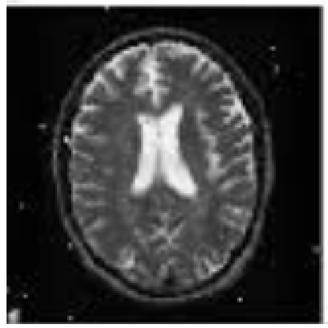
Active contour models for Medical Image Segmentation

Introduction-:

- CNN based image segmentation is divided into two types-:
    - Pixel based, it means to classify each pixel into one of the classes by calculating the SoftMax loss. Here each pixel is taken as input to CNN model for classification. But the problem in doing so is the disparity near the boundary layer. The model gets confused when the boundary passes through a pixel, whether to classify as object or non-object. So it ends up classifying based on number of sample points, if the object covers more part of the pixel than the boundary then that pixel is classified as object pixel. So this problem leads to aliasing
    - Image based, here the image is taken as the input and the output is the segmentation map. The underlying loss functions are minimised to have segmentation map with proper boundaries. Examples are U-NET, FCN which uses skip connection to extract low-level and high level features
- Our model is image based segmentation, which uses DenseNet for image segmentation. Its loss function takes boundary length, inside and outside region area into consideration.
- Why taking length and area of inside and outside region is a good idea?
    - In the image below, we can see white spots around the object. If we don't take the length of the object, area inside and outside the object into consideration, then we will end up segmenting the white spots around the object also. Another advantage of this model over curve evolution is, model based on curve evolution in the process of minimization of curve may end up in local minima which may lead to unsatisfactory results.



Active contour models-:

- Active contour models use image gradients to minimize the energy. But the function may be stuck at local minima, so we will use Active contour without edge models. The internal and external energy of the energy minimization problem are derived from edges of the object, so as we are not using the edges hence the name Active contour without edge (ACWE) model.
- ACWE model formulation is as follows
  -

$$\begin{aligned}\min_{\Omega_c, c_1, c_2} \{E_1^{ACWE}(\Omega_c, c_1, c_2, \lambda) \\ = \int_0^{Length(C)} ds \\ +\lambda \int_{\Omega} (c_1 - f(x))^2 dx \\ +\lambda \int_{\Omega/\Omega_c} (c_2 - f(x))^2 dx\},\end{aligned}$$

Taken from the paper
Chen_Learning_Active_Contour_Models_for_Medical_Image_Segmentation_CVPR_2019_paper

  - the first term is obtained by taking a differential amount (dS) of the length (Euclidean element) of the boundary. So the first term gives the length of the image. C is the total length of the curve
  - the second term, c_1 is the mean value of image to be segmented (f) outside, so it concern about the area of region outside the object. So the mean value of pixel

colour outside the object is c_1

- c_2 is the mean value of f inside. So the mean values of pixel colour inside the image is c_2. /omega_c is the closed subset of the image f, it means that it covers the entire image along with its limiting points.
- But this formulation is time consuming as we have to do this for each image, finding solution for above expression is time taking. So we use Total Variational Energy (TV) to find the solution.
- Total Variational energy, it is the total distance between the successive extremum points which is added to give the final length of the curve. We can also represent it as, if a observer moves along the curve, then the total projection length of the observer on the Y-axis gives the TV energy. So it helps in faster minimization.

Methodology-:

- AC Loss function
  - AC Loss function = Length + (/lambda) * Area
  -

$$Length = \int_C |\nabla u| ds$$

$$Region = \int_\Omega ((c_1 - v)^2 - (c_2 - v)^2) u dx$$

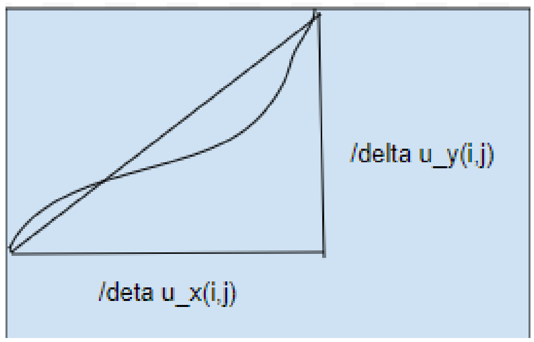
Taken from the paper Chen_Learning_Active_Contour_Models_for_Medical_Image_Segmentation_CVPR_2019_paper

  - Here, c_1 is v,u belongs to [0,1]^mxn. v is the ground truth segmentation, u is the predicted segmentation and they are normalized so the values is [0,1]. c_1 is the area of region inside and c_2 is the area of the region outside.
  - Translating integration into image, we have to do this pixel wise so the length now is given by
  - We have added a stability term /epsilon here to avoid zero initialization and to avoid sqrt(0) term.

$$Length = \sum_{\Omega}^{i=1,j=1} \sqrt{\left|(\nabla u_{x_{i,j}})^2 + (\nabla u_{y_{i,j}})^2\right| + \epsilon}$$

Taken from the paper
Chen_Learning_Active_Contour_Models_for_Medical_Image_Segmentation_CVPR_2019_paper

- So the hypotenuse which gives the approximated length of the boundary inside a pixel is given by the above length function. (i,j) are the index elements of the pixels



- The region is given by
-

$$Region = \left| \sum_{\Omega}^{i=1,j=1} u_{i,j}(c_1 - v_{i,j})^2 \right| + \left| \sum_{\Omega}^{i=1,j=1} (1 - u_{i,j})(c_2 - v_{i,j})^2 \right|$$

Taken from the paper
Chen_Learning_Active_Contour_Models_for_Medical_Image_Segmentation_CVPR_2019_paper

- Intuition behind this is, the first term gives the predicted segment and area of the region inside, similarly the second term gives the segment areas of both predicted and object outside the object. So it tells whether the pixel is inside the object or outside the object.

CNN Architecture-:

- Here two types of CNN backbones are used and compared, DenseNet and U-NET.
    - DenseNet, here the input feature map passes through three a dense layer which has Batch normalization, ReLU convolution of 3x3 and 1x1 convolution to decrease the number of channels. Here the special feature is to add the input layer feature map to each of the above mentioned layer and each layer is connected to all the layers that come after it so that all the layers share lower and higher level features and can co-learn the gradients. To avoid explosion of number of channels 1x1 convolution is used.
    - U-Net uses skip connection after a down layer (which has convolution and max pooling layers), where each layer is added with corresponding up sampled layer.
    - DenseNet are better than U-Net but they take significantly more time to train compared to U-Net for a given loss fucntion

Datasets-:

Here we have 256x256 CMR (Cardiac Magnetic resonance) images because they have huge variations which require high accuracy and they are easily available so the results can be easily generated for reproducible research.

Performance metrics-:

- Hausdorff distance is used here, for given set of input and output points if we find all the distance between a input and all the output points and take the minimum distance. The highest value in those minimum distance is called as Hausdorff distance. So we have to minimize this distance to reduce the loss function.

Conclusion-:

- Compared to energy optimization models, using CNN to find the length and segmenting inside and outside region, decreases the computational time as the energy optimization can be stuck at local minima. And as we are taking shape preservation into account, this model is more robust. So finally here we are trying to segment a given image into object and non-object part and also taking the boundary layer length into consideration.