ANN-based Classification of Urban Road Environments from Traffic Sign and Crossroad Data

Zoltán Fazekas[1], Gábor Balázs[1,2], Péter Gáspár[1]

[1] Institute for Computer Science and Control (MTA SZTAKI), Kende u. 13-17, H-1111 Budapest, Hungary
e-mails: zoltan.fazekas@sztaki.mta.hu, peter.gaspar@sztaki.mta.hu

[2] Zukunft Mobility GmbH, Ruppertswies 14, D-85092 Kösching, Germany
e-mail: gabor.balazs@zf.com

Abstract: A method that distinguishes between urban road environment types, based on traffic sign (TS) and crossroad (CR) data is presented in this paper. The types and the along-the-route locations of the TSs and the CRs – encountered during car trips – are recorded either by a human data entry assistant, or by an advanced driver assistance subsystem that has been enhanced for the purpose. A feed-forward artificial neural network (ANN) – trained in a supervised manner – carries out the classification tasks. ANNs with different topologies and training regimes are considered and tested for the purpose. These ANNs are characterized by different degrees of modularity ranging from fully modular to non-modular networks. The fully modular ANN consists of three functional modules. Two of these three were trained initially as standalone ANNs, to infer the actual road environment type solely from the TS and the CR data, respectively. The outputs of these two modules are combined via the third module. Further synapses supplement the module-level connections in the less modular ANNs. During the training of the full ANN, the TS and the CR modules are kept relatively intact, while the weights and the biases within the merger module can evolve. Test results for the considered ANNs are provided and compared.

Keywords: detection of the road environment; artificial neural networks; traffic sign recognition systems

1 Introduction

The amount of real-time data measured, gathered and processed on-board high-end road vehicles, increases continually from year to year. Such data gathering and data processing are carried out, for instance, by the real-time traffic sign recognition system (TSR) presented in [1]. The system described therein relies on an RGB image sequence, and a depth image sequence and GPS location, odometer

and map data. The depth image sequence is used for the selection of the regions-of interest, while template matching is applied to the color-segmented RGB image sequence. The latter locates the traffic signs within the image sequence. The system repeatedly performs a particle filter based localization of the ego-vehicle based on these data. The real-time implementation is achieved via the use of multicore processors and a number of graphics processing units. As a more recent example of on-board real-time data gathering and high-volume data processing, the object classification system proposed in [2] could be mentioned. The system fuses image data and the LIDAR point cloud data. Furthermore, it uses a deep convolutional neural network that is fed with pixel-level depth data – obtained via point cloud up-sampling – and RGB color image data. In our view, the increasing trend of the real-time data gathered and processed on-board is partly explainable with – but also permits and calls for – a richer set of on-board environment perception capabilities, much richer than it was customary, say, a decade ago.

The need for a more detailed perception of the environment arises in a wide range of road environments and traffic situations. Notably, it arises in urban environments. For instance, the narrow and possibly blocked streets – typical in historical town-centers – require due attention, clear perception and fast decision making [3]. Furthermore, a profound understanding/model of the road environment is indispensable in the ever-growing and increasingly hectic road traffic at major urban junctions for human drivers and for smart/semi-autonomous/autonomous road vehicles alike. In case of smart road vehicles, the information concerning the key components of the road and vehicle environment is made available to the drivers in the form of advanced driver assistance functions; while in case of semi-autonomous/autonomous road vehicles, it is utilized by the vehicles' own control system.

The data describing the surroundings of high-end vehicles – data typically collected and processed on-board – includes the positions, dimensions and velocities of the pedestrians and the vehicles using the road, as well as the positions, shapes and dimensions of the markings, traffic signs, traffic lights and other objects on, along and in the vicinity of the roads. For instance, a computer vision solution that classifies the road environment into urban, rural, or off-road categories based on color and texture features was proposed in [4]. These features were derived from the color and texture distributions extracted from various regions of interest. The features were then combined using a trained classifier approach to resolve two road-type classification problems. The first was the determination of the off-road/on-road situations. The second was the multiclass road environment problem of determining the actual road type, namely off-road, urban, major/trunk road and multilane motorway/carriageway.

From various data mentioned above, computer vision and artificial intelligence units on-board 'guess' (calculate, estimate, determine) current traffic conditions, the actual road and lane geometry, as well as, the traffic control status (e.g., actual speed limit, green light).

Herein, the urban road environment surrounding an ego-car is classified into one of the three predefined road environment categories based on traffic sign (TS) and crossroad (CR) data. These three road environment categories are as follows: downtown (Dt), industrial/commercial (Ind), and residential (Res) areas. The classification is carried out by an artificial neural network (ANN). Though, the collected data is processed in a post-collection manner, the trained ANNs could well be installed on-board of smart cars and operated in a real-time manner within the advanced driver assistance systems (ADAS).

ANNs of different degrees of modularity are proposed and tested for the purpose of urban road environment classification in the present paper. The recognition performances of these ANNs are evaluated and compared.

Modularity is a desired characteristic of systems of any practical purpose. Such systems include computerized systems and computing/ processing machinery, as well. ANNs are no exceptions to this rule. Modularity serves many of the well-founded engineering demands during the life-cycle of systems (i.e., during system design, maintenance, validation and system renewal/update). It also promotes the traceability of the systems, and makes easier for the system developers and users to understand, follow and verify the computations carried out within.

