Medical Image Analysis 51 (2019) 21–45

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Fully convolutional multi-scale residual DenseNets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers

Mahendra Khened[a], Varghese Alex Kollerathu[a], Ganapathy Krishnamurthi[a,*]

Department of Engineering Design, Indian Institute of Technology Madras, Chennai, India

A R T I C L E   I N F O

Article history:
Received 16 January 2018
Revised 11 October 2018
Accepted 18 October 2018
Available online 19 October 2018

Keywords:
Deep learning
Ensemble classifier
Cardiac MRI
Segmentation
Automated diagnosis
Fully convolutional densenets.

A B S T R A C T

Deep fully convolutional neural network (FCN) based architectures have shown great potential in medical image segmentation. However, such architectures usually have millions of parameters and inadequate number of training samples leading to over-fitting and poor generalization. In this paper, we present a novel DenseNet based FCN architecture for cardiac segmentation which is parameter and memory efficient. We propose a novel up-sampling path which incorporates long skip and short-cut connections to overcome the feature map explosion in conventional FCN based architectures. In order to process the input images at multiple scales and view points simultaneously, we propose to incorporate Inception module's parallel structures. We propose a novel dual loss function whose weighting scheme allows to combine advantages of cross-entropy and Dice loss leading to qualitative improvements in segmentation. We demonstrate computational efficacy of incorporating conventional computer vision techniques for region of interest detection in an end-to-end deep learning based segmentation framework. From the segmentation maps we extract clinically relevant cardiac parameters and hand-craft features which reflect the clinical diagnostic analysis and train an ensemble system for cardiac disease classification. We validate our proposed network architecture on three publicly available datasets, namely: (i) Automated Cardiac Diagnosis Challenge (ACDC-2017), (ii) Left Ventricular segmentation challenge (LV-2011), (iii) 2015 Kaggle Data Science Bowl cardiac challenge data. Our approach in ACDC-2017 challenge stood second place for segmentation and first place in automated cardiac disease diagnosis tasks with an accuracy of 100% on a limited testing set (n=50). In the LV-2011 challenge our approach attained 0.74 Jaccard index, which is so far the highest published result in fully automated algorithms. In the Kaggle challenge our approach for LV volume gave a Continuous Ranked Probability Score (CRPS) of 0.0127, which would have placed us tenth in the original challenge. Our approach combined both cardiac segmentation and disease diagnosis into a fully automated framework which is computationally efficient and hence has the potential to be incorporated in computer-aided diagnosis (CAD) tools for clinical application.

Applications in Healthcare

- Data
  - Image data
  - Physiological signals-> ECG from wearable-> Impending heart attack
  - Electronic Health records-> Test Reports -> Risk of death, hospitalization
- Predictions
  - Diagnosis, treatment efficacy
  - Risk of death, hospitalisation
- Models
  - CNNs
  - NLP
  - Conventional ML (Features identified manually)

AI applications in Healthcare

- Too slow or complicated to use in a clinical setting
- AI datasets are very sanitised and might not reflect the diversity of a clinic
    - Randomised controlled trials are necessary
    - Different metrics are required to gauge utility
- Regulatory clearances for clinical adoption

AI in healthcare

- Deep Learning
  - Image analysis/interpretation in radiology, pathology, and ophthalmology
  - Diagnosis risk prediction and treatment - Lung cancer screening
  - Pathology -> Whole Slide Imaging-> survival prediction
  - Improving colonoscopy->Identify irregularities
  - Ophthalmology -> Diabetic Retinopathy

AI in healthcare

- Typical study
  - Involves image classification and segmentation
  - Compare to experts
- Non-typical
  - Uses text, genome data
  - Not necessarily supervised learning
  - Collaborating with humans

AI in healthcare

- Non-image data
  - Alphafold for prediction of protein 3D structure
  - Models for molecular analysis
  - NLP models trained on medical corpus for answering biomedical questions
  - Automatically generating radiology reports
  - Multimodal inputs

AI in healthcare

$\boxed{\text{Data}}$

↳ Annotating

- Supervised Learning
- Unsupervised learning ->clustering algorithms
- Weak supervision,-> Image label but not pixel wise labels
- Super resolution for medical images-> Recover original from synthetically blurred images

AI in healthcare

- Combination of AI and humans -> improved performance in detecting malignant nodules in chest radiographs
- Collaboration with ML models still to be studied
    - How to deal with incorrect predictions
    - Training clinicians

AI in healthcare

- Interpretability/explainability
- Data privacy - > Federated learning
- Data collection to reflect diversity of population
- Data labeling
- Regulations

Outline

1. Introduction and Motivation
2. Proposed Methodology
3. Segmentation Pipeline
4. Diagnosis Pipeline
5. Results on ACDC-2017 Dataset
6. Results on STACOM-2011 Dataset
7. Cross Model Generalization Test
8. Results on Kaggle Data Science Bowl 2016 Dataset
9. Discussion & Conclusion
10. Further Proposed Research Work

Introduction and Motivation

LV

- ➔ Cardiac cine Magnetic Resonance (MR) Imaging is primarily used for assessment of cardiac function and diagnosis of Cardiovascular diseases (CVDs)
- ➔ Cardiac MRI is considered the most accurate method for the estimation of clinical parameters such as ejection fraction, ventricular volumes, stroke volume and myocardial mass.

Cardiac Imaging Planes

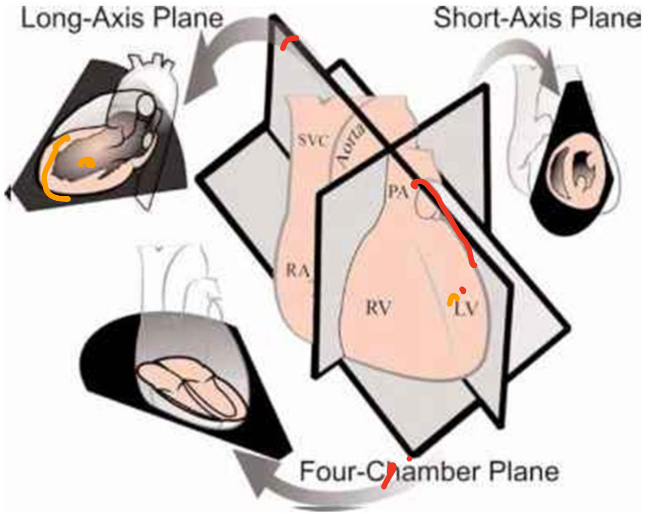
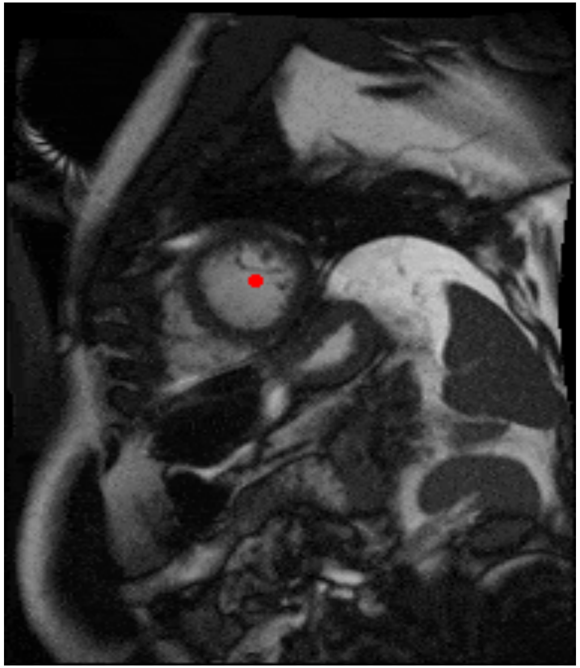


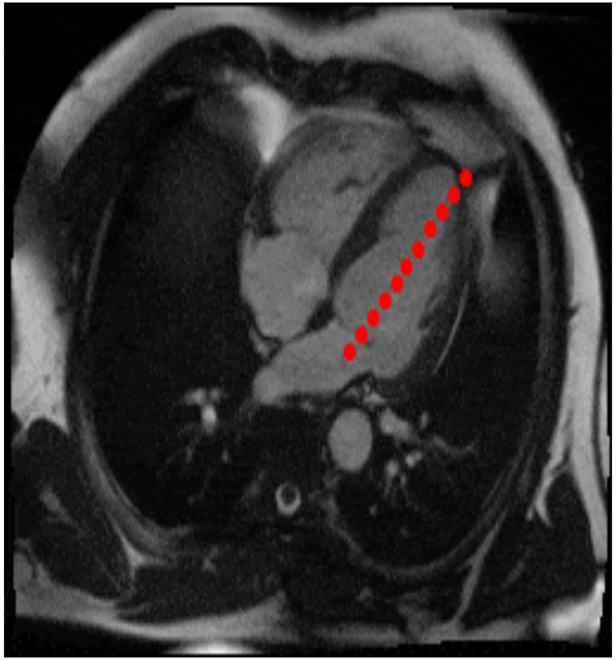
Image Courtesy: https://slideplayer.com/slide/4338139/

Cardiac Cine MRI

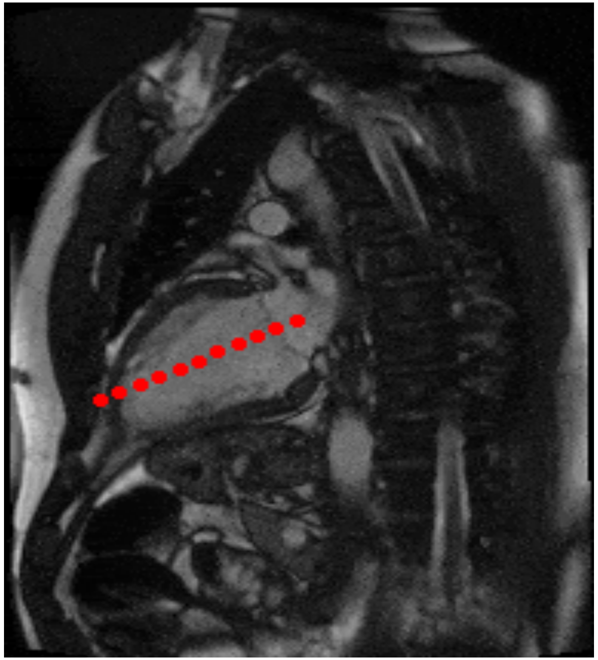
Short-Axis View



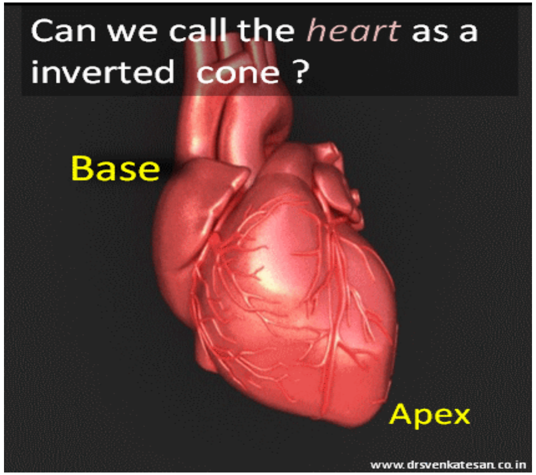
4-Chamber View

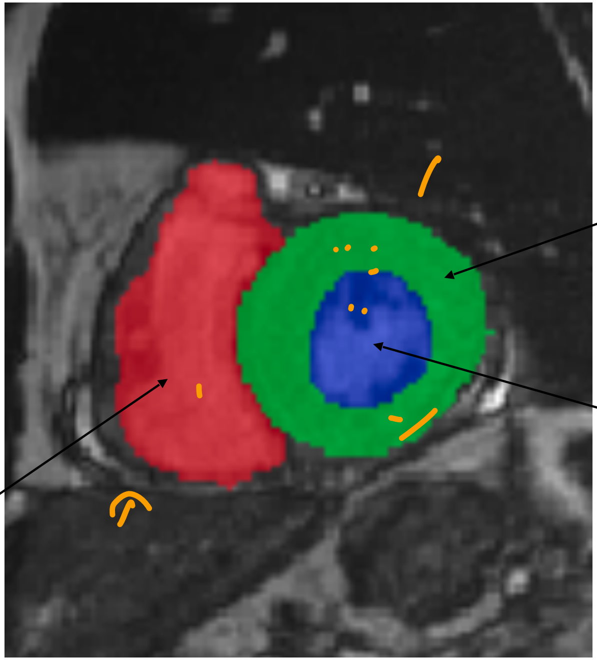


Long-Axis View



Cardiac Structures Labelling in Short-Axis View





Myocardium (**MYO**)

Left Ventricle (**LV**)

Right Ventricle (**RV**)

Proposed Methodology

chest x-ray

Automated Cardiac Segmentation and Disease Diagnosis pipeline

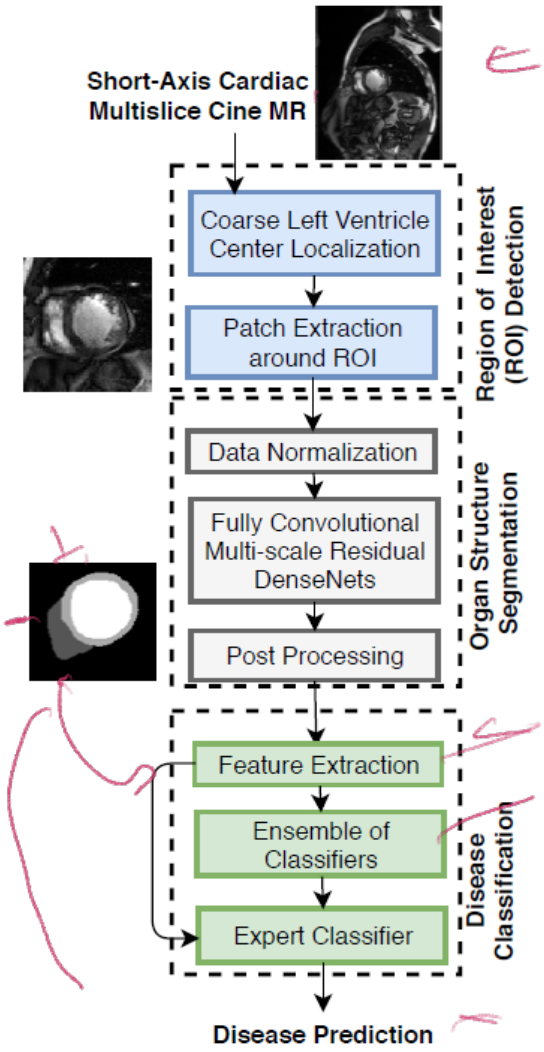
4096 x 4096

(256x256 + 256)

