PRELAT

Precision lateral control for highway automation

Public report

Project within Complex Control
Project nb 2014-06239
Author Martin Sanfridson, Volvo Technology AB
Date 2019-01-27

Content

Currently there are five collaboration programs**:** Electronics, Software and Communication, Energy and Environment, Traffic Safety and Automated Vehicles, Sustainable Production, Efficient and Connected Transport systems.

For more information: www.vinnova.se/ffi

1. Summary

The PRELAT project has addressed robustness of lateral positioning for lateral control of autonomous driving. The target application is automation of trucks for driving on highways. The project is focused on lateral control for which information on drivable road surface and lane markings are of primary interest. In the early autopilot demonstrators, sensors and perception designed for L2 were used. By experience, these proved to have unsatisfactory reliability for a continuous operation such as lane following. When the confidence of a measurement gets low continuously, control has to be suspended and eventually handed over to the driver. For L2 features availability is less of a problem than for L4 since the driver remains in the loop as responsible.

To improve perception for lateral control, the two main ideas pursued in the project have been to add a laser scanner to the camera and to deploy convolutional neural networks for the fusion and classification. The project has also worked on other components in the control loop such as path planning and lateral path following. However, there is not much to be done in a control loop to compensate for unreliable sensor data. The effort should be devoted to reducing the source of the problem.

One result of the project is pointing at the usefulness of lidar as an additional sensor modality to the ubiquitous camera in order to improve robustness. Another result suggests that an intermediate level of fusion, that is, an architecture less demanding than raw sensor data (early fusion) but more informative than composition of previously isolated channels (late fusion). A third result suggests a semi-supervised training scheme with camera and lidar to reduce costly annotation. Much of the work in this project has been carried out by a PhD student in the Applied artificial intelligence group at Chalmers

The PRELAT project has successfully contributed to an area of rapid expansion in the research community and in the automotive industry. The robustness of perception for lateral control has improved. The Volvo Group have explored and refined the results, and will continue to do so for the expected introduction of more automated driving vehicles.

2. Sammanfattning på svenska

PRELAT (precision lateral control for highway automation) startade våren 2015 och har varit ett samarbete mellan Volvo Technology AB och Chalmers tekniska högskola. Målsättningen för PRELAT har varit att förbättra den laterala reglering för automatisering av förare. Det främsta användarfallet har varit autopilot för motorväg. Motorväg erbjuder en kontext som är relativt homogen jämfört stadsmiljö vilket medger att perceptionen och reglering kan nå en högre tillförlitlighet. Inte desto mindre finns det en hel del utmaningar för klassificering av statiska objekt, t ex smutsiga väglinjer, snötäckt vägbana, vägräcken, skarpa skuggor, svag belysning, och trafik som skymmer. Det är åtminstone två saker man vill ha ut från perceptionen: var väglinjerna är relativt fordonet och var det finns körbar vägyta. Det förstnämnda är viktigt för normal lateral filföljning och det andra för planering och beslut vid avvikelse från normal filföljning, t ex för att undvika hinder på vägen eller när väglinjer saknas.

Ansatsen i PRELAT har varit att dels använda lidar för att förbättra perceptionen och dels att använda convolutional neural network för att fusionera data från i första hand kamera och lidar. PRELAT har varit framgångsrikt i ett internationell perspektiv att föreslå och utvärdera förbättringar på dessa två områden.

En sak som skiljer SAE J3016 automationnivå 2 när man tar steget till nivå 4 där fordonet tar ansvar för säkerheten är att kravet på automatiseringen går från att vara tillgängligt till att fungera tillförlitligt. Detta krav påverkar alla komponenter i reglerloopen. Här är vi i första hand inte intresserade av funktionell säkerhet som baserar sig på en felmodell där själva det inbyggda datorsystemet påverkas av en störning. För att klara av ett sådant fel (vanligen antar man enkelfel) bygger man ut med redundans. Vi är istället intresserade av en felmodell som i reglertekniska termer handlar om robusthet mot mätosäkerhet och mätstörningar. Gränsdragningen mellan dessa inte är helt skarp, men åtgärderna för att komma tillrätta med problemen är i huvudsak olika. Applikationsspecifik robusthet är härvidalag viktigare än applikationsoberoende redundans.

Det finns i huvudsak två sätt att komma till rätta med felmodellen robusthet. Den ena är att avgränsa applikationen i ODD (operational design domain) så fordonet inte utsätts för besvärliga situationer, t ex att bara tillåta automation om väg-däck-friktionen är tillräckligt hög. En svårigheten här är emellertid att garantera att man alltid håller sig innanför ODD. Det andra sättet är att förbättra prestandan på den nominella funktionen, till exempel genom att lägga till informationskällor med annan modalitet eller att förbättra algoritmer för perception och reglering.

PRELAT har visat att CNN (convolutional neural networks) som tidigare framgångsrikt använts för klassificering av objekt i kamerabilder, också kan användas för lidardata. Kombinationen av sensordata från kamera och lidar har projektet också arbetat med. Betoningen på lidar motiveras av dess förmåga att erbjuda hög noggrannhet i avståndsmätning i alla typer av ljusförhållanden. Denna alternativa uppfattning av

omvärlden erbjuder ett komplement till kamera som möjliggör att hantera en mycket mer heterogen omgiving.

Genom att arbeta i ett fågelperspektiv med lidardata har PRELAT också visat att det är möjligt att generera banplanering för längre horisonter, i detta fall upp till 30 meter. Planeringen baseras på prediktionen utifrån historiska data och från en hierarkiskt utanförliggande begäran (intention) vart fordonet ska färdas, det vill säga åt höger eller åt vänster eller följa vägen. För att kunna utnytta fördelar hos (fully connected) CNN full ut, kodades intentionerna in i form av pixelvärden i bilder. Träningen baserades på en automatisk generering av ground truth data.

Projektet har också arbetat med multimodal sensor fusion för vägdetektion, som har visat att med en kombination av lidar och kamera är det möjligt att förbättra semantisk segmentering jämför med enbart kamera. Frågan är på vilket sätt lidar och kamera ska fusioneras. En utvärdering av tre olika koncept gjordes: tidig fusion med rådata, sen fusion där data före respektive sensor bearbetats mer först, och ett mellanting där graden av tidig eller sen fusion optimerades i samma träning som de traditionella vikterna. Det sistnämnda fusionsmetoden var klart framgångsrikast av de tre och även mätt i samma benchmark som ovan. Resultatet är intuitivt begripligt.

Slutligen har projekt också tittat på semi-supervised learning som ett komplement till tidigare supervised learning. Här utnyttjas det faktum att man har två sensorer med olika informationsinnehåll. Först tränar man upp klassificerar med hjälp av annoterade bilder. Därefter vidar en iterativ procedur där man klassificerar icke-annoterat material och använder detta (tillsammans med en del redan annoterat) för att träna en ny klassificerar. I nästa steg av iteration byter man roller mellan läraren och studenten.

Sammanfattningsvis har PRELAT avslutats med ett gott resultat i internationellt forskningsområde som har växt explosionsartat i betydelse och storlek under den tid projektet verkat. Detta arbete har i huvudsak utförts av en doktorand vid gruppen Tillämpad artificiell intelligens på Chalmers.

3. Background

PRELAT stands her for PREcision LATeral control for highway automation. The main target has been to improve robustness of lateral positioning on highways. The steering wheel is the primary interface for control of lateral motion, with lane keeping being the basic feature. This actuator gets commands from a path follower. The path follower gets its input from a trajectory planner in combination with a decision maker. For regular autopilot driving on highway the trajectory planner and decision are straight forward. In the beginning of this chain we have sensors followed by sensor fusion. The performance of the sensor and sensor fusion affects the lateral control at both planning level and following level. At Volvo Technology, the demonstrator work in FP7 project AdaptIVe (with participants from most European automotive manufacturer) provided evidence, as indicated above, that the perception needed to be improved.

The focus in the project has been on the perception and especially fusion of sensors. Accuracy and robustness in the perception is related to safety of intended functionality. A typical problem is (partially) unobservable lane markings. The lateral positioning is in this project relative to driveable road surface or lane markings.

Figure 1 A tunnel exit found along highway in southern Sweden

As an example, consider reaching the tunnel's exit in Figure 1 which takes a truck about one second at normal highway speed. The camera is adapted to the relative darkness in the tunnel making the white balance for outdoor sunlight overexpose the landscape outside the tunnel. One way to overcome this problem would be to improve the camera (or have multiple cameras) able to cope with varying exposure. The lane markings are clearly visible inside the tunnel, but from an image classification perspective there are many edges of high contrast inside the tunnel that a typical lane marking detection algorithm would pick up instead of the correct lane marking. An example is the concrete wall at the base of the tunnel wall. Guard rails is a common road side construction and is easily picked up by a vision system instead of lane markings. High contrast can also be

formed by sharp shadows on the road. In short, varying light conditions is a source of performance issues of the perception, forcing a subsequent controller in the loop to be (temporarily) detuned causing a loss of precision and robustness of the lane following. Another typical problem is partially missing or obscured lane markings because of other objects, snow, dirt, dilapidation of the paint, etc.