Modularity also promotes the reusability of programs, applications, methods, modules, subsystems, and even – as it is the case here – the tested and properly functioning weights and biases associated with interneuron synapses within the ANNs and the neurons themselves. These are to be used within modules, or subsystems in a more complex computing network. But modularity comes at a price, e.g., the processing could require several stages/layers and the precision could be slightly impacted. Herein, we look at ANNs exhibiting different levels of modularity, evaluate and compare their performance when these are employed in the given application context.

To cope with the different levels of modularity, furthermore to maintain consistency among the various ANNs used in our experiments, several training regimes – using the analogy of the vibrations of excited nodes within a coupled mechanical network – were devised. These training regimes when applied to an ANN, retain – as much as possible – the weights and biases in certain well-defined parts (e.g., within a module, or within a subsystem) of the network. It is expected that the reasonably good starting values for these weights and biases – if retained or modified with care – shorten the necessary training effort in respect of the network.

The aforementioned three road environment types are rather different from a traffic safety point of view, as it was pointed out by the authors of [5]. Because of these differences, the human drivers, the semi- and the fully- autonomous road vehicles need to look out for very different hazardous traffic situations within these environments. Some important accident and crash data, for different urban road environments in the City of San Antonio, Texas, USA, are given in the cited

paper. The data presented in the article, as well as, similar accident and crash data from other cities, e.g., accident and crash data from Xi'an (China) that was presented and analyzed in [6], underline the need for the ADAS function proposed herein.

In a smart car driven by a human driver, the output of the road environment detection ADAS subsystem could be simply displayed to the driver as a short message (e.g., "You are now probably driving in a downtown area."), so that the driver can adjust to expected traffic conditions and any foreseeable safety hazards. In case of semi-autonomous/autonomous vehicles, for instance, different voluntary speed-limits could be set for different road environments and these limits could be deferred to by the vehicular control system.

The identified road environment type, on the other hand, could be utilized within the mentioned cooperative system architecture. For instance, the customary/standard size of the TSs may vary in different road environments (e.g., nowadays some very small TSs are also deployed in downtown of Budapest), and the TS size for the given environment category, could be beneficial for cross-checking the detected TS candidates. As another example for supporting, with actual road environment information, the processing carried out within a TSR ADAS subsystem, one could mention the different occurrence probabilities of the various TSs. If several candidate TS types are identified for a particular TS encountered along the route, then for choosing the most likely type, these probabilities could be taken into consideration.

In an advantageous implementation, the TS and CR data is gathered, processed, combined and used by ADAS subsystems (e.g., by a camera-based TSR system, or by a lane keeping assistant). These subsystems may rely on data provided by on-board measurement devices (e.g., LiDARs).

The rest of the paper is structured as follows. The first subject in Section 2 is the need for a more robust cooperation amongst the ADAS subsystems in the context of road environment detection. Then achievements in the field of road environment perception, modeling, interpretation and representation are touched upon and some precursors to the proposed ADAS subsystem are mentioned therein. Finally, still in Section 2, the socio-economic relevance of the road and traffic related data is pointed out with reference to an interesting large-scale application of such data. In Section 3, a summary of the car-based TS and CR data collection work carried out in the present research is presented. In Section 4, the ANN-based urban road environment classifiers – exhibiting different degrees of modularity, used in our investigations, are described and their constituent modules/subnetworks, as well as, the two-phase training regime used in conjunction with the classifiers are also described therein. In Section 5, the test results are presented, graphically and also in tabular form, for a particular test route. The results are also discussed therein. Conclusions are drawn and the future work is outlined at the end of this paper.

2 Related Work

2.1 Making Good Use of the Subsystem-Level Cooperation within Advanced Driver Assistance Systems

A biologically inspired system architecture that supports environmental perception capabilities and makes extensive use of subsystem-level cooperation is presented in [7]. The authors of the paper argue that within the majority of the ADAS – at least then, i.e., in 2011 – the ADAS subsystems have clearly defined functionalities, and work fairly independently to complete their specific tasks. In other words, the subsystems do not make extensive use of the results produced by other ADAS subsystems. Although, the loosely integrated ADAS – built from 'individualistic' subsystems – show good performance at the implemented functions, such systems are stuck at a low level of abstraction and are unable to handle complex scenes and traffic situations in a generic way. Furthermore, the extension and the modification of the implemented ADAS capability is far from being straightforward. The biologically inspired system architecture proposed in their paper – which incorporates a module responsible for static domain-specific tasks, pathways for object recognition and location/distance computations, as well as a module for environmental interaction – warrants a higher abstraction level and eliminates the mentioned drawbacks, they claim.

2.2 Road Environment Detection

Road environment detection, perception, modeling, interpretation and representation have been targets of research for some time [7] [8]. The former paper has been discussed above. The authors of the latter paper list a number of current reliable working ADAS subsystems in production cars. In their view, the research activities worldwide have just started to address then, and are definitely addressing now, the driver assistance related environment perception problems for inner-city scenarios.