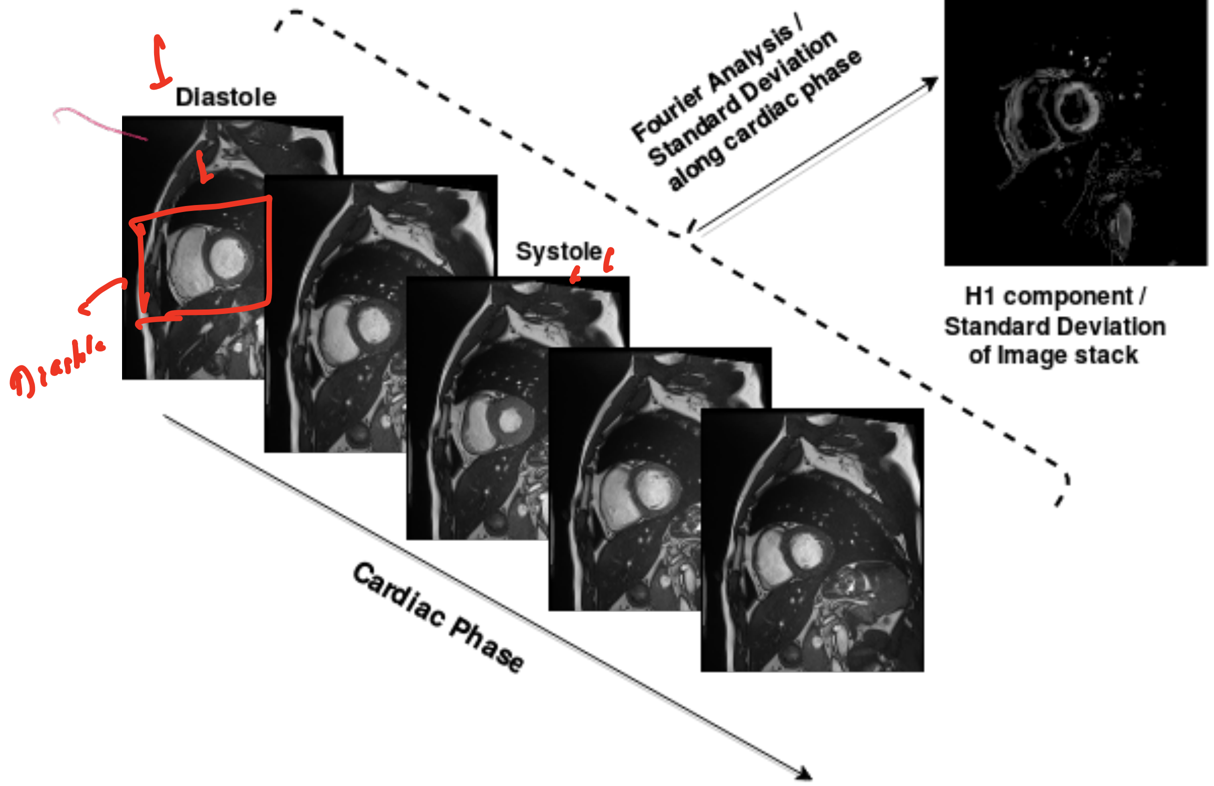
(256^3) → spatial

Segmentation

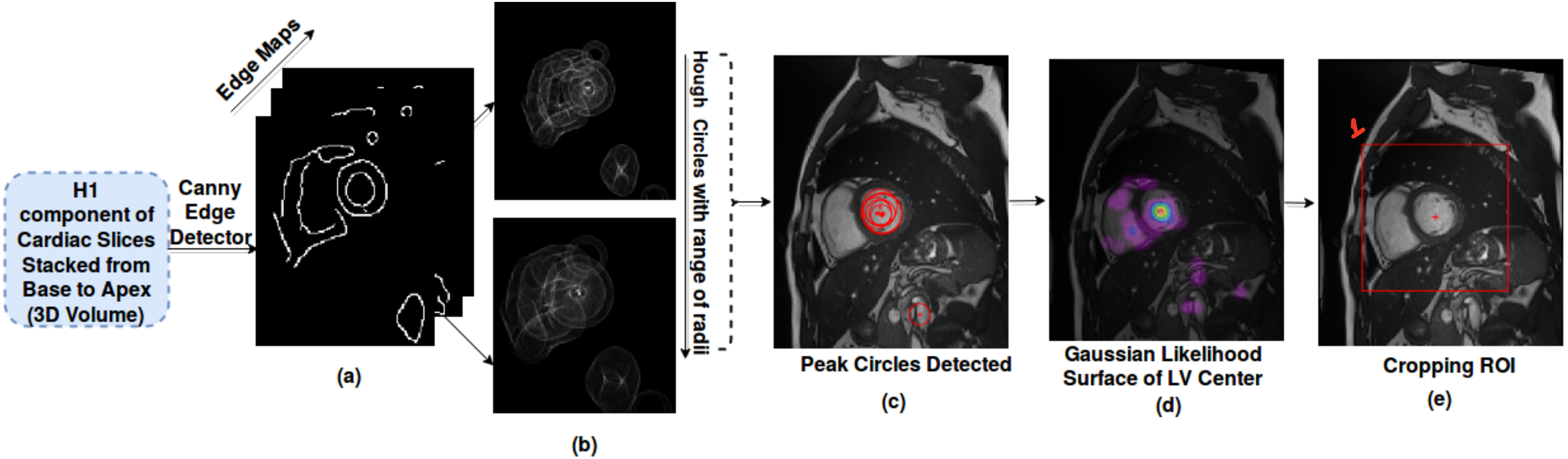
X-ray → Diagnosis



Region of Interest Extraction - Step 1



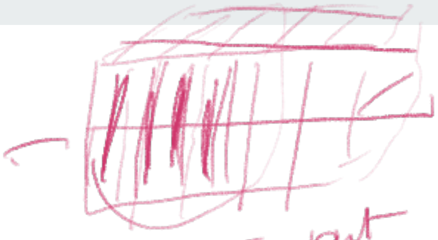
Region of Interest Extraction - Step 2



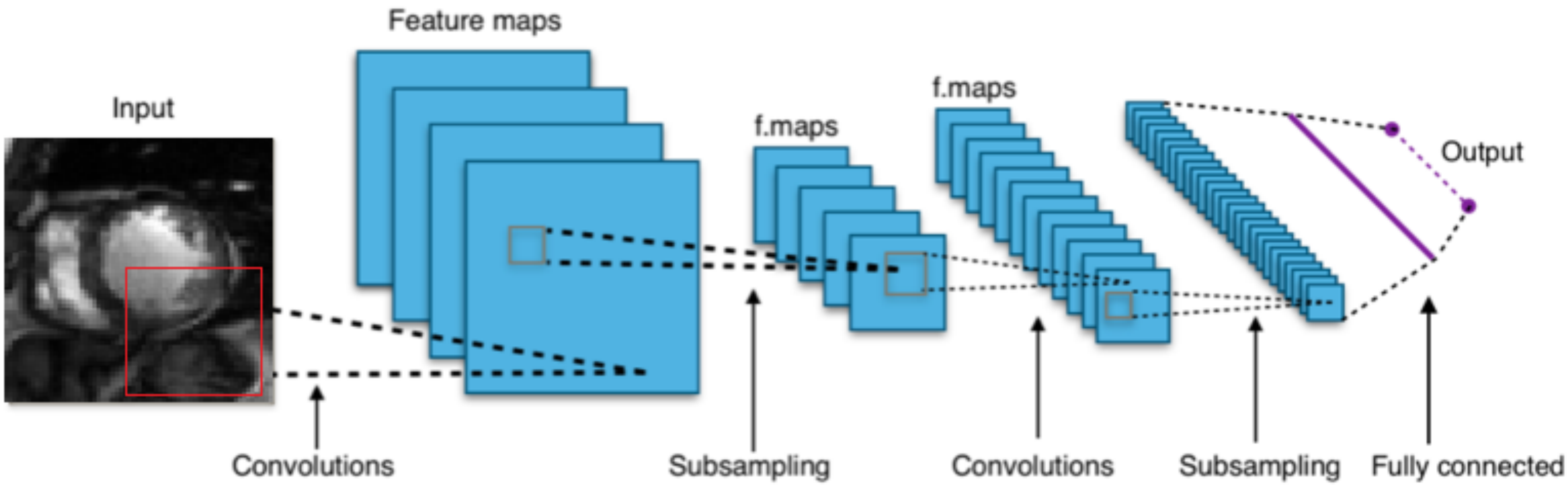
256x256x256 → 3x3x3 → 254x254x254

Typical Convolutional Neural Network (CNN)

