Artificial Intelligence: Paper Critique

Subhani Shaik (MT18117)

Paper Title: DEL- Deep Embedding Learning for Efficient Image Segmentation

Authors: Yun Liu, Peng-Tao Jiang, Vahan Petrosyan, Shi-Jie Li,Jiawang Bian, Le Zhang, Ming-Ming Cheng
venue: IJCAI 2018
Link: https://www.ijcai.org/proceedings/2018/0120.pdf

I. SUMMARY

Image segmentation aims to partition an image into regions of object, which is often used in many computer vision applications. There are many others techniques like HFS,EGB,MCG[1] which are giving good results, but the authors proposed a new method of segmenting the image by using DEL(DeepEmbedding Learning) which can efficiently transform superpixels into image segmentation.Here after converting into SLIC superpixels, they are training a fully convolution network to learn the feature embedding space for each superpixel. The learned feature embedding corresponds to a similarity between two adjacent superpixels. By seeing similarity, we can merge those superpixels into large segments.

A. Superpixel Generation

The pixels in same region have greater similarity than in different region pixels.So,the direct grouping of pixels using distance metrics is time consuming and lacks robustness. Considereing that this algorithm starts with fast superpixel generation by SLIC, which is based on *k*-means clustering algorithm. This superpixel is a small region and thus robust than single pixels.

B. Feature Embedding Learning

After superpixel generation, to learn the feature embedding space the authors trained these on deep convolution network.This performs the pooling operation on the feature embedding space to extract feature vector, where each feature vector is the average of learned deep feature maps in corresponding regions of the superpixel.Each feature embedding vector v_i has 64 dimensions in this design.The authors proposed a backward function of superpixel pooling layer to input x_k, can be written as:

$$\vec{v}_i = \frac{1}{|S_i|}\sum_{k\in S_i}\vec{x}_k \qquad \frac{\partial L}{\partial \vec{x}_k} = \sum_{S_i\in \mathcal{S}} 1_{\{k\in S_i\}}\cdot\frac{1}{|S_i|}\cdot\frac{\partial L}{\partial \vec{v}_i}$$

where I_k is an Indicator function.
The authors proposed a distance metric to measure similarities between two superpixels as

$$d_{ij} = \frac{2}{1+exp(\|\vec{v}_i - \vec{v}_j\|_1)}.$$

where d_{ij} is close to 1 when v_i and v_j are similar, otherwise is close to 0 when they are extremely different. And the authors used loss function as

$$L = -\sum_{S_i\in\mathcal{S}}\sum_{S_j\in\mathcal{R}}[(1-\alpha)\cdot l_{ij}\cdot log(d_{ij}) + \alpha\cdot(1-l_{ij})\cdot log(1-d_{ij})],$$

where l_{ij}=1 where superpixels belongs to same region and 0 where superpixels belongs to different region.
α=parameter to balance +ve and -ve samples.

With this, the similarities between two superpixels of same region will be larger than that of different region. These similarity metrics will be used in merging the superpixels.

To learn the feature embedding space, they used VGG16 net convolution network(Figure 1). In this they divided into 5 convolutional pooling layers and finally got the 64-dimension feature embedding space using convolution layer with kernel size 1x1. Then extract the feature vector from this embedding space corresponding to superpixels.

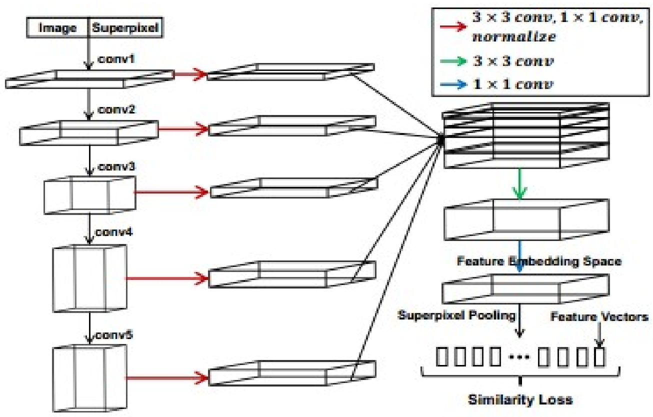

Figure 1: Feature Embedding learning network

From this neural network, we can learn the similarities between adjacent superpixels.

C. Superpixels Merging

After training the deep neural network, we got the dissimilarities between superpixels. A threshold we will know which two adjacen superpixels can be merged or not. For merging the they used EGB(Efficient Graph Based) datastructure. Here all the merging will be done at once so it reduces time consumption.

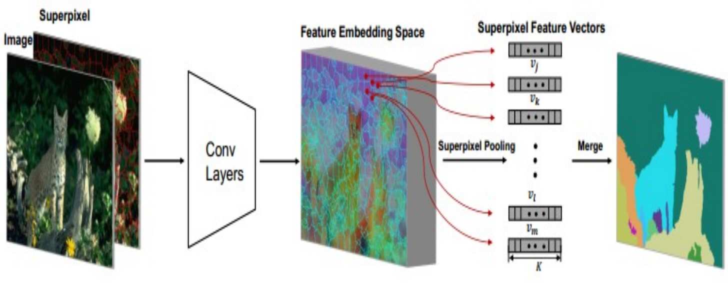

Figure 2: Complete flow of DEL image segmentation

II. LIMITATIONS OF DEL TECHNIQUE

The DEL technique works faster than other techniques like MCG etc. but it has some limitations:

- If the image is noisy which is like object then it will see that as an object.
- If the objects in the image are similar to backgound then it won’t differentiate the object from background(which is a major problem).
- In some cases it is not able to differentiate the different object.By that the different objects are considered as one object.

III. SUGGESTIONS FOR THE PROPOSED TECHNIQUE

- For superpixel generation, they used SLIC method which follows k-means clustering algorithm which is not reliable. They can use other clustering algorithm like Density based spatial clustering.
- By using the GoogLeNet, ResNet and other architectures instead of VGG16 net, so that we can analyse the performance of DEL.
- By using superpixels based on distance metric similarity,it may consider other object as same object.If we use density and intensity based similarity the performance of DEL will be changed.
- By using Deep learning technique it may overfit the data, So the parameters selection should be flexible and the loss function used in this technique is not reliable, we can use other loss functions to improve the segmentation.

REFERENCES

[1].Jordi Pont-Tuset, Pablo Arbelaez, Jonathan T.Barron, Ferran Marques, Jitendra Malik: Multiscale Combinatorial Grouping for Image Segmentation and Object Proposal Generation 2016.