Deep Learning for Self-Driving Cars: Chances and Challenges

Extended Abstract

Qing Rao
BMW Group
Unterschleißheim, Germany
Qing.Rao@bmw.de

Jelena Frtunikj
BMW Group
Unterschleißheim, Germany
Jelena.Frtunikj@bmw.de

ABSTRACT

Artificial Intelligence (AI) is revolutionizing the modern society. In the automotive industry, researchers and developers are actively pushing deep learning based approaches for autonomous driving. However, before a neural network finds its way into series production cars, it has to first undergo strict assessment concerning functional safety. The chances and challenges of incorporating deep learning for self-driving cars are presented in this paper.

CCS CONCEPTS

• **Computing methodologies** → **Artificial intelligence**; **Machine learning**; • **Software and its engineering** → **Software verification and validation**;

KEYWORDS

Deep Learning, Functional Safety, Automotive

ACM Reference Format:
Qing Rao and Jelena Frtunikj. 2018. Deep Learning for Self-Driving Cars: Chances and Challenges: Extended Abstract. In *SEFAIAS'18: IEEE/ACM 1st International Workshop on Software Engineering for AI in Autonomous Systems, May 28, 2018, Gothenburg, Sweden.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3194085.3194087

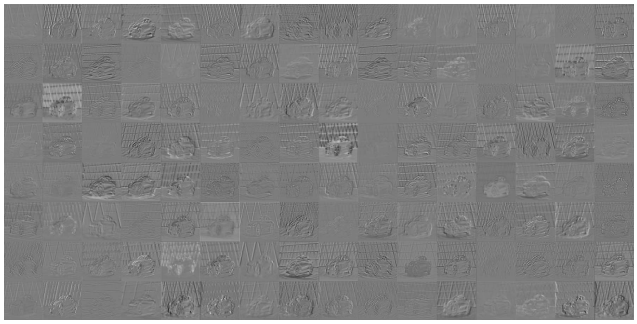
1 INTRODUCTION

A number of self-driving car pilot projects [2, 4, 6, 15] were demonstrated in recent years. One common fact among these pilot projects is, that part of the driving tasks such as environmental perception, path planning, or even steering wheel control is carried out by deep-learning-based approaches. Alone with the successful demonstration of the deep-learning-driven autonomous prototypes, the focus of the automotive industry is gradually moving from building and demonstrating prototype vehicles to series production. Now, the major challenge becomes how to get the neural networks into series production cars, in a safety-conform way.

In 2011, the International Organization for Standardization (ISO) proposed a standard ISO 26262 [10] which regulates functional safety for road vehicles. It includes requirements and recommendations for the entire lifecycle of car manufacturing, from the concept phase to operation and service. The main aim of ISO 26262 was to help the automotive industry address functional safety issues in a more systematical way. However, it was defined without taking deep learning into consideration since the first version of ISO 26262 was published before the boom of AI. This eventually leads to a challenging issue today for car manufacturers and suppliers who are determined to incorporate deep learning for self-driving cars.

In this paper, we present the safety-related issues that we encountered during the development of deep learning solutions for self-driving cars. We identify three concrete problems, and we share our ideas from the point of view of an Original Equipment Manufacturer (OEM) how to approach these problems. Our objective is to initiate the discussion with our industry peers on the basis of

(a) Input image to the neural network. Image source: BMW Group.



(b) Activation of a convulutional layer of VGG16 [13]. Some filters responded to the edges in the image, others to the vehicle parts such as wheels or mirrors. It is not predetermined through programming which filter is activated by which kind of feature. Best viewed in color.

Figure 1: Visualization of layer activation.

SEFAIAS'18, May 28, 2018, Gothenburg, Sweden

ACM ISBN 978-1-4503-5739-5/18/05…$15.00
https://doi.org/10.1145/3194085.3194087

this paper and to develop a common functional safety concept for deep learning approaches together with our partners. We expect the common solution to impact the ISO 26262 standard in the future.

The rest of this paper is structured as follows. Section 2 provides more background information on deep learning and functional safety. Section 3 presents three concrete safety-related issues during our development of deep neural networks, including the training set completeness, network implementation, and transfer learning. Section 4 summarizes this paper.