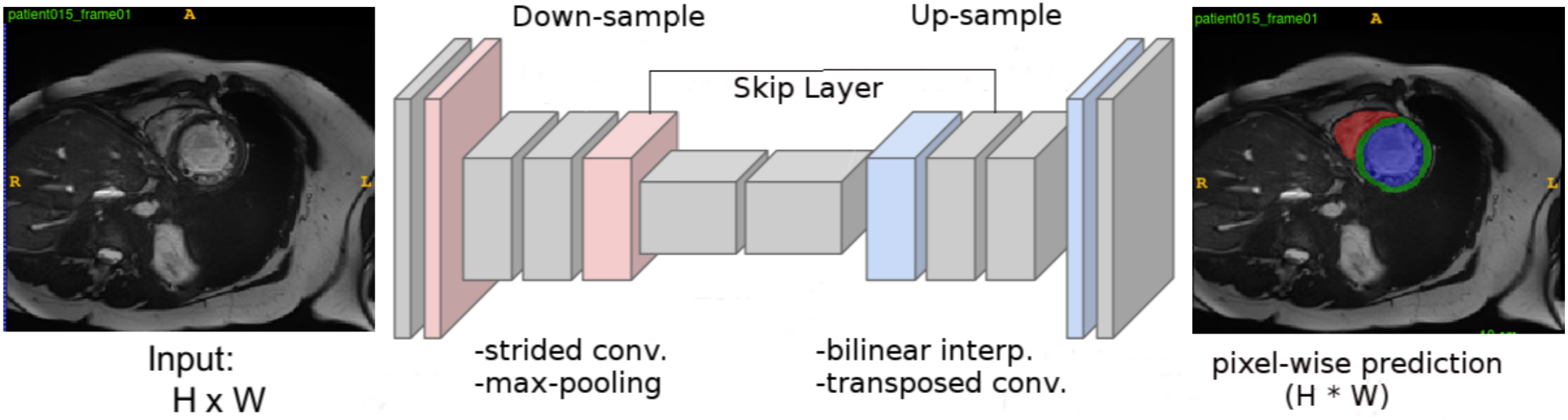
Input



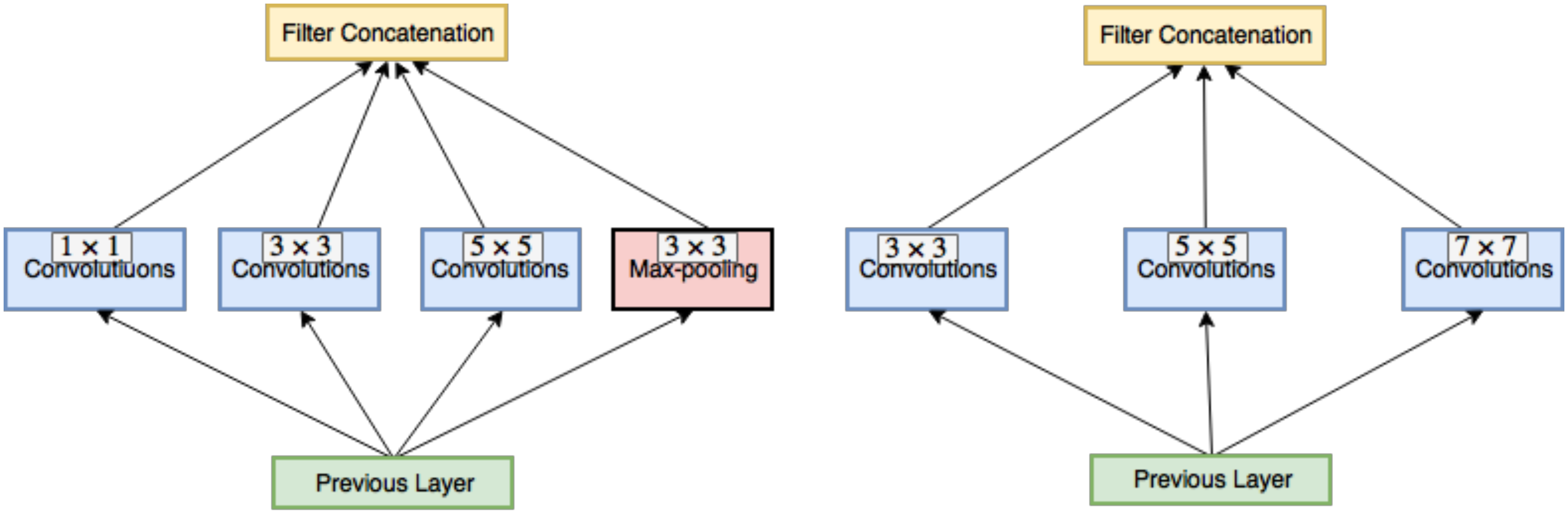
Input



Typical Fully Convolutional Architecture



Overview of Inception architecture

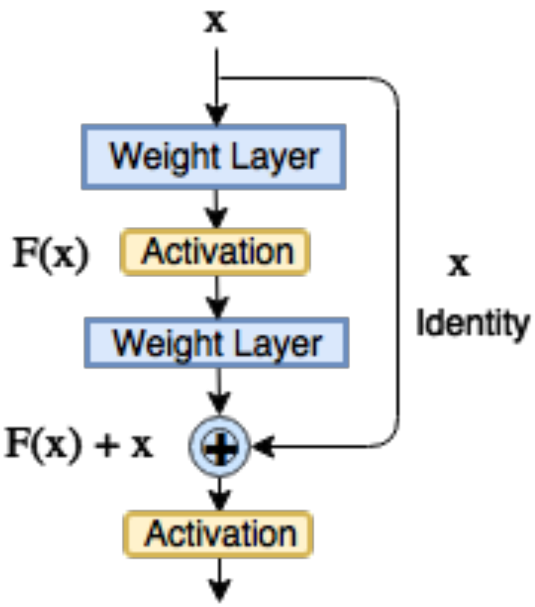


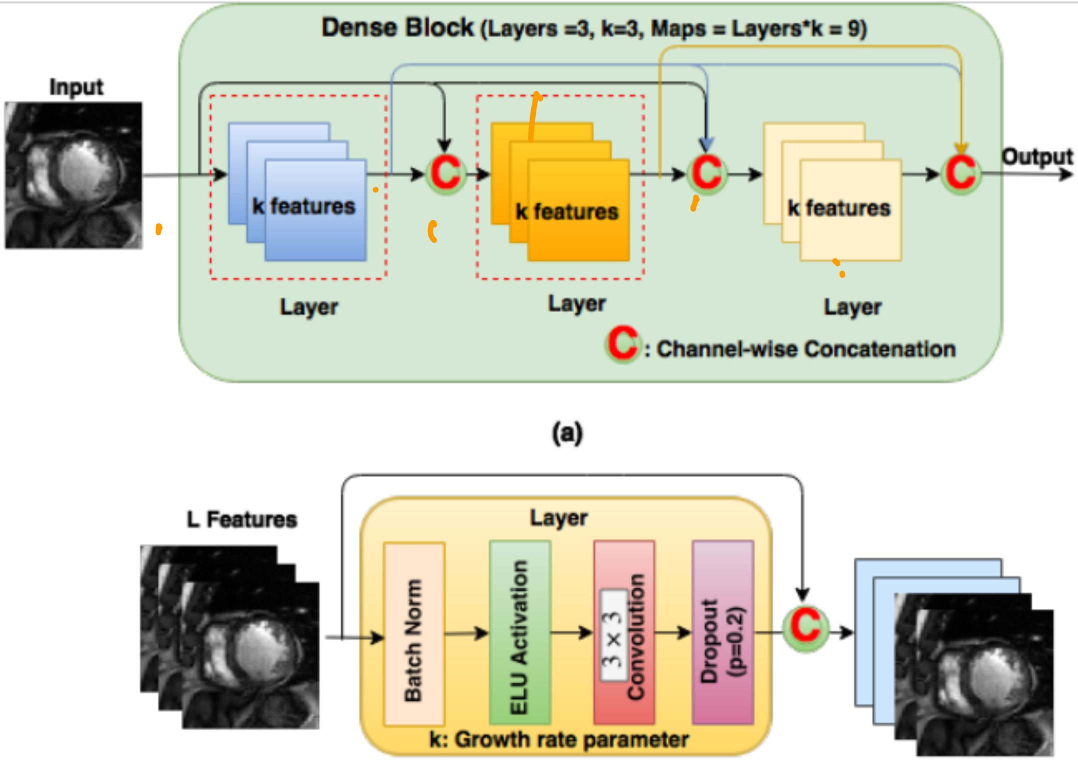
(a) Inception Module
(Naive Version)

(b) Modified
Inception Module
(Inception_X)

Overview of Residual Network and DenseNets

Residual Connection

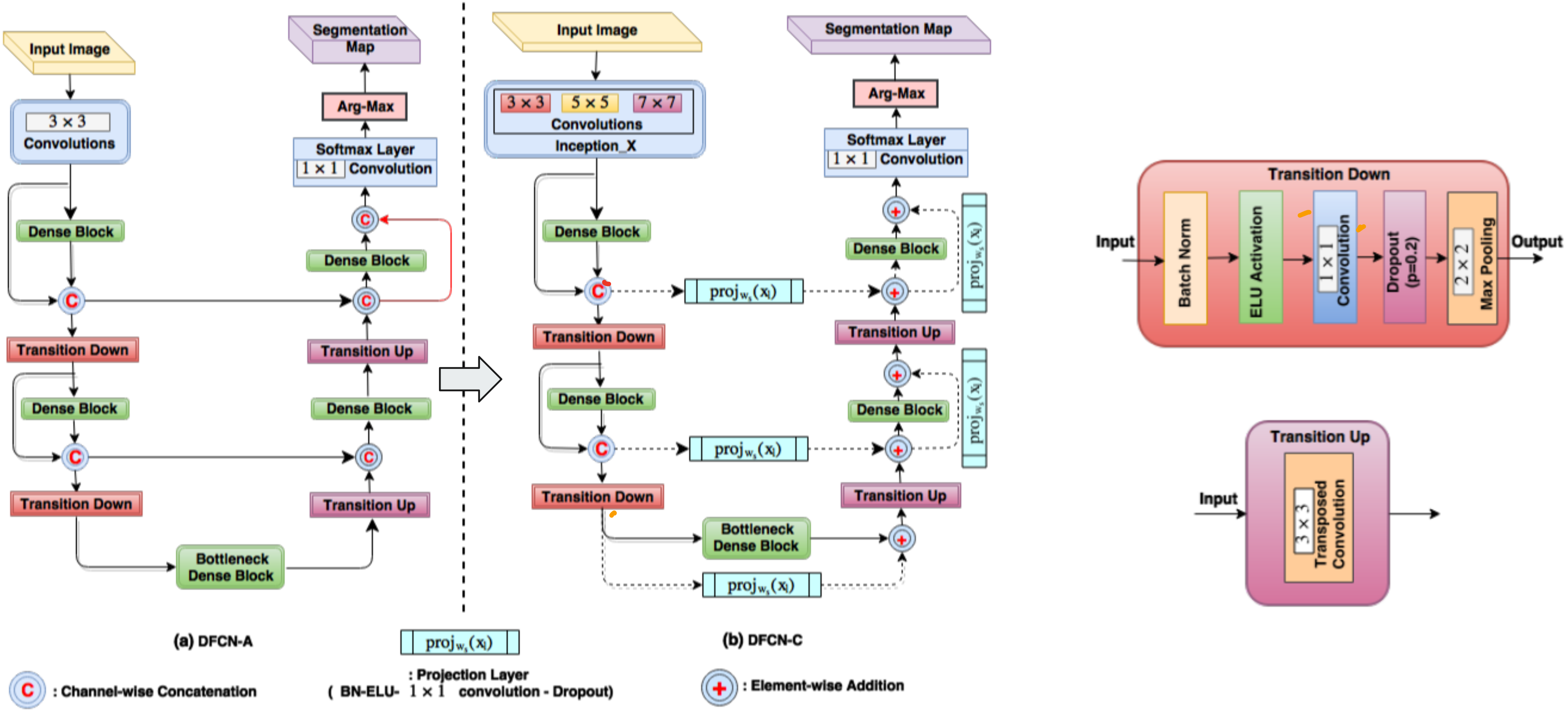




(a)

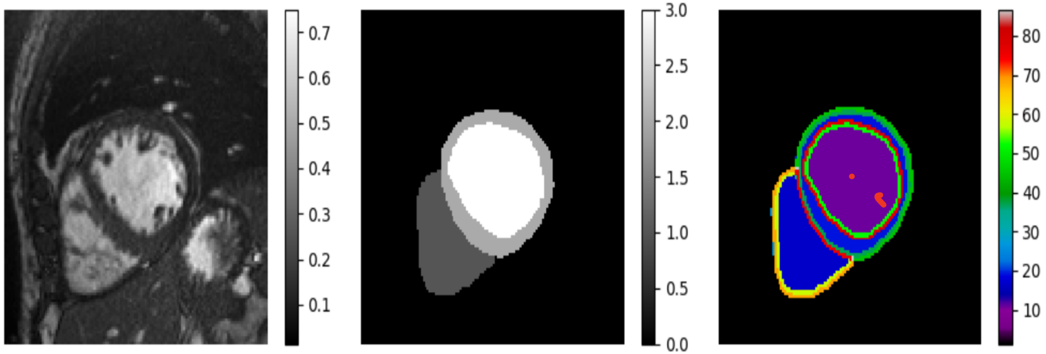
(b)

Proposed Segmentation Architecture



DFCN-A: - Jégou, Simon, et al. "The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation." *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on*. IEEE, 2017.

Proposed Loss Function



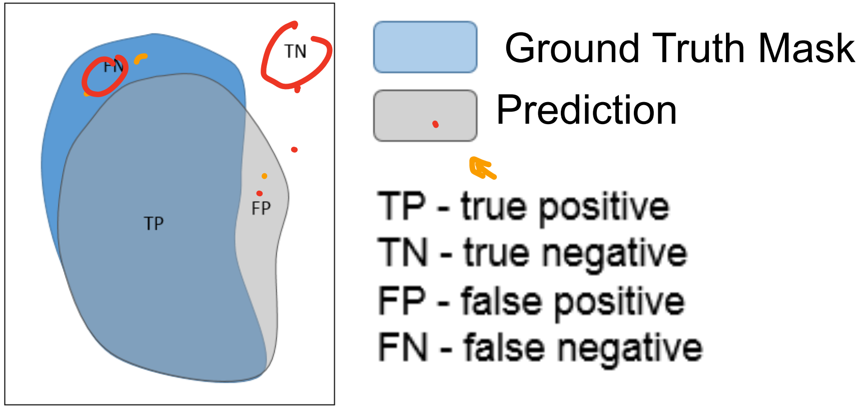
(a) Image (b) Ground Truth (c) Spatial Weight Map

$$L_{CE}(X;W) = -\sum_{x_i \in X} w_{map}(x_i) log(p(t_i|x_i;W))$$

Categorical Cross entropy

pixel wise

DICE Loss

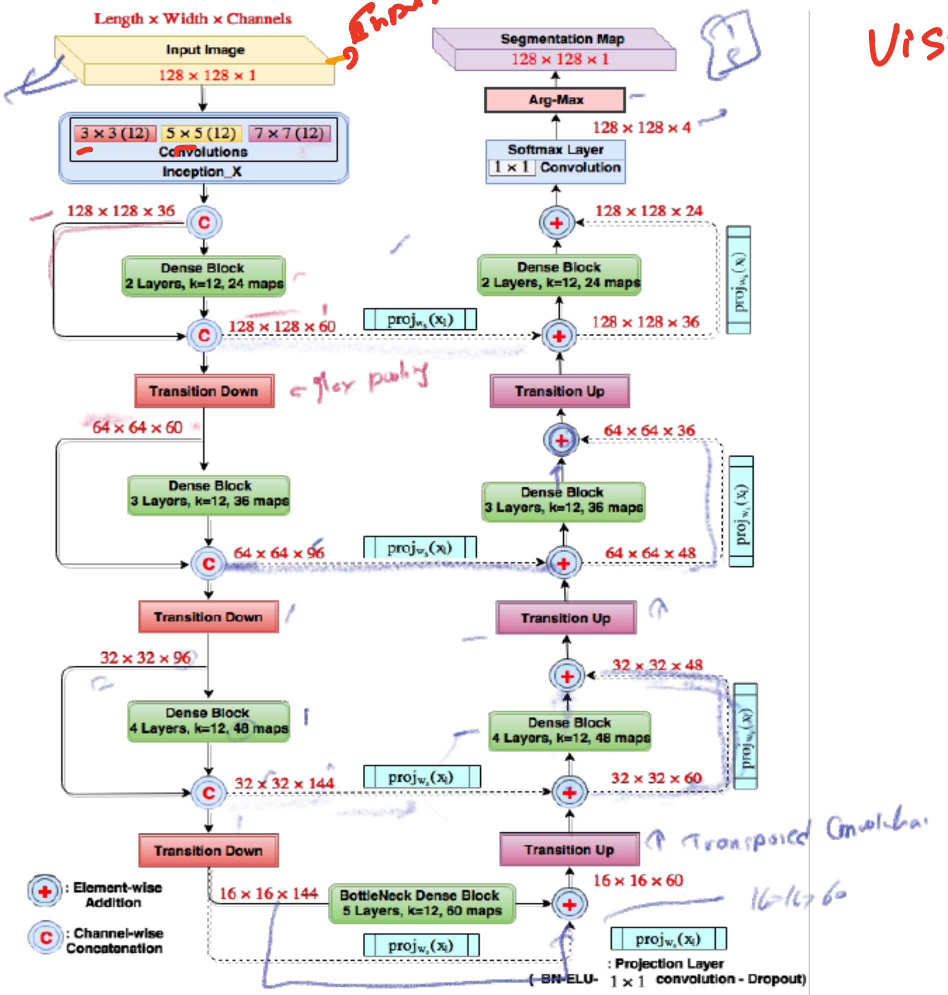


$$Dice = \frac{2 \cdot |mask \cap prediction|}{|mask| + |prediction|}$$

$$l_{DICE}(X;W) = \frac{\sum_{x_i \in X} p(t_i|x_i;W)g(x_i)+\epsilon}{\sum_{x_i \in X}(p(t_i|x_i;W)^2+g(x_i)^2)+\epsilon}$$

$$LOSS = \lambda(L_{CE}) + \gamma(1 - L_{DICE}) + \eta||W||^2$$

Network architecture & hyper-parameter tuning



Vision Transformers, ViT

Training Settings:

