Technical report about Cardiac Volume estimation through Regression CNN on ACDC dataset

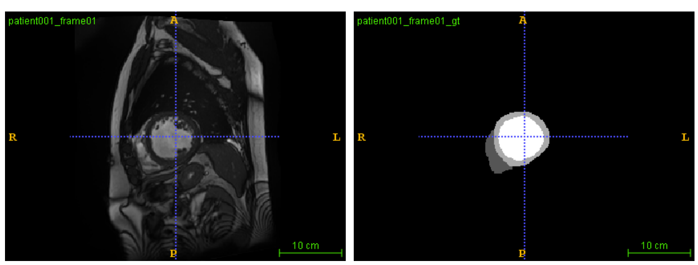
Jing Zhang

April 2021

1 Task description

The task in this work is to predict the volume of cardiac structures using deep learning techniques. Specifically, we want to achieve **a direct multi-class volume estimation model without segmentation intervention.**
This idea is not original, it has been applied in many papers [10, 5].They cover different medical indices prediction and different techniques including traditional machine learning and deep learning in recent decades.

2 Data description



(a) MRI image of a patient (b) Annotations of a patient

Figure 1: One slice of MRI cardiac subject and its ground truth annotated by clinical expert in ACDC dataset. The white color region in (b) is LV (class 3), the light grey region is MYO (label 2), the dark grey region is RV (class 1), the black region is background (class 0), the volume of RV, MYO LV are 139.72 ml, 164.26 ml, 295.51 ml respectively according to Formula 1.

The public "Automatic Cardiac Diagnosis Challenge" dataset (ACDC) dataset [1] is used in this study. It has 100 magnetic resonance images (MRI) subjects in

training set, each subject has 3 manual annotated labels, i.e., left ventricular (LV), myocardium (MYO), right ventricle (RV). Each subject has end diastolic (ED) and end systolic (ES) phase. It also has 4 types of disease, myocardial infarction (MINF), dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM), abnormal right ventricle (ARV), and patients with normal cardiac (NOR). Fig. 1 is the MRI image of a patient and its ground truth annotated by experts.
The volume (ml) of each structure is calculated as below:

$$\text{Volume} = \begin{cases} \text{RV} = \sum_{i}^{N} \text{S\_RV} * \text{px\_x} * \text{px\_y} * \text{space\_z}/1000 \\ \text{MYO} = \sum_{i}^{N} \text{S\_MYO} * \text{px\_x} * \text{px\_y} * \text{space\_z}/1000 \\ \text{LV} = \sum_{i}^{N} \text{S\_LV} * \text{px\_x} * \text{px\_y} * \text{space\_z}/1000 \end{cases} \tag{1}$$

In Formula 1, N is the number of slice in each subject, S_RV is the summary of pixels belong to RV class in one slice (same as S_MYO, S_LV), px_x (mm) is pixel size of x dimension, px_y (mm) is pixel size of y dimenstion, $space\_z$ (mm) is slice thickness.

From Fig. 1 we can tell the features of the dataset are that:
1. The cardiac accounts for a small proportion of the image;
2. The RV and MYO are less clear than the LV;
Besides, according to the author's observation, **for each subject, the shape is different, and the number of slice is also different. Certain slices don't have RV structure or even only have background (black) pixels.** Therefore, it's necessary to preprocess this dataset before using deep learning models on them.

3 Method description

3.1 Data preprocessing

• Data Screening

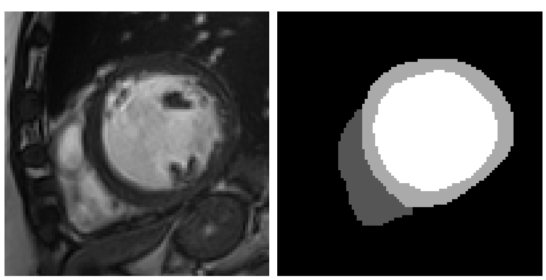
To make sure the ground truth volumes offered by the authors of ACDC dataset are correct, we check the consistence of given volumes and volumes computed through Formula 1. Among 100 pairs subjects (including ED and ES), the given volumes of cardiac of 6 patients are seriously deviated from the computed volumes, the difference are up to hundred level. See Table 1. While the given volumes in left 94 pairs patients are in normal range, because their difference are less than 1. Thus, we will remove these 6 patients in the experiments.

Table 1: Abnormal samples in ACDC dataset, V_g is volume from given data, V_c is volume from calculation (see Formula 1), Diff is the difference between V_g and V_c.