Now, imagine we add a lidar which uses active light emission to sense the distance and reflection intensity in a discrete set of point in an angular space coordinate system. The lidar is largely unaffected by the sunlight. From the resulting cloud of points output from the lidar it is relatively easy to see where the flat road is and where obstacles, such as rail guards, pushes up from the ground. Without knowing how to fuse camera images with lidar point clouds, this reasoning provides evidence that a (mono) camera system can be improved.

4. Purpose, research questions and method

In short, the main purpose of the PRELAT was to improve the lateral control of a level 4 automated truck in terms of robustness, that is, avoiding temporary misinterpretations of the highway.

The research question was: how can we add other sensor modalities and how can we quantify the improvement. The choice was to evaluate lidar point clouds and to combine it with camera based vision.

The main method was to study perception as a component in order to evaluate its performance. The technical approach has been supervised learning, where a human annotates features in an image, e.g. lane markings, free space, cars, etc.

5. Objectives

The goals of the project were formulated to be: 1) Improve definition of logical control architecture for lateral control and 2) Generate redundant sensory modalities for robust lateral control. At the end of the project it can be concluded that we have worked a lot more with goal number 2. The proposed architecture does not differ much from traditional architectures in robotics. We have looked at usage of different sensors when it comes to prediction and path planning. We have looked at selection or omission of information sources (variability of sensors). Any resilience to omitted source is however not demonstrated in vehicle. We have concentrated our effort on nominal driving performance where all sensors are available.

When it comes to the actual precision of later control position, we have not worked on this question. In order to seriously address this issue, we would need a truck to make comparisons with simulations. Instead, we have focused on the perception itself. It was noted when we drove our demonstrator in the AdaptIVe [ADAPT] project, that the vehicle lies very straight on the road at all time. It is questionable if this is necessary at all.

An objective was that models and methodologies for sensor fusion and vehicle control will be improved. The project has proposed architecture for perception but not so much for control. The reason is that we have concentrated our efforts on the parts that we think makes up most of the problems (with disturbances and omissions), and that is the perception system. In the perception system we have worked with camera, lidar and also to some extent IMU and GPS. We have studied a variation where training and inference (usage) is based on camera only, or training and inference is based on lidar only and where training and inference is based on a combination of lidar and camera. Since lidar (still) is a costly device, it makes sense to use camera only in production vehicles, and for this we have also looked at the combination where training is improved by a combination of lidar and camera but the inference is made with the camera alone.

Functional safety has been address in the project. The ISO 26262 standard has been considered and found in need of including topics more related to automated driving. A main reason is that the interesting error model is more related to performance and robustness interpreting the world, rather than to faults affecting any computer control system.

Demonstration of the developed lateral control in a vehicle was promised but not carried out. A reason is that evaluation of learning systems is better done at desk top on a good data set, and we did not have any dataset for our own vehicle until very late in the project. A demonstration at that point would not have contributed sufficiently compared to the cost.

The so called end-to-end approach, an imitation learning where sensor input are trained against signal given by human driver demonstration, was discarded because it was not modular, it would have been difficult to draw conclusions from it, and human driving is maybe not the best teacher.

As a prerequisites of the project it is assumed that lateral positioning is not constrained to be relative to a pre-recorded map of a fixed route. With such a map it is possible to make a feature extraction pre-runtime and then drive according to these features at runtime. For example, instead of camera or lidar, it is possible to use ground radar that penetrates a few meters down into the foundation layers of the road, making feature mapping independent on road conditions. An HD-map, which have road and road side object data with high accuracy as opposed to traditional navigational map, is useful because of its general information contents.

There are a number of other information sources (sensors) that have only partly been considered. Traditional GPS solutions fail to deliver the accuracy and robustness needed for lateral positioning but can be used for initial map matching. High accuracy GPS systems have been discarded for price, availability and robustness reasons. Information over the air (5G, WiFi) from other vehicles, cloud or road side units (V2X), have not been considered in the project either. These non-reliable communication channels will help automated driving by transponder information, collective perception systems (CPS), system that conveys intention of other traffic participants, etc. Lateral control is however essentially a control loop local to the ego vehicle.

The technical approach has been supervised learning, where a human annotates features in an image, e.g. lane markings, signs, pedestrians, cars, etc. Annotating data ourselves in the project was not seen as a viable path, we have used the publicly available KITTI dataset [KITTI]. An alternative would have been to use simulators.

6. Results and deliverables

This section brings up results presented in a top down description starting with architecture, then getting deeper into the three major parts of the control loop: sensor fusion, path planning and lateral control, and finally ending with safety aspects.

6.1 Architecture for L4

The architectural work is concentrated on the interface and behavior between sensor fusion and trajectory planning when it comes to taking lateral actions, primarily to follow the ego lane. The aim is not to suggest a complete architecture for ADS – there has been many such suggested in the literature. Safety aspects is an important part which is treated in another section.

The structure of this section is top down (with models and requirements) from use case combined with assumptions and ideas over conceptual design and landing in suggestions for how components should fit together.

6.1.1 Use cases and User needs

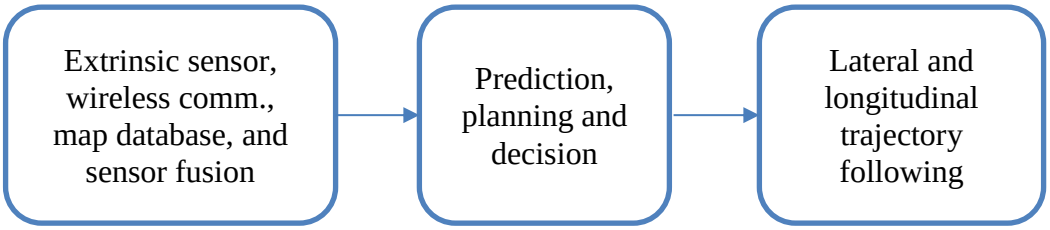

Figure 2 Initial model of the autopilot with perception, planning and actuation; a top level structure of components for driver automation. The flow of data is predominantly from left to right. Prediction and planning are lumped because they are conceptually similar in timeliness and information contents.

The *use case* in this project is lane following on highway for an L4 [J3016] truck. In the following we aim for L4 rather than L3. The reason is that automation passivates the driver [HAN13]: “If you build a system where people are rarely required to respond, they will rarely respond when required.” Given a truck, its L4 system is divided into three subsystems: perception, planning and actuation, see Figure 2.

This is a linear use case: 1) A truck is manually driven up onto the highway. The highway can have any number of lanes in the same direction. Entries and exits are not crossing the

highway. The minimum radius of a lane is bounded. 2) The auto pilot is engaged. This enables the driver to remove feet and hands from pedals and steering wheel. 3) The set point for the lateral position in the lane is assigned either by the driver or by the auto pilot. The speed set point is dynamic given the driver's request, truck constraints and the legal speed constraints. 4) The truck automatically overtakes a vehicle by two subsequent lane changes coming back to the starting lane. 5) Before leaving the highway, the driver takes back direct control over the vehicle.

A *user need* is a more formalized textual statement (requirement) than the example a use case (model) story telling. Several user needs could be derived from the use case above, but since the aim is not to design a complete system, one user need is sufficient: "The lateral control of the truck shall be robust against disturbances that enters through its perception system." The art of the disturbances are related to perceiving the road surface and the lane markings and make a proper classification.

6.1.2 Assumptions, guiding principles and delimitations

There exists many architectural proposals for self-driving robots with ADS as a special case. An architecture is typically defined by a number of views and perspectives such as class diagram, sequence charts, deployment on hardware. In that light, the proposal here concentrates on the information view. Prior to a conceptual design, below is listed a number of guiding principles and assumptions (which are not universally true) on what is a good design.

Layered and modular architecture. The choice of layered and modular is pursued since it really the main stream approach and has a number of advantages. Invocation and communication methodologies supported by any operating system are out of the scope. Specifically we have discarded end-to-end approaches. The main reasons why end-to-end is not advocated are that control lends it well to model based design, which handles variants naturally, and also that little is learned from making an end-to-end system since it is difficult to evaluate other than just end-to-end.

Hardware and software constraints. It is assumed that real-time processing at run-time is made on-board the vehicle. Redundant sensors is costly and it is therefore unlikely in the long run that a sensor will be used exclusively to satisfy a safety goal, however this is more an observation than a goal in itself. Further, it is assumed that the cost of wireless communication is high compared to processing and that wireless communication is not reliable. Since the project does not have any further constraint here, deployment within the vehicle is not discussed.

Choice of technology. As shown in this report, CNN is beneficial for image classification of lane markings. However, there is nothing that suggests that neural networks are advantageous in performance compared to traditional control algorithms when it comes to trajectory planning and following, since these are totally different problems. The

advantage with traditional control theory, such as model predictive control (MPC) is the model based parameter population and the ease of understanding, ease of tuning and to some extent safety assessment.

Information uncertainty. A notion of how self-confident a module is needs to be conveyed to its users/consumers. This is nothing new: a typical pattern in process industry is to add information on the validity of a current signal's value. A characteristic of automotive perception using externally looking sensors is the high degree of uncertainty of the result. Uncertainty can for example be modelled as a probability distribution with a mean value representing the nominal variable. A more problematic type uncertainty can be characterized as sporadic omissions, or sporadic larger deviations, caused for example when the visibility of lane marking suddenly becomes bad. To account for these "worst cases" the tail of the probability distribution need to be long.