- Optimizer: ADAM
- Learning rate: 10^{-3}
- Weight initialization: “He Normal”
- Batch size: 16
- Input image size: 128x128
- Epochs: 200
- Trainable Parameters: **0.4M**

Data Augmentation:

- Random flip, rotate, translate, deform

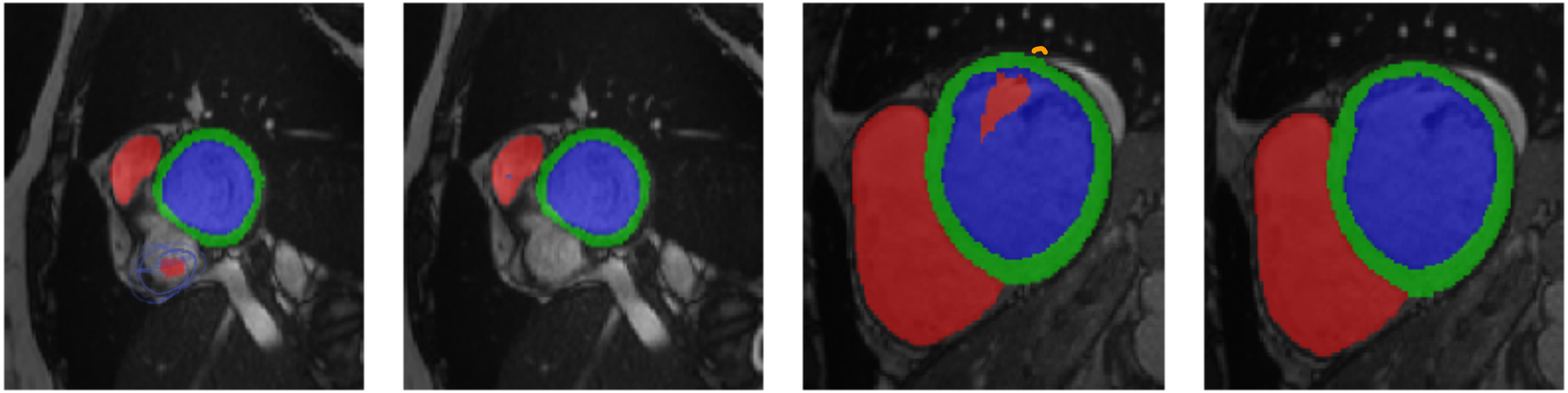
Intensity Normalization:

- Slice-wise Minmax normalization

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Proposed Post-Processing

LCCA :- Largest connected component analysis
MBHF :- Morphological binary hole filling



(a) Before LCCA (b) After LCCA (c) Before MBHF (d) After MBHF

Dataset Description

ACDC-2017 Challenge Dataset

- Training set - 100 cine MRI cases
- Ground truth annotations given for Left Ventricle, Right Ventricle and Myocardium at End Systole (ES) and End Diastole (ED) phases
- 5 Groups of patients categories:
  - **Normal (NOR)**
  - **Dilated Cardiomyopathy (DCM)**
  - **Hypertrophic Cardiomyopathy (HCM)**
  - **Myocardial Infarction (MINF)**
  - **Abnormal Right Ventricle (ARV)**
- Test set - 50 cine MRI cases

STACOM-2011 Challenge Dataset

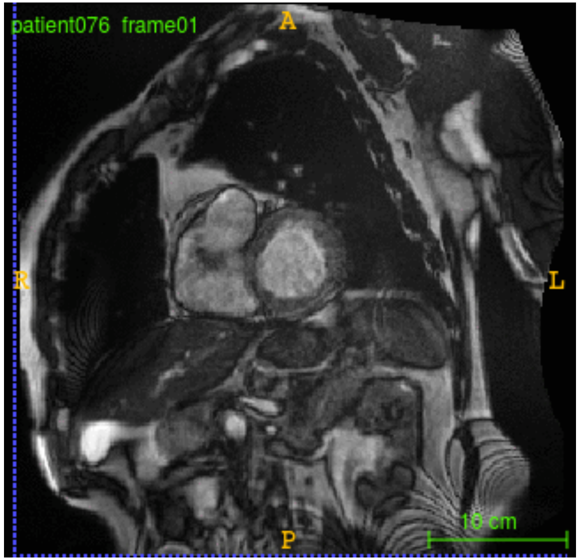
- Training set - 100 cine MRI cases
- Ground truth annotations given for Myocardium at all cardiac phases
- 2 Groups of patients categories:
  - Coronary artery disease
  - Myocardial infarction
- Test set - 100 cine MRI cases

Kaggle Datascience Bowl 2016 Challenge Dataset

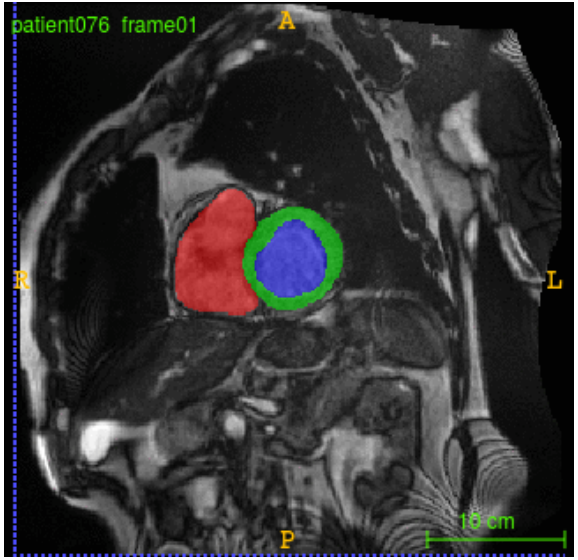
- Training set - 700 cine MRI cases
- No Annotations, only reference volumes at End Systole and End Diastole phases
- Groups of patients categories were Normal and Diseased
- Test set - 440 cine MRI cases