Patient	spacing(x,y,z)	shape	V_g(ml)	S(pixel)	V_c(ml)	**Diff**
P019	(1.445,1.445,10)	(11, 256, 216)	868.59	32269	673.78	194.81
P078	(1.367,1.367,10)	(8, 256, 216)	630.20	25813	482.36	147.84
P079	(1.367,1.367,10)	(9, 256, 216)	455.74	18667	348.83	106.91
P080	(1.758,1.758,10)	(6, 256, 216)	223.95	9173	283.50	59.55
P093	(1.563,1.563,7)	(10, 224, 180)	57.35	23491	401.71	344.35
P099	(1.786,1.786,5)	(16, 224, 154)	866.71	27180	433.49	433.21

• Data Cropping

As Section 2 described, the original data they have different sizes in 3 dimensions, and the cardiac is small in each slice. In order to ensure that there are no basic errors in the model, and to enable the model to fully learn the features of the cardiac, we **crop the data to a fixed size and focus on the cardiac only**. The cropping steps are as follows:
1. Detecting and finding the largest bounding box of cardiac in the whole slices of ground truth subject.
2. Performing Step 1 among all the subjects.
3. Creating a new bounding box that just covers all the detected bounding box of each subject.[1]
4. Cropping the whole ground truths and MRI images based on the new bounding box, the size of a 2D subject is (100,100).
After cropping, the cardiac is preserved to the greatest extent but no other parts compared to Fig. 1, see Fig. 2.



(a) Cropped MRI image (b) Cropped ground truth

Figure 2: One slice of cropped MRI cardiac subject and its ground truth, the size is 100*100.

[1] In this step, we remove 3 extreme examples because the size of them are largely different from others.

• Data Resampling

In original data, the number of slices are different from each other (from 6 to 17 slices). Moreover, there are slices that don't have cardiac labels but only have background pixels. To make sure the model does not go wrong and avoid invalid information, we **fix the number of slice of each subject as 9 and remove blank slices**. The final shape of all the subjects will be (100,100,9). The resampling steps are as follows:
1. Detecting slices that only have background pixels and removing them.
2. Duplicating the tail slice to the bottom of subject until the number of slice comes to 9.
3. Removing the tail slice from the bottom of subject until the number of slice comes to 9.

• Data Normalization & Augmentation

We perform data normalization both in MRI images ($(\mathrm{img} - \mu)/\sigma$) and ground truth volumes ($\mathrm{gt}/\max(\mathrm{gt})$). To enlarge the training set, data augmentation including flipping, translation and rotation are performed. In final, 728 pairs of subjects are used in this experiments.

3.2 Architecture

We design a deep regression CNN architecture shown in Fig. 3. We use VGG16 [7] as backbone to learn the feature from training dataset. For the regression part, we simply use linear regression to predict the volume of 3 structures of cardiac. The networks are optimized by regression loss function such as mean absolute error loss or mean square error loss or huber loss [3].

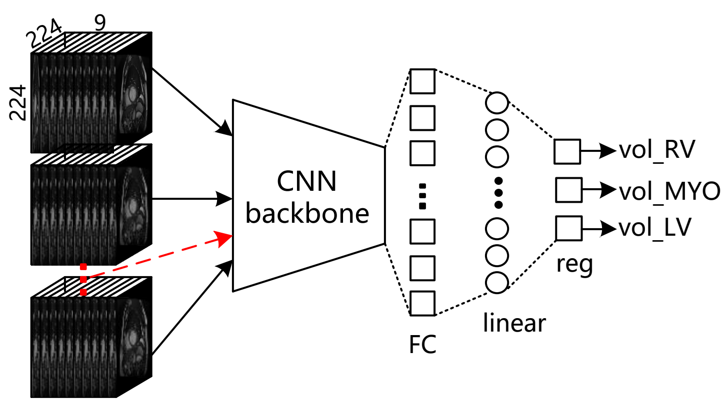


Figure 3: The architecture of regression CNN for predicting the volume of RV, MYO, LV directly. The input data are preprocessed MRI images (224,224,9), the ground truth are the volume of 3 structures of cardiac. FC is fully connected layer after feature representation (the backbone is VGG16), the FC layer goes through linear regression layer, the output is the predicted volumes (vol_RV, vol_MYO, vol_LV).

4 Experiment description

4.1 Configuration

The training set is 600 MRI images (224,224,9), the validation set is 68, the test set is 60. The optimizer is Adam. The learning rate is $1e^{-3}$, the batch size is 16. The algorithm is completed using Python and Keras library with GPU p100. The training epoch is 100. 5-fold cross validation is performed. The structure of regression CNN is: 2D convolutional kernel with multi class volume prediction. For example, the input shape is (N,224,224,3), N is the number of input images. In the experiments, we train two different VGG backbones, which are VGG16 [7] and EfficientNet [8] respectively. See Table 3.

4.2 Results

Table 2 shows the mean absolute error (mae) between ground truth and predicted volumes. From the analysis results we can see that these 2 models have similar performance no matter it is multi class or multi output branches, the mae are all large, they can not well predict the volume of cardiac. The predicting effect is not as good as the expected predicting effect (the best mae is less than 10 ml in [1] Table VIII). Moreover, the mae of RV structure is larger than the other structures, which indicates that it is more difficult to predict than the other two structures.