We are now ready to proceed to from the user need to the system requirements, starting with a conceptual drawing of the very top level architecture.

6.1.3 Conceptual design and System requirements

The top level Perception-Planning-Control can be broken down into subfunctions. A typical stucture is found in Figure 3 [BEH16]. A similar conceptual model is found in for example [WARD18].

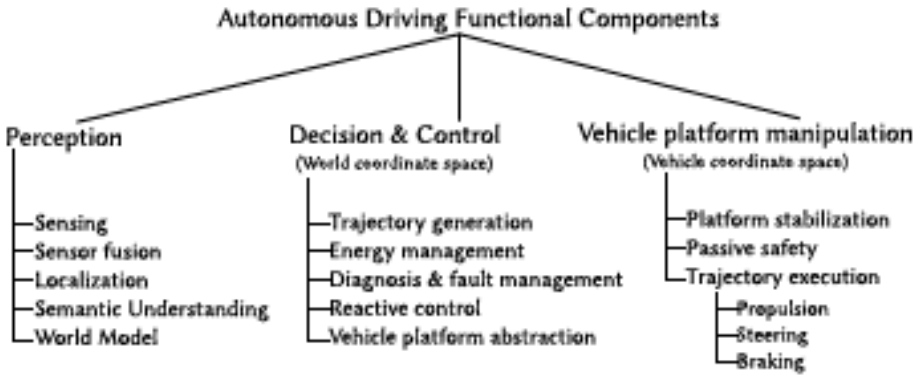

Figure 3 Subsystems hierarchically belonging to respective top component. (Diagram taken from [BEH16])

As an inspiration, [DRE18] has proposed an architecture which is inspired by how the brain works. This architecture involves both run-time processing and training for continuous improvement by defining flows similar to information flow between lobes in the brain. The architecture is nevertheless far away from ideas of how the brain works,

see for example [FREE], a notion for self-organization of the brain related to learning and Bayesian inference.

From a map database, preferably a high definition one, given a position in a suitable coordinate system, the type of data includes: lane structure, lane geometry, road type, lane marking type, drivable area. From ego vehicle sensors and other sources of real-time data, at a high level description the world model includes road condition, drivable area, lane geometry, lane marking type. The (semi) static map data and the real-time data are then fused (or compared) via map matching to make up an enhanced world model for trajectory planning. There could be an advantage of using map data as an input to sensor processing to have a prior on location of static objects such as lane markings, this has however not been investigated in the project.

Leaving the structure of the architecture and concentrating on the interface, we understand that it is important to convey both capabilities and confidence between components along with measurements, set points and other control related signals. This is a first system requirement.

Handling of uncertainty in the lateral control is another topic. In the field of robust control theory, uncertainty of for example sensor reading is modelled and used as an input to the control synthesis. A typical drawback with robust control theory, is that the resulting control laws typically becomes very conservative when taking a wide range of uncertainties into account. This means that the control becomes slow and sluggish. From a performance point of view, it pays off to model a precise amount of uncertainty that is valid in the current context. In that case, other control paradigms can be used as well, since we can always switch between two controllers with different dynamics and characteristics when needed.

Luckily, there is another way to decrease uncertainty. A typical driver response to uncertain lane markings or a narrow driveable surface (in the lateral direction), is to adapt the speed to the current context, that is, an increased uncertainty in perception pertaining to lateral control is not primarily to become conservative in lateral control but to counteract by changing a major contextual factor; in any case, the safety margin will increase. The second system requirement is that a structured way to reason about uncertainties of the top level chain (perception-planning-control) is needed. A system that is self-aware of its uncertainty down to traffic specific harm/risk.

One of our main problem is to understand the world and make decisions and be confident that we can control the vehicle to our decisions. Being self-aware of the uncertainty of the world model we make is a type of analytical redundancy. A tool to structure relations between probabilistic world entities (for example that the current road-tire friction has a certain distribution around a most likely value) is Bayesian network (BN). Bayesian networks is model of conditional stochastic variables that can also be represented by a graph. Nodes represent joint distributions of stochastic variables and directed edges represent relations between nodes. By assigning relations between stochastic variables,

we can break down an initial high combinatorial complexity joint distribution into conditional distributions which is easier to understand and work with.

With a Bayesian network approach we can make a model of the world and introspectively of the ego vehicle to handle information uncertainty. Inference queries can be answered through these relations. For example, what is the maximum speed the vehicle should have given this uncertain friction estimation not to cause any harm of some minor degree and, not the least, given a number of other uncertain observations of the world around us. A third system requirement is to implement a structured self-analysis of the world model to estimate confidences and capabilities. An example is shown in the section on safety of ADS. The sources of information have different quality; ego vehicle sensors are perhaps more reliable than external information over wireless communication. With Bayesian reasoning, also a source with uncertain reliability can issue a warning of an extreme condition that needs to be considered. If all vehicles had collective perceptions systems [ETSI] that would likely improve overall confidence, since more information is brought to the data fusion.

Lastly, we come to safety of embedded system when the error assumption models faults leading to failure communication and processing devices, such as power loss, soft errors, logical code errors, etc. These problems are typically handled by adding redundancy within the vehicle. A cost efficient redundant architecture is proposed in [ADA17]. It addresses functional safety and is modular in the main components: sensor, data fusion, driving function, motion control and motion actuation. It is essentially a so called duo-duplex system for x-by-wire. This architectural pattern for cost efficient redundancy is a strong candidate for any type of L4 system. A duo-duplex system is divided into the hierarchical subsystems “Main” and “Backup”. The “Main” subsystem is the active channel and the “Backup” subsystem is the passive channel, see Figure 4.

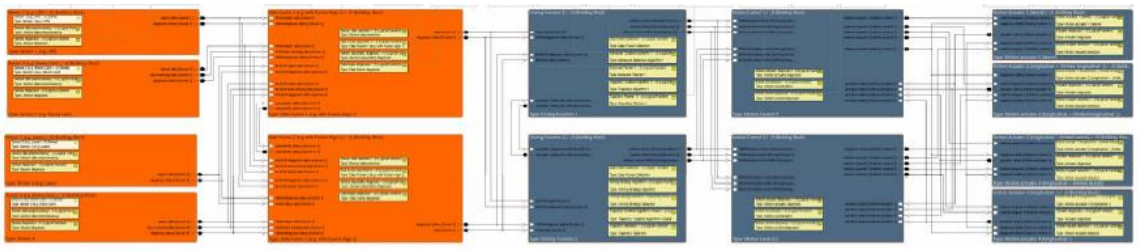

Figure 4 A cost efficient functional safety setup. First and leftmost column is sensors, second column is sensor fusion, third column is decision and control, fourth column is path execution and fifth column is actuators. Upper box in each pair/column is primary subsystem and lower box is corresponding backup subsystem.

The duo-duplex architecture has the following advantages: 1) the architecture supports fail-operational with high reliability, 2) no additional voter is necessary, 3) reliability is as high as a triple modular redundancy, 4) cheaper than a triple modular redundancy architecture, 5) low synchronization effort of the parallel subsystems/paths, 6) different scalability of the parallel subsystems/paths is possible, and finally it also 7) reduces systematic faults if subsystems are developed independently.

6.2 Sensor fusion and perception

This section starts with a problem description, continues with a preamble on fully connected neural networks and concludes with road detection work done in the project.

6.2.1 Road detection problem visualized

A lidar is an active sensor that emits light and detect reflections from it, and thus independent on the lighting condition, which means that the interpretation of a static scene should become similar when lighting condition changes. The camera is a passive sensor where a changing lighting condition can make the road detection task difficult. Precipitation, dirt and other objects can also make road detection more difficult. Figure 5 shows why road detection using a (mono) camera is difficult, and it helps explain why a lidar is useful.

Figure 5 Same place and direction but at different times of the day. What makes road detection challenging is the lighting conditions. (Images from Oxford robocar dataset)

Lidars can detect intensity of reflections (of asphalt versus paint), but cannot see color since they operate at one wavelength. Depending on the number of layers of the rotating lidar device, the returned lidar point data cloud is sparse also when not projected into another angle, e.g. top view. The angular velocity of the rotating lidar used here is about 10 Hz, which for a moving vehicle makes two consecutive 360 degree scans not very similar.

6.2.2 Supervised learning, annotation and semi-supervised learning

Artificial neural networks (ANNs) is a family of computation architectures inspired by biological neural networks. ANN found very little application, or interest outside the research community, until about 2010 when it was discovered image classification could be improved substantially compared to traditional method. The solution had many connected (hidden) layers intersected by non-linear activation functions, and to train and execute the network which had so many parameters, hardware acceleration was deployed. These deep neural networks (DNN) have become ever larger people in general are trying to understand where this technology makes best use. A problem with DNN is that fully connected layers (where each unit in a layer is connected to every artificial neuron unit in the preceding layer) becomes very large. CNN, convolutional neural networks, is a way to decrease the size of the DNN based on biological studies of the brain's visual cortex, which suggested that local connectivity and hierarchical layers might be suitable for image analysis. A typical structure that falls natural to CNN is the encode-decoder structure. The encoder consists of a successive use of convolutional layers and max pooling layers which condenses the representation of the input and also provides translational invariance. The following decoder part inflates the condensed representation by unpooling optionally to a final set of fully connected layers – these are generally called FCN, fully connected convolutional neural networks. When training a FCN an optimal solution is typically searched for in the back propagation by using stochastic gradient descent, first splitting the training data in a number of batches. A problem is overfitting, that is, the network does not generalize to unseen verification data in a good way, but learns unnecessary features of the training data. A drawback with supervised learning is the need for labelled (annotated) data which is labour intense to carry out.

