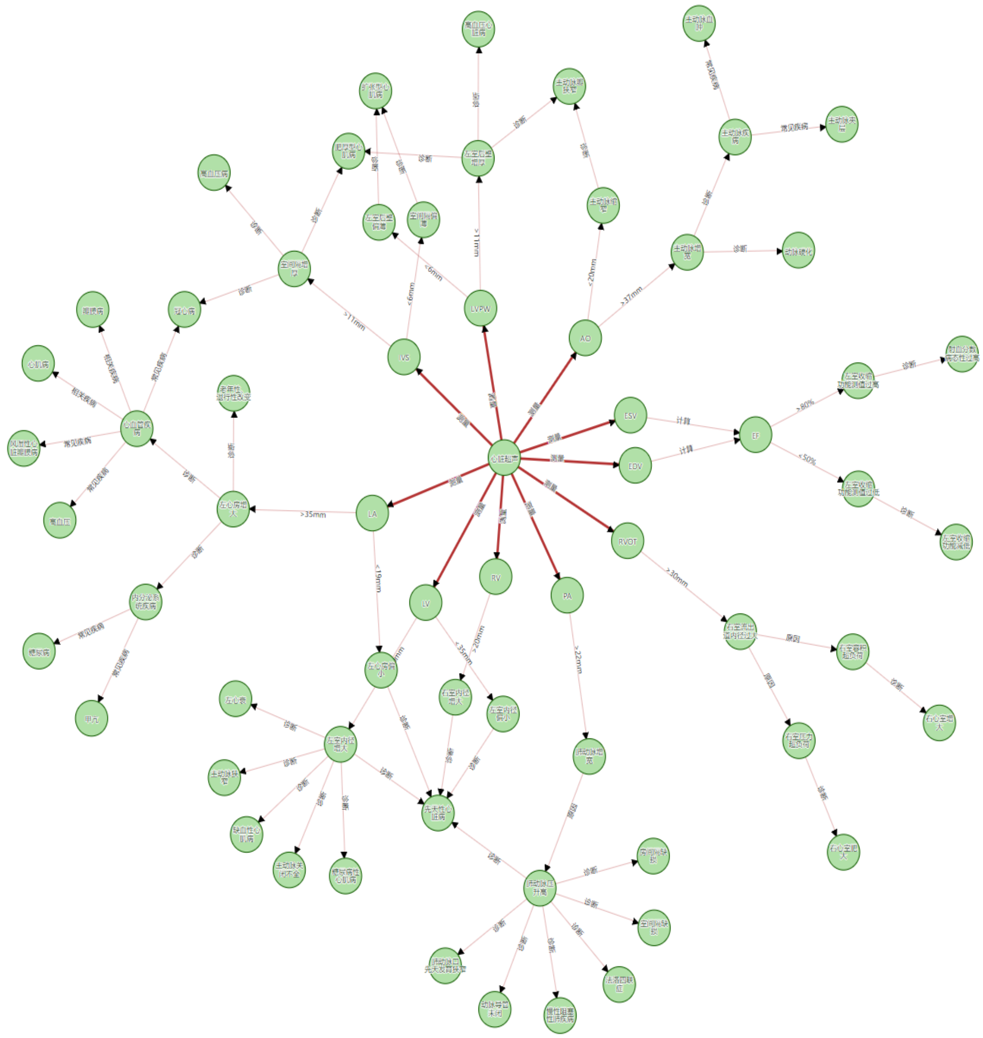

Figure 6. Simple diagnostic knowledge atlas of heart failure

5. Conclusion

We have completed the whole process of cardiac ultrasound diagnosis and the construction of the knowledge graph as the framework of the system. In addition, we have completed the construction of a cardiac ultrasound diagnostic platform and the integration of the eye-brain-heart linkage system platform.

6. Future Work

In the future, we will continue to improve the capabilities of the diagnostic cardiac ultrasound system and expand the content of the knowledge graph. We will extract more parameters based on the information contained in the knowledge graph and perform the corresponding diagnosis and analysis. In addition, we will optimize the convenience of the ELIA platform.

Figure 7. ELIA platform

References

[1] Eirikur Agustsson and Radu Timofte. "Ntire 2017 challenge on single image super-resolution: Dataset and study". In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2017, pp. 126–135.

[2] Maryam Alsharqi et al. "Artificial intelligence and echocardiography". In: *Echo research and practice* 5.4 (2018), R115–R125.

[3] Gregg Belous, Andrew Busch, and David Rowlands. "Segmentation of the left ventricle from ultrasound using random forest with active shape model". In: *2013 1st International Conference on Artificial Intelligence, Modelling and Simulation*. IEEE. 2013, pp. 315–319.

[4] Olivier Bernard et al. "Standardized evaluation system for left ventricular segmentation algorithms in 3D echocardiography". In: *IEEE transactions on medical imaging* 35.4 (2015), pp. 967–977.

[5] Maxime Cannesson et al. "A novel two-dimensional echocardiographic image analysis system using artificial intelligence-learned pattern recognition for rapid automated ejection fraction". In: *Journal of the American College of Cardiology* 49.2 (2007), pp. 217–226.