2 BACKGROUND

In general, there are three major building blocks in the development of deep learning approaches for self-driving cars: *data preparation*, *model generation*, and *model deployment*. Data preparation focuses on getting data ready for training and testing neural networks, covering topics such as data recording, groundtruth labeling, big data storage, etc. Model generation involves developing network architectures, training the networks, and evaluating the trained models. A model is considered "trained" if the difference between its outputs and the corresponding groundtruth labels (the expected outputs) are below a certain threshold. Typically, a trained model is then pruned and optimized for a specific target hardware during the deployment stage, where field tests on testing sites or public roads take place.

As Salay et al. pointed out in [12], the entire pipeline of deep learning development can impact the safety assessment of ISO 26262 through its non-transparency, probabilistic error rate, training-based nature, and instability. First, non-transparency refers to the fact that the knowledge learned by a deep neural network is hardly interpretable by the human (Fig. 1). How to make deep learning more transparent is still an on-going research topic [8]. Second, the error rate and training-based nature are two characteristics of neural networks that must be taken into consideration while developing the safety concept for a deep-learning-based autonomous driving system. Even if the training dataset is 100 percent learned, it is still not guaranteed that the system can operate without error in the real world. There are publications which addressed these issues in more detail, e.g. [14]. Last but not least, instability refers to the same training process producing different training results as the weights are randomly initialized. This is caused by the extremely high dimensionality and complexity of the cost function which contains a massive number of local optima. There is also research [5, 9, 11] addressing the functional safety issues specifically for machine learning, but they are still far away to reach a common industrial solution, let alone to influence ISO 26262.

3 CHALLENGES

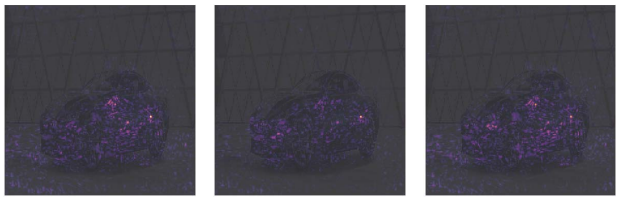
We are currently working on functional safety concepts for deep learning approaches throughout the development of self-driving cars. Here, we present up to three concrete issues to which we do not have answers yet. These include dataset completeness, neural network implementation, and transfer learning. By presenting the issues and posing the questions, we intend to initiate the discussion with our industry peers and work together with them towards a common solution for the automotive industry.

3.1 Dataset Completeness

Today, supervised learning is still the major approach for environmental perception tasks. A supervised model is in general only able to do that which it has been taught. In other words, if a training dataset (which is effectively the set of requirements) only contains cars, a deep neural network trained on that set would not be able to recognize pedestrians. Since the variance of traffic scenes in the real world is near to infinity, it is impossible to ensure a 100 percent coverage of real-world scenes through a single dataset, regardless of its size. Here, the question is rather how to verify that the dataset is "complete" such that a trained model achieves its maximum possible generalization ability.

The criteria are certainly neither the size of the dataset nor the mileage that has been driven by the data collection fleet. Recording eight hours on the Interstate 5 in California for example clearly would not help improve the detection performance of a deep neural network operating in the urban area. In fact, most of the recording data is just "normal" traffic scenes that would contribute very little to updating the weights during the fine-tuning process, given a model that is already pre-trained. It is rather the anomalies in the dataset such as traffic accidents, broken infrastructure, etc. that are of most interest but hard to get. How many anomalies are enough? How to verify that the anomalies are really learned by the network? How to balance the normal scenes and the anomalies during the training? These are the questions that still remain open to us.

Using synthetic data and simulation is one of the possible approaches to the dataset completeness problem. Scenarios that are important for training but rarely happen during a recording session in the real world could be compensated through synthetic data. These include, but are not limited to, hazardous weather conditions (e.g. fog, heavy snow), traffic rule violations (e.g. wrong-way driving, red-light running), animal hazards (e.g. deer), etc. There are on-going projects [1] focusing on generating synthetic training sets for autonomous driving.



(a) Sports car: 43.2% (b) Beach wagon: 14.5% (c) Racer: 13.4%

Figure 2: Visualization of saliency maps through Picasso [7]. The neural network classifies the input image in Fig. 1 as most probably a *Sports Car*, followed by *Beach Wagon* and *Racer*. The most significant areas based on which the network made its decision are highlighted. Best viewed in color.

