Segmentation of biventricle in cardiac cine MRI via nested capsule dense network

Jilong Zhang [Equal first author, 1] , **Yajuan Zhang** [Equal first author, 1] , **Hongyang Zhang** [1] , **Quan Zhang** [Corresp., 2, 3] , **Weihua Su** [1] , **Shijie Guo** [4] , **Yuanquan Wang** [Corresp. 1, 4]

1 School of Artificial Intelligence, Hebei University of Technology, Tianjin, China

2 School of Information and Communication Engineering, North University of China, Taiyuan, China

3 Shanxi Provincial Key Laboratory for Biomedical Imaging and Big Data, North University of China, Taiyuan, 030051, China

4 Hebei Key Laboratory of Robot Perception and Human-Robot Interaction, HeBUT, Tianjin, China

Corresponding Authors: Quan Zhang, Yuanquan Wang
Email address: zhangibmet@nuc.edu.cn, wangyuanquan@scse.hebut.edu.cn

Background. Cardiac magnetic resonance image (MRI) has been widely used in diagnosis of cardiovascular diseases because of its noninvasive nature and high image quality. The evaluation standard of physiological indexes in cardiac diagnosis is essentially the accuracy of segmentation of left ventricle (LV) and right ventricle (RV) in cardiac MRI. The traditional symmetric single codec network structure such as U-Net tends to expand the number of channels to make up for lost information that results in the network looking cumbersome.

Methods. Instead of a single codec, we propose a multiple codecs structure based on the FC-DenseNet(FCD) model and capsule convolution-capsule deconvolution, named *Nested Capsule Dense Network (NCDN)*. NCDN uses multiple codecs to achieve multi-resolution, which makes it possible to save more spatial information and improve the robustness of the model.

Results. The proposed model is tested on three datasets that include the York University Cardiac MRI dataset, Automated Cardiac Diagnosis Challenge (ACDC-2017), and the local dataset. The results show that the proposed NCDN outperforms most methods. In particular, we achieved nearly the most advanced accuracy performance in the ACDC-2017 segmentation challenge. This means that our method is a reliable segmentation method, which is conducive to the application of deep learning-based segmentation methods in the field of medical image segmentation.

1 **Segmentation of Biventricle in Cardiac Cine MRI via Nested Capsule Dense**
2 **Network**

3 Jilong Zhang[1,#] , Yajuan Zhang[1,#] , Hongyang Zhang[1], Quan Zhang[2,3*], Weihua Su[1], Shijie Guo[4],
4 Yuanquan Wang[1,4,*]
5
6 [1] School of Artificial Intelligence, Hebei University of Technology(HeBUT), Tianjin 300401,
7 P.R.China
8 [2] School of Information and Communication Engineering, North University of China, P.R.China
9 [3] Shanxi Provincial Key Laboratory for Biomedical Imaging and Big Data, North University of
10 China, Taiyuan 030051, China
11 [4] Hebei Key Laboratory of Robot Perception and Human-Robot Interaction, HeBUT, Tianjin
12 300401, P.R.China
13
14 [#]:These authors contributed equally to this work.

15 [*]Corresponding Author:
16 Yuanquan Wang[1,4], Quan Zhang[2,3]
17 Hebei University of Technology(HeBUT), Tianjin 300401, P.R.China
18 Email address: wangyuanquan@scse.hebut.edu.cn, zhangibmet@nuc.edu.cn

20 ## Abstract

21 **Background.** Cardiac magnetic resonance image (MRI) has been widely used in diagnosis of
22 cardiovascular diseases because of its noninvasive nature and high image quality. The
23 evaluation standard of physiological indexes in cardiac diagnosis is essentially the accuracy of
24 segmentation of left ventricle (LV) and right ventricle (RV) in cardiac MRI. The traditional
25 symmetric single codec network structure such as U-Net tends to expand the number of channels
26 to make up for lost information that results in the network looking cumbersome.
27 **Methods.** Instead of a single codec, we propose a multiple codecs structure based on the FC-
28 DenseNet(FCD) model and capsule convolution-capsule deconvolution, named *Nested Capsule*
29 *Dense Network (NCDN)*. NCDN uses multiple codecs to achieve multi-resolution, which makes
30 it possible to save more spatial information and improve the robustness of the model.
31 **Results.** The proposed model is tested on three datasets that include the York University Cardiac
32 MRI dataset, Automated Cardiac Diagnosis Challenge (ACDC-2017), and the local dataset. The
33 results show that the proposed NCDN outperforms most methods. In particular, we achieved
34 nearly the most advanced accuracy performance in the ACDC-2017 segmentation challenge.
35 This means that our method is a reliable segmentation method, which is conducive to the
36 application of deep learning-based segmentation methods in the field of medical image
37 segmentation. Code will be available at https://github.com/jk1008611/NCDN.

38 ## 1 Introduction

39   Heart disease causes one-third of all deaths worldwide. The statistics of the World Health
40 Organization in 2016 proved that cardiovascular disease accounted for 31% of the world's total
41 deaths [1]. It is predicted that by the year 2030, a population of 23.3 million will be killed by

42   cardiovascular   diseases   (CVDs)   all   over   the   world   [2,3].
43   With the wide application of modern medical technology, i.e., magnetic resonance imaging (MR
44   I),
45   making a noninvasive qualitative and quantitative evaluation of cardiac anatomical structure an
46   d function has become increasingly convenient. At the same time,
47   researchers have invested a lot of effort in the research of cardiovascular diseases to find out e
48   ffective methods to reduce morbidity and mortality In recent years.
49   The regression method such as direct and simultaneous four-
50   chamber volume estimation by the multioutput sparse latent regression (MSLR)
51   [4], use DAISY feature to train the regression model [5] and Contour-
52   Guided Regression Models [6], has been employed to predict the ventricular functional indices,
53   while
54   the most popular way to estimate the functional indices is based on segmentation, i.e., segmen
55   tation of the ventricles first and then calculating the indices. The calculation of related indicators
56   relies on the manual and accurate depiction of the endocardial and epicardial contours of the left
57   ventricle (LV) and right ventricle (RV)[7]. The continuous optimization of models and
58   segmentation methods has made great contributions to the improvement of accuracy [8,9].
59   Manual rendering is a time-consuming and tedious task and is prone to high variability within and
60   between observers [10–13]. Therefore, it would be very helpful to find a fast, high accuracy,
61   reusable, automatic segmentation.
62   Before the rise of deep learning, some methods, such as threshold-based segmentation
63   [14,15], edge detection-based segmentation [16,17], and genetic algorithm-based segmentation
64   [18] can't compare with deep learning-based segmentation methods in effect.
65   Although deep learning-based cardiac MRI segmentation has made great progress in the past
66   decades, there are still many problems to be solved. So far, a model has not been found that is
67   generally applicable to cardiac MRI segmentation tasks in various scenarios. The existing heart
68   datasets have the problems of a small amount of data and insufficient data distribution so that
69   the trained model does not respond well to the real-world situation, resulting in insufficient
70   generalization ability. It has become the goal of many researchers to construct a model with fast
71   learning speed and strong generalization ability on a limited data set.
72   In this paper, we firstly propose a nested neural network architecture named Nested Capsule
73   Dense Network (NCDN), which combines the FC-DenseNet model [19] and capsule convolution-
74   capsule deconvolution [20]. The Capsule Dense Block (CDB) is an important component module,
75   which consists of a dense connection of multiple Capsule Convolution Units (CCU). Each CCU
76   contains two capsule convolution layers and a capsule deconvolution layer. Because the
77   introduction of the capsule model eliminates the traditional pooling layer for image size scaling
78   and can retain more information for further semantic confirmation, the convolution capsule-
79   deconvolution capsule is used to replace the convolution-deconvolution to implement CCU [21].
80   The correctness of this choice was also proved by subsequent experiments. The nested capsule
81   dense architecture intuitively decomposes a single codec structure into multiple sub-codec
82   structures and uses the dense structure to better integrate feature information. Each step of
83   feature extraction and reconstruction in CDB is accompanied by the abstraction and
84   materialization of features by the network. Through multiple encoding and decoding, the image
85   noise contained in the feature map is filtered out layer by layer, so that the network can learn

