Lund University
Faculty of Engineering
Centre for Mathematical Sciences

Master's Thesis	Instance-level Semantic Segmentation by Deep Normalized Cuts
Student	Alexander Davidson
Supervisor	Cristian Sminchisescu
Examiner	Magnus Oskarsson
Date of Presentation	September 5, 2017

Segmenting Objects in Images Using Deep Learning

To have a computer automatically detect and localize objects in images is a difficult problem. Still, a system able to do so has many potential applications. It could, for example, constitute an important component in a self-driving car. This thesis attempts at a solution based on deep learning.

Unlike humans, computers are poor at reasoning about the meaning of digital images or videos. However, enabling computers to better understand visual information is desirable, as it would pave the way for the automation of many tasks that the human visual system can do.

One such task, referred to as pixel-level semantic segmentation, is to segment the content of an image into visual categories. Consider an image depicting an urban street scene. Each pixel of that image can be associated with one of several categories, such as car, pedestrian, road, and building. While a segmentation of this kind can be useful, it lacks a notion of individual objects. Consequently, it does not answer how many cars appear in the image, nor does it indicate the exact location of any of those cars. Additionally requiring that each object should be outlined gives rise to a similar task called instance-level semantic segmentation, which is the topic of this thesis. Since objects frequently are partially occluded by others, this latter task is more difficult.

Even though the above segmentation tasks are challenging, great progress has been made in the last few years. This is largely due to recent advances in deep learning technology. Deep learning is a particular approach to machine learning that makes use of neural networks. Neural networks have a long history, and have traditionally been viewed as simple models of the human brain.

This thesis combines ideas from deep learning with an established segmentation technique called normalized cuts to segment objects in images. Each pair of pixels is described by a number, with a large number suggesting that they should belong to the same segment, and a small number suggesting that they should not. The normalized cuts method computes a segmentation from these numbers. But rather than manually specifying what they should be, a neural network is used to instead have the computer itself figure out what numbers will result in good segmentations.

An experimental evaluation in which the network is taught to segment cars suggests that the method can produce segmentations of fairly good quality. Unfortunately, the method is computationally expensive and slow, and heavy downsampling of the images need be applied. As can be seen in the predicted segmentation below, this results in car boundaries being coarse at the original resolution.

Input image | Desired segmentation | Predicted segmentation