Segmentation Losses Classification Part-3

ARCHITECTURAL CHANGES TO DETECT SMALLER REGIONS

- So far we have talked about carefully choosing loss function to fit our purpose like to preserve topology or for unbalanced datasets. But CNN architectures is another component in the deep learning framework, which can be modified to fit our purpose. So we will discuss about KiU-Net,

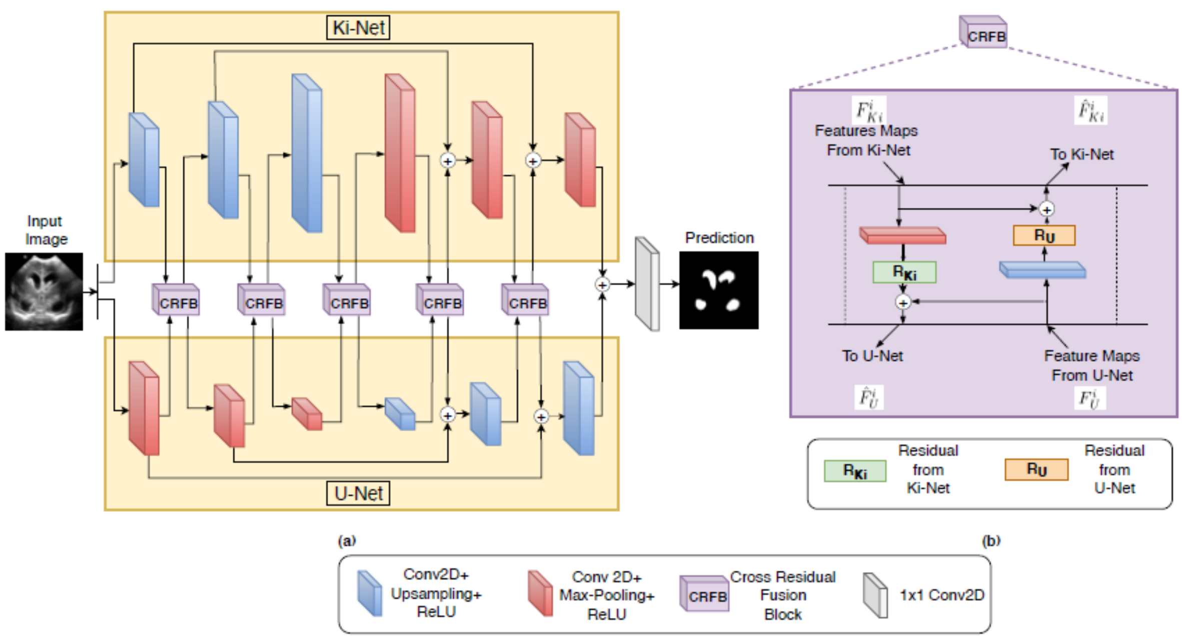
KiU-Net

- In the standard CNN architectures like U-Net we have encoder and decoder layer. Encoder layer or top-down model, has bunch of convolutional layer followed by max-pooling layers and the decoder or down-top model has up-sampling layers to make up for the decrease in dimension due to max-pooling layers.
- Max-pooling layers don't have any parameters to learn, they reduce the dimension of feature map so that the network only concentrates on highlighting or dominating features.
- Max-pooling layers leads to increase in receptive field size
    - Receptive fields can be thought of as the area that the kernel concentrates on. So receptive field is the size of the input which produces one node in the feature map.
    - As our dimension of input image decreases, the subsequent kernels in our network implicitly concentrates on a larger area on the input image.
    - Suppose you are training on MNIST dataset where you have 28x28 input image and kernel size of 3x3. After the first max-pooling layer the dimension reduces to 14x14.
    - So now a 3x3 kernel would concentrate on 9 pixels on 14x14 feature map, implicitly concentrating on 18 pixels on the input image (28x28). Hence, there is

increase receptive field size

- Increasing receptive field size leads to missing smaller regions

- Hence we go for some architectural changes to detect smaller region. As the basic intuition, if max-pooling increases our receptive field size then we need up-sampling layers in the encoder layer which will constrict the receptive field size.
  - And to compensate for this increase in dimension, we will have max-pooling layers in the decoder layer. This is the Ki-Net architecture.
- **So KiU-Net is actually Ki-Net + U-Net**
  - They output feature map of each layer in both networks are combined by CRFB (Cross-Residual Fusion block).
  - CRFB up-samples from U-Net layer and then concatenates with Ki-Net. Similarly, it downsamples by max-pooling Ki-Net layer and then concatenates with U-Net



Taken from the paper KiU-Net : Towards Accurate segmentation of bio-medical images using over-complete representations.