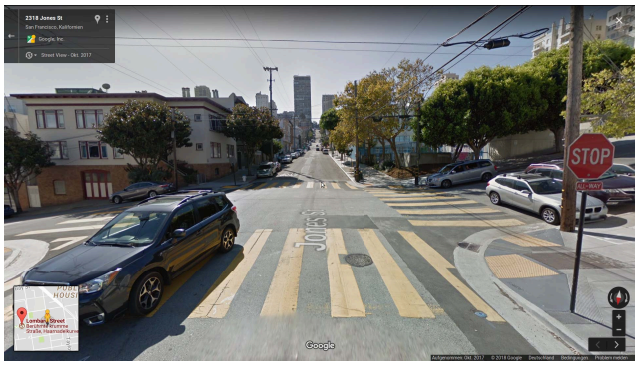
3.2 Neural Network Implementation

Traditionally, functional safety requirements of ISO 26262 are broken down to the code level of the conventional rule-based algorithm and software. It should be clearly identified that a specific functional safety requirement is implemented by a certain piece of code.

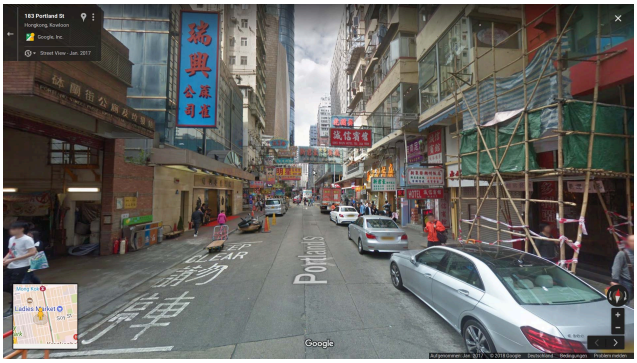
This can hardly be achieved when assessing the implementation of a deep learning approach since building a neural network in the code is nothing more than putting a number of layers together, as demonstrated in the following example using pseudo-code.

```
def build_net_pseudo():
  net = None
  net += conv(inputs, kernel_size, strides, ...)
  net += max_pool(inputs, kernel_size, strides, ...)
  net += ...
  net += fully_connected(inputs, number_outpus, ...)
return net
```

Building the net itself certainly has mostly no influence on the requirements derived from a safety goal, e.g. "Collision with vehicles shall be avoided" or "Collision with pedestrian shall be avoided". Most of the information that is required to assess the safety risk resides in a trained model instead of the code. However, as already addressed in Section 2, the weights of the model and the output of each layer between the input layer and the output layer are not interpretable by the human. This "black box" nature of deep neural networks makes it almost impossible to verify that they are working in a way which they are supposed to work. This is a huge obstacle in the way between deep learning and series production cars which has to be overcome.



(a) Four-way stop in downtown San Francisco. Image source: Google Street View.



(b) *Keep Clear* road markings in Hong Kong. Image source: Google Street View.

Figure 3: Typical traffic scenes in the USA and East Asia.

Visual understanding of convolutional neural networks [16] could be the first step in closing the aforementioned gap. Recently, an open-source framework known as Picasso [7] was published which helps monitor and understand the learning process of neural networks. It supports not only visualizing the feature maps of each layer, but also evaluating partial occlusion, i.e. how the classification changes when parts of the image are occluded, and computing saliency map (Fig. 2), i.e. which input areas matter most to classification. Developers could also implement their own visualizers according to their own use cases.

Another approach that could help interpret the behavior of a neural network is using partial specifications. For example, assume a requirement states that a pedestrian must be less tall than $2.7m$, the constraint on pedestrian height can be used to prove the plausibility of the network output and to filter out false positives. This kind of constraints that are based on the physical properties of the targets could be used during the training process to monitor and influence the behavior of the trained model.

3.3 Transfer Learning

Transfer Learning refers to training a neural network on one task and use part of the pre-trained weights for another related task. For example, we could train a neural network for detecting cars and then retrain it for trucks or possibly even trains. In fact, most of the models in today's deep learning publications are only fine-tuned based on a well-known pre-trained model [3] instead of being trained from the scratch.