The aim of these activities is to provide substantial and reliable information about the actual driving situation even in such complex spatial environments. As the authors point out in their paper, this task necessitates the application of multi-sensor systems and advanced sensor fusion technologies. The most frequently used data fusion methods are either object-based, or occupancy grid-based. The authors present an occupancy grid-based fusion concept that is optimized for the environment perception task facing and posed by road vehicles. Their system initially separates the static and the dynamic information coming from the on-board environment sensors (i.e., stereo cameras, radars).

An interesting view, with considerable insight in the field of data fusion, is presented in [9]. The authors of the paper forecast that future environment perception systems, including those on-board road vehicles, will rely on model-free grid-based representations and, at the same time, on model-based object tracking solutions. This is because only a combination of both will meet the requirements of the complex ADAS.

In regards of our present topic, the multi-layer representation of stationary inner-city intersections presented in [3], is of great interest. The authors of the cited paper, use the term 'multi-layer' both in a geometrical sense (to distinguish between ground and raised features) and in an algorithmic/computer architectural sense (to distinguish between the sensor- and data-specific layer, the tristate abstraction layer and the fusion layer). They differentiate between and/or indicate explicit and implicit free spaces, the ground texture, the curbs, the elevated occupancies, the texture+elevated occupancies, and multiple occupancies within the intersection.

The parametric free space (PFS) map – a novel generic 2D environment representation suitable for automotive applications – was introduced in [10]. The representation proposed there is more compact than those based on common grid-based models, and therefore, it is could be used for the purpose of automotive CAN transmission. The PFS map maintains explicit information about relevant free spaces, while setting aside data describing irrelevant free spaces. Using the PFS map, an arbitrarily fine spatial evaluation can be carried out in a sensor-principle independent and real-time manner. The authors consider radar and stereo camera data streams in their experiments, but claim that additional sensors could be included in the map generation. The generation process, however, is computationally more demanding than pure grid mapping.

In the papers discussed above, the term 'road environment' is used in a fairly narrow, more or less, geometrical sense. The term refers to the immediate/close surroundings of the ego-car, e.g., to the physical extent of an inner-city intersection/roundabout, of a multi-lane road segment, or of a road segment by a construction site. The authors of [11], on the other hand, use a slightly different term, namely 'driving environment', and they study the 'critical changes'. The examples given there are related to the sudden changes of illumination (e.g., tunnel entry, tunnel exit, shadow of an overpass), but still refers to the immediate/close surroundings of the ego-car.

In our view, the above papers tackle the highly relevant and practical spatial perception, modeling, interpretation and representation problems that arise in urban spaces, but miss out on dealing with the urban environment – around the ego-car – at a somewhat larger scale. Herein, the term 'road environment' is used in a more socio-economic sense and refers to larger urban spaces; this interpretation is used in [12]. The authors of the cited paper also define and use a number of socio-economic measures characterizing the urban form and everyday life. These include the residential density, employment density, land use diversity,

intersection density, size of working-age population within a (short) driving distance, number of jobs accessible by car, size of working-age population within commute distance and the number of jobs accessible within a short transit commute. Looking at maps of a given city/town with the above measures noted, one gets a fairly good understanding of how everyday urban activities are conducted in that settlement. One could easily define the road environment categories Dt and Res in a quantitative manner based on these and similar socio-economic measures, and could even combine these measures to come up with practical definitions. Herein, however, a fairly simple and a somewhat subjective categorization of these urban environments are used, which considers observable traits related to the residential and intersection densities.

In respect of the third urban environment type (i.e., Ind), we refer to [13]. The author of this doctoral dissertation examines the spatial distribution of commercial activities in Montréal, Canada. In particular, the impact of spatial determinants pertaining to street permeability and street centrality on the character and spatial distribution of retail activities is investigated. Among a wide range of analyses carried out in respect of the 'urban tissue', the author investigates the spatial characteristics of commercial streets using a morphological approach. To account for the different types of intersections on a commercial street, mainly T-types and +-types, the distances between these intersections were measured for both sides of the street, and the averages are produced. Such an approach could be utilized in precisely defining the Ind road environment category, and also the road environment data collection could be improved in this manner, as presently we do not distinguish between the left-hand-side and right-hand-side environments of the road.

2.3 Statistical Inference and ANN-based Methods for Road Environment Detection

Road environment detection (RoED) – in urban areas and in the above detailed socio-economic sense – was tackled in [14], [15] and [16]. The methods presented in these papers rely on TS and/or CR input data. Two different approaches of RoED, namely statistical inference and ANN-based classification, were presented in these publications.

It should be emphasized that this algorithmic dichotomy is not unique to the problem at hand. The statistical and the ANN-based methods commonly applied in transportation research are surveyed in [17]. The authors look at the differences between and the similarities of these methods and provide insights on how to choose one from the available algorithmic palette.

A shallow ANN was utilized to identify the actual urban road environment based on TS data in [16]. The urban RoED method proposed herein builds upon the results presented there and to a lesser extent, upon those published in [14] and

[15]. In fact, the ANN – described in [16] – that was trained and used for processing TS data has been re-applied herein as a functional module/subnetwork. In the following, this functional unit will be referred to as the TS processing module/ subnetwork. It is augmented with a similar module/subnetwork that inputs CR data and with another module that merges the outputs of the former two.

