BRNO UNIVERSITY OF TECHNOLOGY

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FACULTY OF MECHANICAL ENGINEERING

FAKULTA STROJNÍHO INŽENÝRSTVÍ

INSTITUTE OF SOLID MECHANICS, MECHATRONICS AND BIOMECHANICS

ÚSTAV MECHANIKY TĚLES, MECHATRONIKY A BIOMECHANIKY

SEMANTIC SEGMENTATION OF IMAGES USING CONVOLUTIONAL NEURAL NETWORKS

SÉMANTICKÁ SEGMENTACE OBRAZU POMOCÍ KONVOLUČNÍCH NEURONOVÝCH SÍTÍ

MASTER'S THESIS

DIPLOMOVÁ PRÁCE

AUTOR
AUTOR PRÁCE

Bc. FILIP ŠPILA

SUPERVISOR
VEDOUCÍ PRÁCE

doc. Ing. JIŘÍ KREJSA, Ph.D.

BRNO 2020

Abstrakt
Tato bakalářská práce se zabývá návrhem, výrobou a realizací řízení nestabilního robota, balancujícího na sférické základně, známého také pod názvem ballbot. Předpokládá se kompletní návrh konstrukce, výběr pohonných jednotek, návrh, implementace a testování inteligentního řídícího algoritmu, který udrží robota v metastabilní rovnovážné poloze. Při vývoji budou využity softwarové nástroje MATLAB/Simulink. Práce také počítá s využitím mikrokontroleru dsPIC jako platformy pro finální řízení celého systému. Zadání projektu má interdisciplinární charakter a je realizováno jako týmová práce s jasně vymezenými úkoly pro jednotlivé členy.

Summary
This bachelor's thesis deals with the complete design, manufacture, and control of an unstable robot, balancing on a spherical base, also known as ballbot. The complete design of the construction, motor unit selection, design, implementation testing of an intelligent control algorithm to keep the robot in a meta-stable equilibrium is assumed. Multiple tools such as Matlab/Simulink are used for this approach. It also includes the final implementation of the code in the PIC microcontroller. The project has an interdisciplinary character and is meant to be done as teamwork whereby each team member has a strictly defined role.

Klíčová slova
Ballbot, konstrukce, inteligentní řízení, PID, MATLAB, Simulink

Keywords
Ballbot, construction, intelligent control, PID, MATLAB, Simulink

ŠPILA, F.*Semantic segmentation of images using convolutional neural networks.* Brno: Vysoké učení technické v Brně, Faculty of Mechanical Engineering, 2020. 8 s. Vedoucí doc. Ing. Jiří Krejsa, Ph.D.

Prohlašuji, že jsem svou práci vypracoval samostatně a použil jsem pouze podklady (literaturu, software atd.) citované v práci a uvedené v přiloženém seznamu a postup při zpracování práce je v souladu se zákonem č. 121/2000 Sb.,o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon) v platném znění.

V Brně 1. května 2017

Bc. Filip Špila

Děkuji Ing. Tomáši Spáčilovi, Matěji Rajchlovi a celému týmu z Mechlabu za podnětné připomínky a rady, které mi během práce poskytli. Dále děkuji své rodině za podporu jak během studií, tak během psaní této práce.

Bc. Filip Špila

Contents

1. Introduction

Image segmentation is one of the essential parts of computer vision and autonomous systems alongside with object detection and object recognition. The goal of semantic segmentation is to automatically assign a label to each object of interest (person, animal, car etc.) in a given image while drawing the exact boundary of it and to do this in the most robust and reliable way possible. Speaking in terms of machine learning, each pixel of the input image is intended to belong to a specific class.

We can see a real-world example in Figure 1. Each pixel of the image has been assigned to a specific label and represented by a different colour. Red for people, blue for cars, green for trees etc. This is unlike mere image classification task where we classify the image scene as a whole. It is appropriate to say that semantic segmentation is different from so called instance segmentation, where one not only cares about drawing boundaries of objects of a certain class but also wants to distinguish between different instances of the given class. For instance, all people in the image (each instance of the 'person' class) would all have a different colour.

It turns out that semantic segmentation has many different applications in the fields such as driving autonomous vehicles, human-computer interaction, robotics, and photo editing/creativity tools. The most recent development shows the increasing need for reliable object recognition in self-driving cars because it is crucial for the models to understand the context of the environment they're operating in.

The presented work focuses on research and implementation of one particular segmentation method that uses convolutional neural networks (CNNs). CNNs belong to the family of machine learning algorithms and got under attention mainly due to their groundbreaking success in image classification challenges (ImageNet etc,). They subsequently found their use in segmentation tasks where researchers take the most well-performing CNN architectures and use it as the first stage of the segmentation algorithm.

