Computers and Electrical Engineering 68 (2018) 381–388

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

A road segmentation method based on the deep auto-encoder with supervised learning☆

Xiaona Song[a,b], Ting Rui[a,c,*], Sai Zhang[a], Jianchao Fei[a], Xinqing Wang[a]

[a] College of Filed Engineering, Army Engineering University of PLA, Guanghua Road, Nanjing, China
[b] College of Mechanical Engineering, North China University of Water Resources and Electric Power, Beihuan Road, Zhengzhou, China
[c] State Key Laboratory for Novel Software Technology, Nanjing University, Hankou Road, Nanjing, China

ARTICLE INFO

Keywords:
Image segmentation
Road recognition
Auto-encoder
Semantic segmentation
Unmanned vehicle

ABSTRACT

Road environment perception is a key technique for unmanned vehicles. Segmentation of road images is an important method of determining the driving area. The segmentation precisions of existing methods are not high, and some are not in real-time. To solve these problems, we design a supervised deep auto-encoder (AE) model to complete the semantic segmentation of road environment images. By adding a supervised layer to a classical AE, and using the segmentation image of training samples as the supervised information, the model can learn the useful features to complete the semantic segmentation. Next, the multilayer stacking method of the supervised AE is designed to build the supervised deep AE, since the deep network has more abundant and diversified features. Finally, we verified the method using CamVid. Compared with Convolutional Neural Networks(CNN) and Fully Convolutional Networks(FCN), the road segmentation performance, such as precision and speed were improved.

1. Introduction

The road environment perception of unmanned vehicles has always been a popular research area, and the method based on machine vision is one of the important research methods. It utilizes the camera in the vehicle to acquire the road image of driving vehicle, and uses image process and pattern recognition to segment the image and determine the available driving area.

Traditional image segmentations extracts the low-level visual cues to get the segmentation results. These methods are not complex, but the results are often unsatisfactory on the difficult segmentation tasks such as the segmentation of road environment images. With the continued warming of the deep learning, it is proved that the Deep Convolutional Neural Networks (Deep Convolutional Neural Networks) has a great advantage of image feature extraction, but how to use the CNN to segment the image and improve the segmentation performance still need to make a thorough study. Many researchers [1,2,3,4] used DCNN to segment images. But the DCNN-based methods have to generate proposal which are time-consuming. Long et al. [5] introduced the image semantic segmentation, based on Fully Convolutional Networks (FCN) model in 2015. Shortly thereafter, many FCN-based, semantic segmentation methods were proposed, improving the performance of segmentation. However, the FCN model is very complex, and the fine-tuning of million parameters is a lengthy task, requiring several weeks to train the networks on high-performance GPU. More time is required to train the model, which also needs high-performance system hardware. All of the aforementioned problems affect its application to road detection. To solve these problems, we design a supervised deep auto-encoder (AE) model to complete the

☆ Reviews processed and recommended for publication to the Editor-in-Chief by Associate Editor Dr. Huimin Lu.
* Corresponding authors.
E-mail addresses: songxiaona1@126.com (X. Song), rtinguu@sohu.com (T. Rui).

https://doi.org/10.1016/j.compeleceng.2018.04.003
Received 21 August 2017; Received in revised form 3 April 2018; Accepted 3 April 2018

semantic segmentation of road environment images.

As a classical model of deep learning, auto-encoders (AEs) learn the important features from samples, using unsupervised self-learning, and reconstruct the data information through concise expression. This paper presents the design of the supervised deep AE, which is successfully applied on the semantic segmentation of the road environment.

This paper provides several contributions. First, a new semantic segmentation method is proposed, by adding a supervised layer to a classical AE to learn useful features for image segmentation. Next, because of the supervised layer of the AE, the traditional stacking method is unsuitable. We design the multilayer stacking method for the supervised AE, and build a supervised deep AE that has better feature-extraction performance than single-layer, supervised AE. Finally, the road segmentation performance of the proposed method is proven to be effective and simple, using CamVid. When compared with other methods, the segmentation precision of the road region and the real time are better than others.