Segmentation Results - Normal case at ED

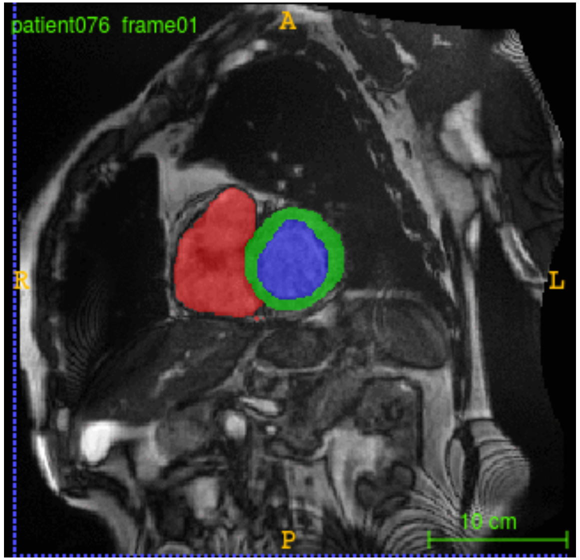
Input Image



Ground Truth

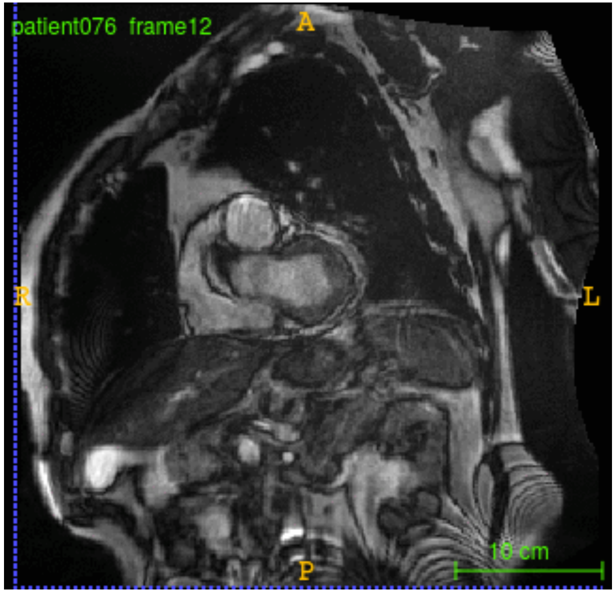


Prediction

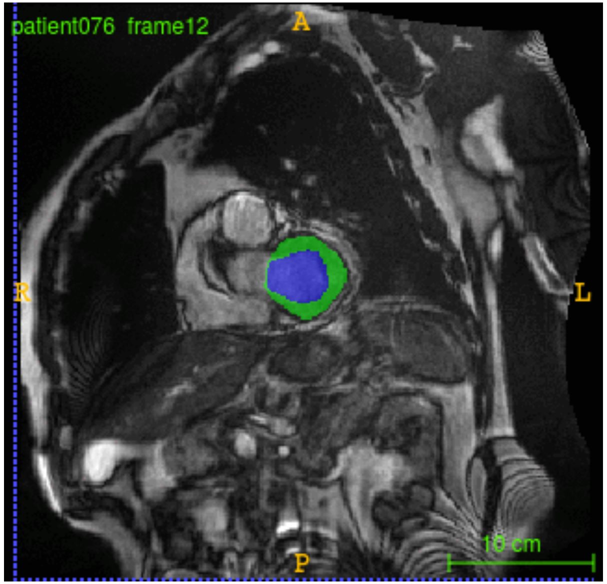


Segmentation Results - Normal case at ES

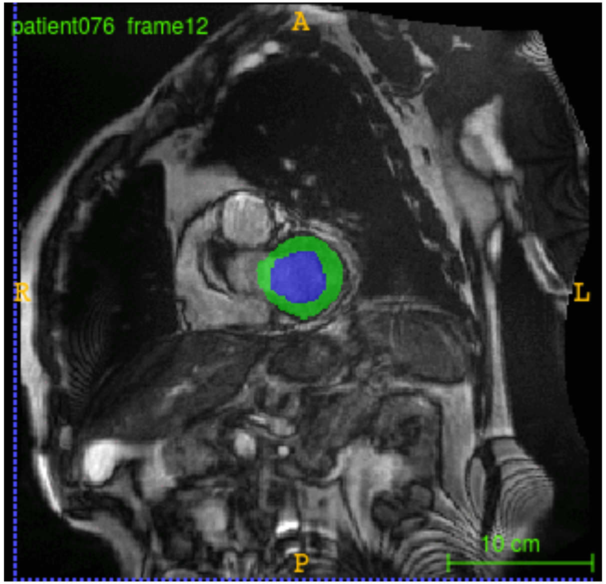
Input Image



Ground Truth

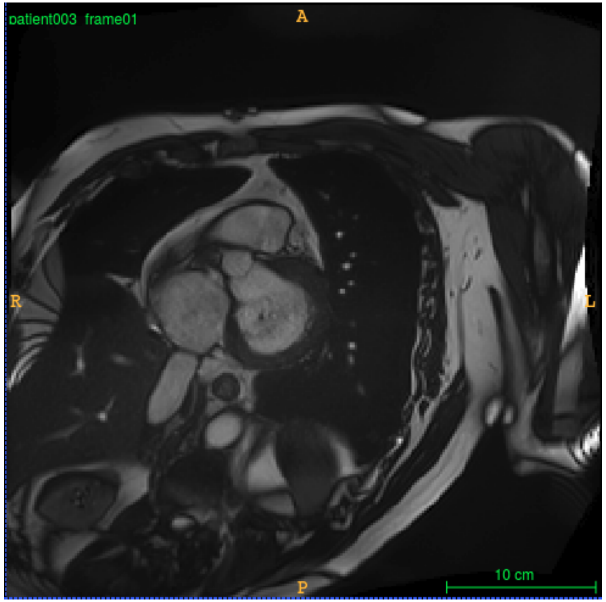


Prediction

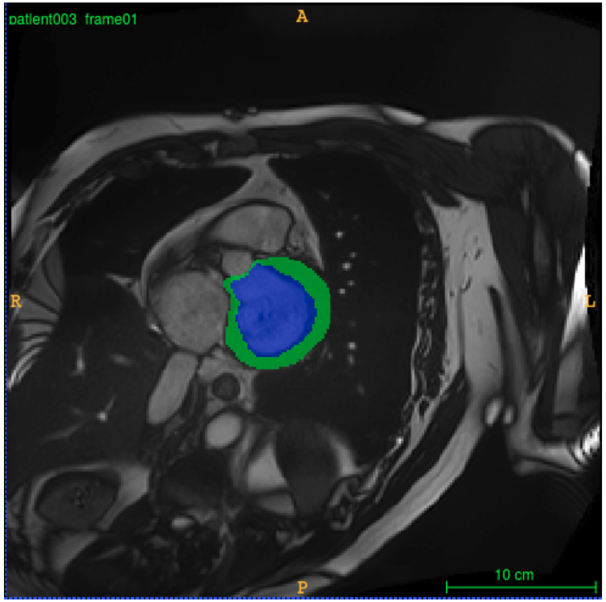


Segmentation Results - End Diastole Basal Slice

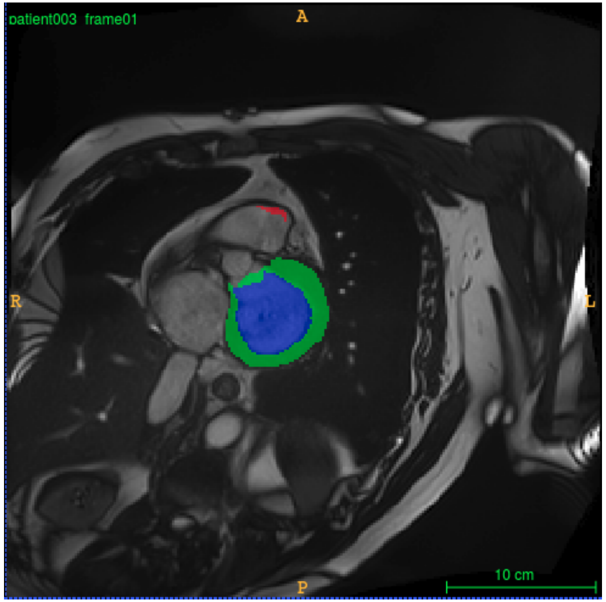
Input Image



Ground Truth

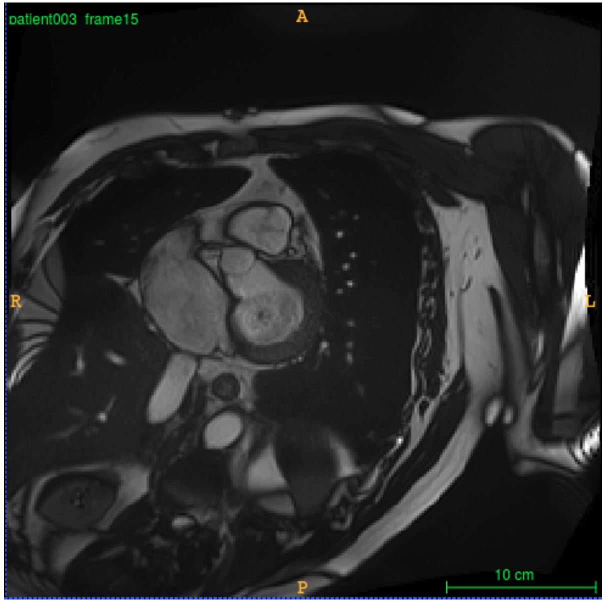


Prediction

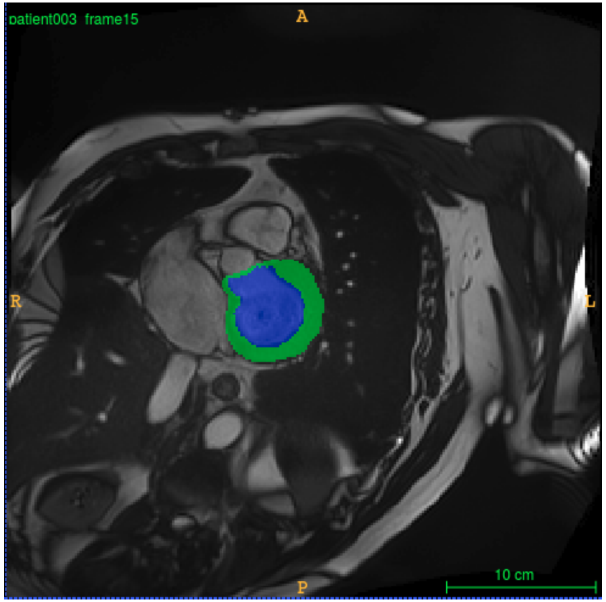


Segmentation Results - End Systole Basal Slice

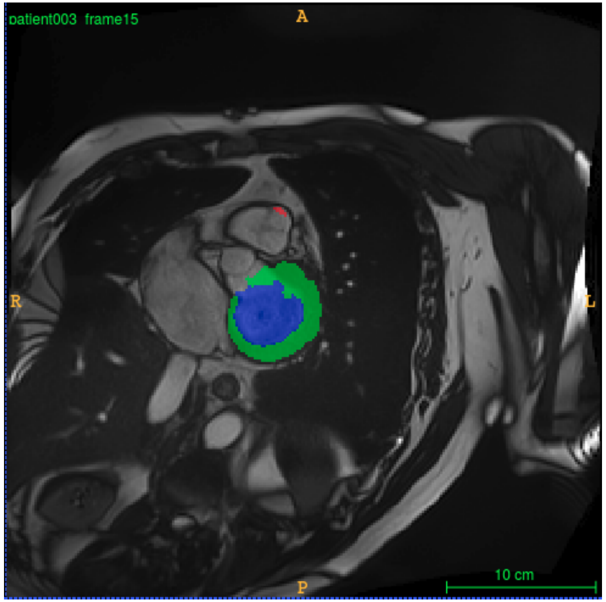
Input Image



Ground Truth

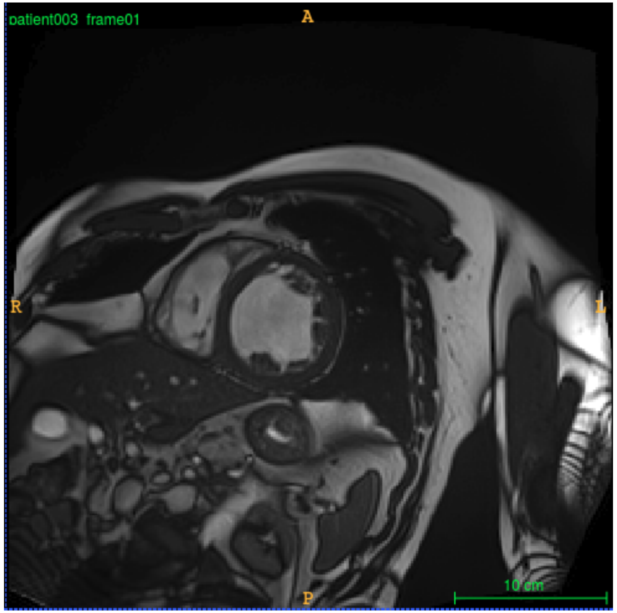


Prediction

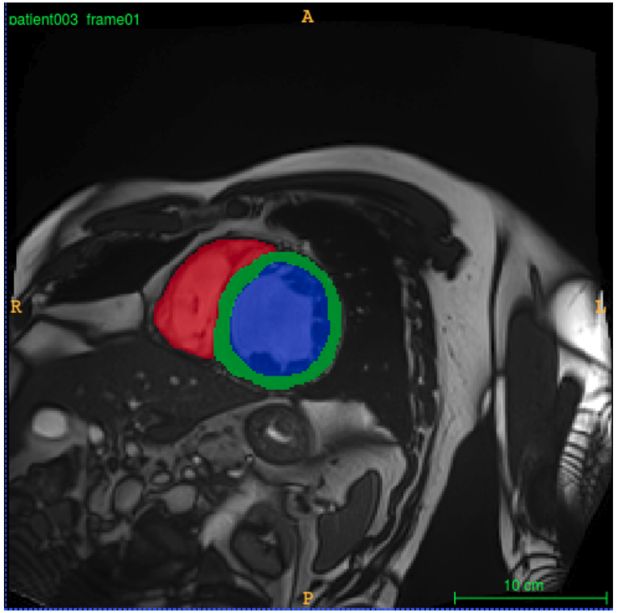


Segmentation Results -End Diastole Mid-Slice

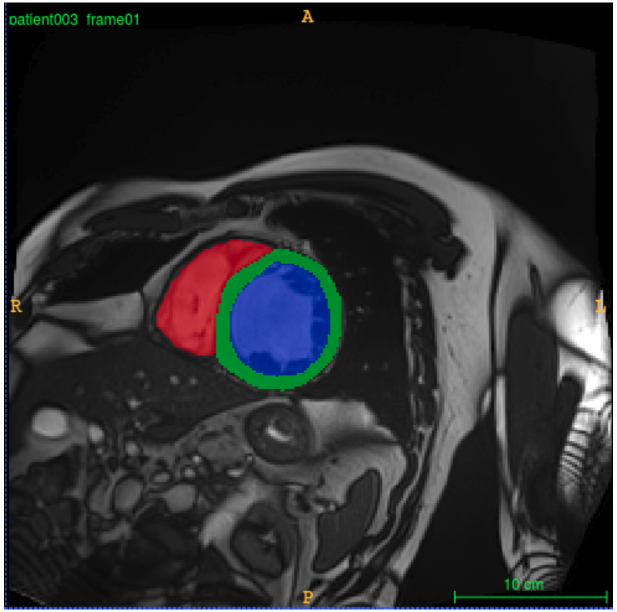
Input Image



Ground Truth

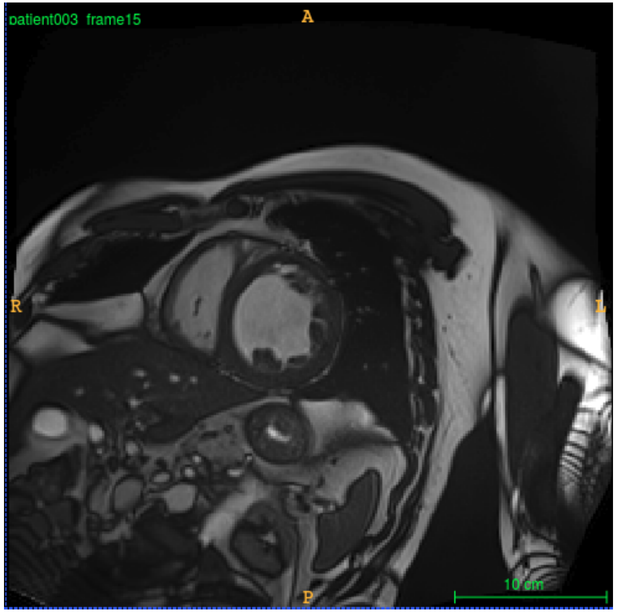


Prediction

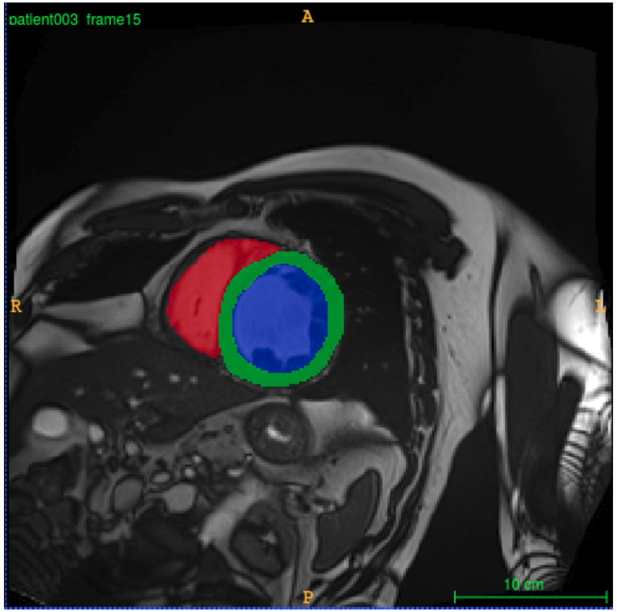


Segmentation Results - End Systole Mid-Slice

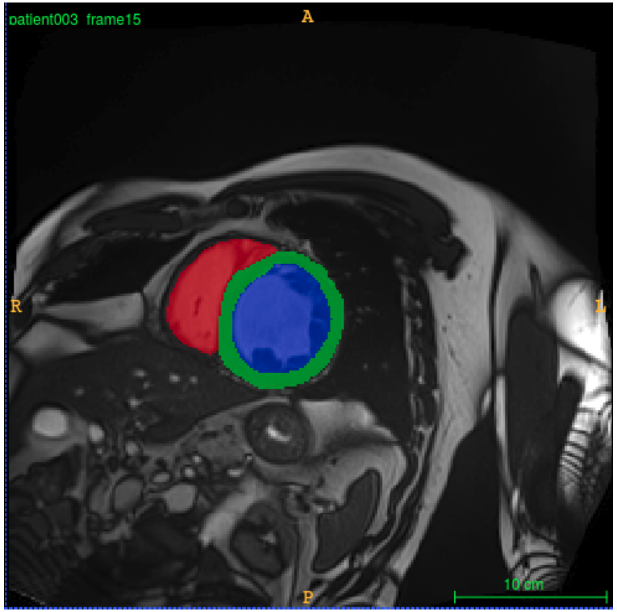
Input Image



Ground Truth

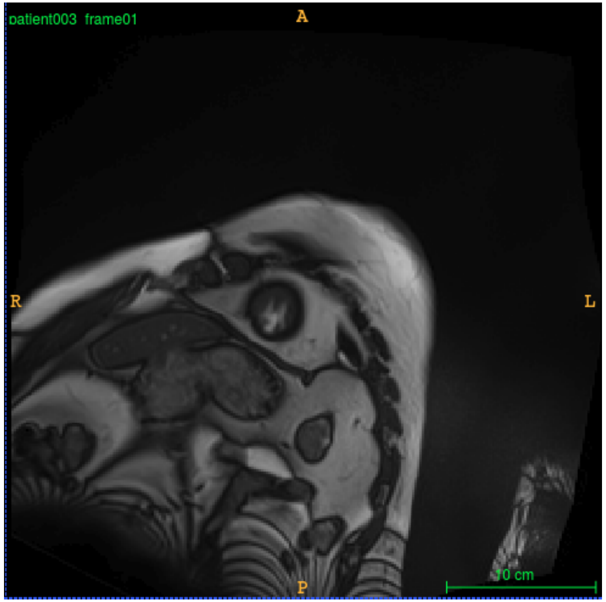


Prediction

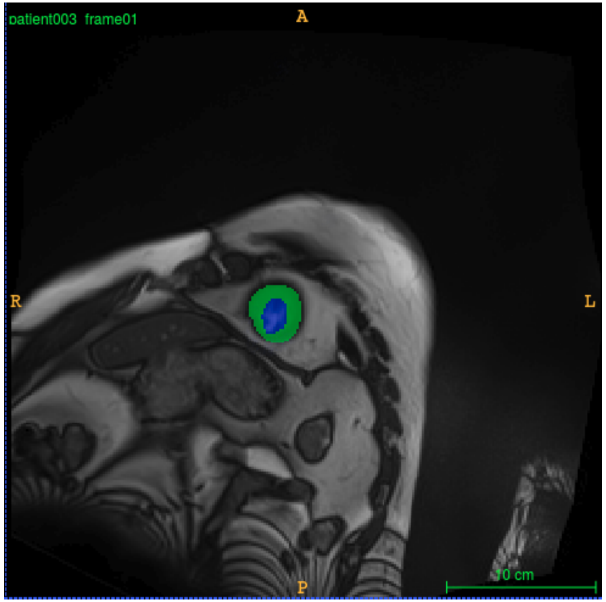


Segmentation Results - End Diastole Apex Slice

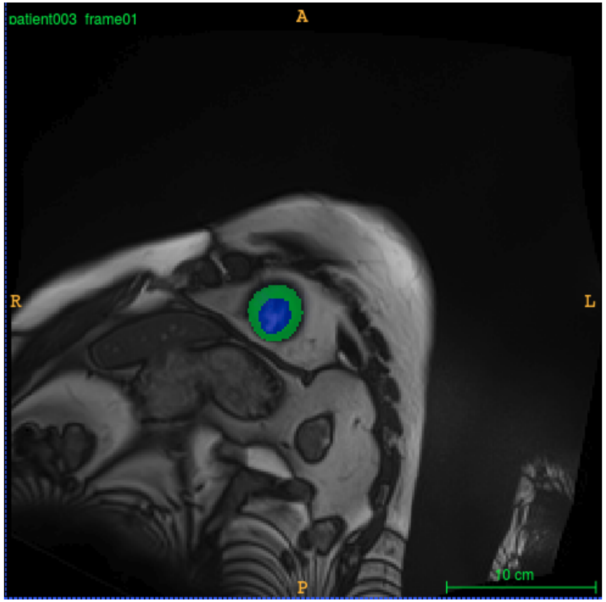
Input Image



Ground Truth

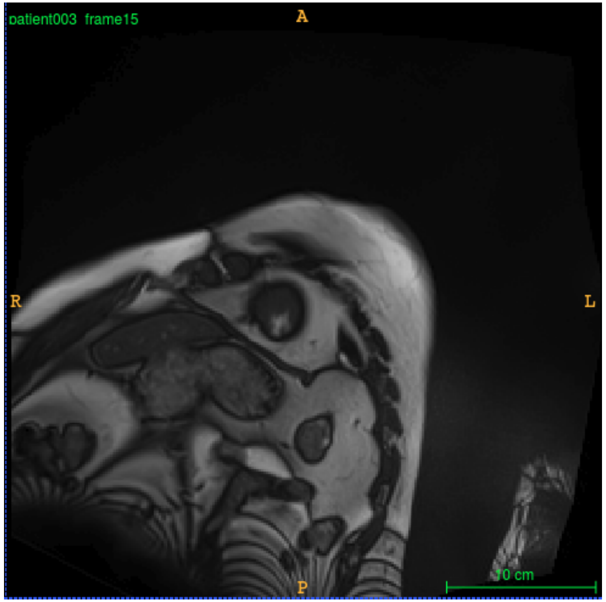


Prediction

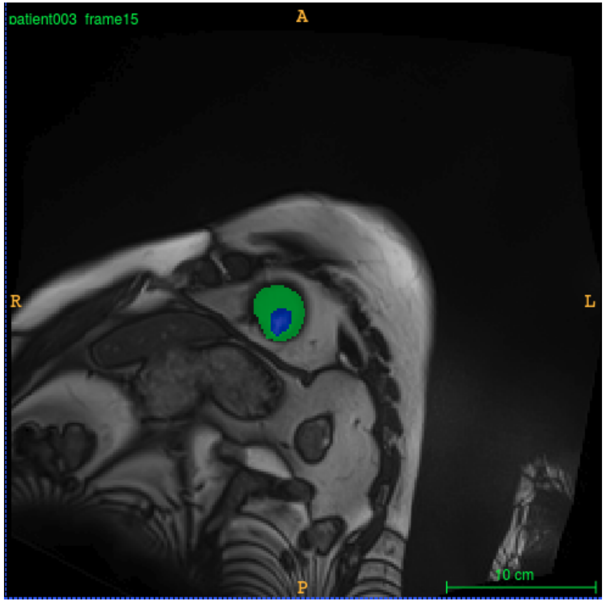


Segmentation Results - End Systole Apex slice

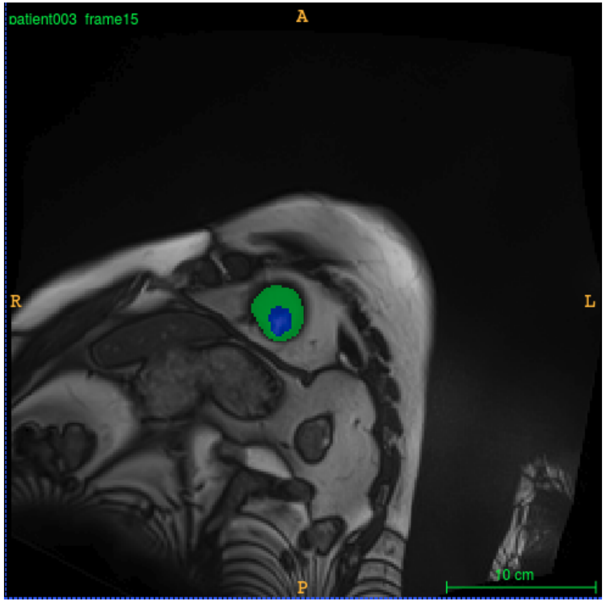
Input Image



Ground Truth



FCN Prediction



ACDC Challenge Test Set(n=50) Segmentation Results

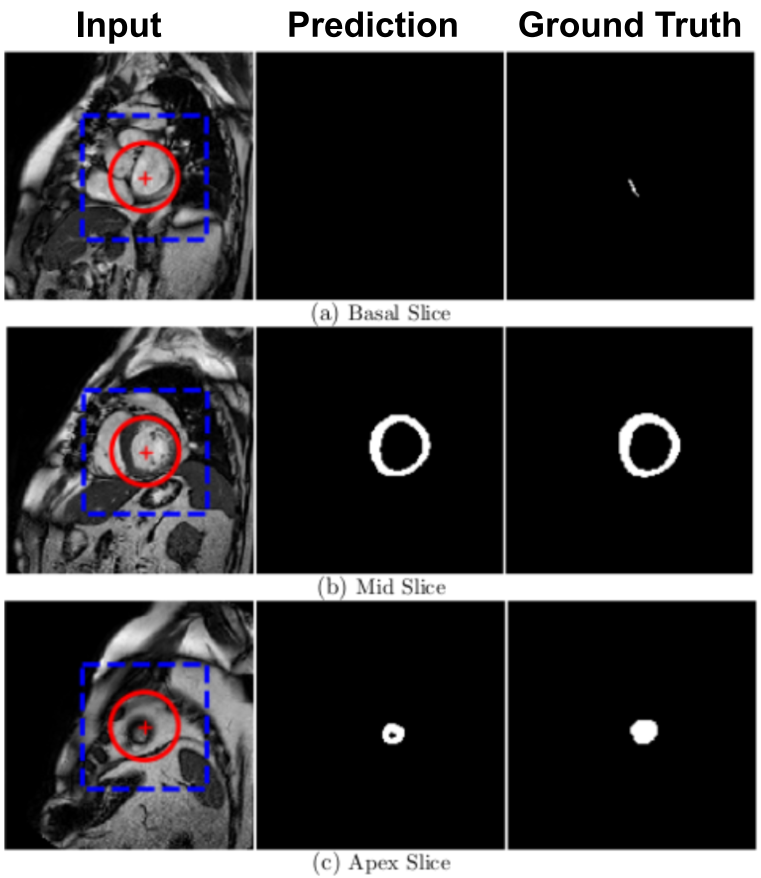
Rank	Method	DC ED	DC ES	HD ED	HD ES	EF cor.	EF bias	EF std.	Vol. ED corr.	Vol. ED bias	Vol. ED std.
LV											
1	Isensee et al. (2017)	0.968	0.931	7.384	6.905	0.991	0.178	3.058	0.997	2.668	5.726
2	**Ours**	0.964	0.917	8.129	8.968	0.989	-0.548	3.422	**0.997**	0.576	**5.501**
3	Jang et al. (2017)	0.959	0.921	7.737	7.116	0.989	-0.330	3.281	0.993	-0.440	8.701