Figure 1.1: Segmentation of an urban road scene

2. Problem statements

The assignment of this thesis consists of several expected achievements. Firstly, a promising segmentation method using CNNs needs to be found and implemented. It is expected that the neural network will be as straightforward as possible while still being likely to be capable of giving satisfactory results for the chosen use case (segmentation of a path in outdoor environment for a robot navigation). The images will be provided by the supervisor of the thesis and used to train and validate the network performance. In addition, the author will pick an appropriate software tool for creating Ground Truths (FOOTNOTE manually created image labels that serve as a reference for the network to validate its current accuracy of prediction and to compute the needed adjustments of its parameters to get closer to the desired output) and use it to create the final training and validating datasets. Lastly, the network should be trained with various sets of hyperparameters (FOOTNOTE: hyperparameter definition) in order to get a closer idea of the network's training behaviour and to ensure the best possible results.

3. Research and theory

First part of this section gives a thorough introduction to neural networks (NN) in general. It begins by definition of fundamental terms needed to fully understand the core principles of NNs. Due to the fact that the research in this area is still heavily ongoing, the more advanced techniques described here may soon be out of date or replaced by better-performing ones and therefore the theoretical background is limited only to the extent relevant for the particular chosen network architecture. Still, it will give a solid foundation needed to understand other similar approaches.

Second part presents some of the main approaches based on machine learning researches have recently used to tackle the semantic segmentation problem. However, not all of them use CNNs as the core algorithm. This part summarizes the main key points from the corresponding papers that contributed to this topic by presenting novel architectures and principles. It finishes by more detailed description of a method that is eventually found the most promising and thus selected for the final implementation.

3.1. Supervised learning

Artificial neural network algorithms are inspired by the architecture and the dynamics of networks of neurons in human brain. They can learn to recognize structures in a given set of training data and generalize what they have learnt to other data sets (supervised learning). In supervised learning one uses a training data set of correct input/output pairs. One feeds an input from the training data into the input terminals of the network and compares the states of the output neurons to the target values. The network trainable parameters are changed as the training continues to minimise the differences between network outputs and targets for all input patterns in the training set. In this way the network learns to associate input patterns in the training set with the correct target values. A crucial question is whether the trained network can generalise: does it find the correct targets for input patterns that were not in the training set?

3.1.1. Feedforward neural networks

The goal of a feedforward neural network is to find a non-linear function that maps the space of the inputs x to the space of the outputs y. In other words, to learn the function [zdroj SANTIAGO]

$$f^* : \mathbb{R}^m \to \mathbb{R}^n, f^*(x; \phi)$$

where ϕ are trainable parameters of the network. The goal is to learn the value of the parameteres that result in the best function approximation, by solving the equation

$$\phi \leftarrow arg\, min\, L(y, f^*(x; \phi))$$

where L is a loss function chosen for the particular task. One can understand the term 'loss function' simply as a metric of 'how happy we are about the output that the network gives us for a given input'. The structure is usually composed of many nested functions. For instance, there might be three functions f(1), f(2) and f(3) connected in a chain that forms into f(x) = f(3)(f(2)(f(1)(x))) (2.4)

These models are called feedforward because information flows through the function being evaluated from x, through the intermediate computations used to define f and finally to the output y. In this case, f(1) is called the first layer of the network, f(2) is called the second layer, and so on. The final layer of a feedforward network is called the output layer. During neural network training, f(x) is driven to match f(x). Each training example x is accompanied by a label y f(x). The training examples specify directly what the output layer must do at each point x; it must produce a value that is close to y. The behavior of the other layers is not specified by the training data, but the learning algorithm must decide how to use those layers in order to produce the desired output. It is for this reason that these layers are called hidden layers.[12] In Figure 2.3 , an image of a four-layer feedforward neural network with two hidden layers can be seen

The neurons are modelled as linear threshold units (McCulloch-Pitts neurons) and are commonly organized in structures called layers. Layer in the network is defined as a set of computational units.

4. Bibliography

[1] ZOUHAR, František. *Návrh konstrukce, řízení a elektroniky pro nestabilní balancující vozidlo* [online]. Brno: Vysoké učení technické v Brně. Fakulta strojního inženýrství, 2011 [cit. 2017-05-21]. Dostupné z: http://hdl.handle.net/11012/21119. Diplomová práce. Vysoké učení technické v Brně. Fakulta strojního inženýrství. Ústav mechaniky těles, mechatroniky a biomechaniky. Vedoucí práce Robert Grepl.

5. Seznam použitých zkratek a symbolů

CMU	Carnegie Mellon University

6. Seznam příloh

- Nastavení režimu External mode: *external.txt*