86   more general features, such as contour features. On this basis, we can improve the
87   generalization ability of the network. This nested network structure is our attempt to improve the
88   FC-DenseNet model, which must be a trade-off between the accuracy of local marking and the
89   determination of semantics. Furthermore, we replace the "concat" operation in the FC-DenseNet
90   model with an "add" operation to reduce the parameters of the model. The verification effect on
91   the test sets proves that our NCDN has better performance and stronger robustness than other
92   models. There are four main contributions to our work, which can be summarized as follows:

93   (1) We firstly propose a nested capsule dense network called NCDN to decompose a single
94   codec into multiple codecs. This structure allows more posture and other information to
95   be retained and the noise in the sample is easier to filter in the early stage of feature
96   learning.
97   (2) The Capsule Dense Block made of Capsule Convolution Units (CCU) designed by us
98   eliminates the traditional pooling layer for image size scaling and can retain more
99   information for further semantic confirmation.
100 (3) We design a new connection structure, which greatly reduces the model parameters.
101 (4) The proposed NCDN model is used to complete the segmentation task of cardiac MRI.
102   The bi-ventricular segmentation and cardiac function diagnosis tasks in the ACDC 2017
103   dataset have shown good results.

104 **2 Related Work**
105   In the past few years, the segmentation of bi-ventricle MR images has received considerable
106 attention. Many scholars have proposed various methods to obtain better accuracy and make
107 the model have stronger generalization ability. The main model types can be summarized as
108 Fully Convolutional Neural Networks (FCNs), Recurrent Neural Networks (RNNs), and
109 Generative Adversarial Networks (GAN) [22].
110   By replacing the fully connected layer in the classification network with a convolutional layer,
111 FCN[23] predicts the category of each pixel in a dense prediction manner, which is successfull
112 y applied to the field of image segmentation and has become the mainstream method of curren
113 t ventricular segmentation.
114 Tran et al.[7] demonstrated the effectiveness of a fully convolutional neural network architectur
115 e for semantic segmentation of cardiac MRI and the utility of FCN to segment complex features
116 ¯¯ of the left and right ventricles. Moreover, in order to reduce the class imbalance problem in
117 ventricle segmentation and reduce the computational cost, Abdeltawab et al. [24] used two FCNs
118 [23] to complete the selection of the region of the interest (ROI) and segmentation of instances.
119 Different from ordinary FCN, Ronneberger et al. (2015) [25] proposed a multi-scale U-shaped
120 network based on channel connections to refine the segmentation results. In order to avoid the
121 loss of information caused by the maximum pooling layer in U-Net, Shen et al.[26] use a purely
122 dilated convolution operation to increase the receptive field to accelerate model convergence
123 and   improve   model   performance.
124 Sun et al.[27] believe that shape information is more meaningful than image texture information
125   and thus add a secondary stream that processes shape features of the image in parallel with t
126 he U-Net to help cardiac ventricular segmentation.
127 Taking advantage of the implicit deep supervision and feature reuse of the dense connection m
128 echanism, Simon et al.[19] extended DenseNet[28] to FC-

129   DenseNet for semantic segmentation problems.

130   Similarly, Marco et al.[29] proposed a dense block-

131   based skip connection structure to reduce the semantic gap of skip connections in ventricular s

132   egmentation.

133   Unlike natural images, many medical images are 3D time series made up of 2D images, such

134   as   CT   and   MRI.

135   3D UNet [30] and 3D VNet [31] extend the 2D segmentation network to the 3D segmentation n

136   etwork, making better use of the temporal and spatial information present in cardiac data to ach

137   ieve accurate segmentation. However, recurrent neural networks, especially LSTM[32] and

138   GRU[33] have more advantages than convolutional neural networks for processing time-series

139   tasks.

140   One of the use cases is Poudel et al. (2016) [34] combined RNN and 2D FCN to exploit the ob

141   served spatial dependencies in adjacent slices, improve the model's ability to identify the borde

142   r regions of the heart, and solve the segmentation problem of multi-slice MRI images in a

143   straightforward manner. GAN network is a competitively aware network structure, which is

144   generally composed of generators and discriminators. In the process of model training, the

145   generator generates images that attempt to deceive the discriminator, and the discriminator aims

146   to identify real images in fake images. In the application of heart image segmentation, the role of

147   the segmentation network is to generate segmentation results, and the discriminator is used to

148   judge the difference between segmentation results and ground truth. Qi et al.[35] adopted an

149   adversarial training approach, where the generator and discriminator optimized the network by

150   competing with each other, which alleviated the class imbalance of the heart, eliminated

151   interference from other organs and tissues, and improved the segmentation accuracy of general

152   "difficult" slices. In this way, a more accurate segmentation map will be generated.

153   Sabour et al. (2017) [21] proposed a capsule network (CapNet) with dynamic routing to use

154   the reconstruction of output capsule instead of maximum pooling. The vector output of CapNet

155   is better than the scalar output of Convolutional Neural Networks (CNN) to discover and save the

156   position and posture information of objects in the image (such as spatial angle magnitude order,

157   etc.). However, although CapNet has obtained good results in digital recognition and small image

158   recognition, it has the problem of large parameters when performing large-scale image

159   segmentation tasks or deep network construction, while images in the medical field are mostly

160   large-size images. Therefore, the original CapNet is not suitable for image segmentation tasks

161   in the medical field. LaLonde and Bagci (2018) [20] modified the capsule network and applied it

162   to the image segmentation task for the first time. They improved the dynamic routing algorithm

163   to reduce the parameters. The dynamic routing in the traditional capsule network is equivalent to

164   the full connection mapping between capsules, which makes the number of parameters huge.

165   The author uses window control and the same type of capsule sharing weight method to reduce

166   the parameters. In addition to changing the dynamic routing algorithm to increase the size of the

167   accepted input picture, a novel capsule convolution-capsule deconvolution network architecture

168   called SegCaps is proposed to perform image segmentation tasks. Based on work[20], Cao et

169   al. proposed [36] to extract low-level image features such as grayscale and texture of the left

170   ventricle of the heart, as well as semantic features such as location and size for ventricular

171   segmentation.

172   Inspired by the FC-DenseNet model [20] and the SegCaps model [20], our model was
173   proposed for the semantic segmentation task of cardiac MRI.

174

3 Materials & Methods

3.1 Network Structure

177   We will introduce our network in detail. As shown in Figure 1, based on FC-DenseNet,

178   (1) We replace the dense block with Capsule Dense Block (CDB) proposed by us, this block will
179   be explained in detail in Section 3.2.

180   (2) We design a new connection structure: Define y_k as the kth layer. When y_{k-1} is Transition
181   Down (TD), y_k is CDB, and y_{k+1} is

182    $$y_{k+1} = y_{k-1} + y_k \qquad\qquad (1)$$

183   where "+" means that the feature maps produced in layers $k-1$, k are added in the last
184   dimension. This means that the shape of the $k-1$th and kth layers must be consistent.

