Final report (Option 1)

Course name: Multimedia Processing Technique (DD026_1594)

Project number: Option 1 – 4

Project name: Survey the paper: Recent progress in semantic image segmentation

link: https://link.springer.com/content/pdf/10.1007/s10462-018-9641-3.pdf

Student name: Yu Jiwon. 201713912

Github of project: https://github.com/abc

Report:

Keywords of this paper are Image semantic segmentation · DNN · CNN · FCN

First, Semantic image segmentation is called pixel-level classification. It is clustering parts of image together which belong to the same object class. Semantic segmentation has been studied for many years. Deep Neural Network (DNN) appears then, segmentation has made a tremendous development.

In general, the methods before DNN are called traditional method of segmentation. This paper want to focus on recent segmentation which is after appear of DNN.

Traditional methods : Before DNN

Features and classification is a piece of information which is relevant for solving the computational tasks. Variety of methods are used for semantic segmentation, such as Pixel color, Histogram of oriented gradients, Local Binary Pattern etc,, Very common method is thresholding methods with gray images. Gray images are very common in medical area where the collection equipment is usually X-ray CT scanner or MRI. Overall, thresholding methods are quite effective in this area.

Recent methods : With DNN

Artificial Neural Network (ANN) is inspired by biologic neurons.

Especially CNN(Convolutional Neural Network), which develop from DNN, mimics visual neurons in animals. So CNN has obtained remarkable results on image classification, segmentation, and detection.

There are methods using CNNs in image semantic segmentation

1 Fully convolutional network (FCN)

The main insight is the replacement of fully connected layer by fully convolutional layer.

Because of Its convolutional layer, takes random output size, it is correspondingly-sized output with efficient inference and learning.

2 Up-sample method

This method has deconvolutional network. Deconvolution is the reverse calculation of convolution. Therefore, it can also recover the input size.

3 FCN joint with CRF and other traditional methods

According to the research of Deeplab, the responses at the final layer of Deep Convolutional Neural Networks (DCNNs) are not sufficiently localized for accurate object segmentation (Chen et al. 2016b). They overcome this poor localization property by combining a fully connected Conditional Random Field (CRF) at the final DCNN layer.

4 Dilated convolution

Most semantic segmentations are based on the adaptations of CNN that had originally been devised for image 'classification' task. However, image semantic segmentation tasks is structurally different from 'classification'. This method focus detailed convolution and it further increase the performance of the network.

5 Progress in backbone network

The backbone network refers to the main structure of the network. As is known to all, the backbone used in semantic segmentation is derived from image classification tasks.

6 Pyramid method in segmentation

Researchers attempt to combine pyramid strategy to CNN.

6-1 Image pyramid

An image pyramid is a collection of images which are successively downsampled until some desired stopping criteria are reached. In semantic image segmentation area, multi-scale image input and sliding pyramid pooling that can effectively improve the performance.

6-2 Atrous spatial pyramid pooling

Inspired by image pyramid strategy, proposes Atrous Spatial Pyramid Pooling to segment objects robustly at multiple scales. It called ASPP, and it can captures objects image context at multiple scales.

6-3 Pooling pyramid

The pyramid pooling adopts different scales of pooling size, then does up-sample process on the outputs to the original size, and finally concatenates the results to form a mixed feature representation

6-4 Feature pyramid

Feature pyramid is a basic component in image tasks for detecting objects at different scales. Feature Pyramid Network (FPN) is developed for building high-level semantic feature maps at all scales.

7 Multi-level feature and multi-stage method

Typically recognition algorithms based on convolutional networks use the output of the last layer as a feature representation. However, the information in this layer is too coarse for dense prediction. On the contrary, earlier layers may be precise in localization, but they will not capture semantics. To get the best of both advantages they propose deep layer cascade (LC) method. That consists of multiple independent models. It not only improves the segmentation performance but also accelerates both training and testing of deep network

8 Supervised, weakly-supervised and unsupervised methods

Till now, more and more methods are emerging to make semantic image segmentation more accurate or faster or both on accuracy and speed.