Reducing Hausdorff Distance

Astarag Mohapatra

May 2020

1 Introduction

- The loss function in image segmentation is important. Earlier Hausdorff distance (HD) was used as a performance metric only because it is the largest segmentation error.

  1. Hausdorff distance is maximum distance among the minimum distance between two sets or closed curves
  2. $$HD_{xy} = max_{x \epsilon X} max_{y \epsilon Y} ||x - y||_2$$

- The authors in the paper are trying to use Hausdorff distance as loss function and they have given 3 ways to do it with their comparison and trade-offs.

- But considering the highest segmentation error as loss function and optimizing the loss function accordingly can lead to instability in training and in complex structures where we have weak or missing edges, artifacts or low S/N ratio it is difficult to achieve high segmentation accuracy

- HD tries to minimize the error at a single point with the largest error, so it does not converge properly. So direct minimization of HD is not a good idea, so we have here HD "inspired" loss functions

- We have here these 3 loss function because they don't penalize the HD based loss extremely, rather they mildly penalize it to avoid any instability

2 Loss Functions

2.1 Notations used here

- q is the output segmentation probability map of CNN with $\bar{q} \epsilon [0, 1]$. After the values go over thresholding of 0.5, $\bar{q}$ values are either 0 or 1

- Similarly the ground truth label we have p and $\bar{p}$

- Boundaries of $\bar{q}$ and $\bar{p}$ are δq and δp respectively.
- The HD distance is measured between $\bar{q}$ and $\bar{p}$.

2.2 Distance based loss function

1. Distance transform represents the minimum distance representation of pixel w.r.t to the object of interest. The magnitude is given by the Euclidean distance.

2. The final loss function is given by

   $\text{LOSS}_{DT}(q, p) = 1/\Omega(\sum_{\Omega}(p - q)^2 * (d_p^\alpha + d_q^\alpha))$

3. We are using p and q which are the direct SoftMax probability output of the network rather than the threshold maps $\bar{q}$ and $\bar{p}$ for better segmentation results. α is the extent of penalization, so we smoothly penalize the largest segmentation error to avoid any instability. α between 1.0 to 3.0 leads to good result

4. d_p and d_q are the Euclidean distance of the distance transform.

5. We use $(p - q)^2$ instead of $|p - q|$ because the l2 norm better results than l1 norm. Also the l2 norm performs slightly better than hinge loss and square hinge loss

- Hinge loss is used in SVM model where the linear regression line has two support vectors at an width w and any point present within the support vector are penalized. Square hinge is the square of the hinge loss term.

6. Drawback of this method is that the computational cost is high so as the training time as we have calculate the distance transform each epoch because q changes after every epoch. d_p is computed once. So to decrease the computational cost we can neglect the d_q term but it will not give better results.

2.3 Morphological operation based loss function

1. Here we use erosion as morphological transformation to find the hausdorff loss. In erosion we have a structuring element with a centred pixel is traversed through the image and if the structuring element is completely inside the image (binary) then the centred pixel is labelled as 1 (foreground) else it is 0 (background).
2. The structuring element B_r is circular with radius r. Here we have morphological erosion of $p\Delta q$ or the symmetric difference between p and q.
3. The hausdorff distance is 2*r, where r is the minimum radius for which erosion between $p\Delta q$ and B_r is ϕ or null.

4. So if the structuring element lies outside the $p\Delta q$ then erosion operation will result ϕ so we need the minimum radius.
5. The final loss function is given by

   $\text{LOSS}_{ER}(q,p) = 1/\Omega((p-q)^2 \ominus_k B) * k^\alpha$

   here α is used for smooth penalization and K denotes the total number of erosions. K is a hyperparameter and increasing it leads to more computational time.

   B is the structuring element and we have to note here that we are using $(p-q)^2$ instead of the threshold map. But erosion operation can only be applied in binary images so convolution kernel is used which takes care of the erosion and soft thresholding of 0.5. It is a cross-shaped kernel with weights summing to 1. So B is ([0,1/5,0], [1/5,1/5,1/5], [0,1/5,0]).