Multiple functionally independent subnetworks (modules), as a part of the whole ANN, were used and experimented with, in [18], the paper also presented various training techniques for such ANNs. Some of the training techniques described were tested in regard to the RoED system proposed herein.

Modularity within ANNs serves a variety of design objectives, see [19] for a good overview; the main motivations for turning to a modular design in this case were the availability of a trained ANN (i.e., reuse of a readily available software component), the expected reduction of the required training effort, and the increased understandability/traceability of the data processing.

2.4   Road Environments, Traffic Patterns, Composition of the Traffic and their Socio-Economic Relevance

Clearly, urban RoED, even in the socio-economic sense of the term as used here, can be carried out in a number of different ways and from different kinds of input data. The most obvious option is to use a navigational device and stored map data, such as provided by the OpenStreetMap [20], for the purpose and to rely on the urban area categorization given there. Nonetheless, a similar argument could be brought forward in conjunction with TSR, and still we see that there are many camera-based TSR systems on the market, and these ADAS subsystems either use, or do not, the ego-car's geographical position and stored map data for data fusion and data corroboration purposes.

We opted for using the TS and CR data as inputs to the RoED ADAS function as the TS data is readily available on-board of high-end road vehicles through the TSR function, while the CR data can be generated from LIDAR point clouds. We note here that the LIDAR-based ADAS solutions have gained popularity in the recent years [21] [22], and we expect to see CR detection ADAS functions in smart cars in the near future. Having said that, it is true that instead of looking at TSs and CRs, or in deed the built environment in general, RoED could be accomplished via observing traffic intensities, traffic patterns and the composition of the traffic (e.g., with respect to road vehicle types, brands, and models).

In this context, the work presented in [23] could be mentioned here. The authors of the cited paper turn to the deep learning methodology for estimating the socio-economic characteristics of approximately two-hundred US regions. Their aim, however, is not related to driver assistance. They extract, using deep learning-

based computer vision techniques, the motor vehicles encountered by the Google cars taking Street View images. In total, several millions of Street View images were processed for the purpose of this major endeavor. From the extracted image segments/blobs, the makes, the models, and the production years of these vehicles were sought and identified. This data is then used for estimating various socio-economic characteristics of the administrative regions of the studied country.

3 Collection of Traffic Sign, Crossroad and Urban Road Environment Data

In this section, a brief description of the data collection work in respect of TSs, CRs and urban road environments is given. More details, including route maps, as well as photographs of typical scenes can be found in [14] [15] [16]. In respect of the TS data, several car-based data collection trips were made to three urban settlements within Hungary, namely to Csepel[1], Százhalombatta[2] and Vác[3]. These locations were chosen to cover urban settlement population sizes characteristic for the country. It was our aim to include industrial, cultural, commercial centers for data collection, while trying to include both historic and modern settlements, as well as settlements with garden suburbs and green residential areas. For convenience reasons, we chose destinations that are not too far from Budapest, where our research institute is located.

An Android application was used for recording the relevant TS data. The application automatically records the car-trajectory, and provides means to log the TSs. The data logged for each TS includes the TS type – the types used by the proposed RoED are shown in Figure 1 – and the TS location along the route covered, and the actual road environment type. As it was mentioned in the Introduction, three predefined road environment types were considered, namely

---

[1]   Csepel is the 21st district of Budapest. Some decades ago, it was a working-class borough with many factories. Today, Csepel contains housing estates, as well as middle-class garden suburbs. It has approximately 85,000 inhabitants. (Excerpts from https://en.wikipedia.org/wiki/Csepel)

[2]   Százhalombatta is situated about 30 km from Budapest, along national motorway No. 6. It has about 18,000 inhabitants. Looking at the modern industrial town today, it is hard to believe that the settlement has about 4,000-year-old history. (Excerpts from http://www.1hungary.com/info/szazhalombatta/)

[3]   Vác is a town in Pest county with about 35,000 inhabitants. The town is located 35 kilometres north of Budapest on the eastern bank of the Danube river. Its history dates back to the days of the Roman Empire. Later, in the middle ages, the town became a Roman Catholic bishopric. Nowadays, Vác is an educational, cultural, commercial, industrial and religous center of Pest county, as well as a popular summer resort. (Excerpts from https://en.wikipedia.org/wiki/Vác)

the downtown (Dt), industrial/commercial (Ind), and residential (Res) areas. A data entry assistant entered the data describing the TSs seen along the route, as well as categorized and recorded the actual road environment according to their best judgment.

Figure 1

The TS types used by the proposed RoED system; the different shades of the background signify the Dt, Ind and Res road environments, respectively, in which the given signs prevail

Eight TS types were identified in [14] as occurring frequently in urban areas and prevailing in one of the three urban environment types mentioned above. In Fig. 1, these TS types are shown in groups corresponding, from left to right, to Dt (dark grey background), Ind and Res road environment types (black and light grey backgrounds, respectively). Frequent occurrences of the two Dt TSs suggest that one is driving in a Dt area, those of the four Ind TSs indicate with some probability an Ind area, while the two Res TSs are indicative, again with some probability, of a Res area.