The idea of *Transfer Learning* could be applied for reducing the amount of required training data for self-driving cars. Due to different local political restrictions, the data collection fleet could not be operated all around the world. In order to still have a deep neural network that is able to operate in different countries, we would collect data in one area of the world and transfer the knowledge to another, e.g. from Europe to Asia or America (Fig. 3). Here, the question is how many data is additionally needed for the fine-tuning? Do we need rather a small dataset and only retrain the last couple of layers of the network, or do we need a considerable amount of additional data and retrain the complete network? How should we verify the fine-tuned network given an already verified pre-trained model? These are some of the on-going topics that we are currently analyzing. Our major focus is on learning domain-invariant representations once we gain an in-depth understanding of how our trained deep neural networks behave.

4 CONCLUDING REMARKS

In this paper, we addressed three specific challenges concerning functional safety during the development of deep learning approaches for self-driving cars. These include verifying dataset completeness for training and testing, tracing safety requirements to the source code level during software development, and incorporating transfer learning for different areas of the world. We expect this paper to awaken the interest of our industry peers in functional safety for deep learning development. In the near future, we expect to devote ourselves to developing a common solution for the automotive industry together with our partners.

REFERENCES

[1] H. Abu Alhaija, S. K. Mustikovela, L. Mescheder, A. Geiger, and C. Rother. 2017. Augmented Reality Meets Computer Vision: Efficient Data Generation for Urban Driving Scenes. (2017). arXiv:1708.01566

[2] Baidu Inc. 2018. Apollo Auto – An Open Autonomous Driving Platform. (2018). Retrieved Feb 5, 2018 from https://github.com/apolloauto

[3] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and F.-F. Li. 2009. ImageNet: A Large-Scale Hierarchical Image Database. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 248–255.

[4] J. Dickmann, N. Appenrodt, J. Klappstein, H.-L. Blöcher, M. Muntzinger, A. Sailer, M. Hahn, and C. Brenk. 2015. Making Bertha See Even More: Radar Contribution. *IEEE Access* 3 (2015), 1233–1247.

[5] F. Falcini, G. Lami, and A. M. Costanza. 2017. Deep Learning in Automotive Software. *IEEE Software* 34, 3 (2017), 56–63.

[6] A. Grzywaczewski. 2017. Training AI for Self-Driving Vehicles: The Challenge of Scale. (2017). Retrieved Feb 5, 2018 from https://devblogs.nvidia.com/training-self-driving-vehicles-challenge-scale

[7] R. Henderson and R. Rothe. 2017. Picasso: A Modular Framework for Visualizing the Learning Process of Neural Network Image Classifiers. (2017). arXiv:1705.05627

[8] L. A. Hendricks, Z. Akata, M. Rohrbach, J. Donahue, B. Schiele, and T. Darrell. 2016. Generating Visual Explanations. (2016). arXiv:1603.08507

[9] M. Henzel, H. Winner, and B. Lattke. 2017. Herausforderungen in der Absicherung von Fahrerassistenzsystemen bei der Benutzung maschinell gelernter und lernenden Algorithmen [Challenges in Securing Driver Assistance Systems for Machine Learning Algorithms]. In *Proceedings of 11th Workshop Fahrerassistenzsysteme und Automatisiertes Fahren (FAS) [Driving Assistance System and Autonomous Driving]*. 136–148.

[10] ISO 26262:2011(E) 2011. *Road Vehicles – Functional Safety*. Standard. International Organization for Standardization.

[11] R. Krutsch and R. Schlagenhaft. 2017. Diagnostic Mechanism and Robustness of Safety Relevant Automotive Deep Convolutional Networks. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 378–385.

[12] R. Salay, R. Queiroz, and K. Czarnecki. 2017. An Analysis of ISO 26262: Using Machine Learning Safely in Automotive Software. (2017). arXiv:1709.02435

[13] K. Simonyan and A. Zisserman. 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. (2014). arXiv:1409.01556

[14] K. R. Varshney. 2016. Engineering Safety in Machine Learning. In *Information Theory and Applications (ITA) Workshop*. 1–5.

[15] Waymo LLC. 2017. Waymo Safety Report: On the Road to Fully Self-Driving. (2017). Retrieved Feb 5, 2018 from https://storage.googleapis.com/sdc-prod/v1/safety-report/waymo-safety-report-2017.pdf

[16] M. D. Zeiler and R. Fergus. 2013. Visualizing and Understanding Convolutional Networks. (2013). arXiv:1311.02901