4	Baumgartner et al. (2017)	0.963	0.911	6.526	9.170	0.988	0.568	3.398	0.995	1.436	7.610
RV											
1	Isensee et al. (2017)	0.946	0.899	10.123	12.146	0.901	-2.724	6.203	0.988	4.404	10.823
2	Zotti et al. (2018)	0.941	0.882	10.318	14.053	0.872	-2.228	6.847	0.991	-3.722	9.255
3	**Ours**	0.935	0.879	13.994	13.930	0.858	-2.246	6.953	0.982	-2.896	12.650
4	Baumgartner et al. (2017)	0.932	0.883	12.670	14.691	0.851	1.218	7.314	0.977	-2.290	15.153
MYO		**DC ED**	**DC ES**	**HD ED**	**HD ES**	**Vol. ES corr.**	**Vol. ES bias**	**Vol. ES std.**	**Mass ED corr.**	**Mass ED bias**	**Mass ED std**
1	Isensee et al. (2017)	0.902	0.919	8.720	8.672	0.985	-3.842	9.153	0.989	-4.834	7.576
2	**Ours**	0.889	0.898	9.841	12.582	0.979	-2.572	11.037	**0.990**	-2.873	**7.463**
3	Baumgartner et al. (2017)	0.892	0.901	8.703	10.637	0.983	-9.602	9.932	0.982	-6.861	9.818
4	Patravali et al. (2017)	0.882	0.897	9.757	11.256	0.986	-4.464	9.067	0.989	-11.586	8.093

DC- Dice score
HD-Hausdorff Distance
cor- correlation

$$H(P,G) = \max(h(P,G), h(G,P))$$

$$h(P,G) = \max_{p_i \in P} \min_{g_i \in G} ||p_i - g_i||$$

STACOM-2011 (n=100) Segmentation Results



Method	FA	Jaccard	Sensitivity	Specificity	PPV	NPV
AU	✗	0.84 (0.17)	0.89 (0.13)	0.96 (0.06)	0.91 (0.13)	0.95 (0.06)
CNR	✗	0.77 (0.11)	0.88 (0.09)	0.95 (0.04)	0.86 (0.11)	0.96 (0.02)
FCN	✓	0.74 (0.13)	0.83 (0.12)	0.96 (0.03)	0.86 (0.10)	0.95 (0.03)
Ours	✓	0.74 (0.15)	0.84 (0.16)	0.96 (0.03)	0.87 (0.10)	0.95 (0.03)
AO	✗	0.74 (0.16)	0.88 (0.15)	0.91 (0.06)	0.82 (0.12)	0.94 (0.06)
SCR	✓	0.69 (0.23)	0.74 (0.23)	0.96 (0.05)	0.87 (0.16)	0.89 (0.09)
DS	✗	0.64 (0.18)	0.80 (0.17)	0.86 (0.08)	0.74 (0.15)	0.90 (0.08)
INR	✓	0.43 (0.10)	0.89 (0.17)	0.56 (0.15)	0.50 (0.10)	0.93 (0.09)

$$Jaccard = \frac{TP}{TP+FN+FP}$$

$$TPR = \frac{TP}{(TP+FN)}$$

$$SPC = \frac{TN}{(TN+FP)}$$

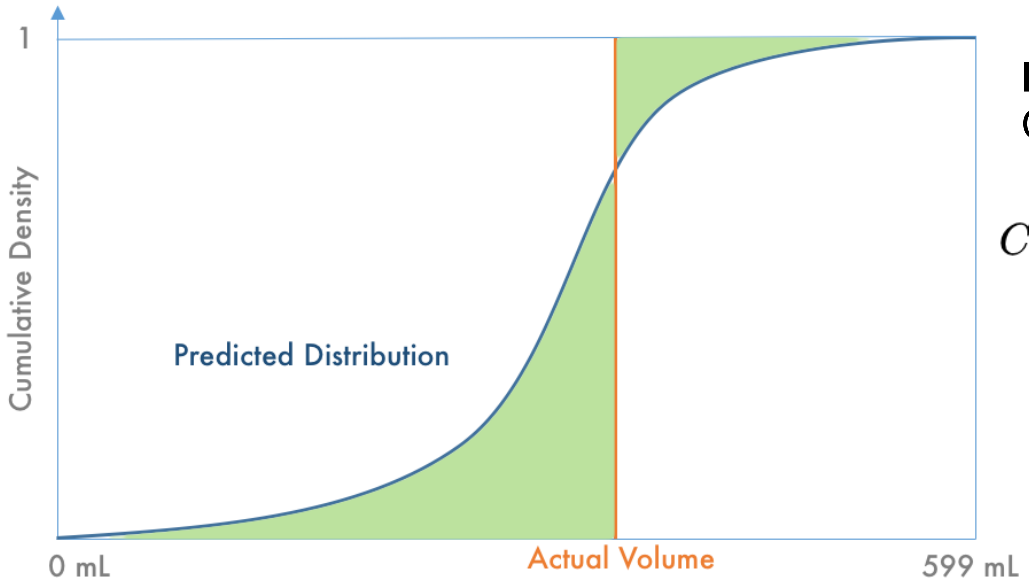
$$PPV = \frac{TP}{(TP+FP)}$$

$$NPV = \frac{TN}{(TN+FN)}$$

Jaccard
Apex: 0.68±0.16
Mid: 0.78±0.13
Base: 0.74±0.18

Note:
Sensitivity (TPR)
Specificity (SPC)
Positive Predictive Value (PPV)
Negative Predictive Value (NPV)
FA- Fully Automated