185   By fusing the output feature map of the CDB and the input feature map again, the
186   relationship between layers can be closer, and the whole semantics cannot be directly connected
187   in the process of capturing, to reduce the omission of image information. The other parts are
188   similar to the idea of FC-DenseNet. Convolution is used to extract feature images, TD reduces
189   the image size to increase the perception range, and Transition Up (TU) performs image
190   reconstruction and precise positioning.

191   The detailed structure is shown in Table 1. We increase its channel number to 32 in the
192   first convolution operation and do not change it in the subsequent process until it is finally
193   transformed into the target number of channels through three convolutions. The purpose of our
194   design is to avoid excessive parameters. The method we adopted is that the input and output
195   of the CDB are the same in shape. If only "concat" is used instead of "add", the parameter
196   increases from 5.5M to 330.8M, which is due to the dense feature of nested structure.

197

3.2 Capsule Convolution Unit

199   The CCU is a constituent element of the CDB. By using the capsule convolution and capsule
200   deconvolution structure proposed by LaLonde and Bagci (2018) [20] to achieve the nested
201   encoding-decoding structure in the CDB. The structure of a single CCU is as shown in Figure 2
202   shown.

203   The specific details are: let the input length, width, number of capsules, and number of
204   channels be K_h, K_w, K_{cap}, K_c. The input [K_h, K_w, K_{cap}, K_c] first passes through 3× 3 capsule
205   convolution with routing number 1, which becomes [$K_h/2$, $K_w/2$, $K_{cap} \times 2$, K_c]. After passing 3 × 3
206   capsule convolution with routing number 3, it becomes [$K_h/2$, $K_w/2$, K_{cap}, K_c]. Finally, the shape of
207   the output is [K_h, K_w, K_{cap}, K_c] after 3 × 3 capsule deconvolution with routing number 3.

208

3.3 Capsule Dense Block

210   Capsule Dense Block (CDB) in Figure 3 consists of a dense capsule connection layer at the
211   front and a regression layer at the rear. The dense capsule connection layer is composed of
212   three Capsule Convolution Units (CCU) densely connected. The regression layer is a
213   convolutional layer and its purpose is to convolve the feature map ($F_{d,4}$) formed by the dense
214   connection into the shape of the CDB input ($F_{d,1}$).

215   Dense connection: in CDB, dense connection is realized by passing the state of the previous
216 layer to the subsequent layers. Let $F_{d,1}$ and $F_{d,5}$ be the input and output of the d-th CDB
217 respectively. The output of d-th CDB can be formulated as
218                    $$F_{d,5} = \sigma\,(w_{d,5}[F_{d,1},\ F_{d,2},\ F_{d,3},\ F_{d,4}])\qquad (2)$$
219 where σ denotes the ReLU [38] activation function. $w_{d,5}$ is the weights of the
220 $F_{d,5}$, where the bias term is omitted for simplicity. $[F_{d,1},\ F_{d,2},\ F_{d,3},\ F_{d,4}]$ refers to the concatenation
221 of the feature maps produced by $F_{d,1}$, $F_{d,2}$, $F_{d,3}$, $F_{d,}4$. The CDB of the former layer and the output
222 of each layer are directly connected with the latter layer, which not only retains the feedforward
223 nature but also extracts the local dense feature.

224 4 Results

225 4.1 DataSet

226   In our previous work, a total of three datasets have been used to evaluate our proposed NCDN
227 model. They are the York University Cardiac MRI dataset, the Automated Cardiac Diagnosis
228 Challenge, and the local dataset.

229   **The York University Cardiac MRI dataset (York)**: York consists of short-axis cardiac MR
230 image sequences of 33 subjects, a total of 7980 2D images, provided by the Department of
231 Diagnostic Imaging of the Hospital for Sick Children in Toronto, Canada[39]. Most data are a
232 variety of cardiac abnormalities, such as cardiomyopathy, aortic regurgitation, ventricular
233 enlargement, ischemia, etc., and a few data are abnormalities related to the left ventricle.
234 Because some of the markers in the dataset are missing or incomplete, the problematic images
235 were removed from the dataset. The original 256×256 pixel images were clipped to form 3020
236 images with a scale of 80×80 pixels that only retained the left ventricle endocardium and
237 epicardium.

238   **The Automated Cardiac Diagnosis Challenge (ACDC)**: The ACDC dataset was created
239 based on real clinical examination results obtained by the University Hospital of Dijon (France)
240 [40]. This dataset is the first and largest fully annotated public MRI cardiac data in the medical
241 imaging community setting. The data consisted of short-axis section sequences of cardiac
242 magnetic resonance images from 150 patients, divided into five subgroups, 30 normal
243 subjects(NOR), 30 patients with previous myocardial infarction(MINF), 30 patients with dilated
244 cardiomyopathy(DCM), 30 patients with hypertrophic cardiomyopathy(HCM), and 30 patients
245 with abnormal right ventricle(RV). The spatial resolution was from 1.37 to 1.68 mm^2/pixel. We
246 obtained 1902 images of 100 subjects from the training set of this dataset. Each slice was center
247 cropped to a resolution of 128px by 128px.

248   **Local dataset**: This dataset has been employed in full left ventricle quantification [41], and d
249 irect multitype cardiac indices estimation [42]. It consists of 2900 images of 145 cases from thre
250 e hospitals belonging to two medical centers (London Healthcare and St. Joseph's Healthcare).
251 Most patients had a variety of pathological manifestations, including regional wall motion abnor
252 malities, myocardial hypertrophy, mildly enlarged LV, atrial septal defect, LV dysfunction, etc.
253

254 4.2 Metrics

255   Let A and M be the corresponding areas enclosed by the predicted (automated) contours a
256   and ground truth (manual) contours m, respectively. The following is our introduction to the main
257   evaluation indicators.
258   **Dice index** The Dice index (DI) [43] is a measure of overlap or similarity between two contour
259   areas and is defined as (3) and Figure 4(a):

260    $$D(A,M) = 2 * \frac{A \cap M}{A + M} \qquad (3)$$

261   The Dice index varies from zero (total mismatch) to unity (perfect match).
262   **Hausdorff distance** Hausdorff distance (HD) [44] is another evaluation metric, as shown in
263   Figure 4(b). $d(p,m)$ means a point (p) on a to its nearest point (p') of m and the converse is $d(p',a)$,
264   as followed:

265    $$d(p,m) = min||p - p'|| \qquad (4)$$

266   then, find the maximum values of $d(p,m)$ and $d(p',a)$ for all the points. HD is the maximum of the
267   two values and always during $[0,\infty]$. HD increases, its performance degrades.

268    $$HD(m,a) = max(max\ d(p,m),\ max\ d(p',a)) \qquad (5)$$

269

270   **4.3 Training Implementation**
271   In the experiment, the deep learning framework is Tensorflow, the GPU is NVidia GTX
272   1080Ti, the optimizer is Adam, the loss function is cross-entropy loss, the learning rate is set to
273   10^{-4}, the batch size is set to 1, and the training is about 30 epochs. Data expansion is applied to
274   image expansion. For each dataset, we first divide the training set, validation set, and testing set.
275   The division ratios are 0.7, 0.1, and 0.2. After that, the divided training set is expanded fourfold
276   by rotating 90°, 180°, 270°.