- KiU-Net requires less number of parameters, as it does not to be deep to detect smaller and larger featues. Hence it has faster convergence with superior results.
- Ki-Net learns smaller features and U-Net learns larger features and they share their knowledge with each other through CRFB layers. IT IS LIKE COMBINING THE BEST OF BOTH WORLDS.

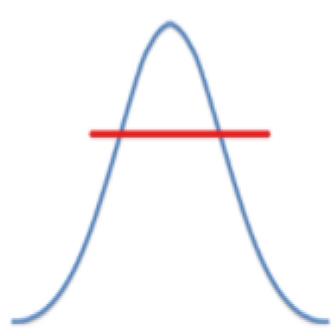
SOME DATA ANNOTATION METHODS

- We discussed under challenges in segmentation, that data annotation is the bottleneck in segmentation problems. It requires intensive man-hours to annotate boundaries accurately. Noisy annotation can lead to sub-optimal results
- So you maybe in two situation
    - You have labelled data but it is noisy and inaccurate. Annotators go for thicker annotation boundaries so that they don't need to worry about precise boundaries.
    - You haven't started labelling

STEAL (Semantically Thinned Edge Alignment Learning)

- If you are in situation - 1 then you can opt for STEAL. It is an add-on in an end-to-end detector model which helps to reduce the thickness of the boundaries.The problems with thick boundaries are
    - It can lead to overlapping boundaries with other objects
    - Network cannot generate accurate boundaries, if it is trained on thick boundaries
- STEAL has 3 components in loss
    - NMS (Non-Max suppression) loss. You can see below that we have a thick edge and we move in the normal direction to the edge (denoted by red line). As we move along the red line the intensity of the edge increase
        - So we keep track of the intensity and clip it, if it exceeds our threshold. It thins our boundary

THICK EDGE



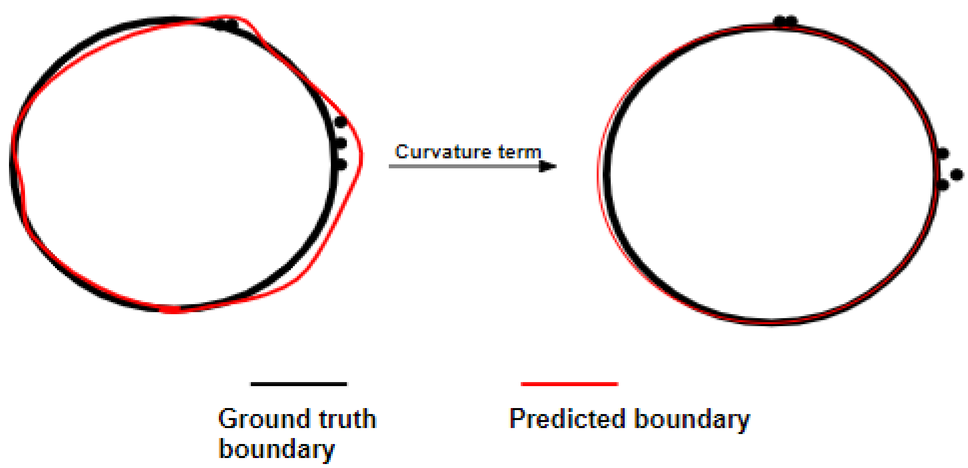
INTENSITY FUNCTION

- BCE loss, is the standard cross-entropy loss to detect segmentation probabilities of pixels
- Direction loss, it is the similarity loss between normal of ground-truth boundary pixel and normal of predicted boundary.
- STEAL helps our model to learn from noisy annotations, as it sort-of does the pre-processing of the noisy annotations to get accurate ground truth boundaries. Thus you can spend less man-hours in annotation and let the network handle it.

DELSE (DEEP EXTREME LEVEL-SET EVOLUTION)

- If you are in situtation-2, then you can opt for DELSE. DELSE uses curve evolution to find boundaries (we discussed curve evolution in Active contour without edges)
- Here the annotators place 4-points around the object of interest, from these 4 points a curve is initialized. There are 3 DELSE components which guide the curve to the object boundary.
    - Motion term
        - It takes care of the velocity vector of each pixel in predicted boundary, we define a unit velocity vector which keeps track of the change in gradient of the level set curve along x and y direction
    - Curvature terms

- It helps the curve to regularize, by regularize we mean that the final curve is displaced slightly towards the true object boundary, so that it does not overfit to noise around the object as we can see in the below diagram. But the regularization needs to be modulated so that the network fits the straight edges properly.



- Regularization term
    - It is a stability loss function which tries to minimize irregularities and numerical errors in final result

CONCLUSION

In the early stages of development in CNN's, people tried to increase number of parameters and built deep architectures to get superior results. This demands a lot of computational power. But carefully devising a loss function, can help us to get superior results with less number of parameters. Also hand-engineering loss functions requires expertise from different fields like topology etc. So it can bring people from different fields under one umbrella, working for advancement in Artificial intelligence