ACDC Model Generalization Test on Kaggle Challenge



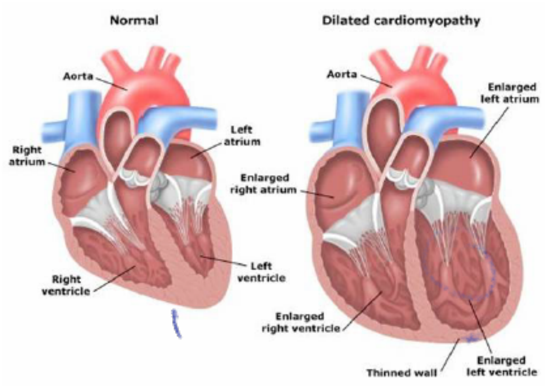
Evaluation Metric:
Continuous Ranked Probability Score (CRPS)

$$CRPS = \frac{1}{600N} \sum_{m=1}^{N} \sum_{n=0}^{599} (P(y \leq n) - H(n - V_m))^2$$

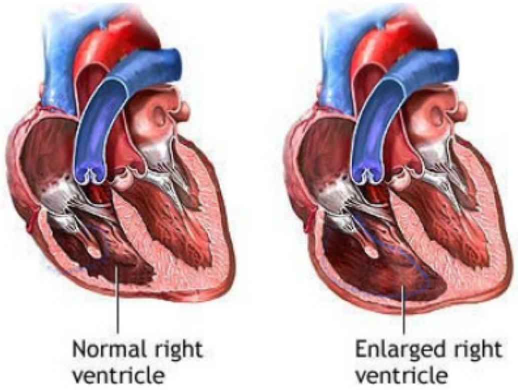
P - Predicted CDF
N - Number of cases (440*2)
V - Actual Volume in mL
H(x) = 1 for x ≥ 0 & 0 otherwise

Our CRPS Score: 0.0127 (10th Position in the actual challenge)

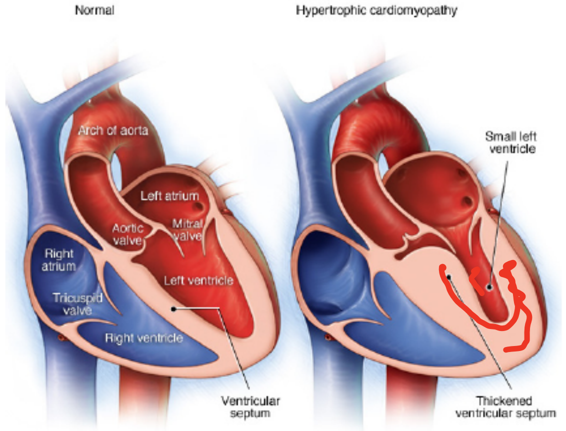
Automated Cardiac Disease Diagnosis



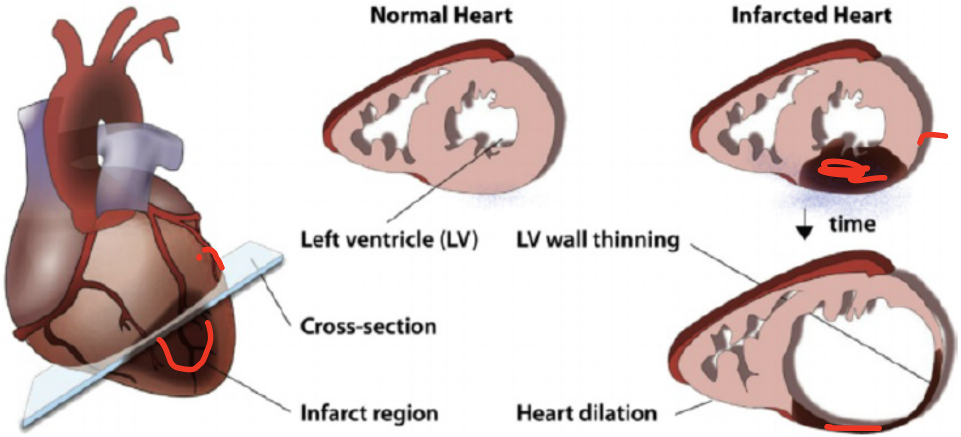
Normal Heart vs DCM [1]



Normal Heart vs ARV [2]



Normal Heart vs HCM [3]



Normal Heart vs MINF [4]

Image Courtesy :

1) http://www.bichonhealth.org/HealthInfo/HereditaryCardiac.htm
2) http://www.nytimes.com/health/guides/disease/tricuspid-regurgitation/overview.html
3) http://www.mayoclinic.org/diseases-conditions/hypertrophic-cardiomyopathy/home/ovc-20122102
4) Hassan Awada et al. "Towards Comprehensive Cardiac Repair and Regeneration after Myocardial Infarction: Aspects to Consider and Proteins to Deliver"

Feature Extraction

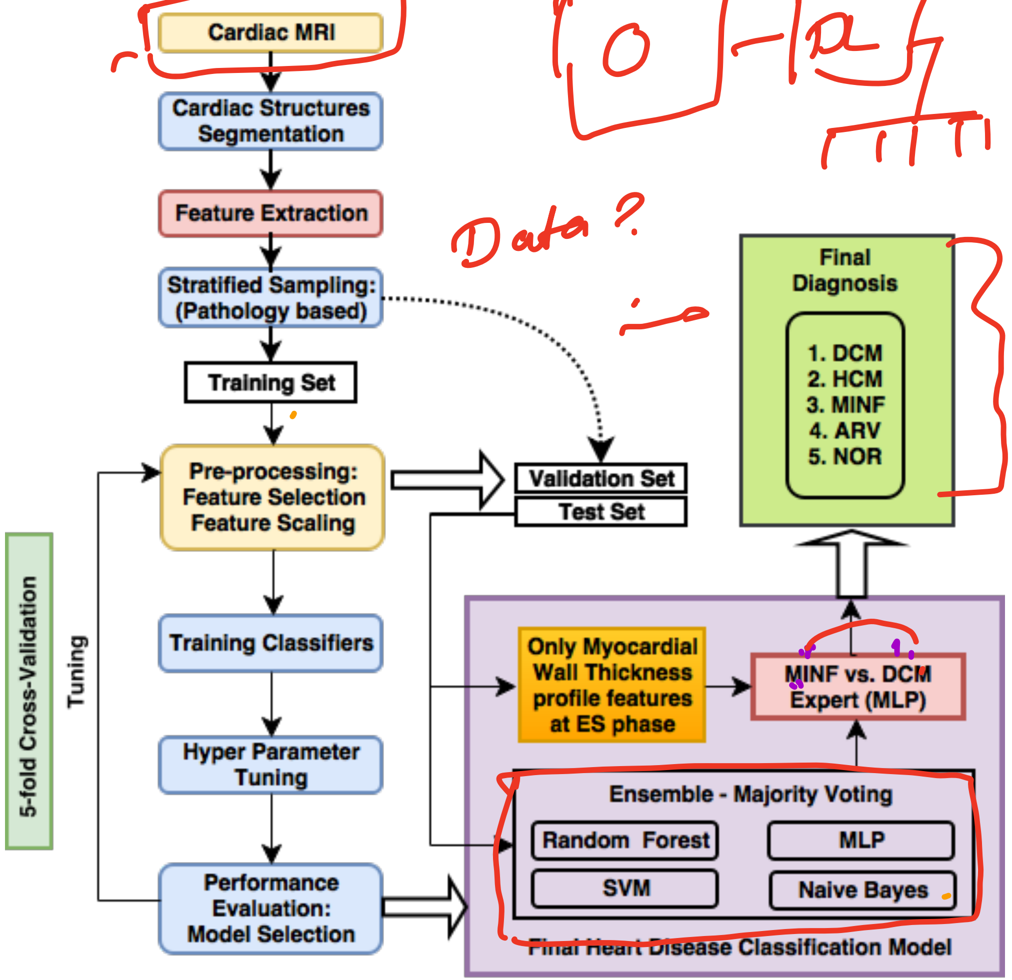
Primary Features:

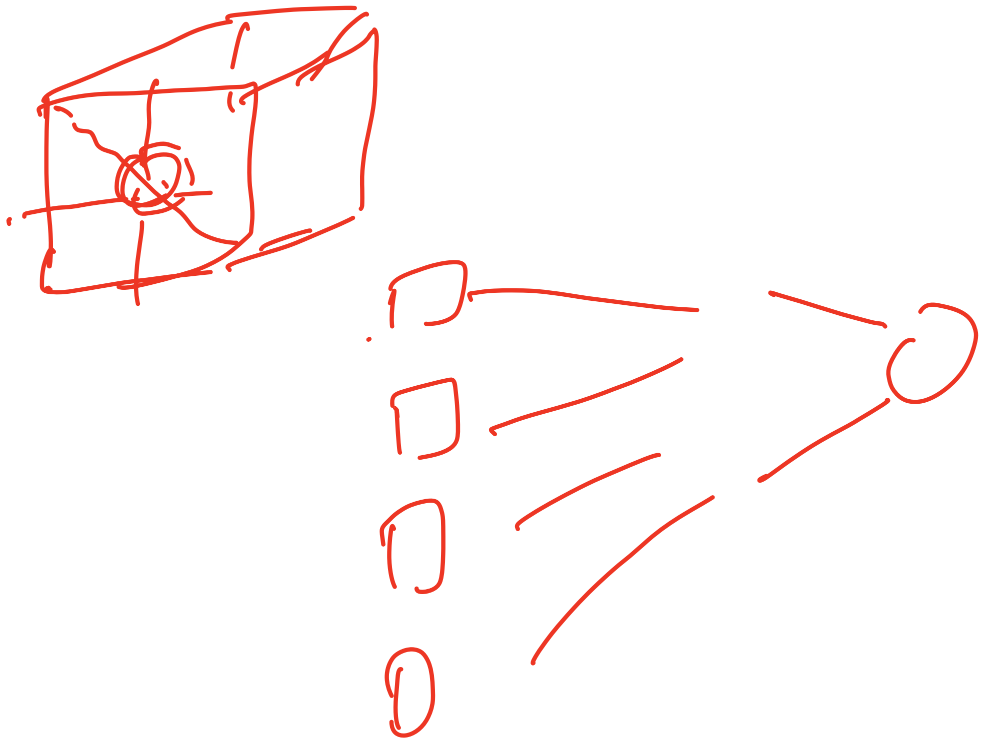
- Volume of LV, RV & MYO at ED and ES
- MYO wall thickness at each slice

Derived Features:

- Ejection Fraction of LV and RV
- RV : LV, MYO : LV at ED and ES
- Standard deviation of MYO wall thickness measures within a slice and across slices.

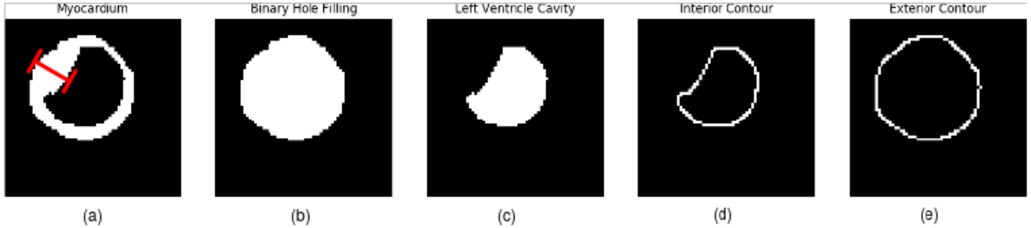
Data ?