277

278   **4.4 Generalization results**
279   In this section, We show the results of the generalization ability of the segmentation model
280   we designed. Table 2 is the test result of the three models trained on the York dataset. Table 3
281   is the result of training on the York dataset and testing on the Local dataset. Compared with the
282   other two models, our model performed well in ordinary segmentation effect evaluation, which
283   its Accuracy, Dice index, and Hausdorff distance all show the best results among the two
284   contrasted models, as shown in Table 2. And as shown in Table 3, our model has better
285   generalization ability than UNet, outperforms the FC-DenseNet on Segmentation of the LV and
286   is on par with the FC-DenseNet on segmentation of the MYO. The model trained on the York
287   dataset is used for testing on the Local dataset. Training with the York data set and then testing
288   with the local data set results in worse segmentation than training and testing with the York data
289   set. However, the NCDN we proposed still has better resistance to image changes and better
290   performance than the other two models on DI, and HD.

291

292   **4.5 Ablation experiments**
293   To verify the performance of the NCDN model, the following models were used for
294   comparison: (1) The origin U-net model [25], (2) the FC-DenseNet introduced in [19], (3) Nested
295   Convolution Dense Network (NConvDN) replaces capsule convolution-capsule deconvolution of
296   NCD-N with convolution-deconvolution. We used the ACDC dataset [40] to train and test these
297   models and separately count the segmentation effects of subjects with different disease types,

298   which distinguished the end-diastolic (ED) and end-systolic (ES). The evaluation indicators
299   include DI and HD. All four models follow the settings in section 4.3 and use five-fold cross-
300   validation. The results of comparing the NCDN model with other models according to the
301   segmentation accuracy are shown in Tables 4, 5, and 6.
302   The right ventricle is the hardest part of ventricular segmentation, yet through Table 4, NCDN o
303   utperforms the other three contrasting models in both Dice and HD metrics. Table 5 lists the se
304   gmentation results of myocardium. FC-
305   DenseNet is slightly higher than NCDN in HD index, but NCDN is still the best performing mode
306   l in Dice index. Table 6 shows the segmentation results of the left ventricle, compared with the
307   other three comparison models, NCDN achieves the best results in both Dice and HD metrics.
308   Figure 5 shows the segmentation effect of the four models in the same image [40]. The
309   images shown cover the ES and ED of different groups of people. From left to right are the
310   original image, ground truth, and the prediction results of the four network models.
311   Using UNet, FC-DenseNet, NConvDN, and the proposed NCDN to conduct ablation exper-
312   iments, the results show that our proposed model has better performance. In the comparison of
313   multiple dimensions, the segmentation accuracy of NConvDN is worse than that of NCDN. This
314   may be because capsule convolution can save more spatial information than convolution, which
315   is beneficial to improve the ability to distinguish things from different perspectives.
316   In order to better prove the performance of the proposed model, the model with the most
317   advanced results was used for comparison, which ranks among the top 10 in ACDC test set
318   segmentation performance. As shown in Table 7(a), on the Dice metric, the NCDN obtained the
319   best results on ED of MYO and ED of LV compared with the other ten models. Besides, ED and
320   ES of RV ranked 6th, ES of MYO ranked 2nd, and ES of LV ranked 4th. As shown in Table7(b),
321   In terms of HD index, ED and ES of RV ranked 8th and 7th respectively, ED and ES of MYO
322   ranked 6th, and ED and ES of LV ranked 6th and 7th respectively. NCDN does not perform very
323   well in HD, using some proper post-processing methods may bring improvements.
324

5. Discussion

326   In this work, a nested network structure is proposed to complete the segmentation task of
327   cardiac MRI, hoping to have better segmentation performance. Local and third-party evaluations
328   have reflected that it has improved segmentation accuracy and robustness relative to the bench-
329   mark model, and it also has the most advanced results in some indicators. The proposed model
330   has good performance on the DI index, but the performance is relatively ordinary on the HD
331   index, especially for the segmentation of RV. Both the DI indicator and HD indicator measure the
332   effect of segmentation but have different focuses. DI can better reflect the consistency of the
333   corresponding pixels of the image, while HD focuses on the consistency of the segmentation
334   edge. This means that an outlier has little effect on DI, but may have a greater impact on HD
335   indicators. NCDN is an end-to-end network with capsule convolution as the kernel. The
336   characteristics of the capsule enable it to better perceive objects in different viewing angles,
337   which effectively avoids the state of failure of recognition in FC-DenseNet, such as incorrectly
338   judging whether a certain category exists on the image. However, it still lacks constraints on the
339   objects to be segmented in the image, which may lead to the appearance of outliers and make
340   the HD index too large. To alleviate this problem, new shape mechanisms such as shape prior
341   [55] can be introduced. In addition, a more reasonable cost function for HD constraints can be

342   discovered or statistical techniques can be used to correct outliers in the segmented image to
343   ensure edge integrity and internal consistency.

6. Conclusions

345   In this work, we propose a nested network structure that decomposes a single codec into
346   multiple codecs to obtain better cardiac MR image segmentation results. This structure based
347   on the FC-DenseNet model and capsule convolution-capsule deconvolution shows a better seg-
348   mentation effect on multiple datasets than each part of the source network. The smaller error
349   and standard deviation further prove the effectiveness of network fusion. The experimental re-
350   sults show that the segmentation effect of our model has better stability than other traditional
351   segmentation models. This makes it possible to apply our method to the automatic segmentation
352   of cardiac MRI systems in the future.

353
354
355

References

357 1. Mozaffarian, D. , et al. "Heart Disease and Stroke Statistics—2015 Update." Circulation
358   131.4(2015):W.H. Organization, Women 47(26), 2562 (2011)
359 2. Null, Who , et al. "Global Status Report on Noncommunicable Diseases 2010." Women
360   47.26(2010):2562-2563.
361 3. Mathers, C. D. , and D. Loncar . "Projections of Global Mortality and Burden of Disease from
362   2002 to 2030." Plos Medicine 3.11(2006).
363 4. Zhen, X. , et al. "Direct and Simultaneous Four-Chamber Volume Estimation by Multi-Output
364   Regression." International Conference on Medical Image Computing & Computer-assisted
365   Intervention-miccai Springer-Verlag New York, Inc. 2015.
366 5. Du, X. , et al. "Deep Regression Segmentation for Cardiac Bi-Ventricle MR Images." IEEE
367   Access 6(2018):3828-3838.
368 6. Wang, W. , et al. "Quantification of Full Left Ventricular metrics via Deep Regression
369   Learning with Contour-guidance." IEEE Access (2019):1-1.
370 7. Tran, P. V. . "A Fully Convolutional Neural Network for Cardiac Segmentation in Short-Axis
371   MRI." arXiv preprint arXiv:1604.00494 (2016).
372 8. Wu, Yuwei , Y. Wang , and Y. Jia . "Segmentation of the left ventricle in cardiac cine MRI
373   using a shape-constrained snake model." Computer Vision & Image Understanding Cviu
374   117.9(2013):990-1003.
375 9. Zz, A , et al. "GVFOM: a novel external force for active contour based image segmentation."
376   Information Sciences 506(2020):1-18.
377 10. Petitjean, C. , and J. N. Dacher . "A review of segmentation methods in short axis cardiac MR
378   images." Medical image analysis (2013).