Table 2: Mean absolute error of test set after 5-fold cross validation on different models and images. The first model predicts three classes of cardiac, the second model have three branches, each branch output one class of cardiac.(these results are old)

Structure	RV(ml)	MYO(ml)	LV(ml)
SingleModel-3Output	41.17(±26.68)	13.85(±11.56)	27.05(±13.23)
3Model-OneOutputPerModel	56.75(±41.01)	**12.62(**±9.03**)**	16.07(±13.60)

Table 3: Configuration of three models

models	inputShape	batchsize	learning rate	# para	time
VGG-1	(224,224,9)	16	1e-3	69.3M	5m
VGG-2	(224,224,9)	8	1e-4	69.3M	11m
VGG-pre	(224,224,3)	8	1e-4	138M	3H31m
EFN-pre-1	(224,224,3)	16	1e-3	9.2M	6h08m
EFN-pre-2	(224,224,3)	8	1e-4	9.2M	7h51m

We have done several supplementary experiments. We change different models, input shape, as well as hyper parameters including batchsize and learning rate during training. See Table 3. The performance of these models are in Table 4.

Table 4: performance of three models

	RV		MYO		LV	
models	mae(ml)	pame(%)	mae(ml)	pmae(%)	mae(ml)	pmae(%)
VGG-1	52.94(±32.86)	28.37(±17.31)	38.12(±20.02)	47.62(±29.15)	60.91(±33.69)	48.83(±35.294)
VGG-2	**8.24**(±7.99)	**5.20**(±5.21)	**6.28**(±6.09)	**5.73**(±7.20)	**8.82**(±9.05)	**5.77**(±5.84)
VGG-pre	16.91(±17.22)	12.22(±8.93)	**6.84**(±7.86)	**4.40**(±3.56)	22.45(±24.12)	13.75(±8.38)
EFN-pre-1	66.55(±47.72)	58.64(±62.91)	45.23(±34.07)	37.96(±29.19)	95.29(±46.89)	88.48(±79.88)
EFN-pre-2	**2.17**(±1.83)	**2.52**(±3.07)	**1.39**(±1.27)	**1.21**(±1.28)	**1.91**(±1.51)	**2.24**(±2.82)

5 Discussion & Conclusion

In this report, we use regression CNN to directly predict the volume of cardiac structures without segmentation on the ACDC dataset.
We spare a lot of effort on preprocessing the MRI images before putting them in the model for training, they are : data screening for consistency verification, data cropping and data resampling for unifying the size of image, as well as data normalization and augmentation.
The idea of regression CNN is quite simple, it consists of a CNN backbone and a regression layer. And this idea has been successfully used in head circumference prediction [9].
However, the prediction effect of the model in this experiment is not ideal. It is mainly reflected in two aspects.

- **The prediction error of this model is large.**
- **The error deviation between cardiac structures is large.**

The author believes that this is due to:

- **the complexity of the data.**
- **the singularity of the model.**

So it is reasonable to start to improve the performance of the model from these two perspectives.

- Performing slice selection. One possible solution is to try to keep the slices with the most obvious features of the cardiac MRI data, for example, we extract the middle slices (1-3 slices) from a MRI image. This similar idea has been used in [5, 6].
- Changing architectures. Change VGG model to a more effective model. Or combine the image segmentation network with regression output as a multi-task model, so that not only useful information can be learned from the input image, but also useful information can be learned from the segmentation results.

6 Ablation experiments

• **Different CNN backbones** - We investigate different CNN backbones in regression CNNs. They are VGG16 [7], ResNet50 [4] and EfficientNet [8]. All of them are pretrained on public dataset ImageNet [2]. So they have good ability of feature representation.
• **Different data slices** - We study the influence of data format on cardiac volume prediction. We compare the cardiac MRI with 1 slice and 9 slices respectively. In 1 slice, we select the top slice of cardiac data which has clear feature of three structures.
• **Data augmentation** - The original ACDC dataset only has 100 subjects. So we will train the regression CNN model with augmented data and original data to see the model performance.
• **Learning rate and batch size** - For the choice of hyper parameters, we set different learning rate and batch size to see the final results.

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] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[3] Ashkan Esmaeili and Farokh Marvasti. A novel approach to quantized matrix completion using huber loss measure. *IEEE Signal Processing Letters*, 26(2):337–341, 2019.

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[5] Gongning Luo, Suyu Dong, Wei Wang, Kuanquan Wang, Shaodong Cao, Clara Tam, Henggui Zhang, Joanne Howey, Pavlo Ohorodnyk, and Shuo Li. Commensal correlation network between segmentation and direct area estimation for bi-ventricle quantification. *Medical image analysis*, 59:101591, 2020.

[6] Gongning Luo, Wei Wang, Clara Tam, Kuanquan Wang, Shaodong Cao, Henggui Zhang, Bo Chen, and Shuo Li. Dynamically constructed network with error correction for accurate ventricle volume estimation. *Medical Image Analysis*, 64:101723, 2020.

[7] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*, May 2015.

[8] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*, pages 6105–6114. PMLR, 2019.

[9] Jing Zhang, Caroline Petitjean, Pierre Lopez, and Samia Ainouz. Direct estimation of fetal head circumference from ultrasound images based on regression cnn. In *Medical Imaging with Deep Learning*, pages 914–922. PMLR, 2020.

[10] Xiantong Zhen, Zhijie Wang, Ali Islam, Mousumi Bhaduri, Ian Chan, and Shuo Li. Direct estimation of cardiac bi-ventricular volumes with regression forests. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 586–593. Springer, 2014.