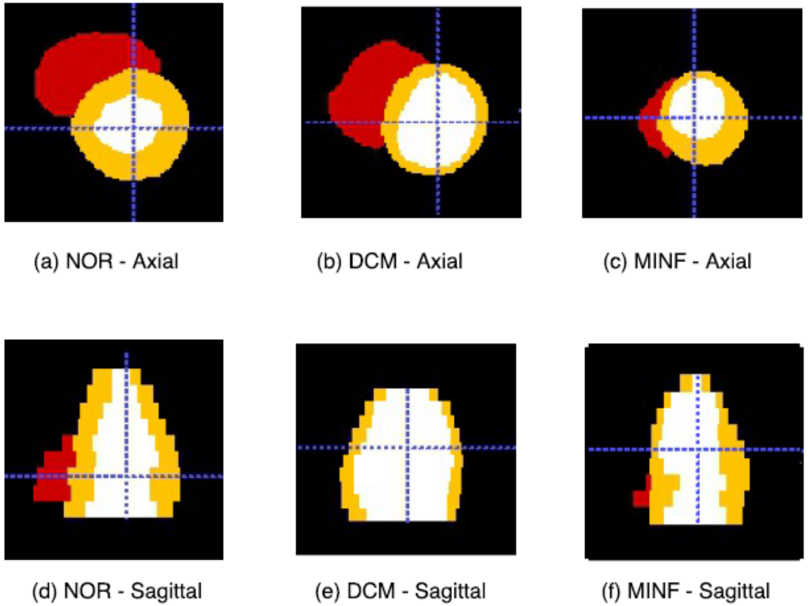




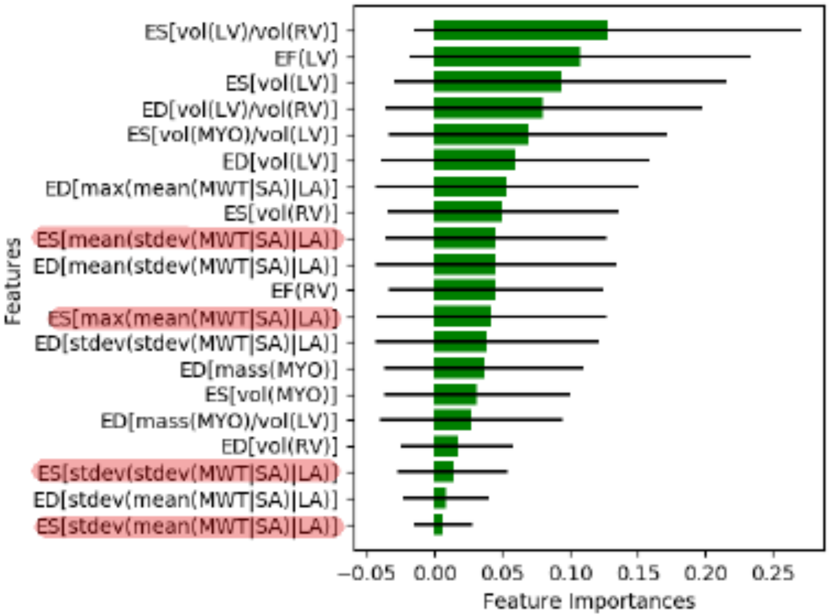
Cardiac Features

Features	LV	RV	MYO
Cardiac volumetric features			
volume at ED	✓	✓	✓
volume at ES	✓	✓	✓
ejection fraction	✓	✓	
volume ratio: $ED[vol(LV)/vol(RV)]$	✓	✓	
volume ratio: $ES[vol(LV)/vol(RV)]$	✓	✓	
volume ratio: $ES[vol(MYO)/vol(LV)]$	✓		✓
volume ratio: $ED[vol(MYO)/vol(LV)]$	✓		✓
Myocardial wall thickness variation profile			
$ED[max(mean(MWT\|SA)\|LA)]$			✓
$ED[stdev(mean(MWT\|SA)\|LA)]$			✓
$ED[mean(stdev(MWT\|SA)\|LA)]$			✓
$ED[stdev(stdev(MWT\|SA)\|LA)]$			✓
$ES[max(mean(MWT\|SA)\|LA)]^*$			✓
$ES[stdev(mean(MWT\|SA)\|LA)]^*$			✓
$ES[mean(stdev(MWT\|SA)\|LA)]^*$			✓
$ES[stdev(stdev(MWT\|SA)\|LA)]^*$			✓

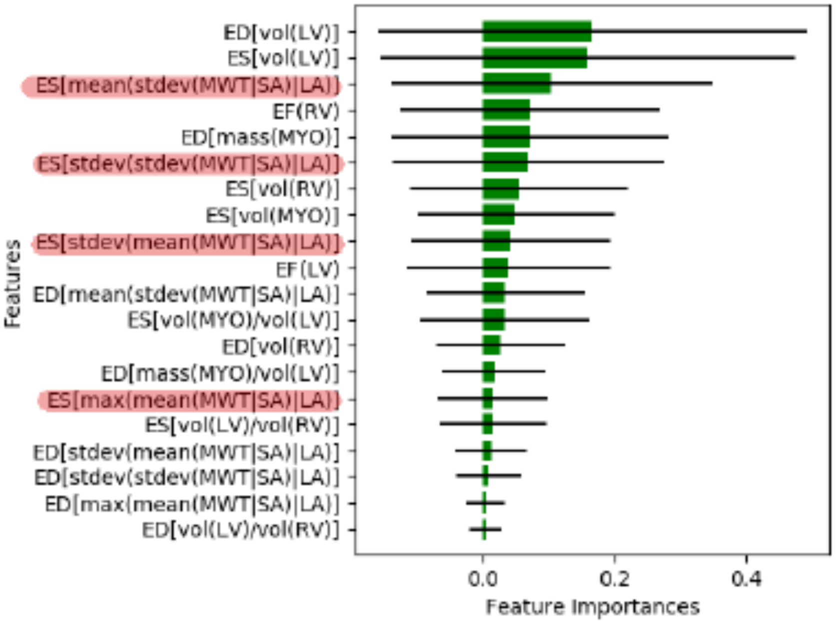




Cardiac Feature Importance Study



(a) 5-Class Feature Importance



(b) 2-Class Feature Importance

ACDC Challenge Test Set(n=50) Diagnosis Results

Results of Combined Stage 1 and Stage 2 Classifiers

Rank	Method	Accuracy
1	Ours	1
2	Isensee et al. (2017) and Cetin et al. (2017)	0.92
3	Wolterink et al. (2018)	0.86

Results of First Stage Classifier

Group →	NOR	DCM	HCM	MINF	ARV
Recall	1	0.8	1	0.8	1
Precision	1	0.8	1	0.8	1
Overall Accuracy	0.92				

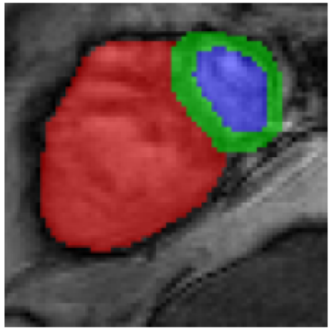
	NOR	DCM	HCM	MINF	RV
NOR	10	0	0	0	0
DCM	0	8	0	2	0
HCM	0	0	10	0	0
MINF	0	2	0	8	0
RV	0	0	0	0	10

Discussion & Conclusion

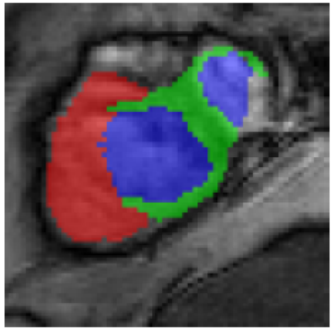
Our work proposed:

- A parameter and memory efficient 2-D multi-scale FCN based on residual DenseNets.
- A novel weighting scheme for combining the benefits of cross-entropy and Dice loss.
- State-of-the-art performance on two challenging cardiac segmentation tasks
.
- A set of novel hand-crafted features for cardiac disease diagnosis.
- State-of-the-art performance on automated cardiac disease diagnosis.

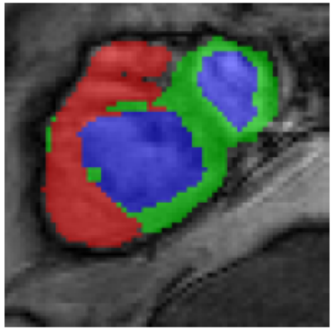
Appendix - Comparison of Loss Function vs Output



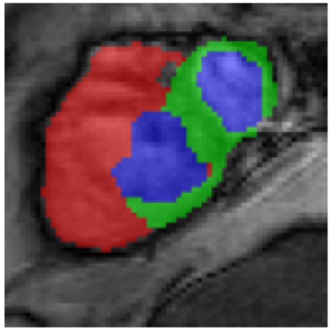
(a) Ground truth



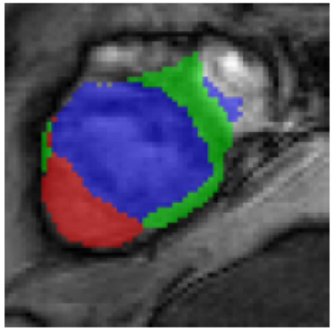
(b) Cross entropy loss



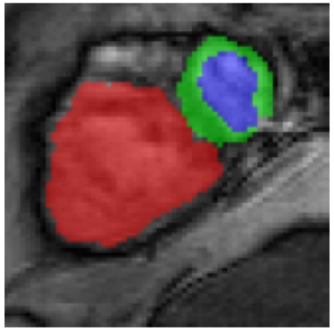
(c) Dice loss



(d) sW-CE loss



(e) mW-D loss



(f) sW-CE + mW-D loss

Method	DICE LV		DICE RV		DICE MYO		Mean Dice
	ED	ES	ED	ES	ED	ES	
CE	0.96 (0.02)	0.91 (0.07)	0.94 (0.02)	0.88 (0.06)	0.87 (0.03)	0.88 (0.03)	0.91 (0.04)
sW-CE	0.96 (0.02)	0.91 (0.06)	0.92 (0.09)	0.85 (0.17)	0.89 (0.03)	0.89 (0.03)	0.90 (0.07)
D	0.96 (0.02)	0.91 (0.09)	0.95 (0.02)	0.88 (0.07)	0.89 (0.03)	0.90 (0.03)	0.91 (0.04)
mW-D	0.96 (0.02)	0.90 (0.08)	0.95 (0.01)	0.87 (0.06)	0.87 (0.02)	0.89 (0.03)	0.90 (0.04)
CE+D	0.96 (0.02)	0.90 (0.08)	0.93 (0.04)	0.86 (0.11)	0.88 (0.03)	0.88 (0.04)	0.90 (0.05)
sW-CE + mW-D	0.96 (0.02)	0.90 (0.08)	0.95 (0.02)	0.87 (0.08)	0.89 (0.03)	0.89 (0.03)	0.91 (0.04)
	HD LV		HD RV		HD MYO		Mean HD
	ED	ES	ED	ES	ED	ES	
CE	2.98 (2.93)	4.73 (3.78)	4.81 (1.94)	6.73 (3.51)	3.57 (2.55)	7.90 (7.12)	5.12 (3.64)
sW-CE	3.82 (4.01)	4.04 (2.51)	5.09 (2.62)	6.19 (3.93)	4.46 (2.82)	4.85 (2.83)	4.74 (3.12)
D	4.70 (6.58)	7.82 (9.72)	4.82 (1.79)	6.41 (3.26)	4.59 (4.93)	6.16 (5.44)	5.75 (5.29)
mW-D	4.49 (8.35)	7.45 (9.01)	10.14 (8.1)*	9.62 (6.65)	6.47 (7.38)	9.54 (10.37)	7.95 (8.31)*
CE+D	7.09 (10.24)	9.47 (10.86)	4.85 (2.08)	6.93 (2.88)	7.67 (8.90)	11.56 (9.27)*	7.93 (7.37)*
sW-CE + mW-D	4.42 (6.39)	6.51 (6.40)	4.20 (1.81)	7.10 (2.95)	4.48 (4.52)	5.87 (4.35)	5.43 (4.40)

Source Code and Implementation Walk Through

https://github.com/mahendrakhened/Automated-Cardiac-Segmentation-and-Disease-Diagnosis

ACDC Data Preparation

1. Register and download ACDC-2017 dataset from https://www.creatis.insa-lyon.fr/Challenge/acdc/index.html
2. Create a folder outside the project with name ACDC_DataSet and copy the dataset.
3. From the project folder open file data_preprocess/acdc_data_preparation.py.
4. In the file, set the path to ACDC training dataset is pointed as: `complete_data_path = '../../ACDC_DataSet/training'` .
5. Run the script acdc_data_preparation.py.
6. The processed data for training is generated outside the project folder named *processed_acdc_dataset*.

Steps to train the model:

1. From the project folder open file estimators/train.py and configure the network hyper-parameters.
2. From the project folder open file estimators/config.py and configure the training hyper-parameters.
3. Run the script train.py.
4. Outside the project in the folder named *trained_models/ACDC/* the model weights and tensorboard summary are saved.
5. While training the training summary can be accessed running: `tensorboard --logdir='path_to/trained_models/ACDC/FCRD_ACDC/summary'` .

Steps to test the model:

1. From the project folder open file estimators/test.py and configure the testing hyper-parameters, path to trained model weights and ACDC-2017 testing dataset.
2. Run the script test.py.
3. The predictions are saved in the path *trained_models/ACDC/FCRD_ACDC/predictions_TIMESTAMP*

Cardiac Diagnosis

1. Extract Features from the training dataset by running: *generate_cardiac_features_train.py*
2. Extract Features from the testing dataset by running: *generate_cardiac_features_test.py*
3. Run the scripts *stage_1_diagnosis.py* and *stage_2_diagnosis.py*
4. The final cardiac disease prediction results on the test set are generated in the prediction folder in a csv file

ACDC & STACOM-2011 Models Generalization Test

ACDC model on STACOM-2011 testing dataset

Training Dataset (No. of images)	Jaccard	Dice	Accuracy	Sensitivity	Specificity	PPV	NPV
LV-2011 (20,360)	0.74 (0.15)	0.84 (0.14)	0.93 (0.04)	0.84 (0.16)	0.96 (0.03)	0.87 (0.10)	0.95 (0.03)
ACDC-2017 (1,352)	0.71 (0.13)	0.82 (0.11)	0.92 (0.04)	0.81 (0.15)	0.91 (0.06)	0.82 (0.12)	0.94 (0.06)

STACOM-2011 model on ACDC testing dataset

Training Dataset (No. of images)	DC ED	DC ES	HD ED	HD ES	Vol. ES corr.	Vol. ES bias	Vol. ES std.	Mass ED corr.	Mass ED bias	Mass ED std
ACDC-2017 (1,352)	0.889	0.898	9.841	12.582	0.979	-2.57	11.04	0.990	-2.873	7.463
LV-2011 (20,360)	0.85	0.86	11.78	11.98	0.913	-7.19	22.29	0.974	-19.22	15.60

Note: The Values are provided as Mean (Stddev) in top table. The Segmentations were evaluated only for the task of myocardium.