The rest of the paper is organized as follows. Section 2 reviews the related work on recent approaches of semantic segmentation. Section 3 introduces our method of road segmentation. In Section 4, we elaborate the model of the supervised Auto-Encoder. Experiments are discussed and evaluated in Section 5. A summary in Section 6 concludes this paper.

2. Related work

There are several methods for road environment perception [6,7]. Those using deep learning [1,8–12] have been of significant interest. Alvarez et al. [8] used the CNN to learn features of road environment. An image patch was classified by CNN and the probability belonging to the sky, surroundings, and road was given to complete the road segmentation. Brust et al. [9] built the Convolutional Patch Networks incorporating spatial information to realize the pixel-wise road detection. These methods mentioned above that detected the entire road image by sliding window was not in real-time. Hariharan et al. [13] first used the DCNN to solve the image segmentation. They proposed a Simultaneous Detection and Segmentation (SDS) method and used the RCNN(Regions with Convolutional Neural Network Features)framework to train the network and improved state-of-the-art in object detection and semantic segmentation at that time. However, this method relied on many proposals which are time-consuming. Long et al. [5] proposed the semantic segmentation method based on the FCN, which transforms fully connected layers into convolution layers. This enables a classification net to output a heatmap and obtain the semantic segmentation through upsampling. The upsampling of the network is then improved by Badrinarayanan [14], who created a convolutional encoder-decoder architecture named Segnet. The performance of Segnet was tested via CamVid and other datasets, achieving good segmentation performance. Teichmann [1] present an approach to joint classification, detection and semantic segmentation via a unified architecture where the encoder is shared amongst the three tasks. The segmentation approach is also based on FCN. In order to achieve better segmentation results, Conditional Random Fields (CRF) is used to optimize the segmentation results by many researchers [15]. Although the methods obtain good segmentation results, the model of the FCN is too complex, and the fine-tuning of parameters is a lengthy task. More time is required to train the model, which also needs high-performance system hardware. All of the aforementioned problems affect its application to road detection.

As a classical deep learning model, AEs [16,17] boost efficiency through automatic feature extraction. They also show the ability to learn essential features from numerous unlabeled samples, and reconstruct the initial image using concise and effective expression. Vincent [18] presents the stacked denoising AE, which can learn data features and reconstruct the initial image. Huang [19] designed an adaptive, deep, supervised AE, to reconstruct the damaged face. This work obtained good reconstruction results. Wang [20] used the convolutional AE to smooth the image, to reduce the difficulty of unsupervised image segmentation and improve the segmentation precision. Because the classical AE only reconstructs the initial image, little research uses AEs to complete the image segmentation. This paper proposes a supervised, deep AE model, and applies it to the semantic segmentation of road environments.

3. The proposed method

A supervised layer is added to the traditional AE, and the segmentation image of the road environment is used as the supervised information. Because of the supervised layer, AEs learn features useful for image segmentation, and complete the semantic segmentation of the road environment. The research shows that deep networks learn more abstract and diversified features. As a result, a deep network is built, to extract deep features of the road environment to complete the semantic segmentation. Because a supervised layer is added to a classical AE, the multilayer stacking form is not applicable. However, the stacking form is studied and a supervised deep Auto-Encoder model is designed. It is shown that the performance of the supervised deep AE is superior to the single-layer. As shown in Fig. 1, a supervised deep AE is trained using training samples and their segmentation images. Testing samples are fed to the model, to obtain the semantic segmentation images. Then the segmentation images are used to obtain the driving area by general image-processing methods.

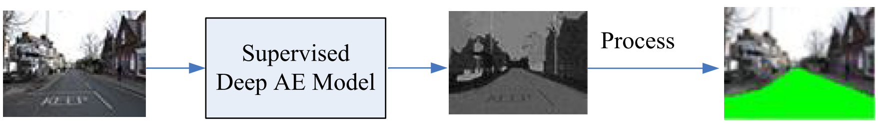

Fig. 1. The proposed method.