For the purpose of our study, the CR data was added in a post-trip manner to the trajectory data. It was done by collating the recorded trajectory with the intersection data extracted from the road layer of a public geographical information system provided by [20].

Five CR categories were considered for the purpose of RoED, namely the T-shaped CRs, the +-shaped CRs, the complex CRs, the roundabouts (all these without traffic lights), and any CRs controlled by traffic lights. Examples of these CRs are shown, from left to right, in Figure 2. These CR categories had been used in [15] for detecting change, with properly tuned Page-Hinkley change detectors, in the character of the urban road environment sweeping past the ego-car collecting data.

Figure 2

Instances of CR categories used by the proposed RoED system, namely instances of T-shaped, +-shaped, complex CRs, roundabouts – all these without traffic lights – and CRs controlled by traffic lights

It should be noted that in an advantageous implementation of the proposed RoED system, the TS and CR data should be gathered, combined and processed by ADAS subsystems. The processing could easily be carried out in real-time, both for the change detection approach and for ANN-based RoED.

4   Urban Road Environment Classifiers Making Use of ANNs Exhibiting Different Degrees of Modularity

4.1   TS-Processing Module/Subnetwork

An ANN with one hidden layer was chosen for detecting the type of the actual urban environment solely from TS data in [14]. The ANN presented therein, as well as, its variants described and used herein, were implemented, trained and tested in the simulation environment Simbrain. The features and capabilities of this simulation environment are presented in [24]. Simbrain has a number of pre-defined ANN-models, including the backpropagation model used in the present work, and instances of these models can be easily created, trained and tested.
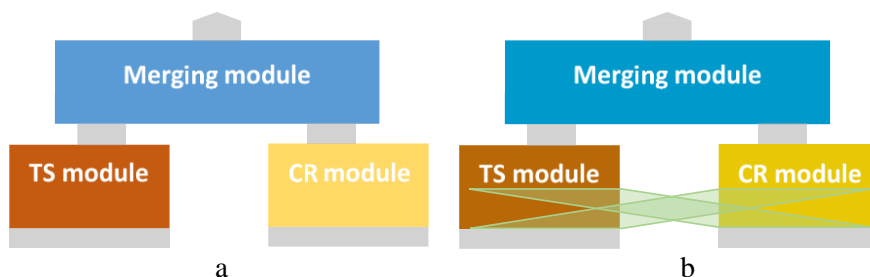

Figure 3

The modular RoED system with separate TS and CR modules (a) and the non-modular RoED system with TS and CR subnetworks connected via a merging module and via additional synopses (b)

The trained ANN is reused herein, as a module and a subnetwork, see Figures 3a and 3b, respectively, for the purpose of TS-processing within the full ANNs that consider both TS and CR data.

The input features of the TS-processing ANN and also of the TS-processing module/subnetwork, within the full ANNs, are the average distances between consecutive relevant TSs (of any sort) over the last 250, 500, 1000 and 2000 meters of the trajectory, and the number of occurrences of the typical TSs pertaining to each of the considered three road environments, again over the mentioned path-lengths. That is, there are in total 16 neurons in the input layer of the TS-processing module as can be verified in Figure 4. (The same number of neurons are used in the non-modular ANN by the TS-processing subnetwork.) Figure 4 shows the inner structure of the (trained) modular ANN for RoED and the three modules shown in Figure 3a, can be identified therein.
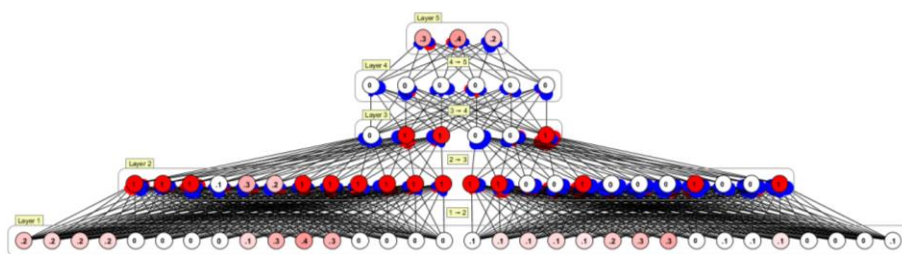
Figure 4

The inner structure of the (trained) modular ANN for RoED

The input features are calculated for consecutive route-segments of 50 meters of the car trajectory. These features are used both in the individual training and testing of the TS-processing module/subnetwork, and in training and testing of the full ANN. The network weights and biases computed in the standalone phase are retained for the full ANN and only a limited and controlled modification is allowed during the final training. After appropriate training, the TS-processing ANN – used herein as module/subnetwork, could achieve a 67.3% agreement with the ground truth data, for a particular route in Csepel.

4.2   CR-Processing Module/Subnetwork

An ANN with the same topology as the TS-processing ANN described above (cf. the left and right modules formed by the lower layers of the full ANN shown in Figure 4) that also relies on the same multiscale approach embodied in the input features was set up for detecting the type of the actual urban environment solely from CR data.

The CR data consists of the type and the location of each intersection along the route. The input features to the CR-processing ANN are the average distances between consecutive CRs (of any sort) over the mentioned path-lengths, and the number of occurrences of the typical intersection types for each of the road environment types again over the above path-lengths.