379   11. Miller, Christopher A , et al. "Quantification of left ventricular indices from SSFP cine
380   imaging: impact of real-world variability in analysis methodology and utility of geometric
381   modeling. " Journal of Magnetic Resonance Imaging Jmri 37.5(2013):1213-1222.
382   12. Tavakoli, V. , and AA Amini. "A survey of shaped-based registration and segmentation
383   techniques for cardiac images." Computer Vision and Image Understanding 117. 9(2013):966-
384   989.
385   13. Suinesiaputra, A. , et al. "A collaborative resource to build consensus for automated left
386   ventricular segmentation of cardiac MR images. " Medical Image Analysis 18.1(2014):50-62.
387   14. Sadeghian, F. , et al. "A Framework for White Blood Cell Segmentation in Microscopic Blood
388   Images Using Digital Image Processing." Biological Procedures Online 11.1(2009):196-206.
389   15. Bazin, P. L. , and D. L. Pham . "Homeomorphic brain image segmentation with topological
390   and statistical atlases." Medical image analysis 12.5(2008):616-625.
391   16. Belaid, L. J. , and   W. Mourou . "IMAGE SEGMENTATION: A WATERSHED
392   TRANSFORMATION ALGORITHM." Image analysis and stereology 28.2(2009):93-102.
393   17. Bellon, Orp , and L. Silva . "New improvements to range image segmentation by edge
394   detection." Signal Processing Letters IEEE 9.2(2002):43-45.
395   18. Ramos, V. , and F. Muge . "Image Colour Segmentation by Genetic Algorithms." arXiv
396   preprint cs/0412087 (2004).
397   19. S Jégou, et al. "The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for
398   Semantic Segmentation." 2017 IEEE Conference on Computer Vision and Pattern
399   Recognition Workshops (CVPRW) IEEE, 2016.
400   20. Lalonde, R. , and   U. Bagci . "Capsules for Object Segmentation." arXiv preprint
401   arXiv:1804.04241 (2018).
402   21. Sabour, S. , N. Frosst , and G. E. Hinton . "Dynamic Routing Between Capsules." (2017).
403   22. Chen, C. , et al. "Deep learning for cardiac image segmentation: A review." Frontiers in
404   Cardiovascular Medicine (2019).
405   23. Long, J. , E. Shelhamer , and T. Darrell . "Fully Convolutional Networks for Semantic
406   Segmentation." IEEE Transactions on Pattern Analysis and Machine Intelligence
407   39.4(2015):640-651.
408   24. Ha, A , et al. "A deep learning-based approach for automatic segmentation and quantification
409   of the left ventricle from cardiac cine MR images." Computerized Medical Imaging and
410   Graphics 81.
411   25. Ronneberger, O. , P. Fischer , and T. Brox . "U-Net: Convolutional Networks for Biomedical
412   Image Segmentation." International Conference on Medical Image Computing and Computer-
413   Assisted Intervention (2015).
414   26. Shen, W. , et al. "Automatic segmentation of the femur and tibia bones from X-ray images
415   based on pure dilated residual U-Net." Inverse Problems and Imaging (2020):1-15.
416   27. Sun, J. , et al. "SAUNet: Shape Attentive U-Net for Interpretable Medical Image
417   Segmentation." (2020).
418   28. Huang, G. , et al. "Densely Connected Convolutional Networks." IEEE Computer Society
419   IEEE Computer Society, 2016.

420 29. Penso, M. , et al. "Automated Left and Right Ventricular Chamber Segmentation in Cardiac
421   Magnetic Resonance Images Using Dense Fully Convolutional Neural Network." Computer
422   Methods and Programs in Biomedicine 42(2021):106059.
423 30. iek, zgün, et al. "3D U-Net: Learning Dense Volumetric Segmentation from Sparse
424   Annotation." Springer, Cham (2016).
425 31. F Milletari, N. Navab , and S. A. Ahmadi . "V-Net: Fully Convolutional Neural Networks
426   for Volumetric Medical Image Segmentation." 2016 Fourth International Conference on 3D
427   Vision (3DV) IEEE, 2016.
428 32. Hochreiter, Sepp , and Jürgen A Schmidhuber. "LSTM can solve hard long time lag
429   problems." Advances in neural information processing systems (1996):476-479.
430 33. Cho, K. , et al. "Learning Phrase Representations using RNN Encoder-Decoder for Statistical
431   Machine Translation." Computer Science (2014).
432 34. Poudel, Rpk , P. Lamata , and G. Montana . "Recurrent Fully Convolutional Neural Networks
433   for Multi-slice MRI Cardiac Segmentation." Lecture Notes in Computer Science
434   3824.1(2016):164-173.
435 35. Qi, L. , et al. "Cascaded Conditional Generative Adversarial Networks With Multi-Scale
436   Attention Fusion for Automated Bi-Ventricle Segmentation in Cardiac MRI." IEEE Access
437   7(2019):172305-172320.
438 36. Cao, Y. J. , et al. "Seg-CapNet:A Capsule-Based Neural Network for the Segmentation of Left
439   Ventricle from Cardiac Magnetic Resonance Imaging." 计算机科学技术学报：英文版
440   36.2(2021):11.
441 37. Zhongyi, et al. "Spine-GAN: Semantic segmentation of multiple spinal structures. " Medical
442   image analysis (2018).
443 38. Glorot, Xavier , A. Bordes , and Y. Bengio . "Deep Sparse Rectifier Neural Networks."
444   Journal of Machine Learning Research 15(2011):315-323.
445 39. Andreopoulos, A. , and J. K. Tsotsos . "Efficient and generalizable statistical models of shape
446   and appearance for analysis of cardiac MRI." Medical Image Analysis 12.3(2008):335-357.
447 40. O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, et al. "Deep Learning Techniques for
448   Automatic MRI Cardiac Multi-structures Segmentation and Diagnosis: Is the Problem Solved
449   ?" in IEEE Transactions on Medical Imaging, vol. 37, no. 11, pp. 2514-2525, Nov. 2018 doi:
450   10.1109/TMI.2018.2837502
451 41. Xue, W. , et al. "Full left ventricle quantification via deep multitask relationships learning."
452   Medical Image Analysis (2017):S1361841517301366.
453 42. Xue, W. , et al. "Direct Multitype Cardiac Indices Estimation via Joint Representation and
454   Regression Learning." IEEE Transactions on Medical Imaging 36.10(2017):2057-2067.
455 43. Dice, L. R. . "Measures of the Amount of Ecologic Association Between Species." Ecology
456   26.3(1945).
457 44. Huttenlocher, et al. "Comparing images using the Hausdorff distance." Pattern Analysis and
458   Machine Intelligence, IEEE Transactions on (1993).
459 45. Isensee, F. , et al. "Automatic Cardiac Disease Assessment on cine-MRI via Time-Series
460   Segmentation and Domain Specific Features." Medical Image Computing, German Cancer

461   Research Center (DKFZ), Heidelberg, Germany;Medical Image Computing, German Cancer
462   Research Center (DKFZ), Heidelberg, Germany;Division of Computer-assisted Medical
463   Interventions, German Cancer Research Center (DKFZ),.

464   46. Zotti, C. , et al. "Convolutional Neural Network with Shape Prior Applied to Cardiac MRI
465   Segmentation." IEEE Journal of Biomedical & Health Informatics PP(2018):1-1.

466   47. Painchaud, N. , et al. "Cardiac Segmentation with Strong Anatomical Guarantees." IEEE
467   Transactions on Medical Imaging PP.99(2020):1-1.

468   48. Khened, M. , V. Alex , and G. Krishnamurthi . "Densely Connected Fully Convolutional
469   Network for Short-Axis Cardiac Cine MR Image Segmentation and Heart Diagnosis Using
470   Random Forest." International Workshop on Statistical Atlases and Computational Models of
471   the Heart (2018):140-151.

472   49. Baumgartner, Christian F , et al. "An Exploration of 2D and 3D Deep Learning Techniques
473   for Cardiac MR Image Segmentation." Springer, Cham (2017).

