Undergraduate Thesis

CurbNet: Semantic segmentation of curbs and curb cuts from street imagery

Yvan Putra Satyawan

Examiner: Prof. Dr. Wolfram Burgard
Advisers: Jannik Zürn

Albert-Ludwigs-University Freiburg

Faculty of Engineering

Department of Computer Science

Chair for Autonomous Intelligent systems

July 22nd, 2019

Writing Period

04. 20. 2019 – 07. 22. 2019

Examiner

Prof. Dr. Wolfram Burgard

Second Examiner

?

Advisers

Jannik Zürn

Declaration

I hereby declare, that I am the sole author and composer of my thesis and that no other sources or learning aids, other than those listed, have been used. Furthermore, I declare that I have acknowledged the work of others by providing detailed references of said work.
I also hereby declare, that my thesis has not been prepared for another examination or assignment, either in its entirety or excerpts thereof.

Place, Date	Signature

Abstract

(TODO: Write this.)

Contents

List of Figures

List of Tables

List of Algorithms

1 Introduction

Semantic scene segmentation is a popular research topic in the field of computer vision, and especially important for autonomous vehicles. The ability to semantically understand a scene is especially important for autonomous vehicles and robots to safely navigate an environment. Generally, most implementations attempt to segment road surfaces but in this thesis, we propose the segmentation of curbs and curb cuts to allow safer sidewalk navigation.

The Europa project has resulted in the Obelix robotic platform, which has already been demoed to successfully perform pedestrian navigation [1][2]. We propose to add to this platform the ability to detect curbs and curb cuts using semantic segmentation. The Obelix platform is **(TODO: Add short description)**.

(TODO: Full description of what obelix is not capable of doing yet.)

Our goal is thus to implement a computer vision algorithm capable of the semantic segmentation of curbs and curb cuts using a single camera image. To do so, we will implement a convolutional neural network (CNN) with a traditional encoder-decoder architecture. We will also include prior knowledge to the training, since we can assume that the camera setup for the Obelix robot will remain relatively similar throughout its lifespan.

We will begin by discussing the motivation behind this thesis, followed by a disucssion of related works and the background. Then the approach will be discussed in detail

along with the experiments and results. Finally, a discussion of potential future research will be presented followed by the conclusion.

1.1 Motivation

(TODO: Do this.)

2 Related Work

There are many works in the field of semantic segmentation in recent years, both discussing object scene segmentation and road segmentation. The field of semantic segmentation using trainable neural network models started in 1989 with the work of Eckhorn et al. and their paper describing how the visual cortex of a cat functions and its implications for network models [3]. This early method used a pulse coupled neural network, which produced synchronous bursts of pulses, effectively grouping the neurons by phase and pulse frequency, which can then be analyzed for feature extraction.

In recent years, models using CNNs have become more commonly used for this task. CNNs were originally designed to be able to encode feature information in an image of object classification. "Very Deep Convolutional Networks for Large-Scale Image Recognition" by Karen Simonyan and Andrew Zisserman was one of the earlier works to show that this network setup could result in very high accuracy image classification [4].

(TODO: resnet improved on this.)[5].

"Fully Convolutional Networks for Semantic Segmentation" by Jonathan Long et al. took these CNNs and added fully connected layers to take the encoded features and use it for semantic segmentation [6]. This paper laid out a novel approach to the problem as a dense image classification problem. In essence, they proposed that image segmentation is nothing more than image classification on a per-pixel basis. As

such, previously developed CNNs for image classification could be used and indeed, they implemented their model based on the ResNet network

From these image segmentation networks, further refinements have been made for networks dedicated to road segmentation. We investigated two of these state-of-the-art models. The first was PSPNet, proposed in "Pyramid Scene Parsing Network" by Hengshuang Zhao et al. [7]. This paper proposed using a pyramid scheme to counter the fact that features would get smaller the further they are from the camera. **(TODO: Include image explaining this.)** The second is DeepLab v3+, which became the basis of the network we used for CurbNet. DeepLab v3+ was proposed in "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation" by Liang-Chieh Chen et al. [8]. **(TODO: explain paper a bit)**.

Thus far, there seems to be no works related directly to the goal of this thesis; the semantic segmentation of curbs and curb cuts using computer vision. Research into curb detection is quite plentiful, with even one work from ETH Zürich being made specifically for the Obelix robotic platform [9]. Curb Detection for a Pedestrian Robot in Urban Environments" by Jérôme Maye, Ralf Kaestner, and Roland Siegwart used the LIDAR sensors that Obelix has to map the world around it and detect the different horizontal planes. A sudden change in the vertical position of the plane would thus be classified as a curb. Unfortunately, this work does not take into account curb cuts.

The only work we were able to find that specifically address curb cuts was "WalkNet: A Deep Learning Approach to Improving Sidewalk Quality and Accessibility" by Andrew Abbott et al. [10]. This paper discusses the use of a deep neural network to classify images in which curb cuts existed. Their goal was the use of Google Street View data to map which intersections in a city already had curb cuts and which didn't. This data would then be supplied to city governments to provide relevant information regarding accessibility and sidewalk quality. Unfortunately, this work

only classified images which contained curb cuts and the neural network architecture they used was not described in depth.