The T-shaped crossings are 'slightly typical' to Res areas, the +-shaped and the *-shaped (i.e., more than four-legged) crossings are 'slightly typical' to Dt areas, while the roundabouts and the traffic light-controlled crossings are 'slightly typical' to Ind areas. The above described CR-processing ANN is used herein as a module and a subnetwork – see Figures 3a and b, respectively – for the purpose of CR-processing within the full ANNs.

Similarly to the training of the TS-processing module/subnetwork, the CR module/subnetwork was trained separately, using a supervised learning approach, via backpropagation. The resultant weights and biases within the module/ subnetwork were retained for the full ANN and only a limited and controlled

modification was allowed during its final training. After appropriate training, the CR-processing ANN, used here as module/subnetwork, could achieve a 59.7% agreement with the ground truth data, for a test route in Csepel.

4.3   ANNs for Processing TS and CR Data Jointly

In Figures 3a and b, the inner structures of the two main types of the full ANNs that were used in our experiments are presented schematically A supervised learning approach was used, as both input data (i.e., TS and CR data) and the desired output (i.e. the actual urban road environment type) were known, and the desired output was available as reference data. In Figure 4, the trained version of the full modular ANN, i.e. the one sketched in Figure 3a, is shown using the Simbrain simulation environment.

Each of the two, full ANNs, comprises three functional modules/subnetworks: two of these feed into the third one. The two feeding modules/subnetworks process, exclusively/primarily, the TS and the CR data, respectively. The input signals, marked collectively with grey rectangles in Figures 3a and b, are fed into the ANNs at the bottom, while the classification results, marked with grey rectangles and arrows, concerning the current urban environment types, appear at the top of the modules/subnetworks.

The TS and CR data logs provide two separate 'views' on the actual urban road environment, while the third module combines the outputs of the other two modules/subnetworks, and produces a final road environment guess. In Figures 3a and 4, the TS and CR modules interact only through the merger module, while in Figure 3b, apart from the interaction via the merger module, there are also synapses between the neurons of the TS and CR processing subnetworks. These synapses are collectively marked by the two green parallelograms in the figure. The detailed graphical representation of the non-modular ANN is omitted from this paper.

Initially, each of the full ANNs in Figure 3a and b comprise only an individually trained TS-processing module and an individually trained CR-processing module. In each of the two networks, the aforementined modules are augmented with a merger module, which is then trained through backpropagation, but without modifying the TS and CR modules. This training of the merger module is referred to as initial training of the full ANN. After this initial training, the full ANN achieved a 63.7% agreement with the ground truth road environment type data on a particular test route in Csepel. This agreement value falls between the agreement values[4] computed for the individual modules; i.e., it falls between 67.3% for the TS-processing module/ANN and 59.7% for the CR-processing module/ANN.

---

[4] These agreement values were given in Subsections 4.1 and 4.2, respectively.

5   Urban Road Environment Detection Results

Following the initial training, a two-phase training regime was carried out. The training parameters, the training errors and the test agreements for the modular full ANN, shown in Figures 3a and 4, trained according to this regime are given in Table 1.

Table 1
First-stiff-then-loose training regime used for the modular full ANN

Parameter setting	Learning rate	Momen-tum	Training error	Agree-ment	Test stripe shown below
A1	0.0250	0.0900	11.7%	56.1%	
A2	0.2500	0.9000	12.1%	56.1%	
B1	0.0050	0.0180	15.5%	49.6%	
B2	0.0500	0.1800	14.2 %	57.9 %	
C1	0.0010	0.0036	13.9%	59.3%	
C2	0.0050	0.0180	14.8%	55.8%	
D1	0.0002	0.0072	3.5%	71.9%	✓
D2	0.0010	0.0036	3.1%	68.7%	

According to this regime, a stiff training phase is followed by a loose training phase. The stiff training phase modifies all three modules (i.e., the TS-processing, the CR-processing and the merger modules) in a controlled manner, which is then followed by a loose training phase that modifies only the weights and biases within the merger.

The stiff training phases with different parameter settings are referred to as A1, B1, C1 and D1. See Table 1 for the concrete parameters. Each of these concrete training phases were then followed by loose training phases, namely A2, B2, C2 and D2, respectively. The parameter settings for these concrete training phases are also given in Table 1.

Similar data for the full ANN shown in Figure 3b trained according to the mentioned two-phase training regime are presented in Table 2. The rows of the table correspond to full non-modular ANNs with increasing percentages (i.e., 20%, 40%, 60% and 80%) of synapses, with the new synapses randomly added to the existing ones, between the neurons of the TS-, and CR-processing subnetworks.

Table 2
First-stiff-then-loose training of the ANN with additional synapses

TS-CR synapses	Stage	Learning rate	Momen-tum	Training error	Agree-ment	Test stripe shown below
20%	stiff	0.0025	0.009	07.9%	69.4%	✓
20%	loose	0.0250	0.090	06.3%	68.3%	
40%	stiff	0.0025	0.009	03.1%	73.4%	
40%	loose	0.0250	0.090	06.5%	73.0%	✓

60%	stiff	0.0025	0.009	05.6%	74.1%	
60%	loose	0.0250	0.090	14.0%	43.2%	
80%	stiff	0.0025	0.009	3.0%	76.3%	✓
80%	loose	0.0250	0.090	19.1%	63.7%	

In Figure 5, the test results for the full ANNs, corresponding to the parameter settings and training stages tagged in Tables 1 and 2, are compared to the ground truth data in respect of a particular test route in Csepel. (Note that all the lower test stripes within the pairs shown in the figure are the same, i.e., the ground truth road environment categorization of the test route. It is just repeated so that the test results are easier compared with the ground truth.)