Much of the work in PRELAT has been based on the open dataset called KITTI, see [KITTI]. This reference and benchmark data set, originating from 2013, has been very valuable. There is high score board for practitioners to compare contributed approaches. The score list uses the measure maximum F1 score, MaxF, to explain accuracy of results, see for example [WMAXF] for an explanation. The higher F1 the better, and absolute maximum is 1.

A drawback with supervised learning is the need for labelling (annotation). The labelling is costly (time consuming) since it is made by hand. An idea explored in [CAL19b] is semi-supervised learning for the case of multi-modal input. It is shown for road detection using camera and lidar, see Figure 6, that for the same size of training data set, it is possible to improve the F1 score with some percentage points by using a much larger unlabelled training set.

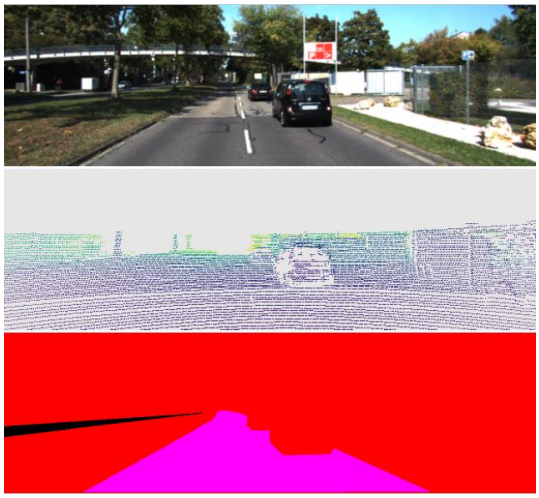

Figure 6 From top to bottom: camera RGB image, lidar height coordinate in cameras view, and semantic segmentation of the road's ground truth annotation

This is an interesting result and very practical result when it comes to robustness, since it opens up for less costly annotation used in training.

6.2.3 Road detection using camera and lidar

The PRELAT project has studied the usefulness of lidar for road detection, which is reported in [CAL17a]. A fully connected convolutional neural network was used on point clouds which were preprocessed to 2D top view representations. Since the KITTI dataset was used, image annotations were also transformed into a top view for the supervised training. Road detection is here defined as a pixel-wise semantic segmentation of the lidar top view, that is, each pixel of the top view lidar image is classified as either road or not. The top view images were spatially resampled into squares with 1 decimeter side and statistics about each square was used as training data. Occupancy images where pixels represents either an obstacle such as a curb or not, was also evaluated with promising result.
A hypothesis was that making a bird eye view projection would be an advantage since in that perspective, the area of ground covered by one pixel or one ray does not vary depending on the distance from the vehicle. Thus, the network does not have to learn different scales. However, it seem that the difference in performance of the two different projections is small.

Figure 7 shows three examples of resulting pixel segmentation of road or not road.

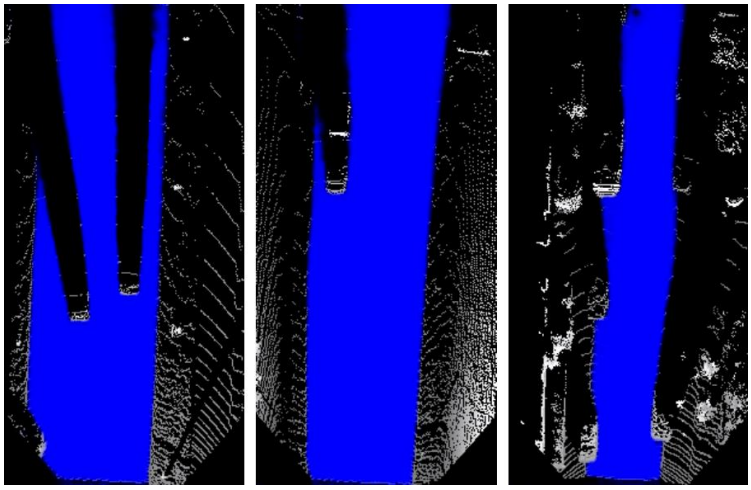

Figure 7 Road detection generated by the proposed FCN. Blue pixels correspond to high probability of road

Figure 8 shows three other examples (pictures originating from [KITTI]) where the segmentation made by lidar image is converted to camera view and superimposed on camera image. Not that false negative classifications are typically found in places where you would not drive your vehicle.

Figure 8 Examples of road detection in images view. Green denotes true positive, red and blue correspond to false negative and false positive respectively.

To use FCN with lidar was new at the time the paper was written and it scored highest on the KITTI bench mark board. The usefulness of the result lies in the fact that a lidar actually contributes to road detection as compared to using mono camera, that is, lidar

should not be neglected when choosing a set of sensor for a vehicle. The on-board execution time of the neural network, in a typical hardware accelerator, is not an obstacle.

A question arises how input from multimodal sensors are best fused. This was addressed in the project and reported in [CAL19a]. Three variations of fusion are compared: early fusion, late fusion and cross fusion, see Figure 9. In the cross fusion, cross connections between a camera branch and a lidar branch are initialized to zero, which corresponds to no connection, and thereafter trained to establish connection data path between the two sensors.

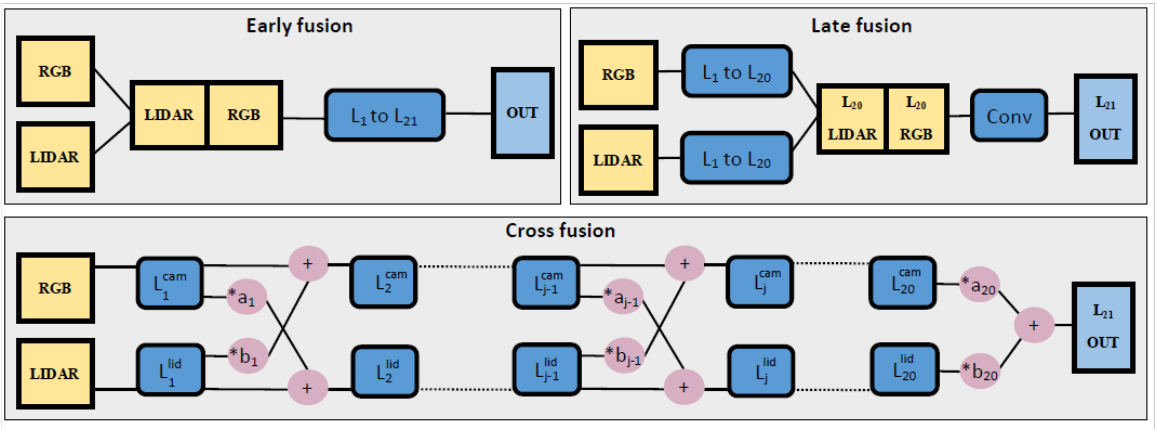

Figure 9 Three different fusion strategies. 1) Early fusion: camera and lidar are concatenated in the depth dimension rendering a tensor of doubled size; this derived sensor is then processed by the base FCN. 2) Late fusion: camera and lidar data are processed in parallel streams up to layer 20, then concatenated in the depth dimension followed by a convolutional layer. 3) Cross fusion: starts with two parallel streams which are interconnected between layers by trainable scalars

It turns out that the cross fusion works best, when comparing the alternatives: lidar only, camera only, early fusion, late fusion and cross fusion. To get a more challenging data set, a set of specific images that was judged to be difficult for a camera only was extracted from the KITTI data set, see Figure 10. For this set, the lidar only performs better than the camera only, but the cross fusion beats all other variants. An important take away from this piece of research is that lidar should be used in order to acquire robust road detection in more difficult lighting scenarios.

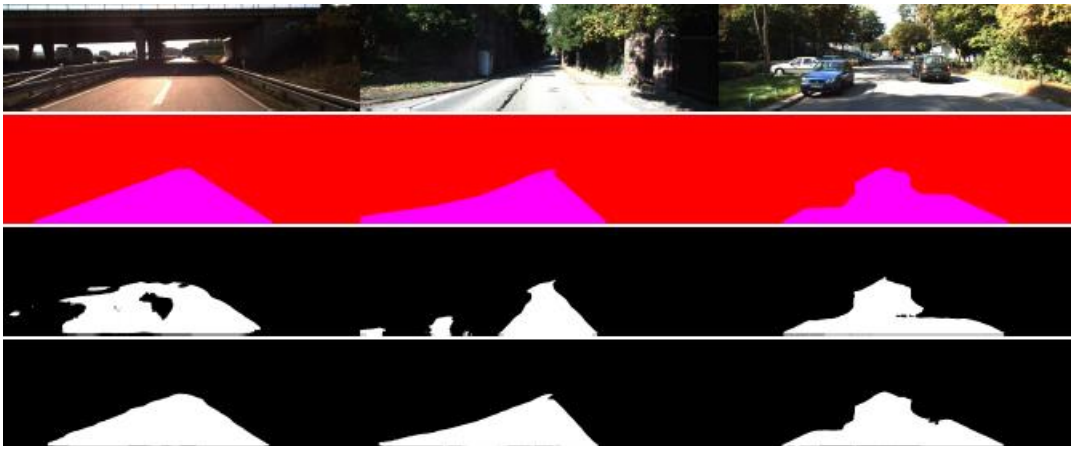

Figure 10 Three examples of difficult lighting conditions. Second row shows respective ground truth, third row shows road segmentation by camera only and fourth row shows cross fusion FCN

6.3 Lateral trajectory planning and following

The aim of this section is to complete the loop from sensing and perception over planning to actuation. The results comes from PRELAT but also from a related project on path following with a truck on running the road.