2.4 Circular convolution based loss function

1. Here we have circular kernel B_r with radius r and elements of B_r are normalized.
2. The final loss function is given by
   $\text{LOSS}_{CV}(q,p) = 1/\Omega(\sum_{rR} r^\alpha \sum_\Omega [f_s(B_r * \bar{p}^C) \bigodot f_{\bar{q}/p} + f_s(B_r * \bar{p}) \bigodot f_{\bar{p}/\bar{q}} + f_s(B_r * \bar{q}^C) \bigodot f_{\bar{p}/q} + f_s(B_r * \bar{q}) \bigodot f_{\bar{q}/\bar{p}}])$

3. - here r is the radius of circular kernel and R is the largest radius that we can take which is a hyperparameter. R is set of radius which are widely spaced like 3,6,9 unlike 3,4,5,6,7,8,9 because it increases computational cost but does not give better result.
4. f_s is a soft thresholding function of convolution between B_r and $\bar{p}^c$ where $\bar{p}^c$ is simply (1-$\bar{p}$)
5. α controls the extent of penalization. $f_(\bar{q}/\bar{p})$ is a relaxed approximation of symmetric difference between $\bar{q}$ and $\bar{p}$ and it is given by (p-q)$^2 * q$.

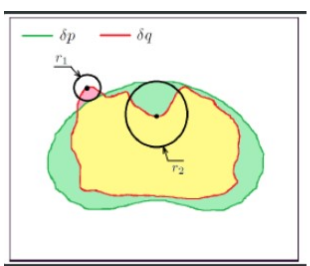

Figure 1: Circular Convolution

6. Here the green colour is the ground truth and red colour is the predicted segmentation. So from r1 and r2 the hausdorff distance is max(r1,r2).

3 CNN Architecture and Training procedure

1. Here we use U-Net for 2D images and 3D U-Net for 3D images
2. We have a extra DSC (dice loss) loss term as DSC loss term along with HD loss term gave better results and we got stable results at the start of the training.
3. Total Loss = $L_{CVorERorDT} + \lambda * L_{DSC}$

4. λ is determined by ratio of mean of HD-based loss and DSC-based loss term. It is high during the initial stages and weighs more the DSC loss term as we have many false positives far from segmentation boundary which increases the HD based loss function (instability).

4 Datasets

We have here 4 datasets

1. 2D ultrasound images of prostates
2. 3D MR images of prostate
3. 3D CT images of liver
4. 3D CT images of pancreas

5 RESULTS/INFERENCES/EXPERIMENTATION

1. Here CV is HD loss function by circular convolution, ER is by morphological operation and DT is by distance transform and DSC is dice loss.
2. According to the time required for training, $L_{DSC} < L_{ER} < L_{CV} < L_{DT}$

As we have to calculate the distance transform of the predicted boundary points in each epoch, the time required to train increases substantially.ER based is the most efficient one. Test times for all are equal as we are not using any loss function for the same.

3. According to performance, $L_{DSC} < L_{ER} < L_{CV} < L_{DT}$

CV and DT based loss function performed similarly. ER based loss function didn't as par other two because it underestimates the HD. Also choosing squares-shaped and cube-shaped structuring element didn't help.

4. According to Pearson correlation coefficient

Pearson correlation coefficient gives the correlation between two sets of data points by the following formulation.

Corr = $[\sum(x * y)/\sqrt{\sum(x^2) * \sum(y^2)}]$,

where x and y are the standard deviation $x - \bar{x}$ of the data points

and x^2 and y^2 is the square of the standard deviation.

$L_{CV} < L_{ER} < L_{DT}$, but we can find many outliers in $L_E R, soitdoesn't fit properly.$

5. For 2D images calculating the distance transform is not a heavy load so it is chosen over other two as it gives more accurate results whereas for 3D images it is difficult to compute the distance transform so CV or ER based loss function is chosen. CV based loss function gives better result if we limit the maximum kernel size (radius R).

6 TL;DR

- Hausdorff distance is the largest segmentation error in image classification. It is primarily used as performance metric.
- Here we tried to use HD as a loss function, but in the process of minimizing the largest segmentation error we may compromise the stability.
- So we try to penalizing the HD based loss function mildly so that our optimization is stable
- Here we propose three loss function which are combined DSC loss function and at the start of optimization the HD based term has less influence but gradually the influence is increased.
- We have compared the three loss functions and found out that circular convolution based loss is more efficient as it gives better results that morphological based error and its computational cost is less than distance transform based error.
- However as distance transform based error has better accuracy, we can use it in 2D images as it is computationally manageable. But in 3D images we would use circular based loss function.