From the top to the bottom, the test stripes for the following parameter settings and training stages appear in the figure: parameter setting D of the modular ANN after the stiff training phase, the non-modular ANN with 20% of the possible TS-CR synapses after the stiff training phase, the non-modular ANN with 40% of the possible TS-CR synapses after stiff training followed by loose training phase, the non-modular ANN with 80% of the possible TS-CR synapses after stiff training phase.

The results shown in Tables 1 and 2 indicate that the agreements with the ground truth road environment types may considerably improve, via the application of the proposed two-phase training regime, compared to the agreements achieved by the full ANNs after just the initial training. Still, one finds that even the best agreement values are around 75%, i.e., they are not that high. Clearly, better results can be gained via using GPS and reliable map data used together. This evident RoED approach was mentioned, together with some other RoED alternatives, in Subsection 2.4.
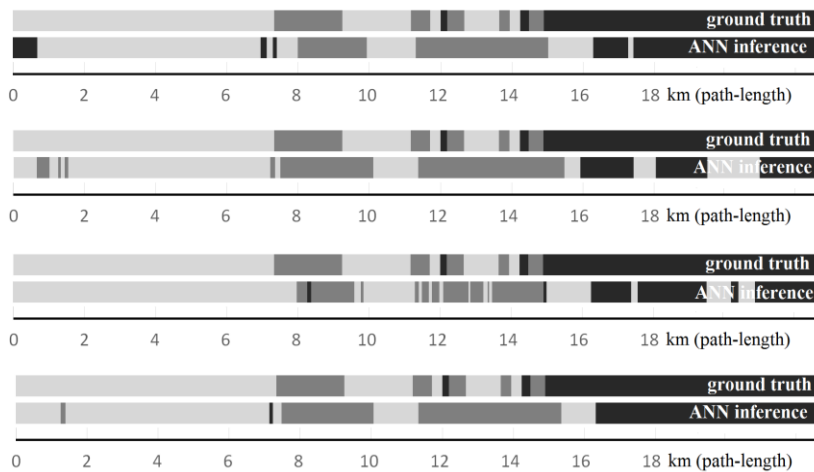

Figure 5
The road environment types manually recorded along a test route in Csepel, and those inferred by the full ANNs in different training phases (see the text for details)

In our view, when one evaluates the agreement values shown in Tables 1 and 2, one should bear in mind what sort of raw input data is available, in the presented application scenario, for the road environment classification. The raw input data series, generated by the car-based TS and CR data collection in respect of the road infrastructure, can be thought of as a realization of a marked random point process with TS and CR types appearing as marks for the along-the-route locations. The locations are characterized only by the path-length covered by the ego-car. Based on the raw input data series, the average distances between TSs and CRs, as well as the number of TS and CR occurrences over certain path-lengths, see Sub-sections 4.1 and 4.2 for details, are calculated by some circuitry (not shown in the figures). This derived/aggregated data are used then as input data by the TS-, and CR-processing modules/subnetworks. The marked random point process generating the raw input data is modeled with a marked Poisson point process in [15].

In our view, the TS and CR data described above is fairly basic, it lacks important road details (e.g., number of lanes and road widths are not known) that could be beneficial in determining the actual road environment type, but which are much more time-, and resource-consuming to collect, and are impractical to use on-board, at the present time. Considering the simplicity of the raw data series, the re-sulting agreement values seem reasonable, perhaps even surprisingly high.

Conclusions and further work

The problem of identifying the actual urban road environment type, sweeping past an ego-car, based on TS and CR data was tackled in this paper. ANNs exhibiting different degrees of modularity and having been trained according to a two-phase regime were considered and tested for the purpose. In the modular case, shown in Fig. 3a, the TS and CR modules were first trained individually and then were put together with the help of a merger module. The aim was to complement the capabilities of the processing modules and bring about improved classification results for the full network.

The considered non-modular ANN layout is shown in Fig. 3b. In this case, some additional synapses between the TS-processing and the CR-processing sub-networks are activated in a random manner and used in the processing. In both cases, improvements of the initial classification performance were achieved for certain parameter settings.

The input data used, was gathered in a small-scale data collection exercise. It is certain that a larger collection of TS and CR data from diverse regions and countries would be essential for real automotive application of the RoED system. Defining and using more urban road environment types could also extend the usability of the approach. Also, adding more CR types and corresponding input features could also be valuable for future research.