Trajectory planning and trajectory prediction of the ego vehicle are related. One difference is that planning implies that a higher level reference exists prior to finding a trajectory. For prediction of surrounding traffic, the intention governing the trajectory is largely unknown, especially for manually driven vehicles or vehicles without V2X. The intention describes what directions to take when there are several options available, for example at intersections. The intention becomes a set point to follow for trajectory planning. A human driver usually has intentions not only to go from A to B but also intentions at a smaller scale in the current traffic situation.

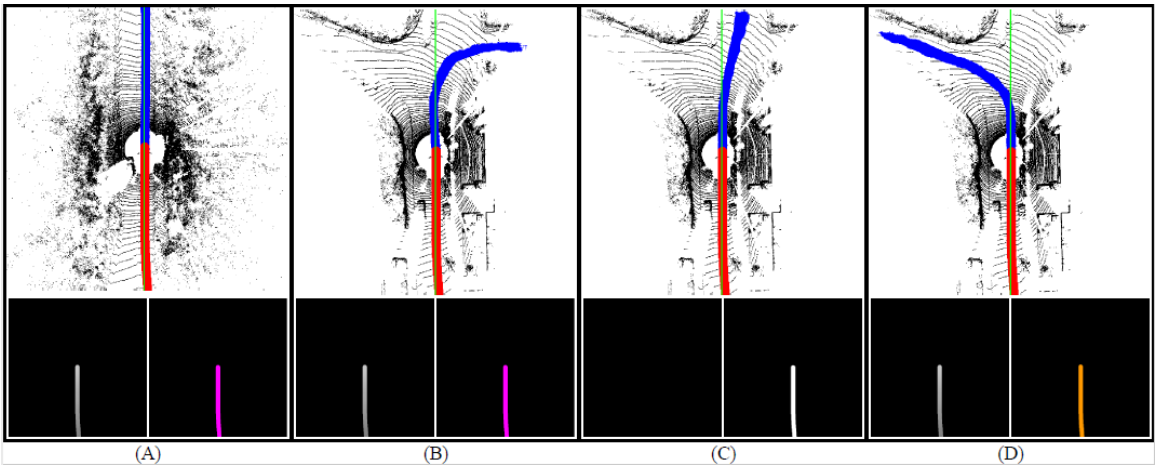

Figure 11 Each pair of bottom panel for column A to D show driving intention proximity (left) coded with intensity and driving intention direction (right) coded with magenta for right turn and orange for left turn. Each lidar map for column A to D shows historic path (red) and predicted path (blue). Column A shows an example where there is disagreement between requested turn and actual prediction. In columns B to D shows the same location where multiple directions are possible and the resulting path prediction in case of three different driving demands

One way of performing planning is by imitation. Imitation learning by Fully Convolutional Neural Network, FCN, is explored in the PRELAT project [CAL17b]. As shown in Figure 11 the intentions are color coded into the convolutional network by the colored historical graphs in the bottom row. Intention of the driver to turn is encoded by a tuple {left, straight, right} and the approximate position (proximity) when the intention should be activated. By this, the FCN is trained to associate {left, straight, right} with examples from historical paths. It should be noted that "straight" means to follow the road, while "left" and "right" divert from it. The accuracy of the prediction is improved by adding other sensor modalities such as IMU (yaw rate and acceleration) and GPS, but without supplying the intention the FCN produced uncertain predictions. Two nice results from this work are 1) knowing the driver's intention improves prediction and 2) this high level set points can be set as an image input to a convolutional neural network.

Planning can also be done by model predictive control (MPC), were a solver is called at runtime to find an optimal solution. In work carried out in collaboration with the PRELAT project, lateral and longitudinal path planning is made for an overtake scenario with traffic, [DOV18]. This requires decision making on a target lane and calculations of smooth paths. Planning is separated from perception; an input to the planning is of course driveable road surface and the location of lane markings. The intention of the ego vehicle is based on the objective function which evaluates a set of high level choices, that is, stay in lane or change lane either left or right. The simulations showed great success in studying different initial settings and driver behaviour related tuning parameters.

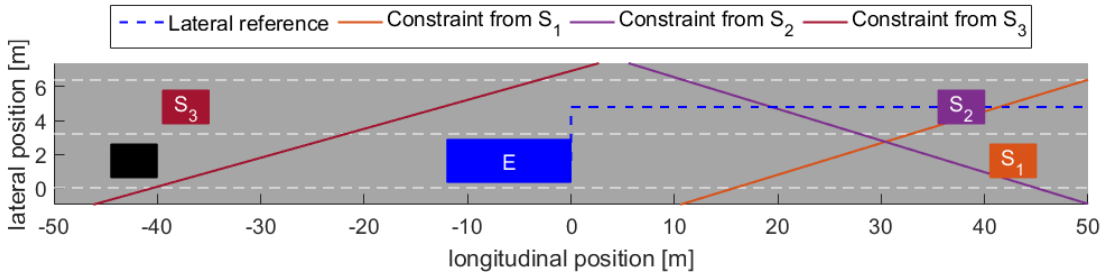

Figure 12 Constraints of ego vehicle E when intention is a lane change to the left. Vehicles move from left to right.

Figure 12 shows the setup with ego vehicle (truck) E, and surrounding vehicles S1 to S3 that constrains lane change from right to left. The trailing black vehicle is not taken into account. Figure 13 shows a typical lane change re-plan succession with this setup.

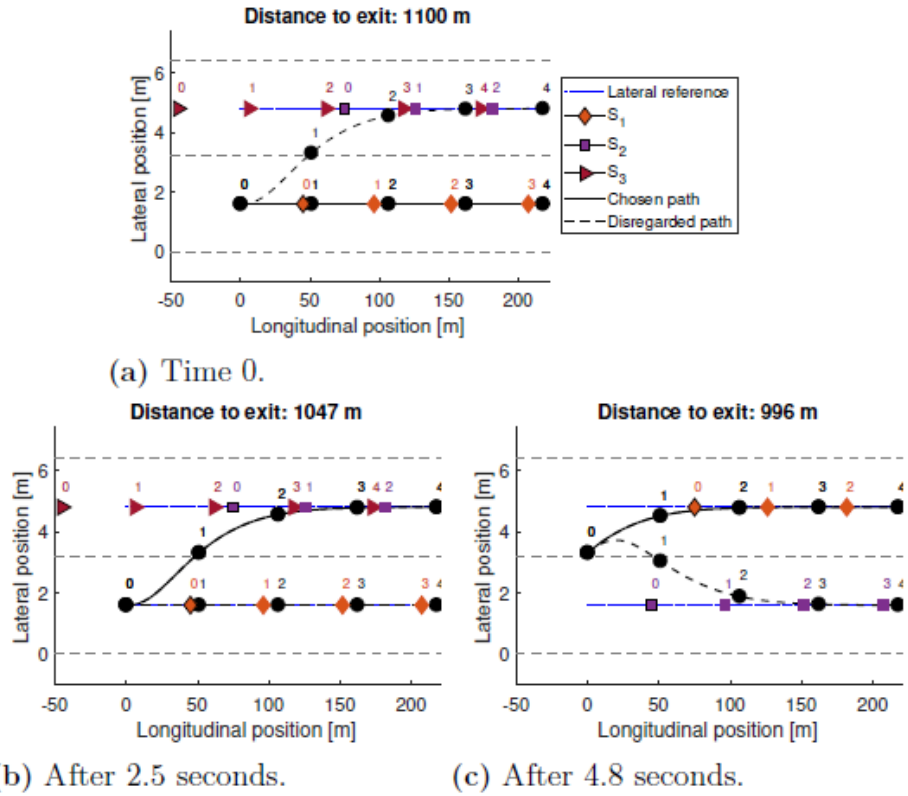

Figure 13 Planned trajectories shown at three different instants (a) to (c) during the execution of a lane change to the left. Dashed line is alternative path, one that is feasible but not selected in the decision making

A lateral control for path following, based on state feedback, was developed in the AdaptIVe [SKO17] and used in several projects, including understanding requirements for PRELAT, see Figure 14 from inside a truck. The lateral error is the measure of the reference path including any offset within the lane (see below) minus lateral position. The

perception system of the forward looking mono camera gives, from a vehicle fixed coordinate system, the distance to left and right lane marking including a description of the curvature of the road. This gives both the lateral position and the heading. In addition to this there is also a notion of uncertainty of the current sensor reading.

