Neural Networks 99 (2018) 158–165

Contents lists available at ScienceDirect

Neural Networks

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

Deep neural network for traffic sign recognition systems: An analysis of spatial transformers and stochastic optimisation methods

Álvaro Arcos-García *, Juan A. Álvarez-García, Luis M. Soria-Morillo

Dpto. de Lenguajes y Sistemas Informáticos, Universidad de Sevilla, 41012, Sevilla, Spain

HIGHLIGHTS

- A Deep Neural Network that is top-1 ranked in the German traffic sign benchmark.
- Effectiveness analysis of Spatial Transformer Networks for traffic sign recognition.
- Quantitative comparison of several stochastic gradient descent optimisation methods.

ARTICLE INFO

Article history:
Received 25 July 2017
Received in revised form 15 November 2017
Accepted 18 January 2018
Available online 31 January 2018

Keywords:
Deep learning
Traffic sign
Spatial transformer network
Convolutional neural network

ABSTRACT

This paper presents a Deep Learning approach for traffic sign recognition systems. Several classification experiments are conducted over publicly available traffic sign datasets from Germany and Belgium using a Deep Neural Network which comprises Convolutional layers and Spatial Transformer Networks. Such trials are built to measure the impact of diverse factors with the end goal of designing a Convolutional Neural Network that can improve the state-of-the-art of traffic sign classification task. First, different adaptive and non-adaptive stochastic gradient descent optimisation algorithms such as SGD, SGD-Nesterov, RMSprop and Adam are evaluated. Subsequently, multiple combinations of Spatial Transformer Networks placed at distinct positions within the main neural network are analysed. The recognition rate of the proposed Convolutional Neural Network reports an accuracy of 99.71% in the German Traffic Sign Recognition Benchmark, outperforming previous state-of-the-art methods and also being more efficient in terms of memory requirements.

1. Introduction

Traffic sign recognition systems (TSRS) are essential in many real-world applications such as autonomous driving, traffic surveillance, driver safety and assistance, road network maintenance, and analysis of traffic scenes. Normally, a TSRS concerns two related subjects which are traffic sign detection (TSD) and traffic sign recognition (TSR). The former focuses on the localisation of the targets in the pictures while the latter performs a fine-grained classification to identify the type of targets detected (De La Escalera, Moreno, Salichs, & Armingol, 1997).

Traffic signs constitute a fundamental asset within the road network because their aim is to be easily noticeable by pedestrians and drivers in order to warn and guide them during both the day and night. The fact that signs are designed to be unique and to have distinguishable features such as simple shapes and uniform colours implies that their detection and recognition is a constrained problem. Nevertheless, the development of a robust real-time TSRS still presents a challenging task due to real-world variability, such as scale variations, bad viewpoints, motion-blur, faded colours, occlusions, and lightning conditions. On top of that, there are more than 300 different traffic sign categories defined by the Vienna Convention on Road Traffic (United Nations Economic Commission for Europe, 1968). This treaty has been signed by 63 countries, although a few minor visual variations of traffic sign pictographs still exist between countries, which can lead to complications in the automated recognition task. Any TSRS must cope well with such issues.

The main contributions of this work are four-fold: (1) A state-of-the-art traffic sign recognition system based on a Convolutional Neural Network (CNN) that includes Spatial Transformer Networks (STN) and outperforms previously published work related with the German Traffic Sign Recognition Benchmark (GTSRB) (Stallkamp, Schlipsing, Salmen, & Igel, 2011); (2) An insight into the proposed CNN capabilities along with the performance impact of spatial transformer layers within the network; (3) Analysis of the effect

* Corresponding author.
E-mail addresses: aarcos1@us.es (Á. Arcos-García), jaalvarez@us.es (J.A. Álvarez-García), lsoria@us.es (L.M. Soria-Morillo).

https://doi.org/10.1016/j.neunet.2018.01.005
0893-6080/

of diverse gradient descent optimisation algorithms on the CNN presented. (4) Multiple publicly available European traffic sign classification datasets are reviewed and evaluated by the CNN. These contributions lead to practical applications, such as self-driving cars and automated inventory and maintenance of vertical signage, since the CNN can perform fine-grained classification once the traffic sign has been detected. Moreover, as the CNN outperforms the human visual system, its inference time is low and can also be deployed as a stand-alone service, it can therefore be used in real-time applications.

The rest of the paper is organised as follows. Section 2 reviews related works of traffic sign recognition systems. Section 3 describes the experiments conducted to analyse the impact of both spatial transformers and stochastic optimisation algorithms on the proposed CNN. Recognition results are then shown in Section 4. Finally, conclusions are drawn and further work is proposed in Section 5.

2. Related work

Chronologically, approaches of published studies on traffic sign recognition systems have evolved from colour and shape-based methods to machine-learning-based methods. In recent times, Deep Neural Networks (DNN) have attracted attention in pattern recognition and computer vision research, and have been widely adopted for both object detection (Liu et al., 2016; Redmon & Farhadi, 2016; Ren, He, Girshick, & Sun, 2015) and recognition (Huang, Liu, Weinberger, & van der Maaten, 2016; Szegedy, Ioffe, Vanhoucke, & Alemi, 2017), thanks to the release of several publicly available datasets composed of millions of images (Everingham, Van Gool, Williams, Winn, & Zisserman, 2010; Krizhevsky, Sutskever, & Hinton, 2012; Lin et al., 2014). Moreover, DNNs have been applied in autonomous driving related challenges such as car (Huval et al., 2015), lane (Li, Mei, Prokhorov, & Tao, 2017), and pedestrian (Tian, Luo, Wang, & Tang, 2015) detection.