474   50. Wolterink, J. M. , et al. "Automatic Segmentation and Disease Classification Using Cardiac
475   Cine MR Images." International Workshop on Statistical Atlases and Computational Models
476   of the Heart Springer, Cham, 2017.

477   51. MM Rohé, M. Sermesant , and X. Pennec . "Automatic Multi-Atlas Segmentation of
478   Myocardium with SVF-Net." Springer, Cham (2017).

479   52. Zotti, C. , et al. "GridNet with automatic shape prior registration for automatic MRI cardiac
480   segmentation." Springer, Cham (2017).

481   53. Patravali, J. , S. Jain , and S. Chilamkurthy . 2D-3D Fully Convolutional Neural Networks
482   for Cardiac MR Segmentation. Springer, Cham, 2017.

483   54. Grinias, E. , and G. Tziritas . Fast Fully-Automatic Cardiac Segmentation in MRI Using MRF
484   Model Optimization, Substructures Tracking and B-Spline Smoothing. 2018.

485   55. Ravishankar, H. , et al. "Learning and Incorporating Shape Models forSemantic
486   Segmentation." MICCAI Springer, Cham, 2017.

Figure 1

The illustration shows the NCDN architecture used for ACDC segmentation tasks, the output image consists of four feature maps that represent background, RV, MYO, and LV, respectively.

The data shown in the figure comes from the ACDC dataset

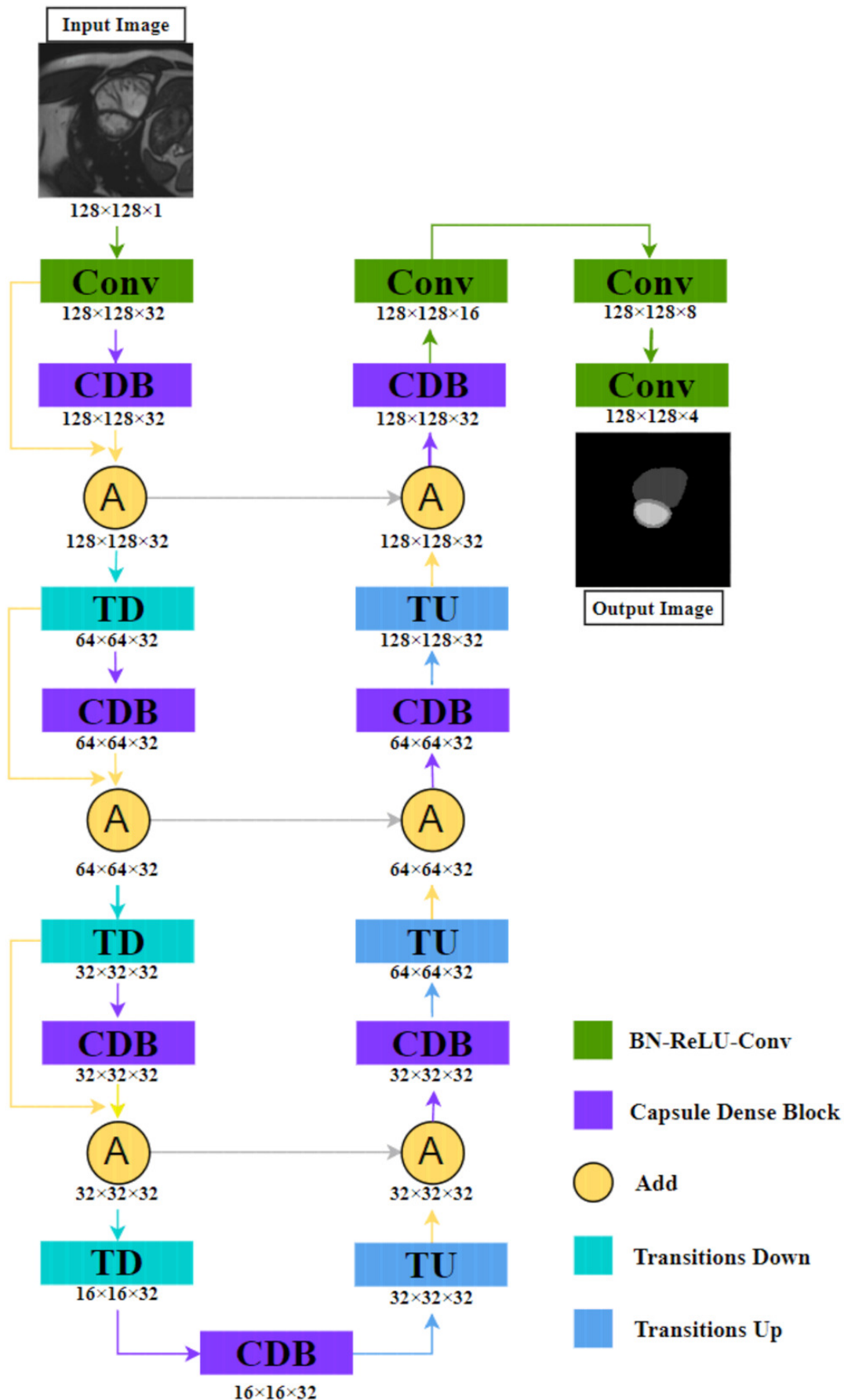



Figure 2

The architecture of our proposed Capsule Convolution Unit (CCU). The input dimensions in the figure are length, width, number of capsules, and number of channels, respectively.

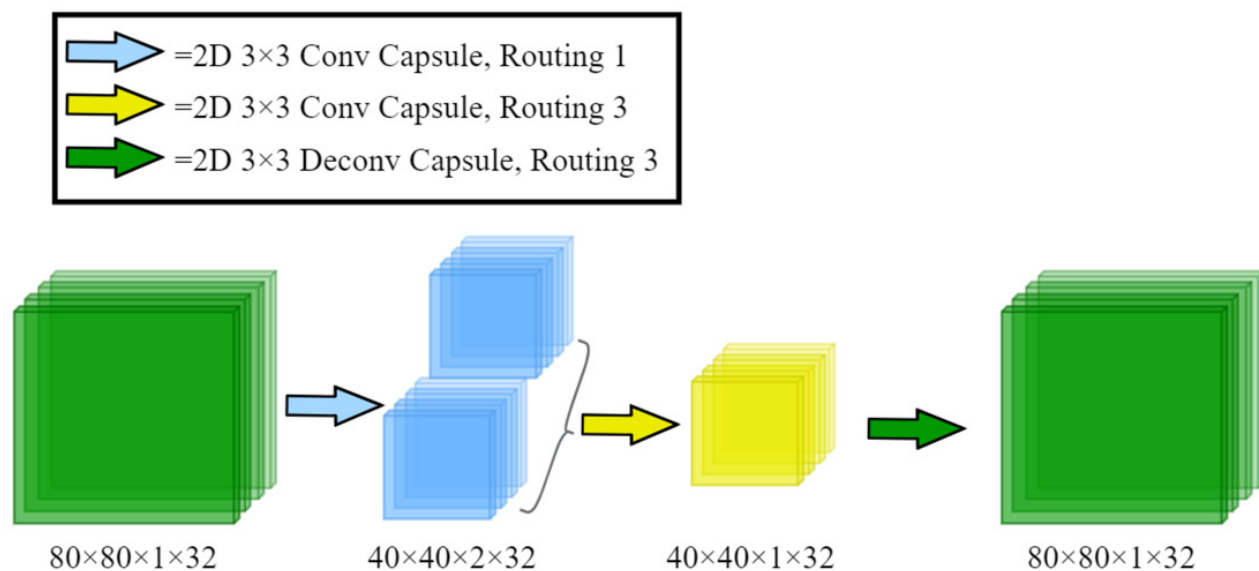



Figure 3

Capsule dense Block (CDB) architecture. Our CDB includes a dense capsule connection layer at the front and a regression layer at the rear. The dense capsule connection layer is composed of three Capsule Convolution Units (CCU) densely connected.