The problem of robustly detecting lane markings is quite separated from following them. A single mono camera system will typically mistake a guard rail for a lane marking especially when the nearby expected lane marking is difficult to see. The lateral position will jump in an irrelevant way, making it unsuitable to continue control on the current measured value. The lateral control cannot work better than what the perception can deliver, or put it another way, the best option to improve this is to work on the perception. A result for PRELAT was that that robustness issues do not arise from the lateral control but from sensor disturbances, and the best way to improve control is to reduce those disturbances or robustness issues of the perception system.

Figure 14 A golden truck at a test track to evaluate lateral control. Right corner: top view of front right wheel.

Following a line exactly between left and right lane marking, assuming parallel lines, is not every driver's choice. There could be several reasons not to, even at highway with high curvature such that cutting is unnecessary. For example, if the lane and also the road shoulder are wide, driving with an offset towards the road shoulder is an option. Another reason to stay off the centre line is that the road surface is dilapidated by tyres, having two parallel tracks that do not fit the width of a truck. A lateral offset positioning function [SKO17] has been implemented and tested in on road (E6 around Göteborg). A path planning, or path generator, based on a single mono camera was designed and used successfully in a highway auto-pilot project and several platooning projects. The lateral reference value is set manually by the driver by pushing two knobs at the steering indicator any time the lateral control is active, see Figure 15.

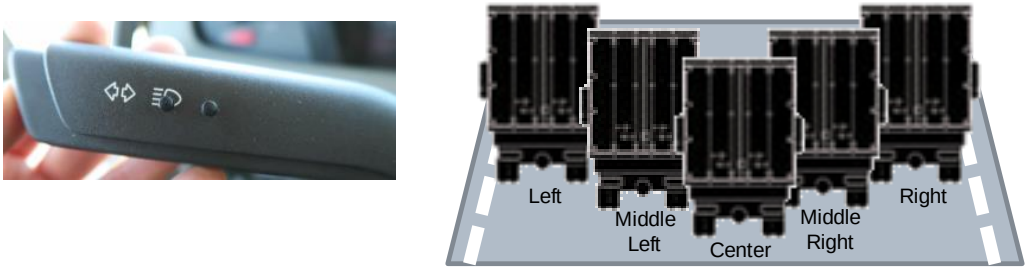

Figure 15 Select one by buttons on the blinker stick (left) one of five inter lane lateral positions (right)

This gives a discrete number of offsets from the centre line and it was found that five steps was enough. Another intuitive way of setting an offset was also implemented: when the lateral control was switched on, the current (filtered) offset was used as reference. Both these ways of setting lateral control set point are analogous to how speed set point is handled in today's cruise control feature. Selecting a god relative position within the lane should be done automatically by an ADS.

6.4 Safety of ADS

Safety is of highest importance when it comes to ADS. This section starts by outlining the relation between functional safety and safety of intended functionality. Since deep neural networks are deployed, a relevant question is: do they bring any additional problem in terms of safety? A literature study of both soft errors and of adversarial attacks on deep neural networks ends the section.

A common terminology for automated driving that is established in [J3016]. A main construction is the ODD, operational design domain, which defines a space that an ADS should stay within. If it is not possible to be within the ODD, the automated driving should not start. If the driving already has started, it should be terminated before leaving the ODD. Constraints in the ODD are typically related to the road, traffic, weather and other environmental conditions.

6.4.1 Error models

To analyse harm caused by a vehicle system we first have to make a model of how the fault enters the system. This section treats the error models of two types of safety issues of an ADS. Besides the well-known functional safety standardization in ISO 26262 there is a not yet mature trial standard ISO/PAS 21448, also called SOTIF – Safety of intended functionality in. Unfortunately, the names "functional safety" and "safety of intended functionality" are a bit misleading.

How can we make the behaviour of the ADS sufficiently safe?

- The problem of sotif can be cast as follows: Assume an embedded computer system with a number of uncertain and partly interdependent information sources as input to perception, decision and control.
- The problem of functional safety can be cast as follows: Assume an embedded computer system subject to (independent) failures of its hardware or software.

Sotif targets automated driving, whereas functional safety has a broader scope; it is more targeting safety analysis for embedded computer systems in vehicles. Functional safety is always relevant for a vehicle but sotif is even more relevant for ADS since it more efficiently addresses robustness issues of the chain perception, decision and vehicle control. An ADS must be safe even when it does not fail in the sense of the functional safety standard. Since "faults" are entering via sensors, we can constrain the problem by writing an ODD to avoid being exposed. As an analogy, the ISO 26262 does not include e.g. fire in computer components, but it does include subsystem failure due to whatever reason, for example fire, which can be met by redundancy. Thus, the standard is strictly speaking possible to use to address fire but not efficient – another method or standard should be used.

6.4.2 Safety of intended functionality

The message in this section is that safety is a difficult thing to prove when it comes to automated driving on public roads. It is often stated that one of the reasons to introduce automated driving is to increase safety by avoiding human errors. It is a well-established idea that when a new technical solution replaces an old one, it must be considered at least as safe.

A process goal of sotif is to increase the volume of known safe scenarios while decreasing both the known and the unknown unsafe scenarios. The standard outlines a work process to meet this and to document the argumentation that the system is sufficiently safe.

Sotif targets failures of the embedded computer system to correctly perceive, decide and handle the environment. The failure is manifested because of an unanticipated condition or event in the environment that is not masked or cancelled out. Archetypical anomalies are sun glare into camera, misclassification of motorway bridges as obstacles, and misunderstanding of driving behaviour of surrounding vehicles, etc. Those are well-known challenges when it comes to driver automation. In this context, the type of error model is different compared to that of functional safety, where a fault is manifested as a error within the embedded computer system. An efficient solution, in terms of redundancy or additional constraints, might also look quite different. A vital part of a solution is to make the environment more homogeneous, for example to constrain operation to highway driving at daytime in high visibility weather conditions.

The universal method to decrease the number of fatalities and wounded in road traffic is to reduce vehicle speed. What work in a statistical sense also works in the individual case – with decreased speed, the time available increases to understand the world, make decisions, and to effectuate the manoeuvre. On roads with high speed the environment is more homogeneous in order to make driving easier; obstacles have been removed and it is easier to spot problems at a longer distance. A top view answer to handle robustness for automated driving is to lower the speed and to make the environment more homogeneous. Of course, massive analysis, simulation, and road testing and driving 11 billion miles [KAL16] will also improve the safety. There is a big difference between demonstrating a feature during a short period and deploying the same feature to work under varying environmental circumstances.

An instructive example of a perception/decision system unable to interpret its environment sufficiently, is found in the accident in Tempe [NTSB19], were a woman leading a bicycle across a driveway was hit. The victim was not classified correctly and classification shifted multiple times, resetting its history. The unknown object was not assigned an intention and thereby not tracked and its path not sufficiently predicted. This type of object, its location and its behaviour was not anticipated in the design. From none of the two parts involved in the accident there was any action taken in the absence of the other parts fault.

To be more specific for the PRELAT project, we can model disturbances and uncertainties with the help of a Bayesian network. The nodes describe joint probabilities and edges denote relations. As an example of road detection, see Figure 16.

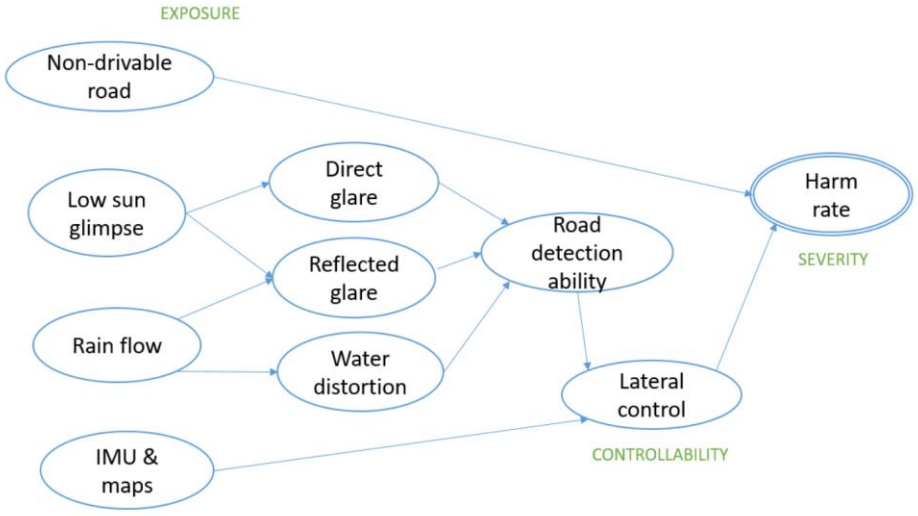

Figure 16 Bayesian network to analyze classification of drivable road in presence of typical disturbances with origin in the environment.

The Bayesian network is a high level model of ADS that gives a systematic approach to reason about risk (rather than hazards). It is possible to add probability distributions to

every node. It is then also possible to calculate expected values and variances. Queries can be posed, such as: given a specific Harm rate (risk), how well do we need to detect the Rain flow in order to stay below the Harm rate for a given confidence level? A difficulty here is to map the ability of the sensors and sensor fusion.

An extension of the above is to add different types of sources, for example prior statistical knowledge, weather forecasting, more ego sensors, and collective perception systems.