3 Background

Explain the math and notation.

Figure 1: Use tikz to draw nice graphs!

Algorithm 1 Stochastic Gradient Descent: Neural Network

Create a mini batch of m samples $\mathbf{x}_0 \dots \mathbf{x}_{m-1}$
foreach sample $\mathbf{x}$ **do**
  $\mathbf{a}^{\mathbf{x},0} \leftarrow \mathbf{x}$ ▷ Set input activation
  foreach Layer $l \in \{1 \dots L-1\}$ **do** ▷ Forward pass
    $\mathbf{z}^{\mathbf{x},l} \leftarrow \mathbf{W}^l \mathbf{a}^{\mathbf{x},l-1} + \mathbf{b}^l$
    $\mathbf{a}^{\mathbf{x},l} \leftarrow \varphi(\mathbf{z}^{\mathbf{x},l})$
  end for
  $\boldsymbol{\delta}^{\mathbf{x},L} \leftarrow \nabla_{\mathbf{a}} C_{\mathbf{x}} \odot \varphi'(\mathbf{z}^{\mathbf{x},L})$ ▷ Compute error
  foreach Layer $l \in L-1, L-2 \dots 2$ **do** ▷ Backpropagate error
    $\boldsymbol{\delta}^{\mathbf{x},l} \leftarrow ((\mathbf{W}^{l+1})^T \boldsymbol{\delta}^{\mathbf{x},l+1}) \odot \varphi'(\mathbf{z}^{\mathbf{x},l})$
  end for
end for
foreach $l \in L, L-1 \dots 2$ **do** ▷ ▷ Gradient descent
  $\mathbf{W}^l \leftarrow \mathbf{W}^l - \frac{\eta}{m} \sum_{\mathbf{x}} \boldsymbol{\delta}^{\mathbf{x},l} (\mathbf{a}^{\mathbf{x},l-1})^T$
  $\mathbf{b}^l \leftarrow \mathbf{b}^l - \frac{\eta}{m} \sum_{\mathbf{x}} \boldsymbol{\delta}^{\mathbf{x},l}$
end for

4 Approach

The approach usually starts with the problem definition and continues with what you have done. Try to give an intuition first and describe everything with words and then be more formal like 'Let g be ...'.

4.1 Problem Definition

Start with a very short motivation why this is important. Then, as stated above, describe the problem with words before getting formal.

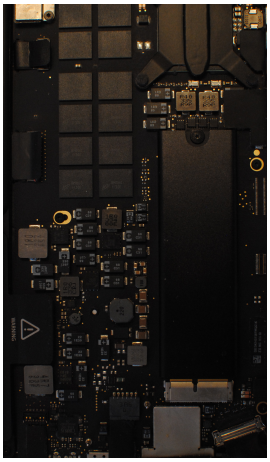
4.2 First Part of the Approach

4.3 N-th Part of the Approach

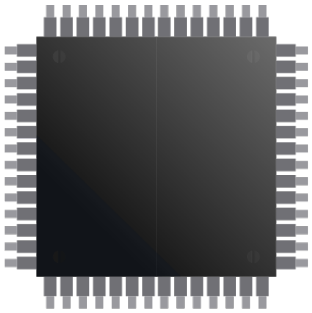
5 Experiments

Type	Accuracy
A	82.47 ± 3.21
B	78.47 ± 2.43
C	84.30 ± 2.35
D	86.81 ± 3.01

Table 1: Table caption. foo bar...



(a) Some cool graphic



(b) Some cool related graphic

Figure 2: Caption that appears under the fig Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

6 Conclusion

7 Acknowledgments

First and foremost, I would like to thank...

- advisers
- examiner
- person1 for the dataset
- person2 for the great suggestion
- proofreaders

Bibliography

[1] "The european pedestrian assistant," http://europa.informatik.uni-freiburg.de 2009.

[2] R. Kümmerle, M. Ruhnke, B. Steder, C. Stachniss, and W. Burgard, "A navigation system for robots operating in crowded environments," in *2013 IEEE International Conference on Robotics and Automation*, IEEE, May 2013.

[3] R. Eckhorn, H. J. Reitboeck, M. Arndt, and P. Dicke, "Feature linking via stimulus - evoked oscillations: Experimental results from cat visual cortex and functional implications from a network model," in *International 1989 Joint Conference on Neural Networks*, IEEE, 1989.

[4] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv 1409.1556*, 09 2014.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[6] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[7] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," 2016.

[8] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," 2018.

[9] J. Maye, R. Kaestner, and R. Siegwart, "Curb detection for a pedestrian robot in urban environments," in *Proceedings - IEEE International Conference on Robotics and Automation*, 05 2012.

[10] A. Abbott, A. Deshowitz, D. Murray, and E. C. Larson, "Walknet: A deep learning approach to improving sidewalk quality and accessibility," *SMU Data Science Review*, vol. 1, no. 1, 2018.