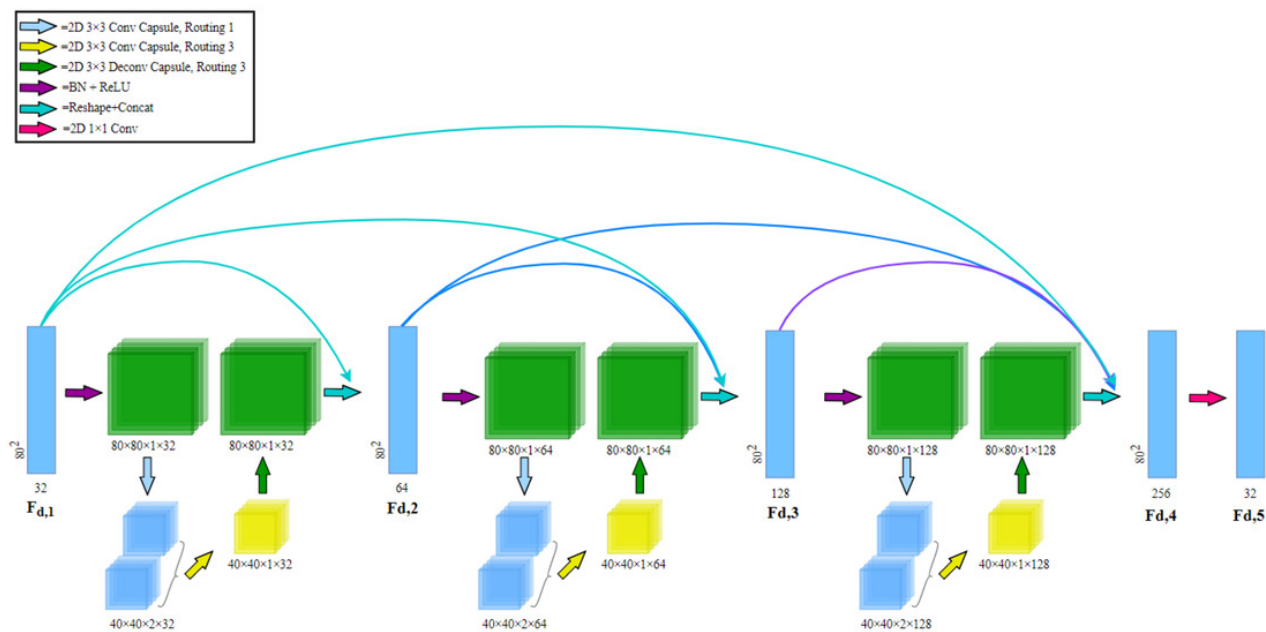



Figure 4

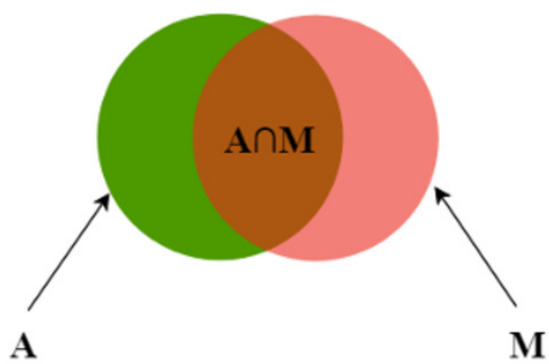
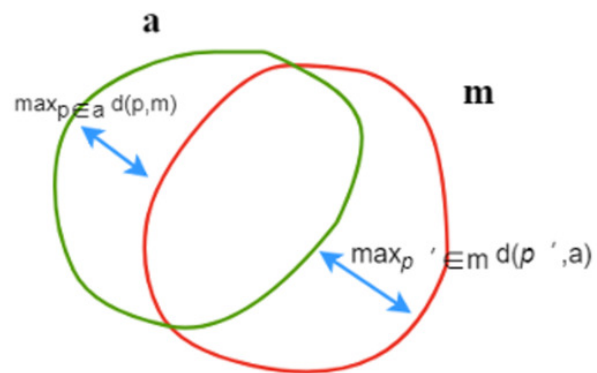
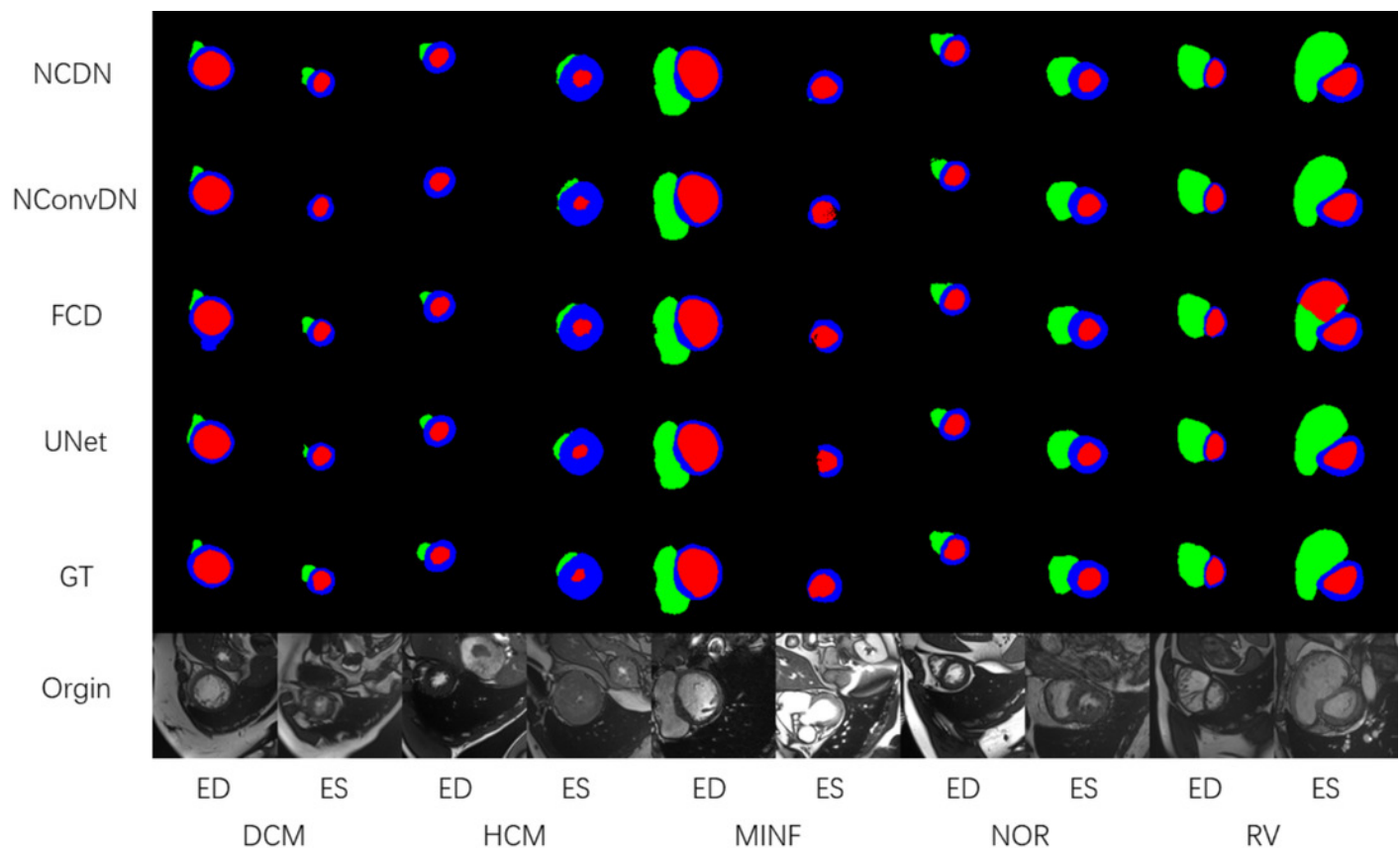
Dice index(a) and Hausdorff distance(b).