6.4.3 Functional safety: soft errors of DNN

With DNN comes a renewed interest to study the resilience to bit flip errors (also called soft errors). This is an error within the computer system and the thus falls under ISO 26262. This standard was written before DNN became ubiquitous in automotive safety critical perception, so there is no specific mitigation strategy mentioned to account for a related safety goal. First, it is of interest to understand how bit flips affect the outcome, and what similarities and differences there are to traditional general-purpose software and processing hardware.

Cosmic high energy neutrons and alpha particles pose an important reliability problem in modern general-purpose processing units as well as graphical processing units (or other specialized hardware accelerators typically used for neural network inference). A single bit flip can propagate through the system and, if unmasked or unmitigated, cause system malfunction. The error rate increases for each new generations of hardware having more transistors per chip and narrower line width. According to [LI17], the failure in time rate (FIT) of accelerators 2019, can exceed safety standards in ISO 26262. A search for particularly vulnerable multiple bits is done in [RAK19]. As an example, flipping 100 random bits decreases classification accuracy in a ResNet-18 (93 million bits) by an average of 1% but in worst case flipping 13 specific bits rendered an accuracy of practically zero. In order to become resilient to bit flips any correction mechanism must not be expensive in terms of hardware or energy consumption. Traditional methods rely on hardware redundancy but more cost-efficient end-to-end approaches have also been suggested, see for example [RAK19] who combines software detection mechanisms with hardware exception.

DNN accelerators have a different structure than general-purpose processing units and the distributed parallel execution and local caching make general low-level protection mechanisms costly. A set of publicly available DNN has been investigated in [LI17] using fault injection techniques. It was found that the topology of the network (the number and type of layers) mattered and in what layer the error occurred. For some network topologies a single error in an early layer was probable to be masked, not affecting the output, whereas for some topologies there was little difference what layer

the error appeared in. Since the error is application specific, the failure rate of a classification application can depend on the number of classes (and layers with only a few neuron outputs). Also, in common with general-purpose software, the data type used and the bit position within the data type matters. Floating point data types (IEEE 745) can have higher dynamic range and are thus more vulnerable to bit-flips in high-order bits compared to fixed point data types. Bit-flips in low-order bit positions are typically masked away due to its insignificant change in value.

Several methods to decrease soft error failure rate in DNN is proposed in [LI17]. A resilience optimization is proposed in [SCH19]. Based on a criticality measure of one layer, weights in a subsequent layer are modified to equalize (partially mask) a bit-flip. The weight adjustment degrades the network's classification accuracy and hence a retraining is performed. It is shown that the failure rate is significantly decreased with no or minor inference execution cost.

6.4.4 Adversarial attack and classification confidence

Deliberate attacks to make classification (or segmentation or detection) fail is called adversarial attacks, see [QIU19] for an overview. It is not a hopeless case and countermeasures have been proposed. What is interesting is that it reveals some properties of the neural network's structure and training.

One type of adversarial attack involves generating slightly perturbed version of the input data that will fool the classifier. What is surprising is that a modification can be imperceptible to the human eye. Another type, and a more realistic one for automatic driving, of adversarial attack is to add stickers to lane markings, traffic signs, or any other place along the road. This patch, or stickers that make up a patch, possibly as an artful camouflage, can cause the classifier to output a random or specific targeted faulty class instead of the correct one. An example with stop sign is found in [EYK18]. This result should be possible to extend to road detection and lane marking. There might be a fuzzy border between deliberate and natural causes, for example with debris, dirt and snow.

The first type of adversarial attacks is primarily a security problem, since a specific noise need to be added to the image and that is difficult to add without tampering with the physical system to control what images are presented to the neural network. Adversarial attacks are typically done on images. It is unclear if it is more difficult to fool a system based on multi modal input.

In white box attacks the attacker has full knowledge how the classification system, which is not a realistic assumption. White box attacks tells more of a specific property of neural networks, than being a real threat. The approach here is to train an adversarial example that makes the original network classify incorrectly. It turns out that adversarial examples for a network of unknown architecture can be constructed by training another network on similar data. That is, in order to figure out how to set up black box attack (i.e. only input

and output are observed), one can practise first on white box attacks. Models trained on input including adversarial examples do not become immune. Also, it seems that adversarial perturbations also transfer between different network architectures, and even between networks trained on disjoint sets of training data. The term universal noise has been used in [MET17], who focused on an adversary which tries to hide all instances of the class pedestrians, while preserving background classes, rendering an output as if pedestrians did not exist in the input image.

The result from adversarial attacks point to a need to make neural networks more stable. It is suggested in [GOO15] that more powerful optimization using more non-linear models will decrease the problem, that the family of models used today are intrinsically flawed.

A realistic adversarial classification approach is to print adversarial patches to be added to a scene. The patch is typically robust to camera angle and lighting conditions. In [BRO17] it is shown how the classification of a banana turns into a toaster by adding a strangely looking patch of approximately the same area as the banana. An explanation why this works is that the adversarial patch exploits learned features by “producing [intermediary] inputs much more salient than objects in the real world”.

7. Dissemination and publications

7.1 Dissemination

How are the project results planned to be used and disseminated?	Mark with X	Comment
Increase knowledge in the field	X	
Be passed on to other advanced technological development projects	X	The rapid expansion of research in this area has been fascinating, there are however a number of reasons to believe that the current way deep neural networks are structured and optimized is still maturing.
Be passed on to product development projects	X	
Introduced on the market		The result is a contribution to an on-going development process we nowadays (as opposed to the general opinion when the project started) understand that this type of technology is a ubiquitous part of perception for an automated driver.
Used in investigations / regulatory / licensing / political decisions		

7.2 Publications

Fast LIDAR-based Road Detection Using Fully Convolutional Neural Networks, [CAL17a]: This work was concerned with road detection using a lidar as the main sensor. The problem was cast as a semantic segmentation task and it was approached using a fully convolutional neural network (FCN). The FCN was designed to have a large receptive field and used high-resolution feature maps in order to achieve higher accuracy. Top-view images encoding several basic statistics were generated to summarize the information content of unstructured point clouds into a format that was suitable for processing with the FCN. The proposed system carried out real-time inference at about 55 Hz using a modern GPU and achieved state-of-the-art performance on the KITTI road benchmark reaching a MaxF score of 94.07% and outperforming by over 3.4 percentage points the second best lidar-only system.

LIDAR based driving path generation using fully convolutional neural networks, [CAL17b]: In this paper, a deep learning approach was developed to carry out driving path generation in lidar top-view images. This was accomplished by fusing lidar point clouds with GPS-IMU information and driving directions using an FCN. The system was trained using as ground truth driving paths followed by human drivers. An

advantage of this approach was that the annotations were obtained automatically and therefore a large data set for supervised learning could be collected with limited effort. Several combinations of inputs were considered to determine the effect of individual information modalities. The main result was that the FCN trained with all the available information (i.e., lidar, GPS-IMU, and driving directions) achieved the highest accuracy thus confirming that the chosen architecture and data representation were suitable for carrying out information fusion and that each modality contributed to the overall system performance.

LIDAR–camera fusion for road detection using fully convolutional neural networks, [CAL19a]: As in [CAL17a], this work was also concerned with the task of road detection. However, in this case, the problem was solved by integrating RGB images and lidar point clouds that were projected onto the image plane. One of the main contributions of this work was the introduction of a novel fusion FCN architecture, called cross fusion, that allowed multimodal information fusion at any depth level in the network by using trainable cross connections. The cross fusion FCN outperformed two other established fusion approaches, early fusion and late fusion, as well as the single modality (RGB or lidar) FCNs. The system was evaluated on the KITTI road benchmark and ranked second with a MaxF score of 96.03%.

Lidar-Camera Co-Training for Road Detection, [CAL19b]: This paper studies a semi-supervised approach for training road detection based on multiple sensor views. An advantage of a semi-supervised learning architecture, as compared to a traditional supervised learning, is that a performance improvement can be made without increasing the amount of costly manually labelled data. A prerequisite is to use input from two complementary sensors such that during semi-supervised training, information from one view, not contained within the other view, can lead to better generalization. The training procedure is initialized by training classifiers with supervised learning, followed by an iterative training of teacher and student roles. It is shown that 36 labelled images together with several thousands of unlabelled images can be compared to supervised training with about twice as many labelled examples. Also shown is an improvement in F1-score of a handful percentage units as compared to only using labelled data.

8. Conclusions and future research

The PRELAT project has successfully contributed to the area of automated driving. The initial ambitious were high and spanned over multiple domains. When looking a research contributions over the five years the project has been run, it is remarkable how rapid the development of ideas in these areas has been. Despite this, many of the observations and questions we made in the start are still valid for further research, maybe because they were broad and basic.

It was shown in the research community around 2010 that convolutional neural networks improved image classification greatly. End-to-end learning of neural networks for self-driving vehicles had been suggested as early as in the 1980s. It was obvious from the literature in 2014 that using deep neural networks was a possible path and PRELAT became a pioneering project for Volvo Technology in this area.

The PRELAT project has concentrated on the lateral control including necessary prediction and trajectory planning. The specific technologies used are lidar and deep learning. A licentiate thesis has been written (2018) and the PhD student at Chalmers is proceeding to a doctoral thesis that will be completed in late 2020 or early 2021.