References

[1] K. Par, O. Tosun: Real-time Traffic Sign Recognition with Map Fusion on Multicore/Many-core Architectures. Acta Polytechnica Hungarica, 9, 231-250, 2012

[2] H. Gao, B. Cheng, J. Wang, K. Li, J. Zhao, D. Li: Object Classification using CNN-Based Fusion of Vision and LIDAR in Autonomous Vehicle Environment. IEEE Transactions on Industrial Informatics, 14, 4224-4231, 2018

[3] J. Rieken, R. Matthaei, M. Maurer: Toward Perception-Driven Urban Environment Modeling for Automated Road Vehicles, In: IEEE Int. Conference on Intelligent Transportation Systems, 731-738, 2015

[4] I. Tang, T. P. Breckon: Automatic Road Environment Classification, IEEE Trans. on Intelligent Transportation Systems, 12, 476-484, 2011

[5] E. Dumbaugh, R. Rae: Safe Urban Form: Revisiting the Relationship between Community Design and Traffic Safety, Journal of the American Planning Association, 75, 309-329, 2009

[6] Y. G. Wang, S. S. Huang, W. S. Xiang, Y. L. Pei: Multipattern Road Traffic Crashes and Injuries: a Case Study of Xi'an City. Acta Polytechnica Hungarica, 8, pp. 171-181, 2011

[7] R. Kastner, T. Michalke, J. Adamy, J. Fritsch, C. Goerick: Task-Based Environment Interpretation and System Architecture for Next Generation ADAS, IEEE Intelligent Transportation System Magazine, 3, 20-33, 2011

[8] T. N. Nguyen, M. M. Meinecke, M. Tornow, B. Michaelis: Optimized Grid-Based Environment Perception in Advanced Driver Assistance Systems, In: IEEE Intelligent Vehicles Symposium, 425-430, 2009

[9] C. Glaser, T. P. Michalke, L. Burkle, F. Niewels: Environment Perception for Inner-City Driver Assistance and Highly-Automated Driving. In IEEE Intelligent Vehicles Symposium, 1270-1275, 2014

[10] M. Schreier, V. Willert, J. Adamy: From Grid Maps to Parametric Free Space Maps – a Highly Compact, Generic Environment Representation for ADAS, In: IEEE Intelligent Vehicles Symposium, 938-944, 2013

[11] C. Y. Fang, S. W. Chen, C. S. Fuh: Automatic Change Detection of Driving Environments in a Vision-Based Driver Assistance System, IEEE Trans. on Neural Networks, 14, 646-657, 2003

[12] K. Ramsey, J. Thomas: EPA's Smart Location Database: A National Dataset for Characterizing Location Sustainability and Urban Form, Draft Technical Report, Washington, DC, USA, Environmental Protection Agency, Office of Sustainable Communities, 1-27, 2012

[13] J. Villain: The Impact of the Urban Form on the Spatial Distribution of Commercial Activities in Montréal, doctoral dissertation, Concordia University, Montreal, Quebec, Canada, 1-203, 2011

[14] Z. Fazekas, G. Balázs, L. Gerencsér, P. Gáspár: Inferring the Actual Urban Road Environment from Traffic Sign Data Using a Minimum Description Length Approach, Transportation Research Procedia, 27, 516-523, 2017

[15] Z. Fazekas, G. Balázs, L. Gerencsér, P. Gáspár: Detecting Change in the Urban Road Environment Along a Route Based on Traffic Sign and Cross-road Data. In: Intelligent Transport Systems - From Research and Development to the Market Uptake, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering (222) Springer, Cham, Switzerland, 252-262, 2018

[16] Z. Fazekas, G. Balázs, P. Gáspár: Identifying the Urban Road Environment Type from Traffic Sign Data Using an Artificial Neural Network. In: the Proceedings of the International Scientific Conference on Modern Safety Technologies in Transportation, Herlány, Slovakia, 42-49, 2017

[17] M. G. Karlaftis, E. I. Vlahogianni: Statistical Methods versus Neural Networks in Transportation Research: Differences, Similarities and Some Insights, Transportation Research Part C: Emerging Technologies, 19, 387-399, 2011

[18] T. Caelli, L. Guan, W. Wen: Modularity in Neural Computing, Proceedings of the IEEE, 87(9), 1497-1518, 1999

[19] R. Rojas: Neural Networks: A Systematic Introduction, Springer Science & Business Media, Berlin, Germany, p. 502, 2013

[20] OpenStreetMap contributors: Road Network in Hungary, URL: http://planet.openstreetmap.org, 2015, last accessed: 23 Feb, 2018

[21] A. Asvadi, C. Premebida, P. Peixoto, U. Nunes: 3D Lidar-based Static and Moving Obstacle Detection in Driving Environments: An Approach Based on Voxels and Multi-region Ground Planes. Robotics and Autonomous Systems, 83, 299-311, 2016

[22] F. Jiménez, J. E. Naranjo, J. J. Anaya, F. García, A. Ponz, J. M. Armingol: Advanced Driver Assistance System for Road Environments to Improve Safety and Efficiency. Transportation Research Procedia, 14, 2245-2254, 2016

[23] T. Gebru, J. Krause, Y. Wang, D. Chen, J. Deng, E. Lieberman Aiden, L. Fei-Fei: Using Deep Learning and Google Street View to Estimate the Demographic Makeup of Neighborhoods across the United States. Proceedings of the National Academy of Sciences, 114, 13108-13113, 2017

[24] Z. Tosi, J. Yoshimi: Simbrain 3.0. Neural Networks, 83, 1-10, 2016