Algorithms for Medical Image Processing
DeepCut: Segmentation of Human Heart Using Bounding Box Annotations
Final Report

Utkarsh Mall: 130050037 **Samarth Mishra:** 130260018

Objective :

- Method to obtain pixelwise object segmentations given an image dataset labelled with bounding box annotations.
- Formulate problem as energy minimization, using densely connected CRFs, with unary term from a CNN.

Brief Overview of the Methods :

Input and Output: The algorithm takes functional MRI image of chest for a patient across different slices at different time. The input MRI image is also annotated with bounding boxes around hearts endocardium. The algorithm finally gives a binary segmentation of heart(endocardium) and rest of the image.

Algorithm: The algorithm alternates between finding label probability of pixel using a patch around it using Convolutional Neural Network(CNN) and regularizing the labels over region using fully connected Conditional Random Field(CRF).

Conditional Random Field:

Using the CRF, we aim to find a labelling f for each pixel i which minimises:

$$E(f) = \sum_i \psi_u(f_i) + \sum_{i<j} \psi_p(f_i, f_j)$$

The CRF has unary potential function for each pixel coming from the CNN output. The pairwise potential function introduces a non-zero penalty only for pairs of pixels with different labels. It has two terms : bilateral potential function penalising difference in both position and intensities of pixel i and j and a spatial smoothing term which penalises just the spatial distance between pixels i and j. These terms are weighted by w_1 and w_2 .

$$\psi_p(f_i) = g(i,j)[f_i \neq f_j]$$

$$g(i,j) = \omega_1 \exp\left(-\frac{|p_i - p_j|^2}{2\theta_\alpha^2} - \frac{|I_i - I_j|^2}{2\theta_\beta^2}\right) + \omega_2 \exp\left(-\frac{|p_i - p_j|^2}{2\theta_\gamma^2}\right) .$$

We used [4] to efficiently infer labels over a fully connected CRF.

CNN Model:

The CNN tries to learn a segmentation of the image from the bounding box provided. It uses a patch of 33x33x3 (note that the patch is 3D) around a pixel to output probability values of the pixel being part of the foreground or the background.
For the CNN, a modification of the LeNet[3] architecture has been used. The CNN consists two layers of convolutions and subsequent max pooling. The final max-pooling layer is followed by a dense fully connected layer, followed by an output layer with softmax function to give the two label probabilities.

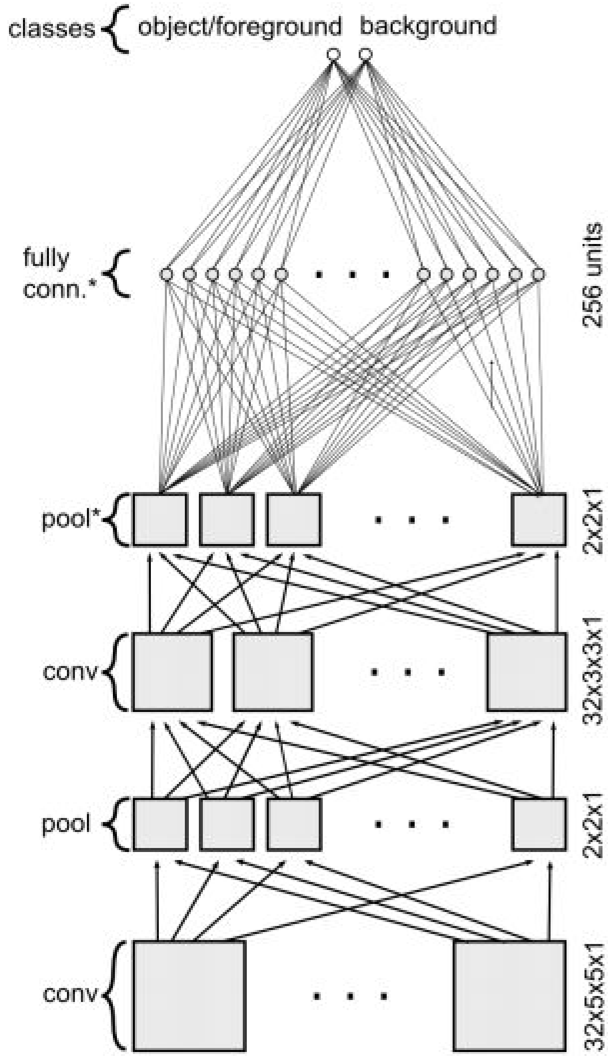

Fig 1. CNN architecture for pixel label probability prediction. Layers with * sign has 50% input dropout for regularization

Naive and DeepCut: The CNN is trained with patches around pixel within bounding box as foreground label and the patches around the bounding box as background pixel. This trained CNN is then used to get probabilities of pixels in the image. Each pixel get a probability between 0-1 of falling in foreground. This soft segmented image is treated as the unary term for CRF, which in turn returns the hard segmentation.