The project has explored several applications of deep learning methods in the context of driver automation. It has been shown that CNNs, which are commonly used for processing camera images, can also be used effectively for working exclusively with lidar data [CAL17a] or in combination with other sensors or information modalities [CAL17b], [CAL19a]. The emphasis on the lidar sensor was motivated by its capacity to provide highly accurate range measurements in any lighting conditions; this alternative view of the environment can then be leveraged for making driving automation possible in a broader spectrum of external conditions. It has been shown that sparse bird's-eye view images can directly be used as input to an FCN for carrying out road segmentation with high accuracy and fast detection. The lidar-only deep neural network, achieved state-of-the-art performance on the KITTI road benchmark.

By working in the same top-view perspective, a novel end-to-end learning system was developed for generating driving paths [CAL17b]. Whereas a behaviour reflex approach simply predicts the next driving action, a driving path specfies the vehicle's future positions over a longer horizon (30 meters, in this work). This makes the output of the proposed approach more interpretable and informative for further planning and decision-making. Additionally, the ground truth annotations for training the FCN are obtained in an automatic fashion thus allowing the generation of large training sets with no additional human effort and the possibility of continuous learning, that is, training the system (off-line) while it is running in order for it to adapt to entirely novel situations. In order to exploit the visual pattern recognition capability of CNNs, information about the vehicle's

past motion and future destination were transformed into a spatial representation that allowed a direct integration with the point cloud top-views.

Multimodal sensor fusion has also been investigated for the task of road detection. It has been shown that by combining camera images and lidar data, an FCN can carry out road segmentation in a wider range of external conditions than possible using only RGB images [CAL19a]. Besides evaluating two established fusion FCNs, early and late fusion, a novel multimodal FCN has been developed that can learn at what abstraction level and to what extent lidar and camera information should be fused. This approach was evaluated on the KITTI road benchmark for which it achieved state-of-the art performance. In fact, it outperformed by almost two percentage points the system introduced in [CAL17a], a result that highlights the importance of developing multimodal approaches in the quest towards full driving automation.

The project has also studied another learning paradigm. A semi-supervised approach for training road detection based on multiple sensor views was studied in [CAL19b]. The advantage is that a performance improvement can be made without increasing the amount of costly manually labelled data. A prerequisite is to use sensors with complementary information contents such as camera and lidar. The training procedure is initialized by training classifiers with supervised learning, followed by an iterative training of teacher and student roles.

What are the technical reasons why L4 driver automation is difficult? The underlying technical difficulties lead to that it is easy to demonstrate a self-driving robot but it is hard to prove it will stay safe and productive for an extended time period. Deep learning has proved better than previous techniques for classification, for example of road signs, and there is reason to believe that this is true for road detection too. With higher sensor modality input, and with a next generation of lidar, radar etc. having increased performance, classification could very well outperform humans, even during more difficult situations. However, if there is a surprise, something the network hasn’t been trained for, the fundamental problem of understanding the environment is still there.

The problem of automated driving on public roads is broader than what has been studied in the PRELAT project. It is difficult to train or program computers to understand the world, an early and famous example is found in Figure 17 [TED15]. When the context of an empty road switches to something completely different, a human with driving license might have no problem at all to understand and how to handle the situation. The problem is that there are so many of these unlikely events.

Figure 17 "Woman in wheelchair chasing ducks with a broom" (Picture from [TED15])

From a societal perspective the automated driver needs to, loosely speaking, be at least as safe as a human driver to be accepted since there doesn't seem to be any clear substantial societal tradeoff. This means an L4 with delimitations that we can guarantee not to violate. Today's perception, decision and planning have an inferior ability to understand the world correctly and react accordingly to be safe, and which is more important, it is difficult to prove it is safe. Since an L5 will take very many years to develop and validate, it leaves us with the only choice of an L4 system with constraints that will make the perception-decision-control task easier. A tricky part here is to, without being conservative, formulate and verify ODD and safety goals such that the vehicle always stays within its ODD. A suggestion is to structure information flow to evaluate the confidence of information necessary to attend to. Future work on safety is something that affects everyone and safety concepts are better public and standardized such that particular product claims becomes easy to verify.

9. Participating parties and contact persons

Luca Caltagirone, PhD student, Chalmers, luca.caltagirone@chalmers.se
Mattias Wahde, Professor of Applied Artificial Intelligence Department of Applied Mechanics, Chalmers, mattias.wahde@chalmers.se
Martin Sanfridson, Project leader and research engineer, Volvo Technology AB, martin.sanfridson@volvo.com

10. References

[ADAPT] AdaptIVe, http://www.adaptive-ip.eu/, EU FP7 project, Grant Agreement no. 610428, 2014-2017.

[ADA17] Etemad A. et al., "D1.7 System architecture and updated system specification", AdaptIVe, 2017

[BRO17] Brown et al., "Adversarial Patch", 31st Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, USA, 2017

[BEH16] Behere S., "Reference architecture for highly automated driving", Doctoral Thesis, KTH, Stockholm, 2016

[CAL17a] Caltagirone L., Scheidegger S., Svensson L, and Wahde M, "Fast LIDAR-based Road Detection Using Fully Convolutional Neural Networks", IEEE Intelligent Vehicle Symposium, Redondo Beach, CA, USA, June 2017

[CAL17b] Caltagirone L., Bellone M., Svensson L., Wahde M., "LIDAR based driving path generation using fully convolutional neural networks", IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), 2017.

[CAL19a] Caltagirone L, Bellone M, Svensson L, Wahde M, "LIDAR–camera fusion for road detection using fully convolutional neural networks", Robotics and Autonomous Systems, Volume 111, January 2019

[CAL19b] Caltagirone L, Wahde M, Svensson L, M Sanfridson, "Lidar-Camera Co-Training for Road Detection", arXiv:1911.12597, November 2019

[DOV18] Dovner G., "Path planning and decision making in a highway exit situation", master thesis 2018:EX044, Chalmers, June 2018

[DRE18] Da Lio, "D2.3 – Report on the Runtime system", Dream-like simulation abilities for automated cars, Horizon2020, No. 731593, 2018

[EYK18] Eykholt et al., "Robust Physical-World Attacks on Deep Learning Visual Classification", CVPR 2018

[ETSI] ETSI, "Analysis of the Collective Perception Service", ETSI TR 103 562, www.etsi.org, accessed 2019.

[FREE], Free energy principle, https://en.wikipedia.org/wiki/Free_energy_principle, accessed 2019

[GOO15] Goodfellow et al., “Explaining and harnessing adversarial examples”, ICLR 2015.

[HAN13] Hancock, P.A., “Automation: how much is too much?”, Ergonomics, vol. 57, no 3, p. 449-454, 2013

[J3016] J3016 SAE standard, “Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems”, J3016_201806, 2018.

[KAL16] Kalra N. and Paddock S., ” How Many Miles of Driving Would It Take to Demonstrate Autonomous Vehicle Reliability?”, Santa Monica, CA, RAND Corporation, https://www.rand.org/pubs/research_reports/RR1478.html, 2016

[KITTI] Geiger et al., “KITTI Vision Benchmark Suite” http://www.cvlibs.net/datasets/kitti/eval_road.php, 2013

[LI17] Li et al., “Understanding error propagation in deep learning neural network accelerators and applications”, ACM Proceedings of SC17, Denver, November 2017.

[MET17] Metzen et al., “Universal Adversarial Perturbations Against Semantic Image Segmentation”, IEEE International Conference on Computer Vision (ICCV), 2017.

[NTSB19] National Transportation Safety Board, ”Vehicle Automation Report – Tempe, AZ”, HWY18MH010, November 2019.

[PERC] PERCEPTRON, FFI project reference number 2017-01942, http://www.perceptron.nu/, 2019

[RAK19] Rakin et al., “Bit-flip attack: crushing neural network with progressive bit search”, arXiv:1903.12269v2 [cs.CV], April 2019.

[SCH19] Schorn et al.,”An efficient bit-flip resilience optimization method for deep neural networks”, DATE19, Florence, March 2019.

[SKA09] Skarin D. and Karlsson J., “Software Mechanisms for Tolerating Soft Errors in an Automotive Brake-Controller”, IEEE/IFIP Conference on Dependable Systems and Networks, 2009

[SKO17] Sköld M, “Implementaiton of lateral control for autopilot and platoon”, internal presentation at Volvo Technology, 2017.

[TED15] Chris Urmson, “Woman in wheelchair” https://www.ted.com/talks/chris_urmson_how_a_driverless_car_sees_the_road/transcript, TED talks, 2015.

[TOM19] TomTom RoadDNA, https://www.tomtom.com/automotive/automotive-solutions/automated-driving/hd-map-roaddna/, August 2019.

[YAV19] Trasiev Yavor, “Performance modelling and simulation of automotive camera sensors”, Master thesis report, Chalmers, September 2019.

[QIU19] Qiu et al., “Review of Artificial Intelligence Adversarial Attack and Defense Technologies”, Journal of Applied Sciences, Volume 9 Issue 5, 2019.

[WARD18] Ward E, “Models Supporting Trajectory Planning in Autonomous Vehicles”, Doctoral Thesis, KTH, Stockholm, 2018

[WMAXF] Wikipedia explains F1 score, https://en.wikipedia.org/wiki/F1_score, December 2019