(a)   (b)

Figure 5

The segmentation results of DCM, HCM, MINF, NOR,and RV. The illustration shows the segmentation effect of NCDN and the other three networks, where GT stands for ground truth [40].

The data shown in the figure comes from the ACDC dataset [40]

The architecture of the Nested Capsule Dense Network. Conv layer in the table represents the BN-ReLU-Conv sequence.

Layers	NCDN	Ouput Size
Input size	128×128×1	-
Convolution	3×3 Conv, stride 2	128×128×32
Capsule Dense Block	[BN-ReLU-CCU-Dropout] × 3	128×128×32
Add	Convolution+Capsule Dense Block	128×128×32
Transition Down	BN-ReLU-2×2 max pool, stride 2	64×64×32
Capsule Dense Block	[BN-ReLU-CCU-Dropout] × 3	64×64×32
Add	Convolution+Capsule Dense Block	64×64×32
Transition Down	BN-ReLU-2×2 max pool, stride 2	32×32×32
Capsule Dense Block	[BN-ReLU-CCU-Dropout] × 3	32×32×32
Add	Convolution+Capsule Dense Block	32×32×32
Transition Down	BN-ReLU-2×2 max pool, stride 2	16×16×32
Capsule Dense Block	[BN-ReLU-CCU-Dropout] × 3	16×16×32
Transition Up	3×3 deconv, stride 2	32×32×32
Add	Deconvolution+Capsule Dense Block	32×32×32
Capsule Dense Block	[BN-ReLU-CCU-Dropout] × 3	32×32×32
Transition Up	3×3 deconv, stride 2	64×64×32
Add	Deconvolution+Capsule Dense Block	64×64×32
Capsule Dense Block	[BN-ReLU-CCU-Dropout] × 3	64×64×32
Transition Up	3×3 deconv, stride 2	128×128×32
Add	Deconvolution+Capsule Dense Block	128×128×32
Capsule Dense Block	[BN-ReLU-CCU-Dropout] × 3	128×128×32
Convolution	3×3 Conv, stride 2	128×128×16
Convolution	3×3 Conv, stride 2	128×128×8
Convolution	3×3 Conv, stride 2	128×128×4

To evaluate the effects of different segmentation techniques, the test results after training on the York data set are presented in the form of average (std.).

	York			
	Dice index		HD(mm)	
	LV	MYO	LV	MYO
U-Net	0.90(0.03)	0.95(0.03)	9.42(3.7)	8.32(3.87)
FC-DenseNet	0.91(0.03)	0.95(0.03)	9.15(3.6)	8.10(3.83)
NCDN	**0.92(0.03)**	**0.95(0.02)**	**9.07(3.49)**	**8.05(3.62)**

To evaluate the generalization ability of different segmentation techniques, use the Local data set to test after training on the York dataset, and the results are presented in the form of average (std.)

	Local			
	Dice index		HD(mm)	
	LV	MYO	LV	MYO
U-Net	0.69(0.09)	0.85(0.07)	22.51(6.58)	21.55(8.41)
FC-DenseNet	0.74(0.07)	0.89(0.06)	18.93(6.87)	**16.74(5.88)**
NCDN	**0.78(0.09)**	**0.89(0.06)**	**14.94(6.32)**	17.13(6.84)

Average DI and HD (std.) on RV in the four models.

	Dice index		HD(mm)	
	ED	ES	ED	ES
U-Net	0.92(0.09)	0.86(0.11)	14.09(11.63)	20.66(19.64)
FC-DenseNet	0.92(0.09)	0.87(0.12)	14.40(13.25)	17.73(17.14)
NConvDN	0.92(0.10)	0.87(0.11)	14.80(16.55)	24.63(34.73)
NCDN	**0.93(0.07)**	**0.88(0.10)**	**13.94(12.23)**	**16.57(14.95)**

Average DI and HD (std.) on MYO in the four models

	Dice index		HD(mm)	
	ED	ES	ED	ES
U-Net	0.87(0.08)	0.89(0.07)	11.90(10.43)	12.32(12.08)
FC-DenseNet	0.88(0.06)	0.89(0.07)	**8.86(4.03)**	**10.29(7.11)**
NConvDN	0.88(0.07)	0.89(0.07)	9.39(6.85)	11.71(10.55)
NCDN	**0.89(0.07)**	**0.90(0.06)**	8.91(5.29)	10.47(6.27)

Average DI and HD (std.) on LV in the four models.

	Dice index		HD(mm)	
	ED	ES	ED	ES
U-Net	0.95(0.06)	0.91(0.09)	9.33(8.15)	9.50(7.37)
FC-DenseNet	0.95(0.05)	0.91(0.07)	7.73(3.56)	9.16(5.01)
NConvDN	0.95(0.08)	0.91(0.09)	7.85(4.89)	9.29(6.56)
NCDN	**0.96(0.04)**	**0.92(0.08)**	**7.71(5.61)**	**8.64(4.86)**

The segmentation effect of different segmentation techniques on the ACDC test set.

	Dice index					
	RV		MYO		LV	
	ED	ES	ED	ES	ED	ES
NCDN	0.932	0.882	**0.899**	0.911	**0.966**	0.916
[45]	**0.946**	**0.904**	0.896	**0.919**	0.965	**0.933**
[46]	0.934	0.885	0.886	0.902	0.964	0.912
[47]	0.933	0.884	0.881	0.897	0.961	0.911
[48]	0.941	0.882	0.889	0.898	0.964	0.917
[49]	0.935	0.879	0.882	0.897	0.963	0.911
[50]	0.932	0.883	0.892	0.901	0.961	0.918
[51]	0.928	0.872	0.884	0.896	0.957	0.900
[52]	0.916	0.845	0.875	0.894	0.957	0.905
[53]	0.911	0.819	0.867	0.869	0.955	0.885
[54]	0.887	0.767	0.799	0.784	0.948	0.848

	HD(mm)					
	RV		MYO		LV	
	ED	ES	ED	ES	ED	ES
NCDN	13.9	14.5	9.7	10.8	7.5	9.6
[45]	**8.8**	**11.4**	**7.6**	**7.1**	**5.6**	**6.3**
[46]	11.0	12.6	9.6	9.3	6.2	8.4
[47]	13.7	13.3	8.6	9.6	6.1	8.3
[48]	10.3	14.0	9.8	12.6	8.1	9.0
[49]	14.0	13.9	9.8	11.3	6.5	9.2
[50]	12.7	14.7	8.7	10.6	7.5	9.6
[51]	11.9	13.4	8.7	9.3	7.5	10.7
[52]	14.0	15.9	11.1	10.7	6.6	8.7
[53]	13.5	18.7	11.5	13.0	8.2	10.9
[54]	19.0	24.2	12.3	14.6	8.9	12.9