- In naive segmentation this CNN segmentation followed by CRF is done once
- In DeepCut after every CRF iteration CNN is retrained with the new labels, followed by CNN, this cycle is iteratively repeated many times.

Modified DeepCut: In DeepCut instead of using CNN predicted label probabilities outside bounding box, we can say that probability of foreground pixel outside bounding box is zero. This improves the accuracy of method.

Parameters Values

	Parameter	Value used
CNN	Patch size	33x33x3
	Learning rate	0.015
	Activation	tanh
	Epochs per DeepCut Iteration	50
	Total patches	100000
	Batch size	5000
CRF	θ_α	1
	θ_α	0.01
	θ_α	0.01
	Iterations	5

Results

We use Dice Similarity Coefficient(DSC) for measuring the accuracy of our segmentations.
The Dice similarity coefficient measuring overlap between two regions A and B is defined as :

$$DSC = \frac{2|A \cap B|}{|A| + |B|}$$

Where |A| is the number of pixels in region A.

Comparison with Ground Truth:

DSC classification accuracy for Modified DeepCut over all patients : 86.82%
Accuracy standard deviation for Modified DeepCut over all patients : 9.87

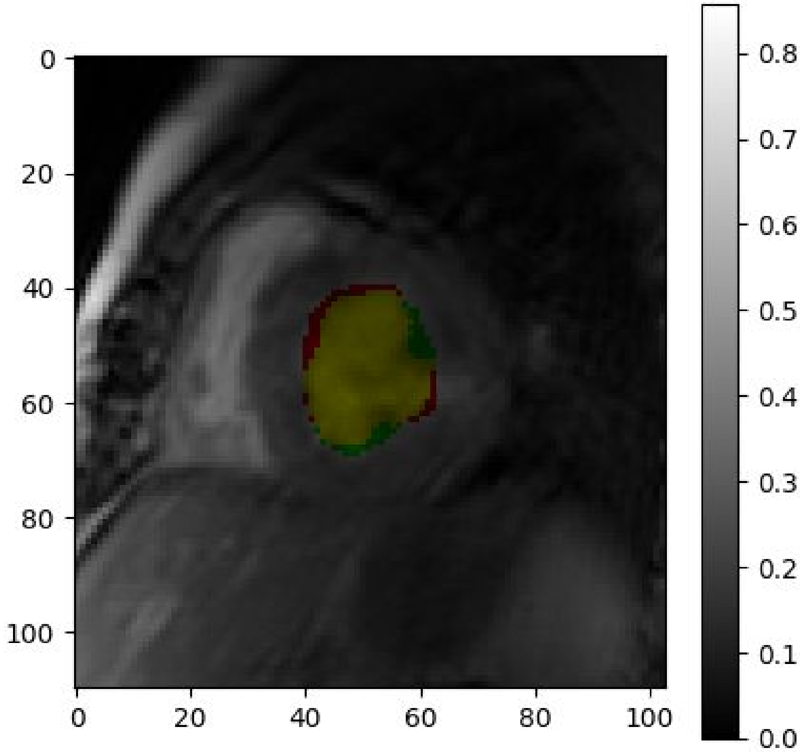

Fig 2. Green + Yellow region manually segmented region. Red + Yellow segmentation using modified deepcut. Yellow is the portion common to both segmentation

Deepcut vs Modified Deepcut:

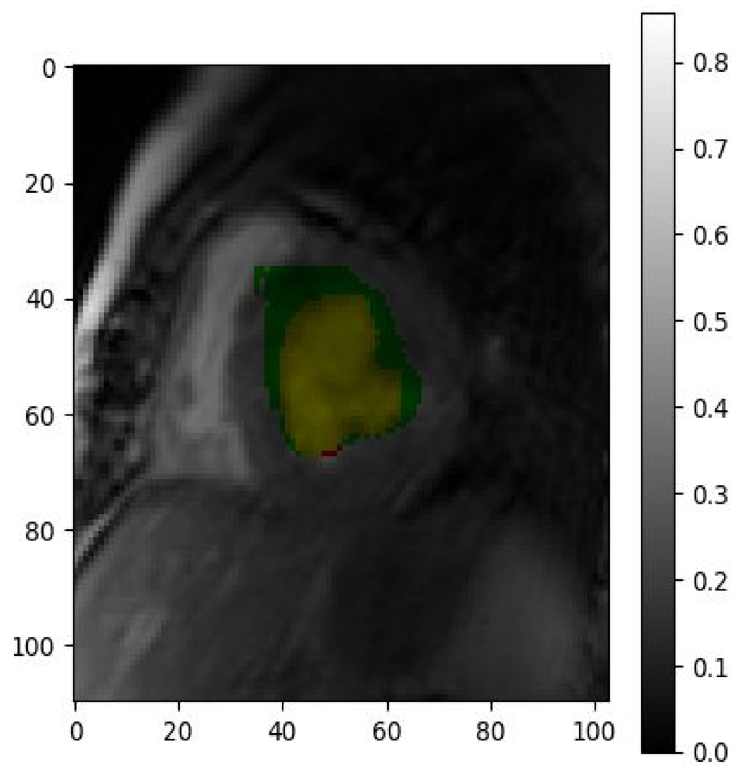

Fig 3. Green + Yellow region DeepCut segmented region. Red + Yellow segmentation using modified deepcut. Yellow is the portion common to both segmentation

DSC classification accuracy for DeepCut over all patients : 80.64%
Accuracy standard deviation for DeepCut over all patients : 9.94

The Accuracy is worse compared to modified DeepCut

Bounding Box vs Naive vs Deepcut:

	Bounding Box	**Naive**	**DeepCut**
DSC Accuracy	81.95%	83.23%	86.82%
Std. Dev	3.18	8.35	9.87

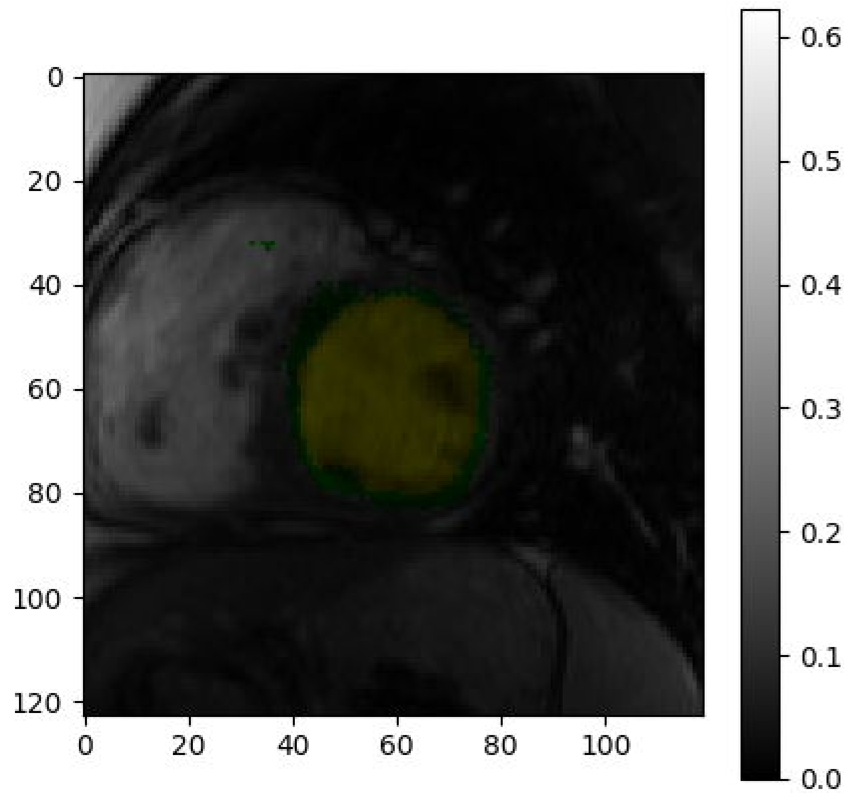

Fig 4. Green + Yellow region Naive segmentation region. Red + Yellow segmentation using modified deepcut. Yellow is the portion common to both segmentation

Segmentation vs Iteration:

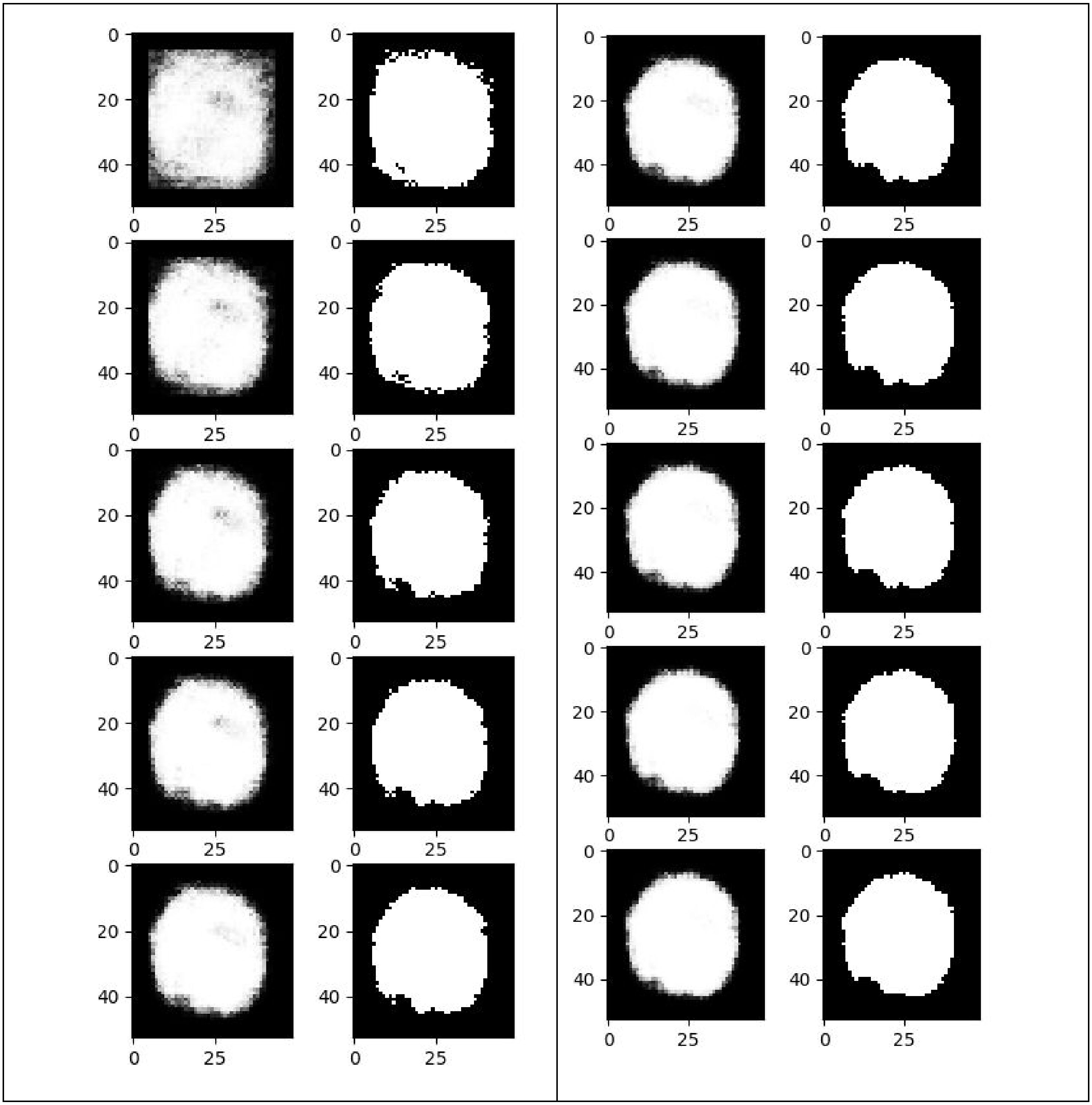

Fig 5. Improvement of segmentation with 10 iteration, the left image is CNN based soft segmentation followed by a CRF based hard segmentation

References :

[1] Dataset : http://www.cse.yorku.ca/~mridataset/
[2] Deepcut Segmentation : https://arxiv.org/abs/1605.07866
[3] LeNet : http://yann.lecun.com/exdb/lenet/
[4] Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials:
https://arxiv.org/abs/1210.5644