Review on DELSE (Deep Extreme Level Set Evolution)

Introduction-:

In semantic segmentation we label each pixels to a class, **but in level sets we label each curve to a certain class. So classifying each curve is extended here in DELSE**. As annotation of objects is the bottleneck in any object segmentation task, the process needs to be expedited. So here the annotators just have to mark 4 points, left-most, right-most, top and bottom. A initial LSF or level steps function is created and it is optimized by curve evolution to fit the boundaries by considering three DELSE components. Also the annotators, after the final curve is generated can correct the erroneous boundaries and this is noted to modify the curve.

Comparison to DEXTR model-:

DEXTR or Deep extreme cut model takes in the above 4 points, but the 4 points by the annotators may not be accurate. So a gaussian function is embedded as fourth channel along with RGB to be trained in the CNN encoder. **In DEXTR we have pixel-wise labelling which is disadvantageous if we have some occlusion. The interdependency of the pixels is not taken into consideration here so we see spurious holes and out-of-place irregular shapes.**

Level sets and curve evolution-:

Level set is the collection of all the points in the domain where the function has a certain constant value. T(x,y) = k. Curve evolution is implemented to minimize the curve to enclose the boundary accurately. In curve evolution, t**he level set function inside the curve is greater than 0, on the curve is equal to 0 and outside the curve is less than 0 (This convention is flexible)**. So if the level set function is initialized outside the curve, the vector field in each pixel tends to point inwards to the object boundary or where the level set function is equal to 0 and vice-versa. The gradient of the pixel is given by

$$\frac{\partial \phi}{\partial t} = -V|\nabla \phi|,$$

The level set function is updated in every iterative step by taking the gradient of each pixel. We thus generate a well behaved curve which has its energy function which has internal (bendiness and stretchiness of curve) and external energy (detection of strong edge) minimized. The initial level set curve is generated by the four points and the gaussian function. Here we use TSDF (truncated signed distance function). **Signed refers to the sign of foreground and background vector field sign and truncated norm thresholds our output to avoid instability.**

DELSE components-:

Here we have to take care of two things-:

1. external terms that attract the curve to the desired location such as edges with strong gradients.
2. internal regularization on the curve's shape.

Motion term-

Here we take care of the velocity vector field of each pixel. The motion branch indicates the direction and magnitude of the motion of the curve so as to minimize the energy. The direction of the vector field which is a unit vector in the gradient direction is given by

$$\vec{U}_{\mathrm{gt}}(x,y) = -\frac{\nabla \phi_{DT}(x,y)}{|\nabla \phi_{DT}(x,y)|}$$

Then the update parameter is given by

$$\left[\frac{\partial \phi_i}{\partial t}\right]_{\mathrm{motion}} = -\langle \vec{V}_\theta, \nabla \phi_i \rangle$$

The update term has subscript as motion term because it allows the curve to expand as well as converge. If the extreme points are initialized inside the curve then converging to a single point will give the least energy curve as the energy is minimized here. This collapsing of points is avoided here. Also hardwiring that the vector field should approach at perpendicular direction is avoided here as the curve as a whole is evolved here so small angular error are tolerable.

Curvature term-

Noise in objects are inevitable and varying, so the curve evolution has to be robust. But the model should not misinterpret the sharp corners to be noisy as it is trained on smooth curves with noises. This may lead to improper fitting so the model should be able to identify where to regularize the curve. This is given by the curvature term, which gives the flexibility to the model to preserve the sharp corners. Small m is the modulation index.

$$\begin{aligned}\left[\frac{\partial \phi_i}{\partial t}\right]_{\text{curvature}} &= m_\theta\ \kappa |\nabla \phi_i| \\ &= m_\theta\ |\nabla \phi_i|\ \mathrm{div}\left(\frac{\nabla \phi_i}{|\nabla \phi_i|}\right).\end{aligned}$$

Divergence means whether the vector field are converging to a point or diverging from a point. A sharp corner at the edge will have vector field converging (divergence is negative) but it will accommodate less vector fields so it will not influence the updated parameter. So we add in this extra term so as to sharp corners fitting can be optimized.

Regularization term-

To tackle instability and avoid numerical errors the gradient unit vector is tried to be close to 0 or 1. So the extra updated parameter term is added along with a double well potential function which has its local minima at 0 and 1. By limiting the value of updated parameter close to 0 or 1 we improve stability.

Updated parameter-:

The updated parameter is given by (the third term is the regularization term).

Architecture-:

Here we use ResNet-101 and PSP module for segmentation. In PSP, the image is sent through a CNN encoder and then pooled to form varied features maps. Then these are convoluted 3x3 filter. The smaller feature maps extract coarse level features and larger feature maps extract fine features. Then all the outputs are up sampled by bilinear interpolation which is concatenated with feature map the pooling layer. These is better than FCN (Fully convolutional network) in segmentation task.