[6] Frederick C Cobey. "Intelligent algorithms in perioperative echocardiography: a new era". In: *Journal of the American Society of Echocardiography* 30.10 (2017), A26–A27.

[7] Frederick C Cobey et al. "The emperor has no clothes: recognizing the limits of current echocardiographic technology in perioperative quantification of mitral regurgitation". In: *Journal of Cardiothoracic and Vascular Anesthesia* 31.5 (2017), pp. 1692–1694.

[8] Bo Dai et al. "Towards diverse and natural image descriptions via a conditional gan". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2970–2979.

[9] Carsten Eickhoff et al. "Overview of ImageCLEF-caption 2017–image caption prediction and concept detection for biomedical images". In: *CLEF 2017 working Notes* 1866.Workin (2017).

[10] Sadid A Hasan et al. *Overview of imageclef 2018 medical domain visual question answering task*. Tech. rep. 10-14 September 2018, 2018.

[11] Jelliffe Jeganathan et al. "Artificial intelligence in mitral valve analysis". In: *Annals of cardiac anaesthesia* 20.2 (2017), p. 129.

[12] Baoyu Jing, Pengtao Xie, and Eric Xing. "On the automatic generation of medical imaging reports". In: *arXiv preprint arXiv:1711.08195* (2017).

[13] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. "Perceptual losses for real-time style transfer and super-resolution". In: *European conference on computer vision*. Springer. 2016, pp. 694–711.

[14] Christian Knackstedt et al. "Fully automated versus standard tracking of left ventricular ejection fraction and longitudinal strain: the FAST-EFs multicenter study". In: *Journal of the American College of Cardiology* 66.13 (2015), pp. 1456–1466.

[15] Franck Levy et al. "Performance of new automated transthoracic three-dimensional echocardiographic software for left ventricular volumes and function assessment in routine clinical practice: Comparison with 3 Tesla cardiac magnetic resonance". In: *Archives of cardiovascular diseases* 110.11 (2017), pp. 580–589.

[16] Yuan Li et al. "Hybrid retrieval-generation reinforced agent for medical image report generation". In: *Advances in neural information processing systems* 31 (2018).

[17] Hanie Moghaddasi and Saeed Nourian. "Automatic assessment of mitral regurgitation severity based on extensive textural features on 2D echocardiography videos". In: *Computers in biology and medicine* 73 (2016), pp. 47–55.

[18] David Ouyang et al. "Video-based AI for beat-to-beat assessment of cardiac function". In: *Nature* 580.7802 (2020), pp. 252–256.

[19] Ryan Poplin et al. "Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning". In: *Nature Biomedical Engineering* 2.3 (2018), pp. 158–164.

[20] Edgard A Prihadi et al. "Feasibility, accuracy, and reproducibility of aortic annular and root sizing for transcatheter aortic valve replacement using novel automated three-dimensional echocardiographic software: comparison with multi–detector row computed tomography". In: *Journal of the American Society of Echocardiography* 31.4 (2018), pp. 505–514.

[21] Sandro Queirós et al. "Fully automatic 3-D-TEE segmentation for the planning of transcatheter aortic valve implantation". In: *IEEE Transactions on Biomedical Engineering* 64.8 (2016), pp. 1711–1720.

[22] Sandro Queirós et al. "Validation of a novel software tool for automatic aortic annular sizing in three-dimensional transesophageal echocardiographic images". In: *Journal of the American Society of Echocardiography* 31.4 (2018), pp. 515–525.

[23] Hind W Rahmouni et al. "Clinical utility of automated assessment of left ventricular ejection fraction using artificial intelligence–assisted border detection". In: *American heart journal* 155.3 (2008), pp. 562–570.

[24] Hoo-Chang Shin et al. "Learning to read chest x-rays: Recurrent neural cascade model for automated image annotation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 2497–2506.

[25] Richard V Stebbing et al. "Data-driven shape parameterization for segmentation of the right ventricle from 3D+ t echocardiography". In: *Medical image analysis* 21.1 (2015), pp. 29–39.

[26] Radu Timofte et al. "Ntire 2017 challenge on single image super-resolution: Methods and results". In: *Proceedings of the IEEE conference on com-*

puter vision and pattern recognition workshops. 2017, pp. 114–125.
[27] Du Tran et al. “A closer look at spatiotemporal convolutions for action recognition”. In: *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 2018, pp. 6450–6459.
[28] Du Tran et al. “Learning spatiotemporal features with 3d convolutional networks”. In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 4489–4497.
[29] Wendy Tsang et al. “Transthoracic 3D echocardiographic left heart chamber quantification using an automated adaptive analytics algorithm”. In: *JACC: Cardiovascular Imaging* 9.7 (2016), pp. 769–782.
[30] Xintao Wang et al. “Real-esrgan: Training real-world blind super-resolution with pure synthetic data”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 1905–1914.
[31] Xintao Wang et al. “Recovering realistic texture in image super-resolution by deep spatial feature transform”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 606–615.