With regard to the traffic sign detection and classification problem domain, colour-based approaches are very common. These methods use different colour spaces for segmentation of the road image, such as RGB (Escalera, Moreno, Salichs, & Armingol, 1997), HIS (Maldonado-Bascon, Lafuente-Arroyo, Gil-Jimenez, Gomez-Moreno, & Lopez-Ferreras, 2007), and HSV (Shadeed, Abu-Al-Nadi, & Mismar, 2003). The shape-based method is another popular approach for traffic sign recognition and detection. Symmetry information of circular, triangular, square and octagonal shapes are used in Loy and Barnes (2004), a radial symmetry detector is proposed in Barnes, Zelinsky, and Fletcher (2008), Hough transforms are investigated in Barnes, Loy, and Shaw (2010) and a circular traffic sign recognition system is studied in Kaplan Berkaya, Gunduz, Ozsen, Akinlar, and Gunal (2016). Hence, neither colour nor shape-based techniques, need any prior knowledge of traffic signs and heavily depend on custom-designed algorithms and feature engineering.

One of the main problems before the year 2011 was the lack of publicly available traffic sign datasets. The Belgian Traffic Sign Dataset (BTSD) (Timofte, Zimmermann, & Van Gool, 2011), the German Traffic Sign Recognition and Detection Benchmark (GTSRB and GTSDB) (Stallkamp et al., 2011), the Croatian traffic sign dataset (rMASTIF) (Jurisic, Filkovic, & Kalafatic, 2015), the Dataset of Italian Traffic Signs (DITS) (Youssef, Albani, Nardi, & Bloisi, 2016) and the Tsinghua-Tencent 100 K benchmark (Zhu et al., 2016) solved this issue and boosted research into TSRS since several of these datasets are commonly used to evaluate the performance of computer vision algorithms for traffic sign detection and recognition. These kinds of datasets are crucial to generate robust machine learning and deep learning models as they contain a huge amount of traffic sign samples of multiple categories, taken by cameras with various weather and lighting conditions, occlusions, bad viewpoints, etc.

More recently, machine learning has started to play a key role in the traffic sign classification task. Mathias, Timofte, Benenson, and Van Gool (2013) propose fine-grained classification by applying different methods through a pipeline of three stages: feature extraction, dimensionality reduction and classification. On GTSRB, they reach 98.53% accuracy by merging grey-scale values of traffic sign images and features based on the Histogram of Oriented Gradients (HOG), reducing the dimensionality through Iterative Nearest Neighbours-based Linear Projections (INNLP) and finally classifying with Iterative Nearest Neighbours (INNC) (Timofte & Van Gool, 2015). Although other machine learning algorithms such as Support Vector Machines (SVM) (Salti, Petrelli, Tombari, Fioraio, & Di Stefano, 2015), Random Forests (Zaklouta, Stanciulescu, & Hamdoun, 2011) and Nearest Neighbours (Gudigar, Chokkadi, Raghavendra, & Acharya, 2017) have been widely used to recognise traffic sign images, Convolutional Neural Networks (Lecun, Bottou, Bengio, & Haffner, 1998), also known as ConvNets or CNNs, showed particularly higher classification accuracies in the competition. Neural networks are data driven self-adaptive methods because they can adjust themselves to the data without any explicit specification of functional or distributional form for the underlying model (Huang, 1996). In addition, there are universal functional approximators in the neural networks that can approximate any function with arbitrary accuracy (Huang, 1999; Huang & Du, 2008). Cireşan, Meier, Masci, and Schmidhuber (2012) won the GTSRB contest (Stallkamp, Schlipsing, Salmen, & Igel, 2012) with 99.46% accuracy thanks to a committee of 25 CNNs by using data augmentation and jittering. Sermanet and LeCun (2011) used a multi-scale CNN and achieved an accuracy of 98.31%, thereby granting them second place in the GTSRB challenge. Later, Jin, Fu, and Zhang (2014) proposed a hinge loss stochastic gradient descent method to train an ensemble of 20 CNNs that resulted in 99.65% accuracy and offered a faster and more stable convergence than previous work. However, these approaches can still be improved through the avoidance of the use of hand-crafted data augmentation and of the application of multiple CNNs in an ensemble or via a committee for the reason that these normally lead to higher memory and computation costs.

3. Methodology

In this work, we propose a traffic sign recognition system that carries out fine-grained classification of traffic sign images through a CNN whose main blocks are convolutional and spatial transformer modules. In order to find an accurate and efficient CNN for such a purpose, the effect of using several STNs and different stochastic gradient descent optimisation methods are researched and discussed.

3.1. Dataset and data pre-processing

Several publicly available traffic sign datasets have been gathered in countries such as the United States (Mogelmose, Trivedi, & Moeslund, 2012), Belgium (Timofte et al., 2011), Germany (Stallkamp et al., 2011), Croatia (Jurisic et al., 2015), Italy (Youssef et al., 2016), Sweden (Larsson & Felsberg, 2011), and China (Zhu et al., 2016).

This paper focuses on both the spatial transformer effectiveness and cost function optimisation experiments on the GTSRB (Stallkamp et al., 2011) dataset. There are multiple reasons for choosing this dataset over the others, including the fact that it is highly accepted and is used for comparing traffic sign recognition approaches in the literature. Moreover, its authors and the organisation behind them held a public competition whereby scientists from different fields contributed with their results and tested the GSTRB dataset. Nowadays, a GTSRB website is maintained where